Multilevel Monte Carlo Methods for UQ

Part II

Robert Scheichl
r.scheichl@uni-heidelberg.de

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Institut für Angewandte Mathematik, Universität Heidelberg

Short Course, **Penn State University**, 21 & 23 April 2021
https://katana.iwr.uni-heidelberg.de/teaching/pennstate2021/

6. Random Fields

Model Elliptic PDE & Random Fields

We return to our model elliptic boundary value problem. In particular, we consider

$$-\nabla\cdot(a\nabla u) = f, \qquad \text{on } D \subset \mathbb{R}^d, \qquad u_{|\partial D} = 0, \tag{6.1}$$

where a and f are random fields defined on D.

Definition 6.1

Let $D \subset \mathbb{R}^d$, $d \in \mathbb{N}$, and let $(\Omega, \mathfrak{A}, \mathbb{P})$ be a probability space (see Appendix A). A (real-valued) random field is a mapping

$$a : D \times \Omega \to \mathbb{R}$$

such that each function $a(\mathbf{x}, \cdot) : \Omega \to \mathbb{R}$, $\mathbf{x} \in D$, is a random variable.

Definition 6.2

For each fixed $\omega \in \Omega$ the associated function $a(\cdot, \omega) : D \to \mathbb{R}$ is called a realization of the random field.

Let $\mathbb{R}^D$ denote the set of all real-valued functions $f : D \to \mathbb{R}$. The mapping $\omega \mapsto a(\cdot, \omega)$ from $(\Omega, \mathfrak{A})$ to $(\mathbb{R}^D, \mathfrak{A}(\mathbb{R}^D))$ is measurable and hence a random variable with values in $\mathbb{R}^D$.

Second-order and Gaussian Random Fields

Similar to a random vector or stochastic process, a random field is a family of random variables indexed by a parameter. Instead of an ordered parameter set (e.g. $\mathbb{N}$ or $\mathbb{R}_0^+$), for random fields the parameter is a spatial coordinate.

Definition 6.3

A random field a on $D \subset \mathbb{R}^d$ is said to be of second order if for all $\mathbf{x} \in D$ there holds $a(\mathbf{x}, \cdot) \in L^2(\Omega; \mathbb{R})$ (see Appendix A). We say a second-order random field a has mean function $\overline{a}(\mathbf{x}) := \mathbb{E}[a(\mathbf{x}, \cdot)]$ and covariance function

$$c(\mathbf{x}, \mathbf{y}) := \mathbf{Cov}(a(\mathbf{x}, \cdot), a(\mathbf{y}, \cdot)), \qquad \mathbf{x}, \mathbf{y} \in D.$$

A sufficient and necessary condition is that $c(\mathbf{x}, \mathbf{y})$ is **symmetric and positive semidefinite**.

Definition 6.4

A random field on $D \subset \mathbb{R}^d$ is called Gaussian if, for any $n \in \mathbb{N}$ and for any $\mathbf{x}_1, \ldots, \mathbf{x}_n \in D$, the random vector $[a(\mathbf{x}_1, \cdot), \ldots, a(\mathbf{x}_n, \cdot)]$ follows an n-variate normal distribution. It is uniquely determined by its mean and covariance function.

Random Fields in $L^2(D)$ – Karhunen-Loève Expansion

Let a be a 2nd-order random field on $D \subset \mathbb{R}^d$ with mean $\overline{a}$. Then the centred field $a - \overline{a}$ can be expanded in any complete orthonormal system $\{\psi_m\}_{m \in \mathbb{N}}$ of $L^2(D)$.

The Karhunen-Loève expansion of a results from choosing as a particular CONS the eigenfunctions of the covariance operator $C : L^2(D) \to L^2(D)$ of a, given by

$$(Cu)(\mathbf{x}) = \int_D u(\mathbf{y}) c(\mathbf{x}, \mathbf{y}) \, \mathrm{d}\mathbf{y}, \quad \mathbf{x} \in D. \tag{6.2}$$

Theorem 6.5 (Karhunen-Loève (KL) Expansion)

Let $a \in L^2(\Omega; L^2(D))$ (see Appendix A) with mean function $\overline{a}(\mathbf{x})$ and denote by $(\lambda_m, a_m)_{m \in \mathbb{N}}$, $\|a_m\|_{L^2(D)} = 1$, the sequence of eigenpairs of the covariance operator C in descending order. Then

$$a(\mathbf{x}, \omega) = \overline{a}(\mathbf{x}) + \sum_{m=1}^{\infty} \sqrt{\lambda_m} \, a_m(\mathbf{x}) \, \xi_m(\omega), \tag{6.3}$$

where the random variables $\xi_m(\omega) = \frac{1}{\sqrt{\lambda_m}} (a(\cdot, \omega) - \overline{a}, a_m)_{L^2(D)}$ have mean zero, unit variance and are pairwise uncorrelated. The series converges in $L^2(\Omega; L^2(D))$. If the random field is, in addition, Gaussian, then $\xi_m \sim \mathrm{N}(0, 1)$ are i.i.d.

One-Dimensional Example [Ghanem & Spanos, 1991]

Example. For $d = 1$, $D = [-1, 1]$ and the exponential covariance function

$$c(x, y) = e^{\frac{-|x-y|}{\ell}}, \qquad \ell > 0,$$

the eigenvalues of the associated covariance operator are given by

$$\lambda_m = \frac{2\ell}{\ell^2\omega_m^2 + 1}, \ (m \text{ even}), \qquad \lambda_m = \frac{2\ell}{\ell^2\tilde{\omega}_m^2 + 1}, \ (m \text{ odd})$$

where ω_m and $\tilde{\omega}_m$ denote the solutions of the transcendental equations

$$1 - \omega\ell\tan(\omega) = 0 \quad \text{and} \quad \tilde{\omega}\ell + \tan(\tilde{\omega}) = 0, \quad \text{respectively.}$$

The associated eigenfunctions are given by

$$f_m(x) = \sqrt{\tfrac{2\omega_m}{1+\sin(2\omega_m)}}\cos(\omega_m x), \qquad \tilde{f}_m(x) = \sqrt{\tfrac{2\tilde{\omega}_m}{1+\sin(2\tilde{\omega}_m)}}\sin(\tilde{\omega}_m x).$$

However, in general it is not possible to compute the KL-expansion analytically.

Practical Application – Truncated KL Expansion

- The KL expansion suggests a convenient approach for approximating a random field to a specified accuracy by truncation:

$$a(\mathbf{x}, \omega) \ \approx \ a_s(\mathbf{x}, \omega) := \overline{a}(\mathbf{x}) + \sum_{m=1}^{s} \sqrt{\lambda_m}\, a_m(\mathbf{x})\, \xi_m(\omega). \tag{6.4}$$

- The truncated RF a_s has the same mean as a and the covariance function

$$c_s(\mathbf{x}, \mathbf{y}) = \sum_{m=1}^{s} \lambda_m a_m(\mathbf{x}) a_m(\mathbf{y}), \qquad \mathbf{x}, \mathbf{y} \in D\,, \tag{6.5}$$

  converges uniformly to c as $S \to \infty$.

- For the variance of the truncated KL expansion, we have DIY

$$\mathbf{Var}(a(\mathbf{x}, \cdot)) - \mathbf{Var}(a_s(\mathbf{x}, \cdot)) = \sum_{m=s+1}^{\infty} \lambda_m a_m(\mathbf{x})^2 \geq 0.$$

  Hence, a_s always underestimates the variance of a. Moreover, this implies

$$\|a - a_s\|^2_{L^2(\Omega; L^2(D))} = \sum_{m=s+1}^{\infty} \lambda_m = \int_D \mathbf{Var}\, a(\mathbf{x})\, \mathrm{d}\mathbf{x} - \sum_{m=1}^{s} \lambda_m\,,$$

  i.e. the truncation error in $L^2(\Omega; L^2(D))$ is explicitly computable.

Stationary and Isotropic Random Fields

Definition 6.6

(a) A random field a is stationary or homogeneous if it is invariant under translation, i.e. if the multivariate distributions of $(a(\mathbf{x}_1,\cdot),\ldots,a(\mathbf{x}_n,\cdot))$ and $(a(\mathbf{x}_1+\mathbf{h},\cdot),\ldots,a(\mathbf{x}_n+\mathbf{h},\cdot))$ are the same, for any $\mathbf{x}_1,\ldots,\mathbf{x}_n$ and $\mathbf{h}$.

(b) A stationary random field a is isotropic if its covariance function is invariant under rotations, i.e.,

$$c(\mathbf{x},\mathbf{y}) = c(r), \qquad r = \|\mathbf{x}-\mathbf{y}\|_2.$$

Example (Isotropic Gaussian covariance).
A simple and widely used example of an isotropic covariance function is the *Gaussian covariance* $c(r) = \sigma^2 e^{-r^2/\rho^2}$, where σ^2 and ρ are two constants defining the *variance* and the *correlation length* of the field.

The Matérn Class

A family of isotropic covariance functions that is very popular in spatial statistics or machine learning, is the Matérn class with covariance function given by

$$c(r) = \frac{\sigma^2}{2^{\nu-1}\,\Gamma(\nu)} \left(\frac{2\sqrt{\nu}\,r}{\rho}\right)^{\nu} K_\nu\left(\frac{2\sqrt{\nu}\,r}{\rho}\right), \qquad r = \|\mathbf{x}-\mathbf{y}\|_2, \tag{6.6}$$

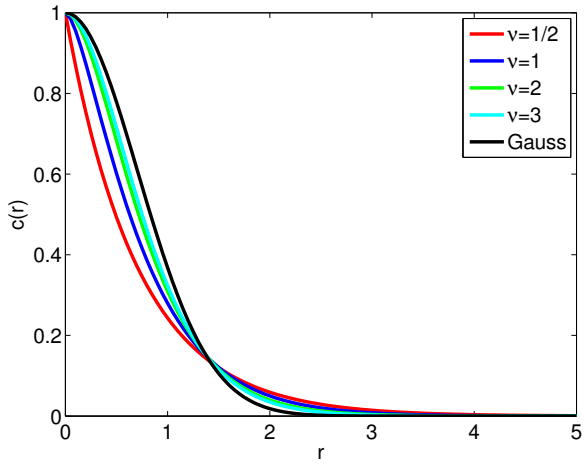
where

K_ν	is the modified (second-kind) Bessel function of order ν,
Γ	denotes the Gamma-function,
ν	is known as the smoothness parameter,
σ^2	is the variance parameter,
ρ	is the correlation length parameter.

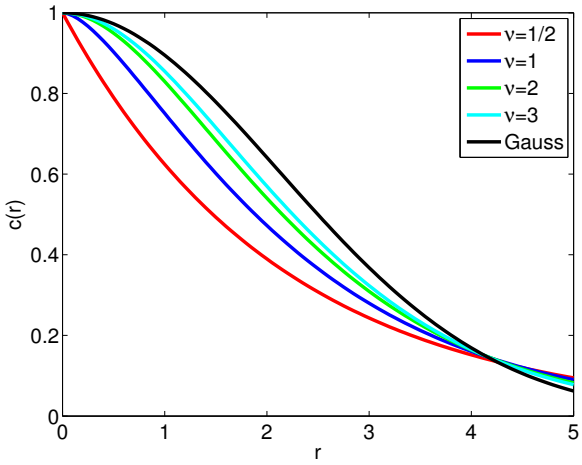
It contains exponential, Gaussian, as well as Bessel covariance functions as special cases:

$\nu = \frac{1}{2}$:	$c(r) = \sigma^2 \exp(-\sqrt{2}r/\rho)$	exponential covariance
$\nu = 1$:	$c(r) = \sigma^2 \left(\frac{2r}{\rho}\right) K_1\left(\frac{2r}{\rho}\right)$	Bessel covariance
$\nu \to \infty$:	$c(r) = \sigma^2 \exp(-r^2/\rho^2)$	Gaussian covariance

The Matérn Class



$\rho = 1$



$\rho = 3$

- By reducing the correlation length ρ the Matérn covariance function can be concentrated more strongly near $r = 0$.
- By increasing the smoothness parameter ν the Matérn covariance function becomes smoother at $r = 0$. (It is analytic everywhere else.)
- Flexible parametrisation allows its application to many statistical situations.
  (Parameters may be estimated from observed data using statistical techniques.)

Eigenvalue Decay for the Matérn Class

A result by H. Widom from 1963 allows us to analyse the decay rate of the eigenvalues of the covariance operator of isotropic random fields:

Theorem 6.7 (Widom, 1963)

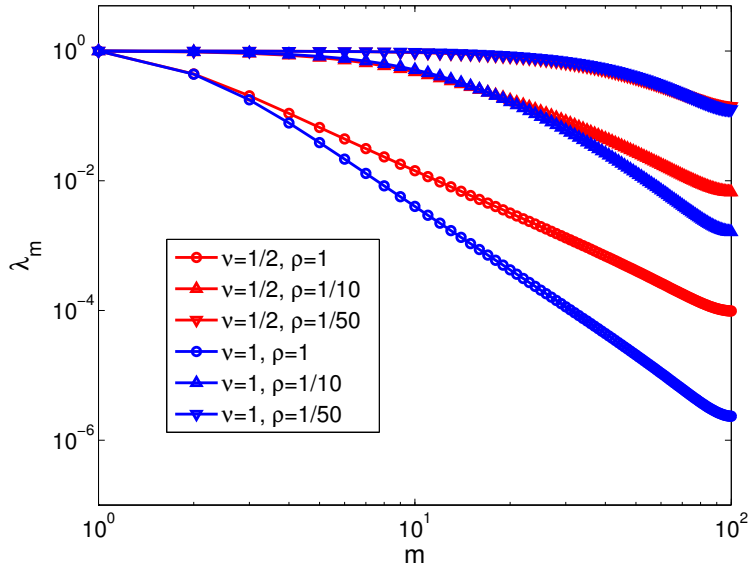
Let $c = c(r)$ be the (isotropic) Matérn covariance function with parameters ν, σ^2 and ρ. Let D be a bounded domain in $\mathbb{R}^d$ and let $\{\lambda_m\}_{m\in\mathbb{N}}$ denote the (nonincreasing) eigenvalues of the covariance operator C given by (6.2).

$$\lambda_m \eqsim m^{-(1+2\nu/d)}, \qquad \text{for } m \to \infty.$$

- Allows to estimate truncation error and thus dimensionality of the problem.
- Rate of convergence of the eigenvalues is crucial to obtain dimension-independent QMC and sparse grid quadrature and approximation results.
- The (spatial) smoothness of realizations is also linked directly to the parameter ν: in particular, a random field with Matérn covariance function is k-times mean-square differentiable if and only if $\nu > k$.

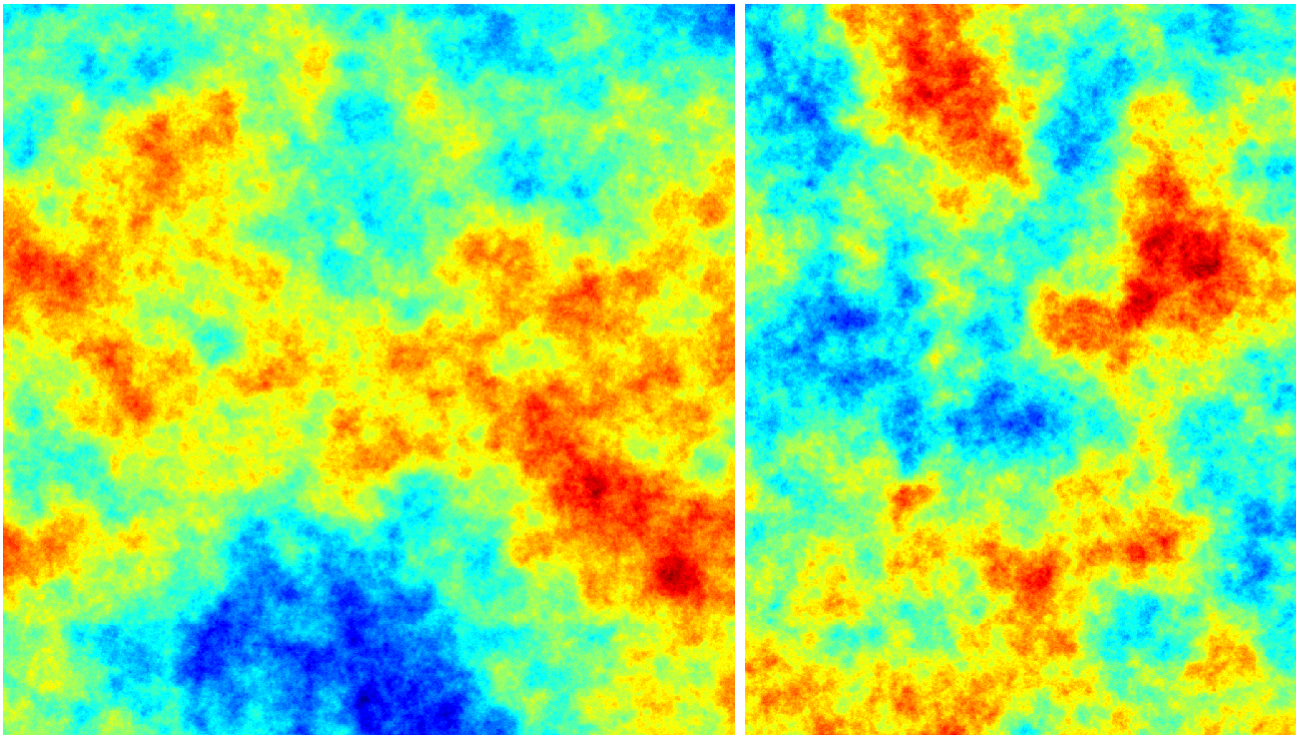
Asymptotic Eigenvalue Decay & Plateau (Matérn)

Before asymptotic decay sets in (rate determined by smoothness parameter ν), there is a preasymptotic plateau. Its length is determined by the correlation length ρ.



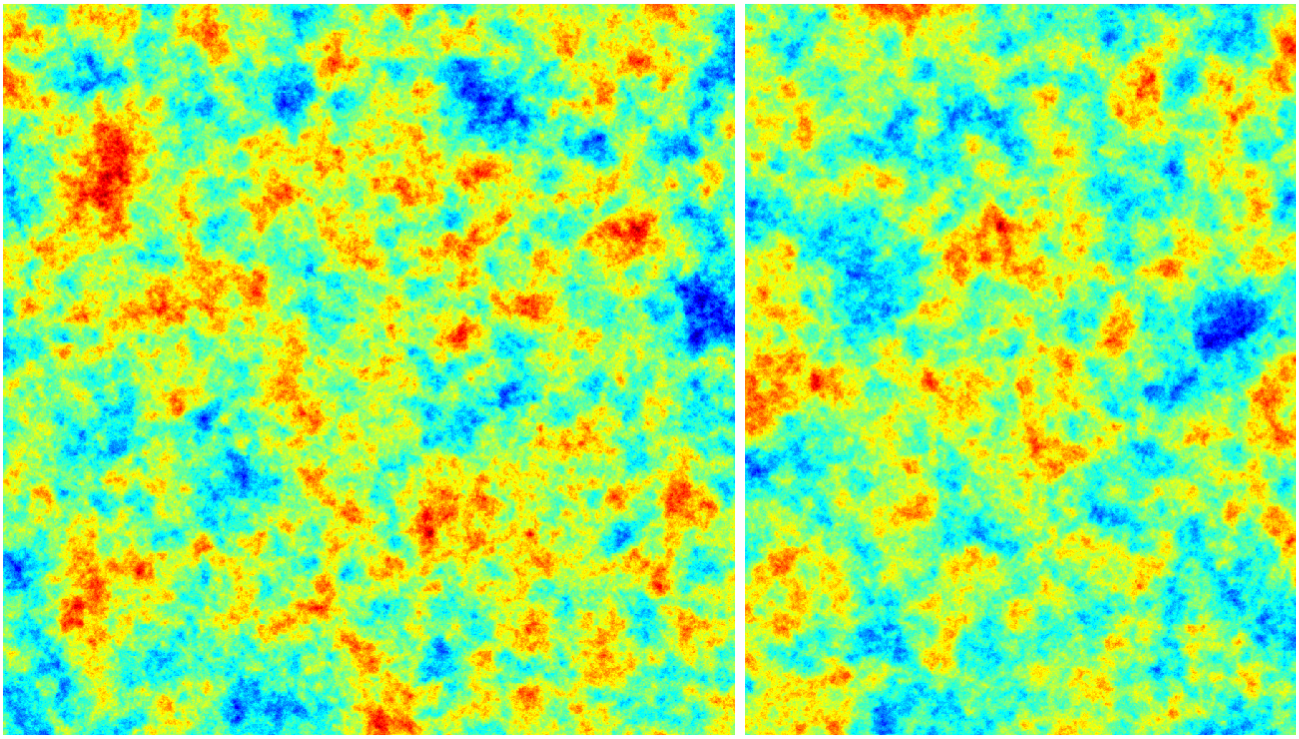
Eigenvalue decay, Matérn covariance kernel, $D = [-1, 1]$.

Realizations of Gaussian Random Fields



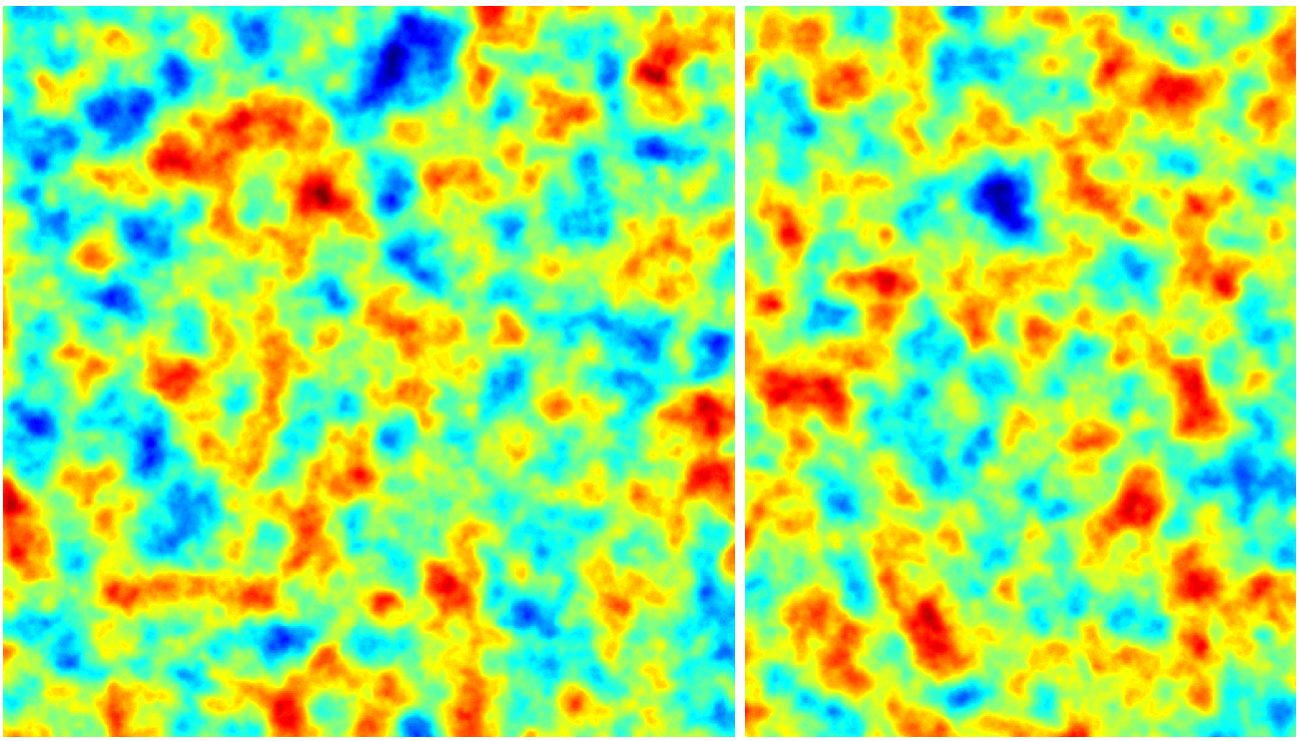
Matérn covariance: $\nu = 1/2$, $\sigma = 1$, $\ell = 0.5$

Realizations of Gaussian Random Fields



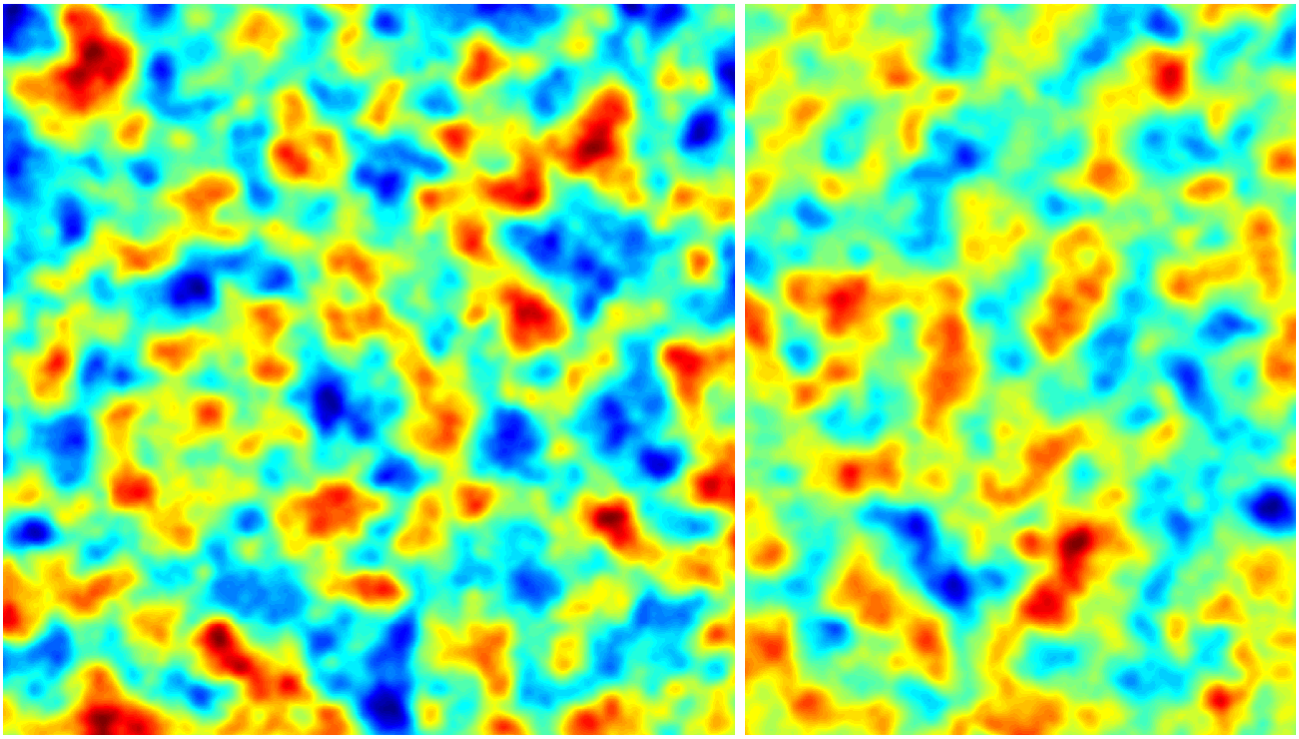
Matérn covariance: $\nu = 1/2$, $\sigma = 1$, $\ell = 0.05$

Realizations of Gaussian Random Fields



Matérn covariance: $\nu = 3/2$, $\sigma = 1$, $\ell = 0.05$

Realizations of Gaussian Random Fields



Matérn covariance: $\nu = 5/2$, $\sigma = 1$, $\ell = 0.05$

Further Reading on Random Fields

- KL expansion is widely used (especially in theoretical NA literature), **but** especially for rough fields (e.g. $\nu < 1$), cost can grow very quickly.
- For isotropic RF more efficient: circulant embedding and other FFT methods:
  - Dietrich & Newsam, Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix, *SIAM J Sci Comput* **18**, 1997
  - Graham, Kuo, Nuyens, **RS** & Sloan, Analysis of circulant embedding methods for sampling stationary random fields, *SIAM J Num Anal* **56**, 2018
  - Bachmayr, Graham, Nguyen & **RS**, Unified analysis of periodization-based sampling methods for Matérn covariances, Preprint arXiv:1905.13522, 2019
- Exploiting a link between the inverse C^{-1} of the covariance operator and stochastic PDEs, e.g. Matérn fields a can be sampled by solving the sPDE

$$(\kappa^2 - \Delta)^\beta a(\mathbf{x}, \omega) \;=^d\; \mathcal{W}(\mathbf{x}, \omega) \quad \text{in } \mathbb{R}^d,$$

  where Δ is the Laplacian and $\mathcal{W}$ is Gaussian white noise on $\mathbb{R}^d$.
  (The parameters are related by $\nu = 2\beta - \frac{d}{2}$, $\rho = 2\frac{\sqrt{\nu}}{\kappa}$ and $\sigma^2 = \sigma^2(\kappa, \beta)$.)
  - Lindgren, Rue & Lindström, An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic PDE approach, *J Roy Statist Soc B* **73**, 2011
  - Bolin, Kirchner, Kovács, Numerical solution of fractional elliptic stochastic PDEs with spatial white noise, *IMA J Num Anal* **40**, 2020
  - Drzisga, Gmeiner, Rüde, **RS** & Wohlmuth, Scheduling massively parallel multigrid for multilevel Monte Carlo methods, *SIAM J Sci Comput* **39**, 2017

7. Monte Carlo Finite Element Methods

Elliptic Boundary Value Problems with Random Data

We return again to our model elliptic boundary value problem with random data

$$-\nabla\cdot(a\nabla u) = f, \qquad \text{on } D \subset \mathbb{R}^d, \qquad u_{|\partial D} = 0, \tag{7.1}$$

where a and f are random fields on D with respect to a probability space $(\Omega, \mathfrak{A}, \mathbb{P})$.

- If f is random, we assume $f(\cdot,\omega) \in L^2(D)$ for (almost) all $\omega \in \Omega$.
- **Could** require coefficient a to satisfy Assumption 1 in Appendix B **uniformly** to ensure existence & uniqueness of $u(\cdot,\omega) \in H^1_0(D)$ with $\|\cdot\|_{H^1_0(D)} = |\cdot|_{H^1(D)}$. But in many situations too restrictive! The following assumption suffices:

Assumption 1

For almost all $\omega \in \Omega$ ($\mathbb{P}$-a.s.), realizations $a(\cdot,\omega)$ of the coefficient function a are strictly positive and lie in $L^\infty(D)$, i.e.

$$0 < a_{\min}(\omega) \le a(\mathbf{x},\omega) \le a_{\max}(\omega) < \infty \quad \textit{almost everywhere (a.e.)} \quad \textit{in } D, \tag{7.2}$$

where

$$a_{\min}(\omega) := \operatorname*{ess\,inf}_{\mathbf{x}\in D} a(\mathbf{x},\omega), \qquad a_{\max}(\omega) := \operatorname*{ess\,sup}_{\mathbf{x}\in D} a(\mathbf{x},\omega). \tag{7.3}$$

Realization-Wise Solvability

For any realization ω for which Assumption 1 holds and $f(\cdot,\omega) \in L^2(D)$, we may apply the Lax-Milgram Lemma (Lemma B.5) and obtain a unique solution of (7.1).

Theorem 7.1

Let Assumption 1 hold and $f(\cdot,\omega) \in L^2(D)$ $\mathbb{P}$-a.s. Then (7.1) has a unique solution $u(\cdot,\omega) \in H_0^1(D)$ and $|u(\cdot,\omega)|_{H^1(D)} \leq C a_{\min}^{-1}(\omega)\|f(\cdot,\omega)\|_{L^2(D)}$ $\mathbb{P}$-a.s.

Recall Definition A.21, of Banach space-valued L^p-spaces over a probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ – so-called *Bochner spaces*. These spaces provide a generalisation of standard Lebesgues spaces. A result that we will use throughout is:

Lemma 7.2 (Hölder's Inequality)

Let $p, q, r \in [1,\infty]$ be such that $\frac{1}{p} = \frac{1}{q} + \frac{1}{r}$. Then

$$\|XY\|_{L^p(\Omega,W)} \leq \|X\|_{L^q(\Omega,W)}\|Y\|_{L^r(\Omega,W)}, \quad \text{for all } X \in L^q(\Omega,W), Y \in L^r(\Omega,W).$$

Note that the case of $q = \infty$ is explicitly included; in that case $p = r$.
For $p = 1$ & $q = r = 2$, Hölder's Inequality reduces to the Cauchy-Schwarz inequality.
The inequality holds over any measure space Ω; in particular, also in standard Lebesgues spaces.

Summability

The following theorem provides sufficient conditions for u to have finite p-th moments, i.e., to lie in $L^p(\Omega; H_0^1(D))$.

Theorem 7.3

Let Assumption 1 hold. Assume further that the mappings $a : \Omega \to L^\infty(D)$ and $f : \Omega \to L^2(D)$ are measurable and that $a_{\min}^{-1} \in L^q(\Omega;\mathbb{R})$ for some $q \in [1,\infty]$.

(a) *If $f \in L^2(D)$ deterministic (i.e. a degenerate constant RF), then*

$$\|u\|_{L^p(\Omega;H_0^1(D))} \leq C\|a_{\min}^{-1}\|_{L^p(\Omega;\mathbb{R})}\|f\|_{L^2(D)}, \quad \text{for all } p \leq q.$$

(b) *If $f \in L^r(\Omega; L^2(D))$ with $r \in [1,\infty]$ and $\frac{1}{p} = \frac{1}{q} + \frac{1}{r} \leq 1$, then*

$$\|u\|_{L^p(\Omega;H_0^1(D))} \leq C\|a_{\min}^{-1}\|_{L^q(\Omega;\mathbb{R})}\|f\|_{L^r(\Omega;L^2(D))}.$$

Proof. Follows directly from Theorem 7.1 (using Hölder's Inequality for Part (b)).

Finite Element Discretization

- Let $V_h \subset H_0^1(D)$ denote a closed subspace, e.g., the finite element (FE) space of piecewise polynomial functions with respect to a triangulation $\mathscr{T}_h$ of D with mesh width $h > 0$ (see Appendix B).
- **FE system:** Suppose $u_h : \Omega \to V_h$ satisfies $\mathbb{P}$-a.s.

$$\int_D a(\mathbf{x},\omega)\nabla u_h(\mathbf{x},\omega)\cdot\nabla v_h(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_D f(\mathbf{x},\omega)v_h(\mathbf{x})\,\mathrm{d}\mathbf{x} \quad \forall v_h \in V_h\,. \tag{7.4}$$

- Since V_h is a closed subspace of $H_0^1(D)$ with norm $|\cdot|_{H^1(D)}$ all the above results hold in an identical form also for u_h:

Theorem 7.4

The results about solvability and summability, as well as the norm bounds in Theorems 7.1 and 7.3 hold under the same assumptions on a and f also for (7.4) and its solution u_h.

H^2 Regularity Assumption & Error Analysis

The regularity assumption, which is necessary to bound the finite element error (cf. Assumption 2 in Appendix B), is again made only realization-wise.

Assumption 2

For almost all $\omega \in \Omega$, there exists a constant $C_2(\omega) > 0$ such that, for every $f(\cdot,\omega) \in L^2(D)$, we have $u(\cdot,\omega) \in H^2(D)$ and

$$|u(\cdot,\omega)|_{H^2(D)} \le C_2(\omega)\|f(\cdot,\omega)\|_{L^2(D)}.$$

- For Assumption 2 to hold, it suffices that D is convex, $a(\cdot,\omega)$ is Lipschitz continuous and Assumption 1 holds.
- A careful derivation how $C_2(\omega)$ depends on $\|a(\cdot,\omega)\|_{C^{0,1}(D)}, a_{\min}(\omega), a_{\max}(\omega)$ can be found in [Charrier, **RS**, Teckentrup, *SIAM J Num Anal*, 2013].
- In particular, it is shown there that for lognormal a with Matérn covariance, we have $C_2 \in L^p(\Omega;\mathbb{R})$ for all $p < \infty$.

The constant C in the interpolation result on Slide 84 of Appendix B is independent of ω.

Finite Element Convergence Results

Theorem 7.5 (Deterministic or L^∞ RHS)

Let Assumptions 1 and 2 hold, and let $V^h \subset H_0^1(D)$ be the space of piecewise linear FEs with respect to a shape-regular triangulation $\mathscr{T}_h$ (see Appendix B). Furthermore, suppose that $f \in L^\infty(\Omega; L^2(D))$ (in particular includes deterministic f), $a_{\min}^{-1/2} a_{\max}^{1/2} \in L^q(\Omega;\mathbb{R})$ and $C_2 \in L^r(\Omega;\mathbb{R})$ with $q, r \in [1,\infty]$ s.t. $\frac{1}{p} = \frac{1}{q} + \frac{1}{r} \leq 1$, then

$$\|u - u_h\|_{L^p(\Omega;H_0^1(D))} \leq ch\|f\|_{L^\infty(\Omega;L^2(D))}.$$

Proof. Demonstrated on tablet.

- The general case of $f \in L^r(\Omega; L^2(D))$, $r < \infty$ can be proved similarly.
- Via duality arguments it is possible to show faster convergence in the (spatial) $L^2(D)$-norm and for suff. smooth functionals $G(u)$ on $H_0^1(D)$, i.e.

$$\|u - u_h\|_{L^p(\Omega;L^2(D))} = \mathcal{O}(h^2) \quad \text{and} \quad \|G(u) - G(u_h)\|_{L^p(\Omega;\mathbb{R})} = \mathcal{O}(h^2). \quad (7.5)$$

Monte Carlo Finite Element Method

- Our goal now is to use the MC method to estimate a quantity of interest that depends on the (random) solution u. This could be the mean $\mathbb{E}[u(\mathbf{x},\cdot)]$, the variance $\mathbf{Var}[u(\mathbf{x},\cdot)]$ or the expected value of a functional $G(u)$.
- Consider N i.i.d. realizations $a^{(j)} = a(\cdot,\omega_j)$ and $f^{(j)} = f(\cdot,\omega_j)$ and let $u^{(j)} = u(\cdot,\omega_j) \in H_0^1(D)$ and $u_h^{(j)} = u_h(\cdot,\omega_j) \in V_h$ be the associated unique solution and its FE approximation, respectively.
- Compute the ($H_0^1(D)$-valued) MC estimates

$$\overline{u}_{h,N} := \frac{1}{N}\sum_{j=1}^N u_h^{(j)}, \quad s_{h,N}^2 := \frac{1}{N-1}\sum_{j=1}^N \left(u_h^{(j)} - \overline{u}_{h,N}\right)^2,$$

  and the (scalar-valued) estimate

$$\widehat{Q}_{h,N} := \frac{1}{N}\sum_{j=1}^N G(u_h^{(j)}),$$

  for $Q := G(u)$ with $G : H_0^1(D) \to \mathbb{R}$ bounded or Fréchet differentiable.
- To estimate the complexity of these estimators we can use the abstract Theorem 5.1. We simply have to verify Assumptions (5.1) and (5.2).

Let us first consider **Assumption** (5.1):

- For a scalar functional $Q = G(u)$ with $G : H_0^1(D) \to \mathbb{R}$ suff. smooth, using Jensen's inequality (Thm. A.20), it follows from (7.5) that

$$|\mathbb{E}[Q - Q_h]| \leq \mathbb{E}[|G(u) - G(u_h)|] = \mathcal{O}(h^2).$$

  Thus, Assumption (5.1) holds with $\alpha = 2$.

- For $Q = u \in H_0^1(D)$, measuring the bias error in $|\cdot|_{H^1(D)}$, we get again using Jensen's inequality (noting that norms are convex functions) and Theorem 7.5 that

$$|\mathbb{E}[u - u_h]|_{H^1(D)} \leq \mathbb{E}\big[|u - u_h|_{H^1(D)}\big] = \mathcal{O}(h).$$

  Thus in that case, Assumption (5.1) holds with $\alpha = 1$.

Next consider **Assumption** (5.2):

- If the meshes $\mathscr{T}_h$ are (quasi-)uniform (not only shape-regular), then the number of unknowns M_h in the resulting FE system (B.8) satisfies $M_h = \mathcal{O}(h^{-d})$.
- Using a multigrid iterative method it is possible to solve the FE system (B.8) in linear complexity, i.e.

$$\text{Cost}(Q_h^j) \;=\; \mathcal{O}(M_h) \;=\; \mathcal{O}(h^{-d}).$$

  Thus, Assumption (5.2) holds with $\gamma = d$.

Monte Carlo Finite Element Complexity Result

Corollary 7.6

Consider the **Monte Carlo FE method** *with p.w. linear FEs applied to the* **elliptic BVP** (7.1) *in* $\mathbb{R}^d$ *to estimate* $\mathbb{E}[u]$ *or* $\mathbb{E}[G(u)]$, *with* $G : H_0^1(D) \to \mathbb{R}$ *suff. smooth. For any* $\varepsilon > 0$ *and* $\theta \in (0,1)$ *there exist* $h > 0$, $N \in \mathbb{N}$, *such that*

Case $Q = G(u)$: $\|\mathbb{E}[Q] - \widehat{Q}_{h,N}\|_{L^2(\Omega;\mathbb{R})} < \varepsilon$ *or* $\mathbb{P}\{|\mathbb{E}[Q] - \widehat{Q}_{h,N}| < \varepsilon\} > \theta$ *and*

$$\text{Cost}(\widehat{Q}_{h,N}) \;=\; \mathcal{O}(\varepsilon^{-2-d/2}).$$

Case $Q = u$: $\|\mathbb{E}[u] - \overline{u}_{h,N}\|_{L^2(\Omega;H_0^1(D))} < \varepsilon$ *or* $\mathbb{P}\{|\mathbb{E}[u] - \overline{u}_{h,N}|_{H^1(D)} < \varepsilon\} > \theta$ *and*

$$\text{Cost}(\overline{u}_{h,N}) \;=\; \mathcal{O}(\varepsilon^{-2-d}).$$

Proof. For $Q = G(u)$, we can simply apply Theorem 5.1 with $\alpha = 2$ and $\gamma = d$.

For $Q = u$, the bias-variance decomposition also works in the $|\cdot|_{H^1(D)}$-norm (both in mean squared and in probability). To bound the sampling error, we only require square-summability of $u_h : \Omega \to H_0^1(D)$, which is guaranteed by Theorem 7.4 (under suitable conditions on a and f). □

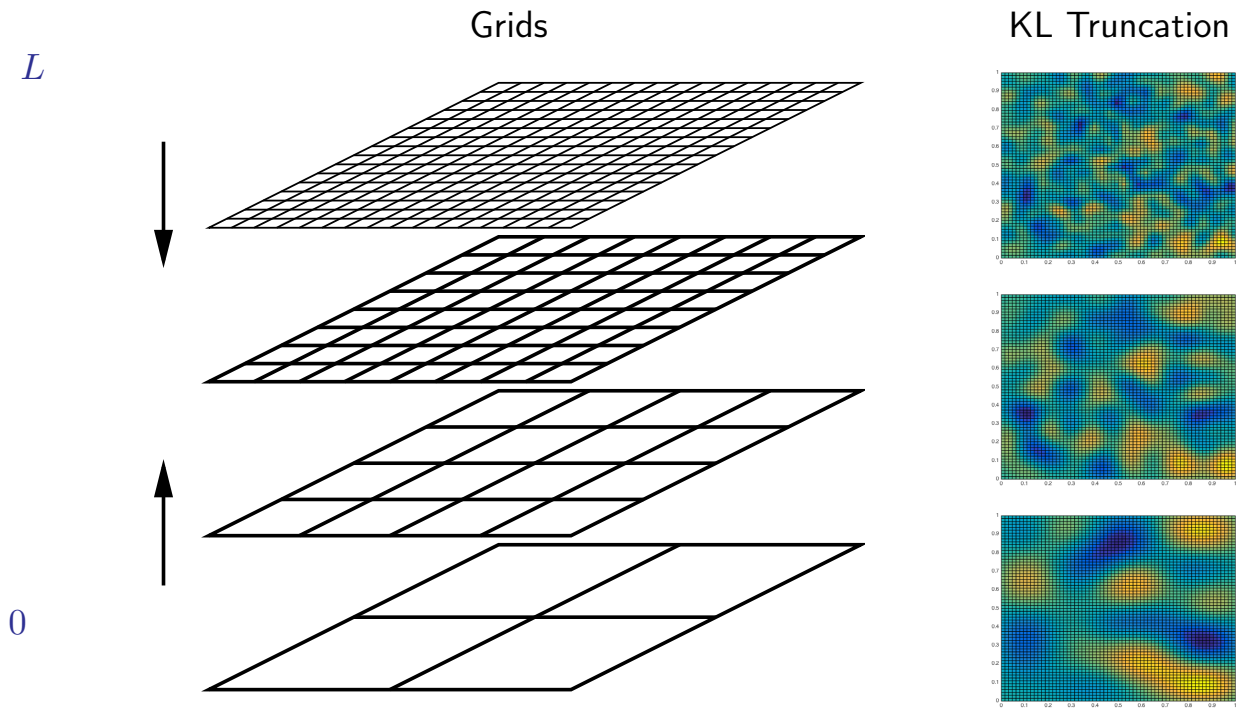
Multilevel Acceleration

- Especially in 2D or 3D this is a very high complexity, but it is straightforward again to accelerate the Monte Carlo FE method via a **multilevel approach.**
- Consider a hierarchy of FE meshes $\mathcal{T}_0, \ldots, \mathcal{T}_L$, for simplicity using uniform grid refinement of an (arbitrary) coarsest grid $\mathcal{T}_0$, i.e. $h_\ell = h_{\ell-1}/2$ $(m = 2)$
  (These grids are also needed in the MG solver assumed above, so no extra overhead!)
- We now use the **abstract complexity theorem**, Theorem 5.2, to estimate the complexity of a multilevel MC-FE estimator for (7.1).
- **Assumptions (M1)** and **(M3)** in Theorem 5.2 have already been verified above. So it only remains to prove **Assumption (M2)**.
- For scalar (smooth) $Q := G(u)$, using (7.5)

$$\begin{aligned}\textbf{Var}\,[Y_\ell] &\leq \mathbb{E}\left[(Q_\ell - Q_{\ell-1})^2\right] \\ &\leq 2\mathbb{E}\left[(G(u) - G(u_{h_\ell}))^2\right] + 2\mathbb{E}\left[(G(u) - G(u_{h_{\ell-1}}))^2\right] = \mathcal{O}(h_\ell^4)\end{aligned}$$

  Thus, **Assumption (M2)** in Theorem 5.2 holds with $\beta = 4$.
- For $Q := u$ we can show similarly that $\beta = 2$.

Grid & Model Hierarchy for Elliptic BVP



Have not really discussed how to sample the field or how to also change the truncation dimension across the levels.

Multilevel Complexity Theorem for the Elliptic BVP

Corollary 7.7 (Case of scalar functional $Q := G(u)$)

Consider the **Multilevel Monte Carlo FE method** *with p.w. linear FEs (uniform refinement) applied to the* **elliptic BVP** (7.1) *in* $\mathbb{R}^d$ *to estimate* $\mathbb{E}[G(u)]$, *with* $G : H_0^1(D) \to \mathbb{R}$ *suff. smooth. For any* $0 < \varepsilon < \exp(-1)$ *and* $\theta \in (0,1)$ *there exist* $L, N_\ell \in \mathbb{N}$, *such that* $\|\mathbb{E}[Q] - \widehat{Q}_L^{ML}\|_{L^2(\Omega;\mathbb{R})} < \varepsilon$ *or* $\mathbb{P}\{|\mathbb{E}[Q] - \widehat{Q}_L^{ML}| < \varepsilon\} > \theta$ *and*

$$\mathit{Cost}(\widehat{Q}_L^{ML}) \;=\; \mathcal{O}(\varepsilon^{-2}).$$

- For $Q = u$ (see above), for less smooth functionals, or for less smooth data, we often obtain only $\alpha = 1$ and $\beta = 2$, so that for $d = 2, 3$ the other regimes in the MLMC complexity theorem become important.
- Also, for rough coefficients often only $\gamma > d$ is possible (even with a MG solver).
- Thus, we can make the following very important observation (for $d = 2, 3$):

Optimality of MLMC (for $\gamma > \beta = 2\alpha$)

In that case, the MLMC cost is asymptotically the same as **one deterministic solve** to accuracy ε, i.e. $\mathrm{Cost}(\widehat{Q}_L^{\mathrm{ML}}) = \mathcal{O}(\varepsilon^{-2-(\gamma-\beta)/\alpha}) = \mathcal{O}(\varepsilon^{-\gamma/\alpha})$ **!!**

Comparison of Complexities

We compare MLMC-FE and MC-FE for (7.1) in the two regimes discussed above:

Case $\alpha = 2$, $\beta = 4$, $\gamma = d$:

d	**MC**	**MLMC**	**Gain**	One Sample Q_L^j
1	$\mathcal{O}(\varepsilon^{-5/2})$	$\mathcal{O}(\varepsilon^{-2})$	$\mathcal{O}(\varepsilon^{-1/2})$	$\mathcal{O}(\varepsilon^{-1/2})$
2	$\mathcal{O}(\varepsilon^{-3})$	$\mathcal{O}(\varepsilon^{-2})$	$\mathcal{O}(\varepsilon^{-1})$	$\mathcal{O}(\varepsilon^{-1})$
3	$\mathcal{O}(\varepsilon^{-7/2})$	$\mathcal{O}(\varepsilon^{-2})$	$\mathcal{O}(\varepsilon^{-3/2})$	$\mathcal{O}(\varepsilon^{-3/2})$

Case $\alpha = 1$, $\beta = 2$, $\gamma = d$:

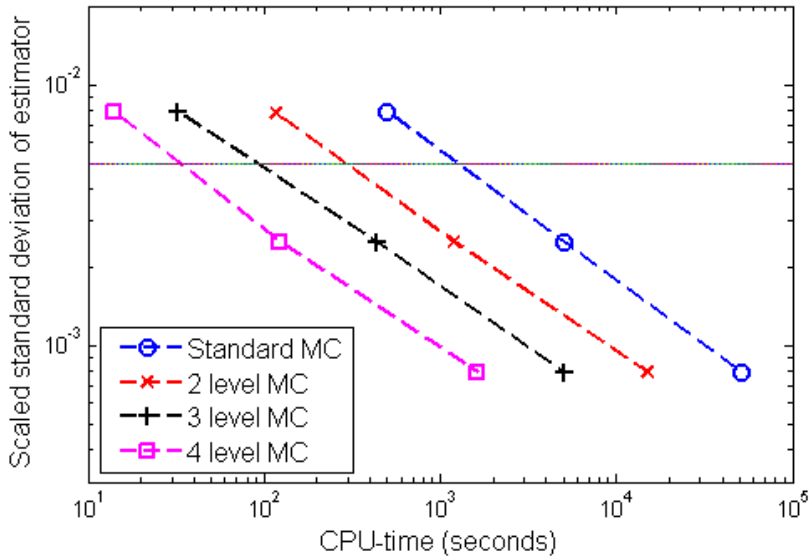
d	**MC**	**MLMC**	**Gain**	One Sample Q_L^j
1	$\mathcal{O}(\varepsilon^{-3})$	$\mathcal{O}(\varepsilon^{-2})$	$\mathcal{O}(\varepsilon^{-1})$	$\mathcal{O}(\varepsilon^{-1})$
2	$\mathcal{O}(\varepsilon^{-4})$	$\mathcal{O}(\varepsilon^{-2})$	$\mathcal{O}(\varepsilon^{-2})$	$\mathcal{O}(\varepsilon^{-2})$
3	$\mathcal{O}(\varepsilon^{-5})$	$\mathcal{O}(\varepsilon^{-3})$	$\mathcal{O}(\varepsilon^{-2})$	$\mathcal{O}(\varepsilon^{-3})$

(ignoring log-factors)

Can we achieve such huge gains in practice?

Multilevel MC-FE Method for Radioactive Waste Disposal Problem

$D = (0,1)^2$; lognormal a w. exponential covariance; $Q = \|u\|_{L_2(D)}$; p.w. linear FE

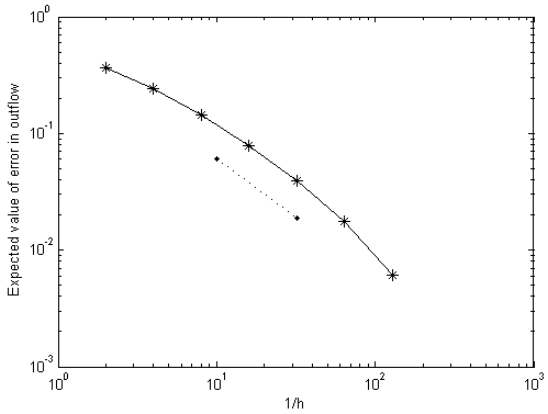


$h_L = 1/256$ (solid line is FE-error)

`Matlab` implementation on 3GHz Intel Core 2 Duo E8400 processor, 3.2GByte RAM, with **sparse direct solver**, i.e. $\gamma \approx 2.4$

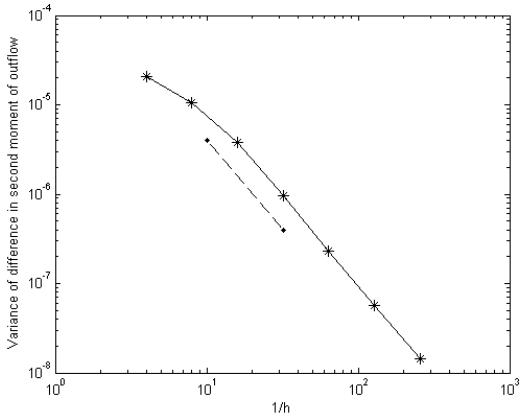
Verifying Assumptions in Complexity Theorem Numerically

Lognormal a with exponential covariance (i.e. $\nu = 1/2$), $\sigma^2 = 1$ and $\lambda = 0.3$.



$\left|\mathbb{E}[G_1(u) - G_1(u_h)]\right|$

where, given $\Psi(x) = x$,
$G_1(u) := (f,\Psi)_{L^2(D)} - (a\nabla u, \nabla\Psi)_{L^2(D)}$
(average flow through D).



$\mathbb{V}\left[G_2(u_h) - G_2(u_{2h})\right]$

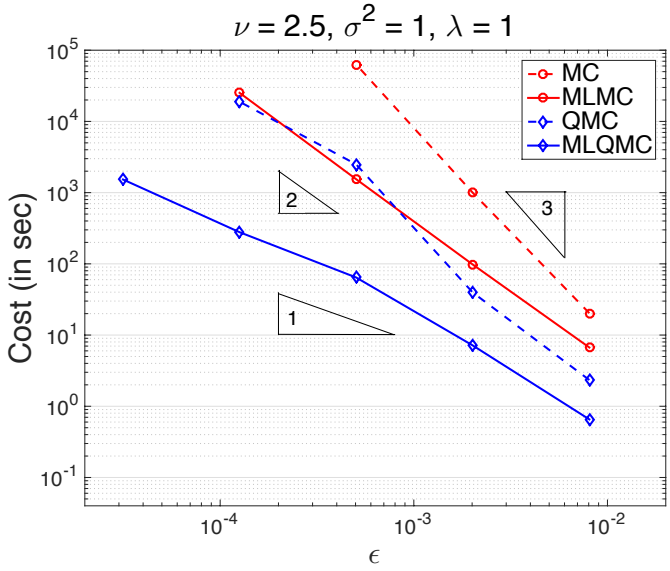
where
$G_2(u) := \left(\frac{1}{|D^*|}\int_{D^*} u(x)\,\mathrm{d}x\right)^2$
(i.e. 2nd moment of u over patch D^*)

$\Longrightarrow \quad \alpha = 1$ and $\beta = 2$

Can be proved rigorously! [Teckentrup, **RS** Giles, Ullmann, *Numer Math* **125**, 2013]

Smoother Coefficients & Outlook to Multilevel QMC

$Q = \frac{1}{|D^*|} \int_{D^*} u \, \mathrm{d}x$ & lognormal a with Matérn covariance and



For QMC using a randomised lattice rule with product weights $\gamma_j = 1/j^2$.

[Kuo, **RS**, Schwab, Sloan, Ullmann, *Math Comput* **86**, 2017]

Further Reading on Multilevel Monte Carlo

- Analysis simplifies considerably for uniformly bounded, affine coefficients, i.e.,

$$0 < a_{\min} = \text{const} < a(\mathbf{x}, \omega) < a_{\max} = \text{const} < \infty \qquad \mathbb{P} - \text{a.s.}$$

  - Barth, Schwab & Zollinger, Multi-level Monte Carlo Finite Element method for elliptic PDEs with stochastic coefficients, *Numer Math* **119**, 2011

- The MLMC-FE method has been applied to many other PDEs. For a comprehensive list see Mike Giles' MLMC Community Webpage
  - `http://people.maths.ox.ac.uk/~gilesm/mlmc_community.html`

- Particular current interest in adaptive FEs and sample-adaptive hierarchies:
  - Kornhuber & Youett, Adaptive Multilevel Monte Carlo Methods for Stochastic Variational Inequalities, *SIAM J Numer Anal* **56**, 2018
  - Detommaso, Dodwell & **RS**, Continuous Level Monte Carlo and Sample-Adaptive Model Hierarchies, *SIAM/ASA J Uncertain Q* **7**, 2019

- In the latter, we have also extended the concept of MLMC to allow for a continuous level parameter ℓ.

Other Multilevel Quadrature Methods in UQ

- As stated above, it is not essential to use **Monte Carlo** estimators to estimate the contributions $\mathbb{E}[Y_\ell]$ from each level.
- **Multilevel quasi-Monte Carlo** uses quasi-MC quadrature rules, i.e. special deterministic point sets (can be unbiased through randomisation):
  - Kuo, Schwab & Sloan, Multi-level quasi-Monte Carlo finite element methods for a class of elliptic PDEs with random coefficients, *Found Comput Math* **15**, 2015
  - Dick, Kuo, Le Gia & Schwab, Multilevel higher order QMC Petrov–Galerkin discretization for affine parametric operator equations, *SIAM J Numer Anal* **54**, 2016
  - Kuo, **RS**, Schwab, Sloan & Ullmann, Multilevel quasi-Monte Carlo methods for lognormal diffusion problems, *Math Comput* **86**, 2017

  with rigorous theory proving almost $\mathcal{O}(\varepsilon^{-1})$ complexity (or better).
- **Multilevel sparse grid approximation/quadrature** uses sparse grid polynomial quadrature rules, with rigorous complexity theory:
  - Teckentrup, Jantsch, Webster & Gunzburger, A multilevel stochastic collocation method for PDEs with random input data, *SIAM/ASA J Uncertain Q* **3**, 2015
  - Zech, Dung & Schwab, Multilevel approximation of parametric and stochastic PDEs, *Math Mod Meth Appl Sci* **29**, 2019
  - Lang, **RS** & Silvester, A fully adaptive multilevel stochastic collocation strategy for solving elliptic PDEs with random data, *J Comput Phys* **419**, 2020

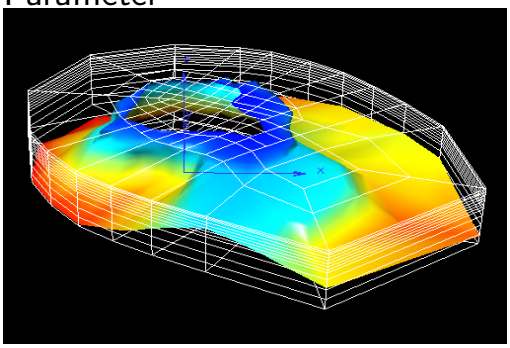
  Under strong regularity conditions allows significantly better complexity.

8. Conditioning on Data – Bayesian Inverse Problems

Inverse Problems

Data

Parameter



$$y = F(x) + \eta$$

forward model (PDE) ← → observation/model errors

$y \in \mathbb{R}^m$	Data y are limited in number, noisy, and indirect.
$x \in H$	Parameter x often a function (discretisation needed).
$F : H \to \mathbb{R}^m$	Continuous, bounded, and sufficiently smooth.

Examples of Inverse Problems

- **Deblurring a noisy image:**
  y: image; F: blurring operator
- **Seismic inversion**
  y: reflected wave image; F: wave equation
- **Computer tomography**
  y: radial x-ray attenuation; F: line integral of absorption
- **Weather forecasting**
  y: satellite data, sparse indirect measurem.; F: atmospheric flow
- **History matching in oil reservoir simulation**
  y: well pressure/flow rates; F: subsurface flow
- **Predator-prey model**
  y: state of $u_2(T)$; F: dynamical system

Classically [Hadamard, 1923]: Inverse map "F^{-1}" ($y \to x$) is typically ill-posed, i.e. lack of (a) **existence**, (b) **uniqueness** or (c) **boundedness**

Linear Inverse Problems & Least Squares

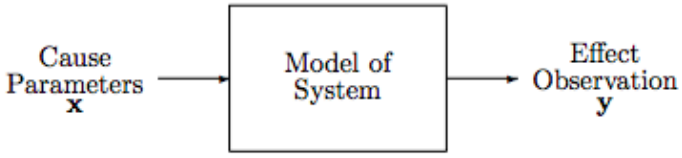
- Consider the linear forward operator $F(x) = Ax$ from $\mathbb{R}^s$ to $\mathbb{R}^m$ with $A \in \mathbb{R}^{m\times s}$ and assume that $\eta \sim \mathrm{N}(0, s_\eta^2 I)$.
- *Least squares minimisation* seeks "best" solution $\widehat{x}$ by minimising residual norm
$$\mathrm{argmin}_{x\in\mathbb{R}^s} \|y - Ax\|^2$$
- In the case of full rank (for $m > s$), this actually leads to a unique map
$$\widehat{x} = (A^T A)^{-1} A^T y$$
which also minimises the mean-square error $\mathbb{E}\left[\|\widehat{x} - x\|^2\right]$ and the covariance matrix $\mathbb{E}\left[(\widehat{x} - x)(\widehat{x} - x)^T\right]$ and satisfies
$$\mathbb{E}\left[\widehat{x}\right] = x \quad \text{and} \quad \mathbb{E}\left[(\widehat{x} - x)(\widehat{x} - x)^T\right] = s_\eta^2 (A^T A)^{-1}$$

Using *singular value decomposition* of $A^T A = U\Sigma V^T$ with $U = [u_1, ..., u_m]$, $V = [v_1, ..., v_n]$ unitary and $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_m^2)$ we have in fact
$$\widehat{x} = \sum_{k=1}^{m} \frac{u_k^T y}{\sigma_k} v_k = x + \sum_{k=1}^{m} \frac{u_k^T \eta}{\sigma_k} v_k$$

Error Amplification & Tikhonov Regularisation

- In typical physical systems $\sigma_k \ll 1$, for $k \gg 1$, and so the "high frequency" components $u_k^T \eta$ in the error get amplified with $1/\sigma_k$.
- In addition, if $m < s$ or if A is not full rank, then $A^T A$ is not invertible and so $\widehat{x}$ is not unique (what is the physically best choice?)
- An approach that guarantees uniqueness of the least squares minimiser and prevents amplification of high frequency errors is *regularisation*, i.e solving instead
$$\underset{x\in\mathbb{R}^m}{\mathrm{argmin}}\ s_\eta^{-2}\|y - Ax\|^2 + \alpha\|x - x_0\|^2$$
α is called the *regularisation parameter* and controls how much we trust the data or how much we trust the a priori knowledge about x.
- In general, with $\eta \sim \mathrm{N}(0, Q)$ and $F : H \to \mathbb{R}^m$ we solve
$$\underset{x\in H}{\mathrm{argmin}}\ \|y - F(x)\|_{Q^{-1}}^2 + \|x - x_0\|_{R^{-1}}^2$$

Bayesian Interpretation (Conditional Parameter Distribution)



(Physical) model gives $\pi(y|x)$, the *conditional probability of observing y given x*, but to predict, control, optimise or to do UQ we are really interested in $\pi(x|y)$, the *conditional probability of possible causes x given the observed data y*.

Bayes' rule states:

$$\pi(x|y) = \frac{\pi(y|x)\pi(x)}{\pi(y)}$$

- $\pi(x) =$ **prior density**: what we know/believe about x prior to observing y
- $\pi(x|y) =$ **posterior density**: what we know about x after observing y
- $\pi(y|x) =$ **likelihood**: (physical) model or how likely it is to observe y given x
- $\pi(y) =$ **evidence**: marginal of $\pi(x, y)$ over all possible x

(scaling factor that can be determined by normalisation)

Link between Bayes' Rule and Tikhonov Regularisation

- Bayesian interpretation of the least squares solution $\widehat{x}$, is to find the *maximum likelihood estimate*.
- Bayesian equivalent of the regularisation term is the prior distribution $\pi(x)$: for Tikhonov $x \sim \mathrm{N}(x_0, R)$ (could be different distribution).
- Bayes interpretation of the regularised least squares solution is the *maximum a posteriori (MAP) estimate*. In the simple linear case it is

$$\widehat{x}^{\mathrm{MAP}} = (A^T A + \alpha s_\eta^2 I)^{-1}(A^T y + \alpha s_\eta^2 x_0)$$

However, in the Bayesian setting, the **full posterior** contains **more information** than the MAP estimator alone, e.g. the posterior covariance matrix $P^{-1} = (A^T Q^{-1} A + R^{-1})^{-1}$ reveals more or less certain components of x..

- In the linear Gaussian case, posterior mean (MAP) and covariance matrix describe the entire distribution. What about general case? Can we do better?

Metropolis-Hastings Markov Chain Monte Carlo

YES. We can sample from the posterior distribution and/or compute posterior expectations $\mathbb{E}_{\pi(x|y)}[G(x)]$ using

- importance sampling
- rejection sampling
- variational inference methods
- filtering
- Markov chain Monte Carlo methods:

ALGORITHM 1 (Metropolis-Hastings Markov Chain Monte Carlo)

- Choose initial state $x^0 \in X$.
- At state n generate proposal $x' \in X$ from distribution $q(x' \,|\, x^n)$
  (e.g. via a random walk $x' \sim \mathrm{N}(x^n, \varepsilon^2 \mathrm{I})$)
- Accept x' as a sample and set $x^{n+1} = x'$ with probability

$$\boldsymbol{\alpha}(x'|x^n) \;=\; \min\left(1, \frac{\pi(x' \,|\, y)\, q(x^n \,|\, x')}{\pi(x^n \,|\, y)\, q(x' \,|\, x^n)}\right)$$

  Otherwise set $x^{n+1} = x^n$.

Links to what I have told you so far and to Machine Learning

- What does this all have to do with UQ?
- In context of what I said so far, we want to "condition" our uncertain models on information about input data (prior) and output data (likelihood).
- Again we have to distinguish whether we are interested
  - only in statistics about some QoI (**quadrature w.r.t. the posterior**) or
  - in the whole posterior distribution of the inputs and/or of the state
- Allows to **learn** something about a model parameter or physically relevant, derived quantity from **noisy, indirect measurements**.
- Outcome **crucially** depends on **choice of prior** (curse and blessing):
  - If nothing is known use non-informative prior!
  - If we have solid/complicated prior knowledge can use it!
- Updating prior belief given measured data. In that sense optimal and theoretically rigorous (**Bayes optimality**).
- **Most importantly**: can rigorously **quantify uncertainties !**

9. Model Problems & Markov Chain Monte Carlo

Example 1: Predator-Prey Problem

In the predator-prey model, a typical variation on the problem studied so far that leads to a Bayesian UQ problem is:

1. **Prior:** $\mathbf{u}_0 \sim \mathrm{U}\Big(\overline{\mathbf{u}}_0 + [-\delta, \delta]^2\Big)$
2. **Data:** $y = u_2^{\mathsf{obs}}$ at time T with measurement error $\eta \sim \mathrm{N}(0, s_\eta^2)$
3. **Likelihood**: (with bias due to the numerical approximation of F):

$$\pi_h(y|\mathbf{u}_0) \;\approx\; \exp\left(\frac{-|y - u_{M,2}(\mathbf{u}_0)|^2}{s_\eta^2}\right)$$

4. **Posterior:** $\pi_h(\mathbf{u}_0|y) \propto \pi_h(y|\mathbf{u}_0)\, \underbrace{\pi_{\mathsf{pr}}(\mathbf{u}_0)}_{=\mathsf{const}}$
5. **Statistic:** $\mathbb{E}_{\pi_h(\mathbf{u}_0|y)}\left[u_{M,1}(\mathbf{u}_0)\right]$ ($\approx$ expected value of $u_1(T)$ under the posterior)

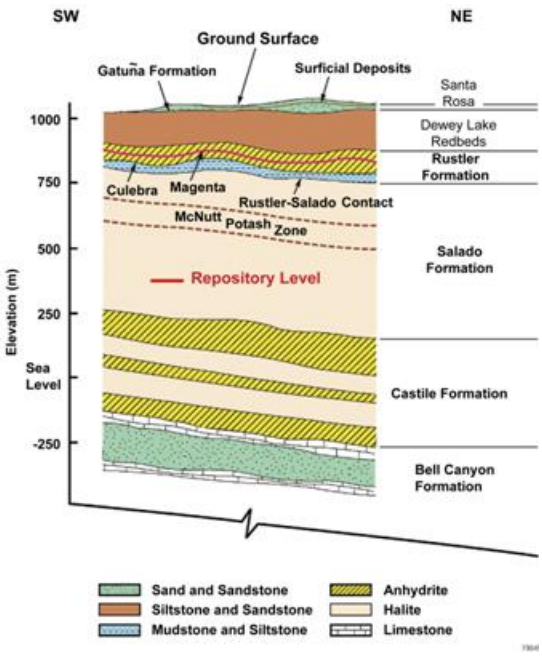
Depending on size of s_η^2 uncertainty in expected value of $u_1(T)$ is vastly reduced.
(can be computed, e.g., with Metropolis-Hastings MCMC).

Example 2: Deep Geological Disposal of Radioactive Waste

- Area where UQ has played central role in past 25 years.
- Deep geological disposal favoured by nearly all countries with a radioactive waste disposal programme.
- Storage in containers in tunnels, hundreds of meters deep in stable geological formations. No human intervention required after sealing repository.
- Several barriers: chemical, physical, geological.
- Ccontainment must be assured for at least 10,000 years.
- Main escape route for radionuclides: groundwater pathway.
- **Assessing safety** of potential sites of utmost importance
  long timescales $\rightarrow$ **modelling essential!**
- **Key aspect:** How to quantify uncertainties in the models?

WIPP – Waste Isolation Pilot Plant

- US DOE repository for radioactive waste situated near Carlsbad, NM.
  (Fully operational since 1999.)
- Extensive site characterisation and performance assessment since 1976, also in course of compliance certification and recertification by US EPA (every 5 years).
- Lots of publicly available data at
  http://www.wipp.energy.gov
- Repository located at 655m depth in bedded evaporites (mainly halite, a salt).
- Most transmissive rock layer in the region is the Culebra Dolomite: principal pathway for transport of radionuclides in the event of an accidental breach.



Groundwater Flow Model

Stationary Darcy flow	$\mathbf{q} = -K\nabla p$	$\mathbf{q}$: Darcy flux
		K : hydraulic conductivity
		p : hydraulic head
mass conservation	$\nabla\cdot\,\mathbf{u} = 0$	$\mathbf{u}$: pore velocity
	$\mathbf{q} = \phi\mathbf{u}$	ϕ : porosity
transmissivity	$k = Kb$	b : aquifer thickness
particle transport	$\dot{\mathbf{x}}(t) = -\frac{k(\mathbf{x})}{b\phi}\nabla p(\mathbf{x})$	$\mathbf{x}$: particle position
	$\mathbf{x}(0) = \mathbf{x}_0$	$\mathbf{x}_0$: release location

Quantity of interest: $\log_{10}$ of particle travel time to reach boundary

UQ Problem – PDE with Random Coefficient

Primal form of Darcy equations is our "fruit fly" with $a = k$ and $u = p$:

$$-\nabla\cdot[a(\mathbf{x})\nabla u(\mathbf{x})] = 0, \quad \mathbf{x} \in D, \qquad u = u_0 \text{ along } \partial D.$$

Model transmissivity as a random field (RF) $a = a(\mathbf{x}, \omega)$, $\omega \in \Omega$, with respect to underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Modeling Assumptions (standard in 2D hydrogeology):

- finite mean and covariance

$$\overline{a}(\mathbf{x}) = \mathbb{E}\left[a(\mathbf{x}, \cdot)\right], \qquad \mathbf{x} \in D,$$
$$\mathbf{Cov}_a(\mathbf{x}, \mathbf{y}) = \mathbb{E}\left[\left(a(\mathbf{x}, \cdot) - \overline{a}(\mathbf{x})\right)\left(a(\mathbf{y}, \cdot) - \overline{a}(\mathbf{y})\right)\right], \qquad \mathbf{x}, \mathbf{y} \in D.$$

- a is lognormal, i.e., $Z(\mathbf{x}, \omega) := \log a(\mathbf{x}, \omega)$ is a Gaussian RF.
- $\mathbf{Cov}_Z$ is stationary and isotropic, i.e., $\mathbf{Cov}_Z(\mathbf{x}, \mathbf{y}) = c(\|\mathbf{x} - \mathbf{y}\|_2)$

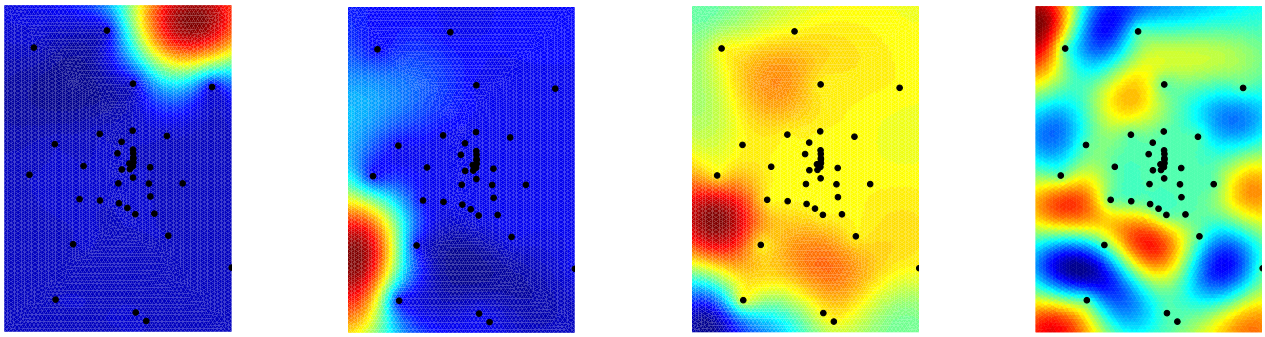
Data for Radioactive Waste Example (WIPP)

Prior Model [Ernst et al, 2014]

$$\log a \approx \sum_{j=1}^{s} \sqrt{\mu_j}\, \phi_j^{\text{cond}}(x)\theta_j(\omega) \text{ with i.i.d. } \theta_j \sim \mathrm{N}(0,1)$$

Karhunen-Loeve modes ($j = 1, 2, 9, 16$) conditioned on 38 transmissivity observations

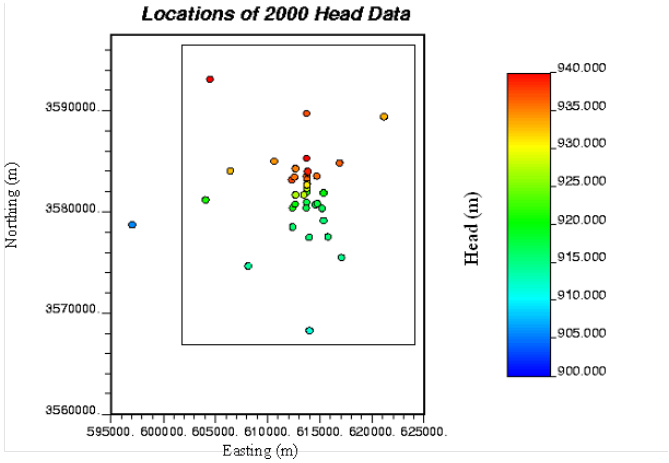
(via kriging (Gaussian process regression): a simple low-rank change to covariance operator)



Prior model: $\pi_{\text{pr},s}(\boldsymbol{\theta})$ is the multivariate standard Gaussian density for $\boldsymbol{\theta} \in \mathbb{R}^s$.

Data for Radioactive Waste Example (WIPP)

Likelihood Model [Ernst et al, 2014]



- Data y are pressure measurements.
- $F_h(\boldsymbol{\theta})$ is the model response.

Likelihood model: assuming Gaussian errors with covariance Σ

$$\pi_{h,s}(y|\boldsymbol{\theta}) \;\asymp\; \exp(-\|y - F_h(\boldsymbol{\theta})\|^2_{\Sigma^{-1}})$$

Posterior through **Bayes' rule**: $\pi_{h,s}(\boldsymbol{\theta}\,|\,y) \;\asymp\; \pi_{h,s}(y|\boldsymbol{\theta})\,\pi_{\text{pr},s}(\boldsymbol{\theta})$

Markov Chain Monte Carlo (Metropolis-Hastings Algorithm)

(for the discretised fruit fly problem)

ALGORITHM 1 (Standard Metropolis Hastings MCMC)

- Choose $\boldsymbol{\theta}^0 \in \mathbb{R}^s$.
- At state $\boldsymbol{\theta}^n$ generate a $\boldsymbol{\theta}' \in \mathbb{R}^s$ from the proposal distribution $q(\boldsymbol{\theta}' \,|\, \boldsymbol{\theta}^n)$
  (e.g. basic or preconditioned Crank-Nicholson random walk [Cotter et al, 2012])
- Accept sample $\boldsymbol{\theta}'$ and set $\boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}'$ with probability

$$\boldsymbol{\alpha}_{h,s}(\boldsymbol{\theta}' \,|\, \boldsymbol{\theta}^n) \;=\; \min\left(1, \frac{\pi_{h,s}(\boldsymbol{\theta}' \,|\, y)\, q(\boldsymbol{\theta}^n \,|\, \boldsymbol{\theta}')}{\pi_{h,s}(\boldsymbol{\theta}^n \,|\, y)\, q(\boldsymbol{\theta}' \,|\, \boldsymbol{\theta}^n)}\right)$$

  Otherwise $\boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}^n$.

Samples $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^N$ used as usual for inference (even though not i.i.d.):

$$\mathbb{E}_{\pi(\cdot|y)}[Q] \;\approx\; \mathbb{E}_{\pi_{h,s}(\cdot|y)}[Q_{h,s}] \;\approx\; \frac{1}{N}\sum_{i=1}^{N} Q_{h,s}^{(n)} =: \widehat{Q}^{\mathsf{MH}}$$

where $Q_{h,s}^{(n)} = G(\boldsymbol{\theta}^n) = \Psi\big(u_h(\boldsymbol{\theta}^n)\big)$ is nth sample of the QoI using Model(h, s).

Markov Chain Monte Carlo Theory (for simplicity only finite dimensional)

Theorem 9.1 (Metropolis et al. 1953, Hastings 1970, . . .)

The Markov chain simulated by the Metropolis-Hastings algorithm is **reversible** *with respect to $\pi(\cdot|y)$. If we also have*

$$\pi(x'|y) > 0 \quad \Rightarrow \quad q(x'|x^n), \quad \textit{for all } n \in \mathbb{N}$$
$$\mathbb{P}\big(\boldsymbol{\alpha}(x'|x^n) = 1\big) < 1, \quad \textit{for all } n \in \mathbb{N},$$

then it defines a ***geometrically ergodic*** *Markov chain with unique equilibrium density $\pi(\cdot|y)$ (for any initial state x^0) and the* ***Central Limit Theorem*** *gives*

$$\sqrt{N}\left(\widehat{Q}^{\mathrm{MH}} - \mathbb{E}_{\pi(\cdot|y)}[G(X)]\right) \xrightarrow[N\to\infty]{d} \mathcal{N}(0, \sigma_G^2) \tag{9.1}$$

with ***asymptotic variance***

$$\sigma_G^2 := \mathbf{Var}(G(X_1)) + 2\sum_{j=1}^{\infty} \mathbf{Cov}\left(G(X_1), G(X_{1+j})\right). \tag{9.2}$$

Crudely speaking geometrically ergodic means that there exists an $r \in (0,1)$ s.t. the TV-distance between the target distribution and the distribution of the nth state converges with $\mathcal{O}(r^{-n})$.

Markov Chain Monte Carlo

Comments in the context of our UQ problem

Pros:

- Produces a Markov chain $\{\boldsymbol{\Theta}^n\}_{n\in\mathbb{N}}$ with $\boldsymbol{\Theta}^n \sim \pi_{h,s}(\cdot|y)$ as $n \to \infty$.
- Can be made dimension independent (e.g. via pCN sampler).
- Therefore often referred to as **"gold standard"** (Stuart et al)

Cons:

- Evaluation of $\boldsymbol{\alpha}_{h,s}(\boldsymbol{\theta}'|\boldsymbol{\theta}^n)$ **very expensive** for small h (cost/sample $\geq \mathcal{O}(h^{-d})$)
- Acceptance rate $\boldsymbol{\alpha}_{h,s}$ can be very low for large s ($< 10\%$)
- $\mathrm{Cost}(\widehat{Q}^{\mathrm{MH}}) = \mathcal{O}(\varepsilon^{-2-\frac{\gamma}{\alpha}})$ as above, **but** the constant is multiplied by the **integrated autocorrelation time** (= relative asymptotic variance)

$$\tau_G := \frac{\sigma_G^2}{\mathbf{Var}(G(X_1))} = 1 + 2\sum_{j=1}^{\infty} \mathbf{Corr}\left(G(X_1), G(X_{1+j})\right) \qquad (9.3)$$

  (which depends on stepsize in q and on $\alpha_{h,s}$)

- In addition, require **burn-in** to reduce the initiation bias.

Prohibitively expensive – significantly worse than standard MC w. iid. samples!

10. Multilevel Markov Chain Monte Carlo

Multilevel Markov Chain Monte Carlo – Idea

[Dodwell, Ketelsen, RS, Teckentrup, JUQ 2015] & [Dodwell et al, SIAM Rev. 2019]

- What were the **key ingredients** of "standard" multilevel Monte Carlo?
  - **Telescoping sum:** $\mathbb{E}[Q_L] = \mathbb{E}[Q_0] + \sum_{\ell=1}^{L} \mathbb{E}[Q_\ell - Q_{\ell-1}]$
  - Models on coarser levels **much cheaper** to solve ($h_0^{-d} \ll h_L^{-d}$).
  - $\mathbb{V}[Q_\ell - Q_{\ell-1}] \overset{\ell\to\infty}{\longrightarrow} \to 0$ as $\implies$ much **fewer samples** on finer levels.
- **But Important!** Now target distribution $\pi_\ell := \pi_{h_\ell, s_\ell}(\cdot \,|\, y)$ **depends on** ℓ:

$$\mathbb{E}_{\pi_L}[Q_L] = \underbrace{\mathbb{E}_{\pi_0}[Q_0]}_{\text{standard MCMC}} + \sum_\ell \underbrace{\mathbb{E}_{\pi_\ell}[Q_\ell] - \mathbb{E}_{\pi_{\ell-1}}[Q_{\ell-1}]}_{\text{multilevel MCMC (NEW)}}$$

$$\widehat{Q}_{h,s}^{\text{MLMH}} := \frac{1}{N_0}\sum_{n=1}^{N_0} Q_0(\boldsymbol{\Theta}_{0,0}^n) + \sum_{\ell=1}^{L} \frac{1}{N_\ell}\sum_{n=1}^{N_\ell} \left(Q_\ell(\boldsymbol{\Theta}_{\ell,\ell}^n) - Q_{\ell-1}(\boldsymbol{\Theta}_{\ell,\ell-1}^n)\right)$$

with **correlated** Markov chains $\{\boldsymbol{\Theta}_{\ell,\ell-1}^n\}$ and $\{\boldsymbol{\Theta}_{\ell,\ell}^n\}$ (see below).

- For simplicity we describe only the case $s_\ell = s_{\ell-1} = \ldots = s_0$.
  (In practice, useful to reduce also number $s_{\ell-1}$ of random parameters on coarser levels.

Multilevel Markov Chain Monte Carlo – Algorithm

Choose **subsampling rates** $t_0, \ldots, t_L \in \mathbb{N}$ (see below) and set $T_{\ell,k} := \prod_{j=k}^{\ell-1} t_j$.

ALGORITHM 2 (Multilevel Metropolis Hastings MCMC for $Q_\ell - Q_{\ell-1}$)

Given realisations $\boldsymbol{\theta}_{\ell,0}^n, \ldots, \boldsymbol{\theta}_{\ell,\ell}^n$ at state n of Markov chains on levels $k = 0, \ldots, \ell$.

1. $k = 0$: Set $\mathbf{x}_0^0 := \boldsymbol{\theta}_{\ell,0}^n$. Use **Algorithm 1** (standard Metropolis-Hastings) to generate samples $\mathbf{x}_0^i \sim \pi_0$, $i = 1, \ldots, T_{\ell,0}$. Set $\boldsymbol{\theta}_{\ell,0}^{n+1} := \mathbf{x}_0^{T_{\ell,0}}$.
2. $k > 0$: Set $\mathbf{x}_k^0 := \boldsymbol{\theta}_{\ell,k}^n$. Generate samples $\mathbf{x}_k^i \sim \pi_k$, $i = 1, \ldots, T_{\ell,k}$ as follows:
   (a) Propose $\mathbf{x}_k' = \mathbf{x}_{k-1}^{(i+1)t_{k-1}}$ — Subsample to reduce correlation!
   (b) Accept $\mathbf{x}_k'$ and set $\mathbf{x}_k^{i+1} = \mathbf{x}_k'$ with probability
   $$\boldsymbol{\alpha}_k^{\text{ML}}(\mathbf{x}_k' | \mathbf{x}_k^i) = \min\left(1, \frac{\pi_k(\mathbf{x}_k')\pi_{k-1}(\mathbf{x}_k^n)}{\pi_k(\mathbf{x}_k^n)\pi_{k-1}(\mathbf{x}_k')}\right)$$ JS Liu, 2001
   Otherwise set $\mathbf{x}_k^{i+1} = \mathbf{x}_k^i$.
   (c) Set $\boldsymbol{\theta}_{\ell,k}^{n+1} := \mathbf{x}_k^{T_{\ell,k}}$ with $T_{\ell,k} := \prod_{j=k}^{\ell-1} t_j$.
3. Set $Y_\ell^n := Q_\ell(\boldsymbol{\theta}_{\ell,\ell}^n) - Q_{\ell-1}(\boldsymbol{\theta}_{\ell,\ell-1}^n)$.

MLMCMC – Comments

- Each $\{\Theta_{\ell,k}^n\}_{n\geq 1}$, $k = 0, \dots, \ell$, is a **Markov chain** with $\Theta_{\ell,k}^n \sim \pi_k$ as $n \to \infty$ and $t_\ell \to \infty$.
- Theoretically need $t_\ell \to \infty$ to guarantee **consistency** of multilevel algorithm (no bias between levels)
- In practice, it suffices to choose $t_\ell \approx C\tau_{G,\ell}$ with $C = 1$ or 2.
- States may differ between level ℓ and $\ell - 1$:

State $n+1$	Level $\ell - 1$	Level ℓ
accept on level ℓ	$\boldsymbol{\theta}_{\ell,\ell-1}^{n+1}$	$\boldsymbol{\theta}_{\ell,\ell-1}^{n+1}$
reject on level ℓ	$\boldsymbol{\theta}_{\ell,\ell-1}^{n+1}$	$\boldsymbol{\theta}_{\ell,\ell}^{n}$

but this does not happen often for larger ℓ since **acceptance probability** $\boldsymbol{\alpha}_\ell^{\mathsf{ML}} \overset{\ell\to\infty}{\longrightarrow} 1$.

Lemma 10.1 (Dodwell, Ketelsen, RS, Teckentrup, '15)

$$\mathbb{E}_{\pi_\ell,\pi_\ell}\Big[1 - \boldsymbol{\alpha}_\ell^{ML}(\cdot|\cdot)\Big] \;=\; \mathcal{O}\Big(\mathbb{E}_{\pi_{pr}}\left[|F(\boldsymbol{\theta}) - F_\ell(\boldsymbol{\theta})|\right]\Big) \;=\; \mathcal{O}(h_\ell^\alpha)$$

- Note that this also implies $\tau_{G,\ell} \overset{\ell\to\infty}{\longrightarrow} 1$.

Complexity Theorem for Multilevel MCMC (Dodwell et al. '15)

Suppose there are constants $\alpha, \beta, \gamma, \eta > 0$ such that, for all $\ell = 0, \dots, L$,

M1 $\big|\mathbb{E}_{\pi_\ell}[Q_\ell] - \mathbb{E}_{\pi(\cdot|y)}[Q]\big| \;=\; \mathcal{O}(h_\ell^\alpha)$ (discretisation and truncation error)

M2' $\mathbf{Var}_{\mathsf{alg}}[\widehat{Y}_\ell] + \Big(\mathbb{E}_{\mathsf{alg}}[\widehat{Y}_\ell] - \mathbb{E}_{\pi_\ell,\pi_{\ell-1}}[\widehat{Y}_\ell]\Big)^2 = \mathbf{Var}_{\pi_\ell,\pi_{\ell-1}}[Y_\ell]\,\mathcal{O}(N_\ell^{-1})$ (MCMC-error)

M2 $\mathbf{Var}_{\pi_\ell,\pi_{\ell-1}}[Y_\ell] \;=\; \mathcal{O}(h_\ell^\beta)$ (multilevel variance decay)

M3 $\mathsf{Cost}(\widehat{Y}_\ell^{\mathsf{MC}}) \;=\; \mathcal{O}(N_\ell\, h_\ell^{-\gamma})$. (cost per sample)

Then there exist L, $\{N_\ell\}_{\ell=0}^L$ s.t. $\mathsf{MSE} < \varepsilon^2$ and

$$\mathcal{C}_\varepsilon(\widehat{Q}_{h,s}^{\mathsf{MLMH}}) \;=\; \mathcal{O}\left(\varepsilon^{-2-\max\left(0,\frac{\gamma-\beta}{\alpha}\right)}\right) \quad (+\text{ log-factor when } \beta = \gamma)$$

(This is totally **abstract** & applies not only to our subsurface model problem!)

- Proof of Assumptions **M1** and **M3** similar to i.i.d. case.
- **M2' not** specific to multilevel MCMC; first steps in [Hairer, Stuart, Vollmer, '11].

Proof of Assumption **M2** for lognormal diffusion & linear FEs (Dodwell et al '15)

$$\mathbf{Var}_{\pi_\ell,\pi_{\ell-1}}\left[Q_\ell(\Theta_{\ell,\ell}^n) - Q_{\ell-1}(\Theta_{\ell,\ell-1}^n)\right] \;=\; \mathcal{O}(h_\ell^\alpha) \quad (\text{unfortunately } \beta = \alpha \text{ not } 2\alpha)$$

More Comments – Related Literature

- Typically also increase number of parameters s_ℓ from level to level and use standard proposal kernel for new parameters (see paper).
- Subsampling essential (exact only in limit of infinite subsampling), but small bias for sampling rates with $C = 1$ or 2.
- **New ("multiplicative") version**: Current work with Colin Fox (Otago, NZ).
- Algorithm 2 is a special case of a **surrogate transition method** [Liu, Monte Carlo Strategies in Scientific Computing, 2001, §9.4.3]
- and of **delayed acceptance Metropolis-Hastings** [Christen, Fox, '05]

But crucially exploiting variance reduction & proved rates in MLMCMC are **new**!

(Corollaries on adaptive error estimates using the Markov chains, current work with Colin Fox)

- Other references on related **mutlilevel Monte Carlo methods** recently developed for Bayesian inverse problems:
  - Hoang, Schwab & Stuart, Complexity analysis of accelerated MCMC methods for Bayesian inversion, *Inverse Prob* **29**, 2013
  - Beskos, Jasra, Law & Zhou, Multilevel sequential Monte Carlo samplers, *Stoch Proc Appl* **127**, 2017

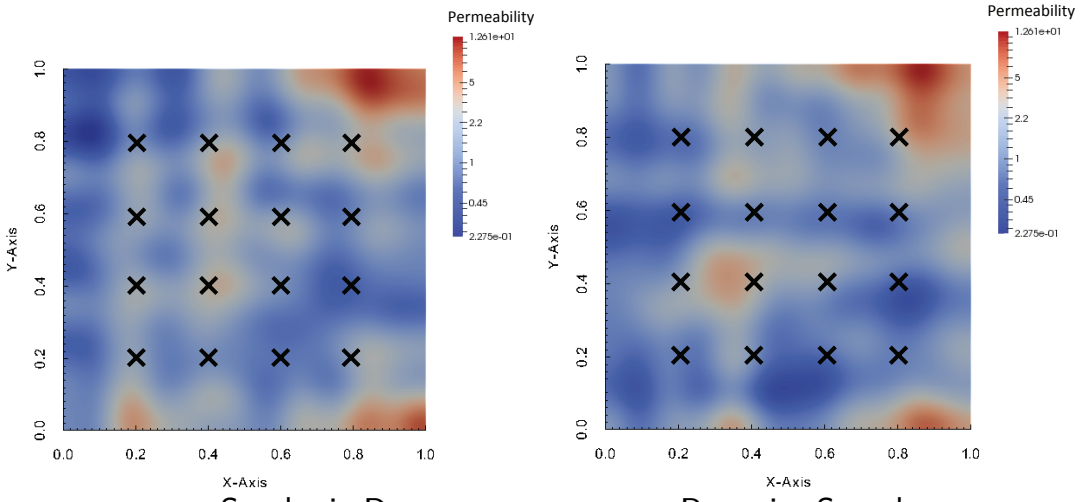
Numerical Example

Fruit fly (2D lognormal diffusion) on $D = (0,1)^2$ with linear FEs

- **Prior:** Separable exponential covariance with $\sigma^2 = 1$, $\lambda = 0.5$.

  i.e. $\mathbb{E}[Z(x)Z(x')] = \sigma^2 e^{-\frac{|x-x'|}{\lambda} - \frac{|y-y'|}{\lambda}}$

- **"Data"** y: Pressure at 16 points $x_j^* \in D$ and covariance $\Sigma = 10^{-4} I$.

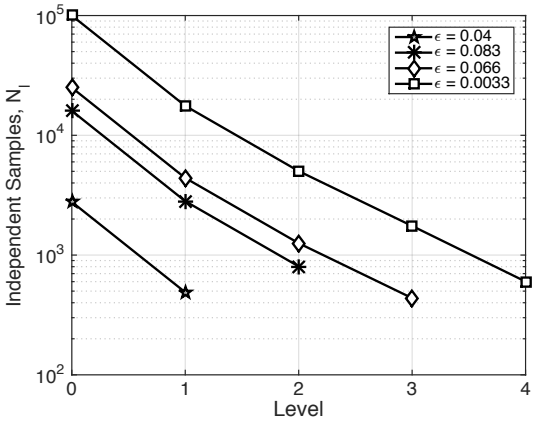


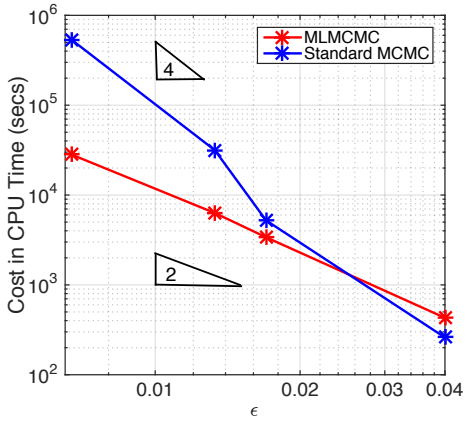
Synthetic Data | Posterior Sample

Numerical Example

Quantity of interest: $Q = \int_0^1 k \nabla p \, dx_2$; coarsest mesh size: $h_0 = \frac{1}{9}$

- 5-level method with #KL modes increasing from $s_0 = 50$ to $s_4 = 150$





- #independent samples = $\dfrac{N_\ell}{\tau_\ell}$ (τ_ℓ . . . integrated autocorrelation time)

Level ℓ	0	1	2	3	4
i.a.c. time τ_ℓ	136.23	3.66	2.93	1.46	1.23

Choice of Proposal Distribution

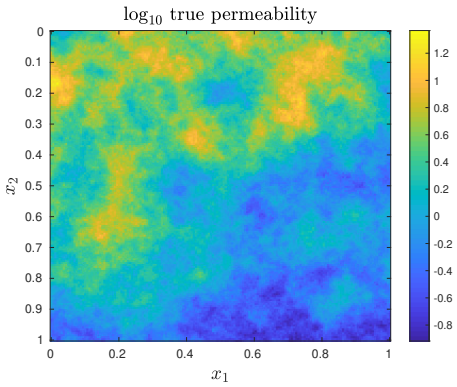
Multilevel DILI (recent preprint with T Cui & G Detommaso)

- **So far:** pCN random walk proposal (uses no gradient/Hessian info) [Cotter, Dashti, Stuart, '12]
- **Problem:** Dimension independent but very high IACT for $s \to \infty$!
  $\tau_0 \approx 136$ above, i.e. need 136 samples to obtain **one independent** sample!!
- **However**, can use any other proposal (e.g. MALA, stochastic Newton)
- **DILI MCMC** [Cui, Law, Marzouk, '16]:
  (DILI = dimension-independent likelihood-informed)
  samples from preconditioned Langevin equation using **low-rank** approximation of data-misfit Hessian at some points (incl. MAP point)
- **New** multilevel construction of DILI (with T Cui and G Detommaso) ...

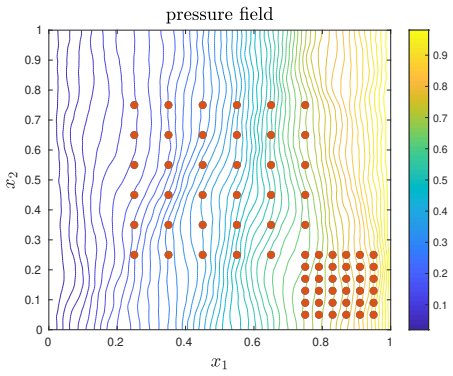
Cui, Detommaso, **RS**, Multilevel dimension-independent likelihood-informed MCMC for large-scale inverse problems, submitted, 2019 [arXiv:1910.12431]

Testing on a Much Harder Example

True parameter



Pressure



Model:

$$-\nabla \cdot \left(e^{z(x)} \nabla u(s) \right) = 0, \quad x \in [0,1]^2$$

Top/bottom: zero Neumann b.c.; left/right: Dirichlet b.c. zero/one, respectively.

Gaussian process prior for $z = \log a$ with covariance fct. $k(x, x') = \exp(-5|x - x'|)$

71 sensors; signal to noise ratio 50.

QoI: $Q^{(\text{flux})}$ = average flux over the left boundary

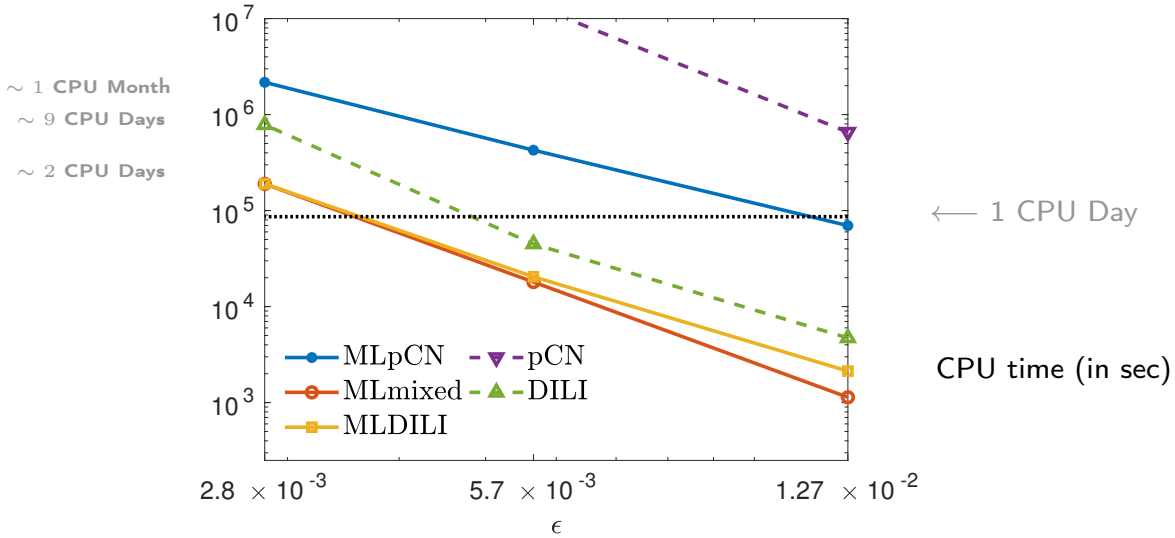
Numerical Comparison: IACTs & CPU Times

Refined parameters

Level ℓ	0	1	2	3
iact(pCN)	4300	45	48	24
iact(DILI)	34	11	3.6	2.0

$Q_\ell(\boldsymbol{\theta}^n_{\ell,\ell}) - Q_{\ell-1}(\boldsymbol{\theta}^n_{\ell,\ell-1})$

Level ℓ	0	1	2	3
iact(pCN)	4100	4.9	2.8	1.9
iact(DILI)	9.0	4.6	2.4	1.8



Key References for Multilevel Bayesian Inference

1. JS Liu, *Monte Carlo Strategies in Scientific Computing*, Springer-Verlag, NY, 2001
2. JA Christen & C Fox, MCMC using an approximation, *J Comput Graph Statist* **14**, 2005
3. VH Hoang, C Schwab & AM Stuart, Complexity analysis of accelerated MCMC methods for Bayesian inversion, *Inverse Prob.* **29**, 2013
4. TJ Dodwell, C Ketelsen, **RS** & AL Teckentrup, A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow, *SIAM/ASA J Uncertain Quant* **3**, 2015
5. **RS**, AM Stuart, and AL Teckentrup, Quasi-Monte Carlo and multilevel Monte Carlo methods for computing posterior expectations in elliptic inverse problems, SIAM/ASA J Uncertain Quantif **5**, 2017
6. A Beskos, A Jasra, KJH Law & Y Zhou, Multilevel sequential Monte Carlo samplers, *Stoch Proc Appl* **127**, 2017
7. TJ Dodwell, C Ketelsen, **RS** & AL Teckentrup, Multilevel Markov chain Monte Carlo, *SIAM Review (SIGEST)* **61**, 2019
8. MB Lykkegaard, G Mingas, **RS**, C Fox & TJ Dodwell, Multilevel Delayed Acceptance MCMC with an Adaptive Error Model in PyMC3, to appear in *NeurIPS 2020*
9. T Cui, G Detommaso & **RS**, Multilevel dimension-independent likelihood- Informed MCMC for large-scale inverse problems, submitted [arXiv:1910.12431]

11. Conclusions

Conclusions

- I hope the course gave you a basic understanding of the questions & challenges in modern uncertainty quantification.
- The focus of the course was on the design of computationally tractable and efficient multilevel Monte Carlo methods for high-dimensional and large-scale UQ problems in science and engineering.
- Of course it was only possible to give you a snapshot of the methods and we went over some parts too quickly.
- Finally, I apologise that the course was of course also strongly biased in the direction of my research and my expertise and was probably not doing some other methods enough justice.
- But I hope I managed to interest you in the subject and persuade you of the huge potential of multilevel sampling methods.

Thanks!