Die Ressourcenuniversität. Seit 1765.

Institut für Numerische Mathematik und Optimierung

Covariance Eigenproblems and their Numerical Treatment

Oliver Ernst
(joint work with Ingolf Busch)

AMCS Seminar, KAUST
September 26, 2012

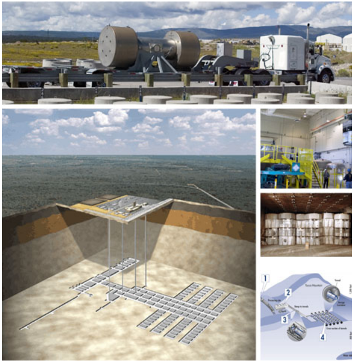
. . . where again?

. . . where again?

PDEs with Random Data

Typical UQ Application: Radioactive Waste Repository Site Assessment

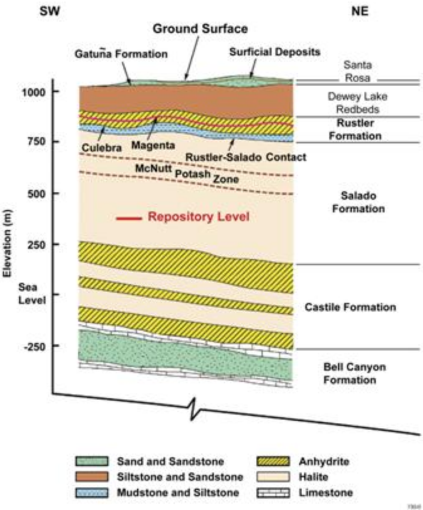
- Waste Isolation Pilot Plant (WIPP) Carlsbad, NM
- Groundwater transport of radionuclides
- Uncertainty in hydraulic conductivity
- Quantity of interest: travel time
- Approach: Model uncertainty (lack of knowledge) stochastically. Propagate random input data to travel time.
- Requires solution of PDE with random data + post-processing.



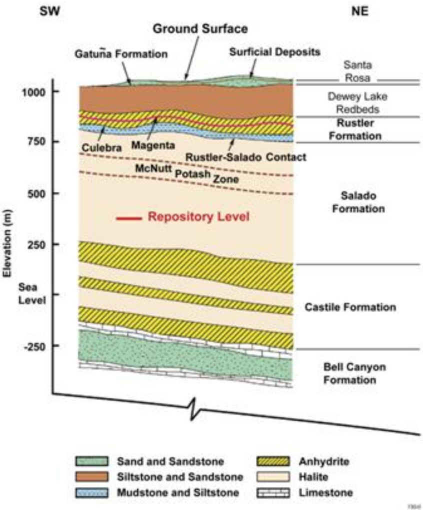
PDEs with Random Data

Typical UQ Application: Radioactive Waste Repository Site Assessment

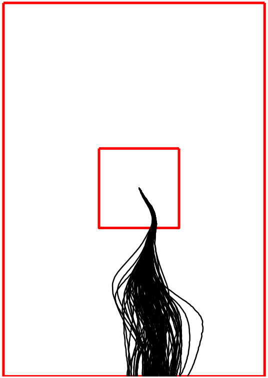
- Waste Isolation Pilot Plant (WIPP) Carlsbad, NM
- Groundwater transport of radionuclides
- Uncertainty in hydraulic conductivity
- Quantity of interest: travel time
- Approach: Model uncertainty (lack of knowledge) stochastically. Propagate random input data to travel time.
- Requires solution of PDE with random data + post-processing.



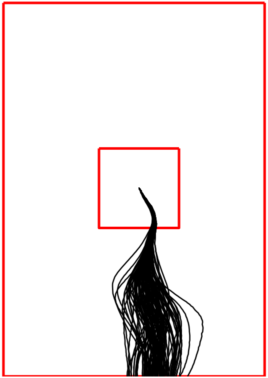
- Waste Isolation Pilot Plant (WIPP) Carlsbad, NM
- Groundwater transport of radionuclides
- Uncertainty in hydraulic conductivity
- Quantity of interest: travel time
- Approach: Model uncertainty (lack of knowledge) stochastically. Propagate random input data to travel time.
- Requires solution of PDE with random data + post-processing.



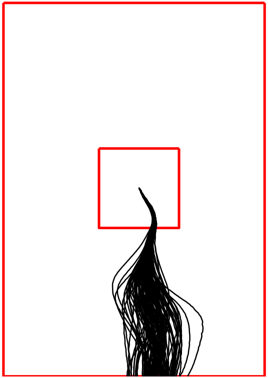
- Waste Isolation Pilot Plant (WIPP) Carlsbad, NM
- Groundwater transport of radionuclides
- Uncertainty in hydraulic conductivity
- Quantity of interest: travel time
- Approach: Model uncertainty (lack of knowledge) stochastically. Propagate random input data to travel time.
- Requires solution of PDE with random data + post-processing.



- Waste Isolation Pilot Plant (WIPP) Carlsbad, NM
- Groundwater transport of radionuclides
- Uncertainty in hydraulic conductivity
- Quantity of interest: travel time
- Approach: Model uncertainty (lack of knowledge) stochastically. Propagate random input data to travel time.
- Requires solution of PDE with random data + post-processing.



- Waste Isolation Pilot Plant (WIPP) Carlsbad, NM
- Groundwater transport of radionuclides
- Uncertainty in hydraulic conductivity
- Quantity of interest: travel time
- Approach: Model uncertainty (lack of knowledge) stochastically. Propagate random input data to travel time.
- Requires solution of PDE with random data + post-processing.



Outline

Next...

1. Expansions of Random Fields
   - Random Fields and Covariance
   - RKHS
   - Karhunen-Loève Expansion

2. Numerical Approximation
   - Galerkin Discretization
   - Adapted Quadrature
   - Lanczos Eigenpair Approximation
   - Hierarchical Matrix Approximation

3. Numerical Examples

Random Fields

Given: compact domain $D \subset \mathbb{R}^d$, probability space $(\Omega, \mathfrak{A}, \mathbf{P})$.

A real-valued random field (RF)

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

is a stochastic process whose index variable is a spatial coordinate.

Thus, for each $\boldsymbol{x} \in D$,

$$a(\boldsymbol{x}, \cdot) \quad \text{is a random variable (RV).}$$

Alternatively: for each $\omega \in \Omega$,

$$a(\cdot, \omega) \quad \text{is a random function defined on } D.$$

Second-order RF: $a(\boldsymbol{x}, \cdot) \in L^2_{\mathbf{P}}(\Omega) = L^2(\Omega, \mathfrak{A}, \mathbf{P})$ for all $\boldsymbol{x} \in D$.

Mean of RF at $\boldsymbol{x} \in D$:

$$\overline{a}(\boldsymbol{x}) := \mathbf{E}\left[a(\boldsymbol{x},\cdot)\right].$$

Covariance of RF at $\boldsymbol{x}, \boldsymbol{y} \in D$:

$$\begin{aligned} c(\boldsymbol{x},\boldsymbol{y}) :&= \operatorname{Cov}(a(\boldsymbol{x},\cdot), a(\boldsymbol{y},\cdot)) \\ &= \mathbf{E}\left[(a(\boldsymbol{x},\cdot) - \overline{a}(\boldsymbol{x}))\,(a(\boldsymbol{y},\cdot) - \overline{a}(\boldsymbol{y}))\right] \end{aligned}$$

For $\tilde{a} := a - \overline{a}$, we have $\mathbf{E}\left[\tilde{a}\right] = 0$ (centered RF).

Moreover, for any selection $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in D$,

$$0 \leq \operatorname{Var}\left(\sum_{i=1}^{n} \alpha_i \, a(\boldsymbol{x}_i, \cdot)\right) = \sum_{i,j=1}^{n} \alpha_i \alpha_j \, c(\boldsymbol{x}_i, \boldsymbol{x}_j),$$

i.e., covariance functions are positive definite. This is also sufficient for $c(\boldsymbol{x}, \boldsymbol{y})$ to be the covariance function of a second-order RF.

Note: if a covariance function $c : D \times D \to \mathbb{R}$ is continuous along the diagonal set $\{(\boldsymbol{x}, \boldsymbol{x}) : \boldsymbol{x} \in D\}$, then it is continuous on all of $D \times D$.

- Translation invariance:

$$c(\boldsymbol{x}, \boldsymbol{y}) = c(\boldsymbol{x} - \boldsymbol{y})$$

  (RF stationary, homogeneous).

- Rotation invariant:

$$c(\boldsymbol{x}, \boldsymbol{y}) = c(\|\boldsymbol{x} - \boldsymbol{y}\|)$$

  (RF isotropic).

- RF Gaussian: each finite collection $\{a(\boldsymbol{x}_i, \cdot)\}_{i=1}^n$ has multivariate Gaussian distribution.
- For now: assume RF Gaussian, centered, with strictly positive definite, continuous covariance function.

Expansion of Random Fields

Goal: Representation of second-order centered Gaussian RF as

$$a(\boldsymbol{x},\omega) = \sum_{j=1}^{\infty} \xi_j(\omega)\, a_j(\boldsymbol{x}), \qquad \xi_j \in L^2(\Omega, \mathfrak{A}, \mathbf{P}),$$

$a_j : D \to \mathbb{R}$ suitable functions.

Convenient Setting: Introduce separable Hilbert space structure.

Set

$$\mathscr{S} := \left\{ f : D \to \mathbb{R} : f(\cdot) = \sum_{j=1}^{n} \alpha_j\, c(\boldsymbol{x}_j, \cdot),\ \alpha_j \in \mathbb{R}, \boldsymbol{x}_i \in D, n \in \mathbb{N} \right\}$$

with inner product (note $c(\cdot,\cdot)$ strictly pos. def.)

$$(f,g) = \left(\sum_{i=1}^{n} \alpha_i\, c(\boldsymbol{x}_i, \cdot), \sum_{j=1}^{m} \beta_j\, c(\boldsymbol{x}_j, \cdot) \right) := \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j\, c(\boldsymbol{x}_i, \boldsymbol{x}_j).$$

This inner product on $\mathscr{S}$ has reproducing kernel property w.r.t. c:

$$(f, c(\boldsymbol{y}, \cdot)) = \left(\sum_{i=1}^{n} \alpha_i c(\boldsymbol{x}_i, \cdot), c(\boldsymbol{y}, \cdot) \right) = \sum_{i=1}^{n} \alpha_i \, c(\boldsymbol{x}_i, \boldsymbol{y}) = f(\boldsymbol{y}). \qquad (*)$$

For sequence $\{f_n\}_{n \in \mathbb{N}}$ in $\mathscr{S}$, if $\|\cdot\|$ denotes associated norm,

$$\begin{aligned} |f_n(\boldsymbol{x}) - f_m(\boldsymbol{x})| &= |(f_n - f_m, c(\boldsymbol{x}, \cdot))| \\ &\leq \|f_n - f_m\| \, \|c(\boldsymbol{x}, \cdot)\| = \|f_n - f_m\| \, c(\boldsymbol{x}, \boldsymbol{x}), \end{aligned}$$

i.e., $\{f_n\}$ Cauchy in $\|\cdot\| \Rightarrow \{f_n\}$ Cauchy pointwise.

Define reproducing kernel Hilbert space (RKHS) $\mathscr{H}_c$ of c (or a) as closure of $\mathscr{S}$ w.r.t. $\|\cdot\|$. Reproducing property $(*)$ for all $f \in \mathscr{H}_c$ follows from separability of compact set D.

RKHS in General

Hilbert space $\mathscr{H}$ of functions $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^d$, for which all evaluation functionals

$$\delta_{\boldsymbol{x}} : \mathscr{H} \to \mathbb{R}, \quad \langle \delta_{\boldsymbol{x}}, f \rangle = f(\boldsymbol{x}), \quad \forall\, \boldsymbol{x} \in D,\, f \in \mathscr{H}.$$

are continuous.

Reproducing kernel $k : D \times D \to \mathbb{R}$ such that $k(\boldsymbol{x}, \cdot) \in \mathscr{H}$ and

$$(f, k(\boldsymbol{x}, \cdot)) = f(\boldsymbol{x}) \qquad \forall f \in \mathscr{H}, \forall \boldsymbol{x} \in D$$

i.e., $k(\boldsymbol{x}, \cdot) = \delta_{\boldsymbol{x}}$.

- Long history dating back to [Mercer, 1909], [Aronsajn, 1944].
- Popularized as setting for optimal prediction/estimation of time series by E. Parzen in the 1960s.
- Recent monograph [Berlinet & Thomas-Agnan, 2007].
- Generalizations to Hilbert spaces of distributions [Meidan, 1979], [Bogachev, 1998]

For $\mathscr{V} := \operatorname{span}\{a(\boldsymbol{x}, \cdot) : \boldsymbol{x} \in D\} \subset L^2_{\mathbf{P}}(\Omega)$, define linear mapping

$$\Xi : \mathscr{S} \to \mathscr{V}$$
$$f = \sum_{j=1}^{n} \alpha_j \, c(\boldsymbol{x}_j, \cdot) \mapsto \sum_{j=1}^{n} \alpha_j \, a(\boldsymbol{x}_j, \cdot).$$

Clearly: $\Xi(f)$ Gaussian $\forall f \in \mathscr{S}$ and

$$(f, g) = (\Xi(f), \Xi(g))_{L^2_{\mathbf{P}}(\Omega)} \quad \forall f, g \in \mathscr{S}.$$

Extend Ξ to all of $\mathscr{H}_c$:

- range equal to all of $\mathscr{V}$
- limits again Gaussian

Canonical isomorphism between the RKHS and the space of RV associated with RF a.

$\mathscr{H}_c$ separable, therefore $\mathscr{V}$ separable.

Orthonormal (ON) basis $\{f_n\}_{n\in\mathbb{N}}$ of $\mathscr{H}_c$ yields ON basis

$$\xi_n := \Xi(f_n), \quad n \in \mathbb{N}$$

of $\mathscr{V}$ where $\xi_n \sim N(0,1)$.

ON expansion in $\mathscr{V} \subset L^2_{\mathbf{P}}(\Omega)$:

$$a(\boldsymbol{x},\cdot) = \sum_{n=1}^{\infty} \mathbf{E}\left[a(\boldsymbol{x},\cdot)\xi_n\right]\xi_n.$$

Isometry property of Ξ and reproducing property yield

$$\mathbf{E}\left[a(\boldsymbol{x},\cdot)\xi_n\right] = \Big(c(\boldsymbol{x},\cdot), f_n\Big) = f_n(\boldsymbol{x}).$$

Result: given an ON basis $\{f_n\}_{n\in\mathbb{N}}$ of $\mathscr{H}_c$, the RF a has the expansion

$$a(\boldsymbol{x},\cdot) = \sum_{n=1}^{\infty} \xi_n \, f_n(\boldsymbol{x}), \qquad \boldsymbol{x} \in D,$$

where ξ_n is a sequence of uncorrelated Gaussian RVs with unit variance given by $\xi_n = \Xi(f_n)$.

Note: If a has a.s. continuous realizations, then convergence is uniform on D with probability one.

Karhunen-Loève expansion: use scaled eigenfunctions of Fredholm integral operator with kernel function $c(\boldsymbol{x}, \boldsymbol{y})$ as the ON basis $\{f_n\}$.

Denote by $\{(v_m, \lambda_m)\}_{m\in\mathbb{N}}$ the sequence of eigenpairs of the (compact, selfadjoint) covariance operator

$$C : L^2(D) \to L^2(D), \qquad (Cu)(\boldsymbol{x}) = \int_D u(\boldsymbol{y}) c(\boldsymbol{x}, \boldsymbol{y}) \, d\boldsymbol{y}, \quad \boldsymbol{x} \in D,$$

with $\|v_m\|_{L^2(D)} = 1 \, \forall n$.

Theorem (Mercer, 1909)

The continuous covariance kernel c has the expansion

$$c(\boldsymbol{x}, \boldsymbol{y}) = \sum_{n=1}^{\infty} \lambda_n v_n(\boldsymbol{x}) v_n(\boldsymbol{y})$$

which converges absolutely and uniformly on $D \times D$.

Easy to prove: $\{\sqrt{\lambda_n} v_n\}_{n\in\mathbb{N}}$ is a complete ON system of $\mathscr{H}_c$.
Therefore

Theorem (Karhunen, 1947; Loève, 1945)

A second-order Gaussian random field $a : D \times \Omega \to \mathbb{R}$ with continuous covariance function c and mean field $\overline{a}$ has the expansion

$$a(\boldsymbol{x}, \omega) = \overline{a}(\boldsymbol{x}) + \sum_{n=1}^{\infty} \xi_n(\omega) a_n(\boldsymbol{x})$$

with uncorrelated RVs $\xi_n \sim N(0, 1)$ and the scaled eigenfunctions $a_n(\boldsymbol{x}) = \sqrt{\lambda_n} v_n(\boldsymbol{x})$. The convergence is in quadratic mean in $L^2_{\mathbf{P}}(\Omega)$ and uniform on D.

Next...

Covariance Eigenvalue Problem

- Find $(\lambda, u) \in \mathbb{R} \times L^2(D)$ such that

$$Cu = \lambda u, \qquad \|u\|_{L^2(D)} = 1$$

- with covariance operator $C : L^2(D) \to L^2(D)$ defined by

$$(Cu)(x) = \int_D c(\boldsymbol{x}, \boldsymbol{y}) u(\boldsymbol{y}) \, d\boldsymbol{y}$$

- $c(\boldsymbol{x}, \boldsymbol{y})$ covariance function (kernel) of RF defined on $D \subset \mathbb{R}^d$.

- Variational Formulation: Find $(\lambda, u) \in \mathbb{R} \times L^2(D)$, such that

$$(Cu, v) = \lambda(u, v) \qquad \forall v \in L^2(D),$$

$$(Cu, v) = \int_D \int_D u(\boldsymbol{y}) c(\boldsymbol{x}, \boldsymbol{y}) v(\boldsymbol{x}) \, d\boldsymbol{y} \, d\boldsymbol{x}$$

$$(u, v) = \int_D u(\boldsymbol{x}) v(\boldsymbol{x}) \, d\boldsymbol{x}$$

- Galerkin approximation on finite dimensional subspace $\mathscr{U}_N = \operatorname{span}\{\phi_1, \phi_2, \dots, \phi_N\} \subset L^2(D)$

e.g.: $\mathscr{U}_N$ space of discontinuous piecewise polynomials on a FE triangulation of D

- No inter-element continuity needed for conforming discretization, basis function have small support.

Theorem (Todor, 2006)

Let $\{\mathscr{T}_h\}_{h>0}$ be a family of admissible triangulations of D with meshwidth h and define S_h to be the space of discontinous piecewise polynomials of degree p on $\mathscr{T}_h$.

Then for any $s > 0$ there exists a constant $K = K(C, \mathscr{T}, p, s) > 0$ such that the Galerkin approximations $\lambda_m^{(h)}$ of the eigenvalues λ_m of the covariance operator C satisfy

$$0 \le \lambda_m - \lambda_m^{(h)} \le K(h^{2p+2}\lambda_m^{1-s} + h^{4p+4}\lambda_m^{-2s}) \quad \forall m \in \mathbb{N}, \forall h > 0,$$

implying

$$0 \le \lambda_m - \lambda_m^{(h)} \le K h^{2p+2}\lambda_m^{\frac{1}{2}-s} \qquad \forall m \in \mathbb{N}, \forall h > 0.$$

- Coefficient vector $\boldsymbol{u} \in \mathbb{R}^N$ for $u = \sum_{j=1}^N u_j \phi_j$
- Galerkin projection leads to generalized eigenvalue problem

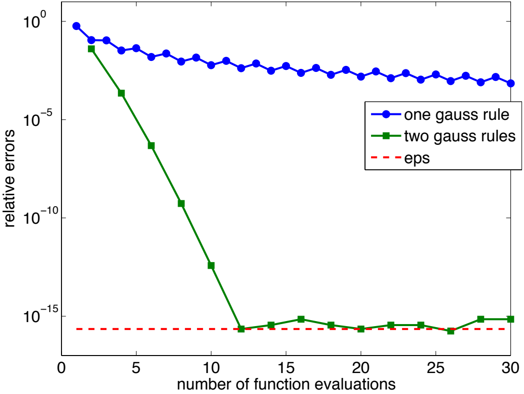
$$\boldsymbol{C}\boldsymbol{u} = \lambda \boldsymbol{M}\boldsymbol{u}$$

where

$$[\boldsymbol{C}]_{i,j} = (C\phi_j, \phi_i) \qquad \text{(discrete integral operator)}$$
$$[\boldsymbol{M}]_{i,j} = (\phi_j, \phi_i) \qquad \text{(mass matrix of basis)}$$
$$i, j = 1, \ldots, N.$$

- $\boldsymbol{M}$ can be made diagonal (orthogonalize basis elementwise), but $\boldsymbol{C}$ is in general full.

- High-order quadrature assumes smoothness.
- Example: $\int_{-1}^{1} e^{-|x|}\,dx$ with a single Gauss rule.
- Better: same Gauss rule on 2 subintervals.



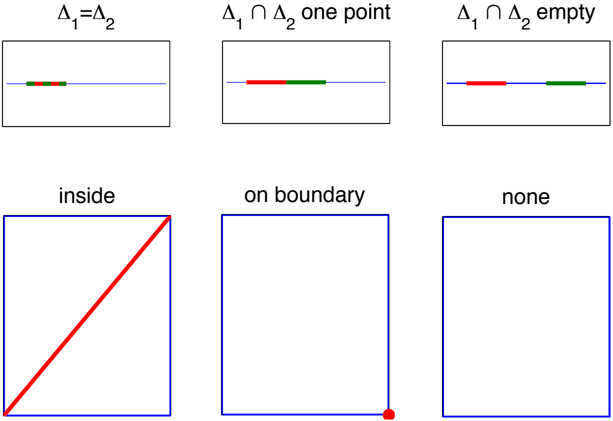
Adapted Quadrature
Assembly of C

- For piecewise constant approximation matrix entries are

$$[\boldsymbol{C}]_{ij} = \int_{\Delta_i} \int_{\Delta_j} \phi_i(\boldsymbol{y}) c(\boldsymbol{x}, \boldsymbol{y}) \phi_j(\boldsymbol{x}) \, d\boldsymbol{y} \, d\boldsymbol{x}$$

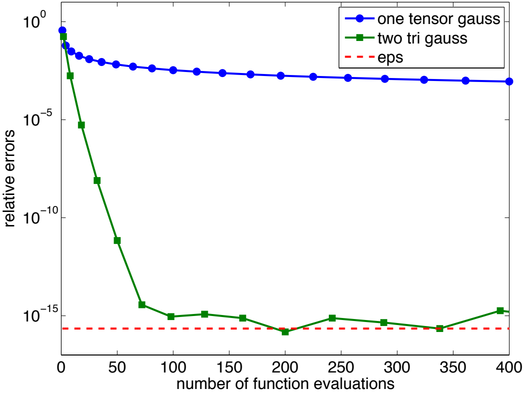
- Typical covariance functions have low smoothness for $\boldsymbol{x} = \boldsymbol{y}$.
- Same trick as in previous example: divide the integration region (subset of $D \times D$) into subregions such that no points with $\boldsymbol{x} = \boldsymbol{y}$ lie in the interior of a subregion.

- Integration region Cartesian product of intervals Δ_i and Δ_j.
- Three possible cases for points $x = y$:



Adapted Quadrature
Qudrature rule for identical case

- Only need to worry about identical intervals case.
- Subdivision obvious: divide square into two triangles.
- Compare product Gauss quadrature over square with two triangular Gauss formulas over the two triangles:



- Integration region Cartesian product $\Delta_i \times \Delta_j$ of two triangles
- After transformation of Δ_i and Δ_j to reference triangle integration domain is fixed.
- Possible cases in 2D:
  - identical triangles
  - common edge
  - common point
  - disjoint

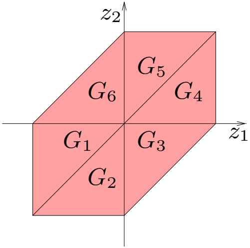
- Similar quadrature problems in 3D-BEM, but there kernels have stronger singularities in $\boldsymbol{x} = \boldsymbol{y}$.
- Adapt 3D-BEM quadrature techniques [Sauter & Schwab, 2004]
- Three basic steps::
  1. Change of variables to shift singularity to origin.
  2. Divide domain of integration leaving singularity on subdomain boundary.
  3. Apply standard quadrature on subdomains.
- Consider case of identical triangles.

  reference triangle:

$$R = \left\{ (x_1, x_2) \in \mathbb{R}^2 : \quad 0 \le x_1 \le 1, 0 \le x_2 \le x_1 \right\}$$

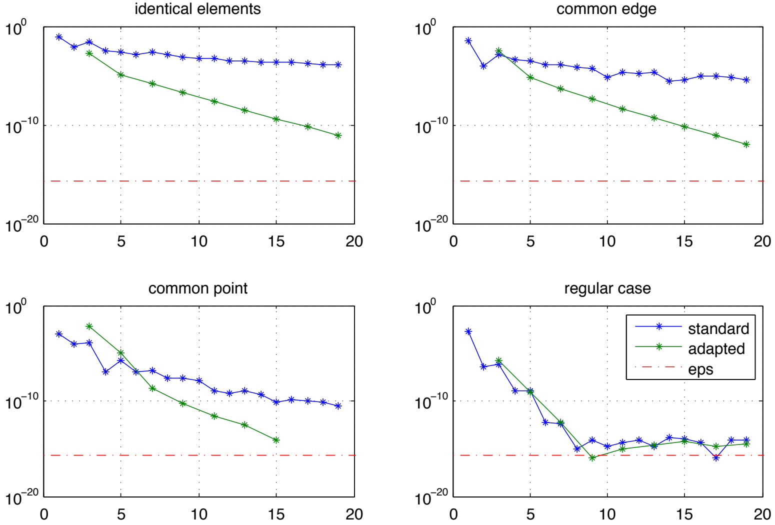
Adapted Quadrature
Identical triangles in 2D

- Reference triangle:
  $R = \{(x_1, x_2) \in \mathbb{R}^2 : 0 \le x_1 \le 1, 0 \le x_2 \le x_1\}$
- Difference coordinate $\boldsymbol{z} = \boldsymbol{y} - \boldsymbol{x} \Rightarrow$ points with $\boldsymbol{x} = \boldsymbol{y}$ fixed at $\boldsymbol{z} = 0$.
- Projection of the domain of integration on the $\boldsymbol{z}$-plane



- 6 subdomains (all 4-simplices) $\Rightarrow$ quadrature rules for 4-simplices or transformation to $[0, 1]^4$

Adapted Quadrature
Example: $c(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\|\boldsymbol{y} - \boldsymbol{x}\|)$



- Require M largest approximate eigenvalues & associated eigenvectors of generalized eigenvalue problem.
- Krylov projection methods avoid computing all eigenpairs; require only matrix-vector products
- Covariance operators selfadjoint, hence short recurrence Krylov methods like Lanczos applicable.
- Thick-Restart variant of Lanczos [Simon & Wu, 2000] allows iterative improvement of desired eigenspace by efficient restarting scheme.
- Extended to generalized eigenvalue problem and block version (multiple eigenvalues)

- Lanczos-decomposition after m (standard) steps:

$$\boldsymbol{A}\boldsymbol{Q}_m = \boldsymbol{Q}_m\boldsymbol{T}_m + \beta_m\boldsymbol{q}_{m+1}\boldsymbol{e}_m^T \tag{L}$$

- $k < m$ Ritz values $\vartheta_1, \vartheta_2, \dots, \vartheta_k$ to be refined in next restart cycle
- Ritz pairs $(\vartheta_j, \boldsymbol{y}_j)$ satisfy

$$\boldsymbol{T}_m\boldsymbol{Y} = \boldsymbol{Y}\operatorname{diag}(\vartheta_1, \vartheta_2, \dots, \vartheta_k) =: \boldsymbol{Y}\hat{\boldsymbol{T}}_k \quad \text{with} \quad \boldsymbol{Y}^T\boldsymbol{Y} = \boldsymbol{I}$$

Multiply (L) from right by $\boldsymbol{Y}$

$$\boldsymbol{A}\hat{\boldsymbol{Q}}_k = \hat{\boldsymbol{Q}}_k\hat{\boldsymbol{T}}_k + \beta_m\hat{\boldsymbol{q}}_{k+1}\boldsymbol{s}^T$$

with $\hat{\boldsymbol{Q}}_k = \boldsymbol{Q}_m\boldsymbol{Y}$, $\hat{\boldsymbol{q}}_{k+1} = \boldsymbol{q}_{m+1}$ and $\boldsymbol{s} = \boldsymbol{Y}^T\boldsymbol{e}_m$

but: this is not a Lanczos-decomposition (trailing rank-1 matrix)

- Next Lanczos vector $\hat{\boldsymbol{q}}_{k+2}$ by full orthogonalization:

$$\begin{aligned}\hat{\beta}_{k+1}\hat{\boldsymbol{q}}_{k+2} &= (\boldsymbol{I} - \hat{\boldsymbol{Q}}_{k+1}\hat{\boldsymbol{Q}}_{k+1}^T)\boldsymbol{A}\hat{\boldsymbol{q}}_{k+1} \\ &= (\boldsymbol{I} - \hat{\boldsymbol{q}}_{k+1}\hat{\boldsymbol{q}}_{k+1}^T - \hat{\boldsymbol{Q}}_k\hat{\boldsymbol{Q}}_k^T)\boldsymbol{A}\hat{\boldsymbol{q}}_{k+1} \\ &= \boldsymbol{A}\hat{\boldsymbol{q}}_{k+1} - \hat{\alpha}_{k+1}\hat{\boldsymbol{q}}_{k+1} - \hat{\boldsymbol{Q}}_k\beta_m\boldsymbol{s}\end{aligned}$$

- $\hat{\boldsymbol{Q}}_k^T\boldsymbol{A}\hat{\boldsymbol{q}}_{k+1} = \beta_m\boldsymbol{s}$
- Obtain decomposition with right structure

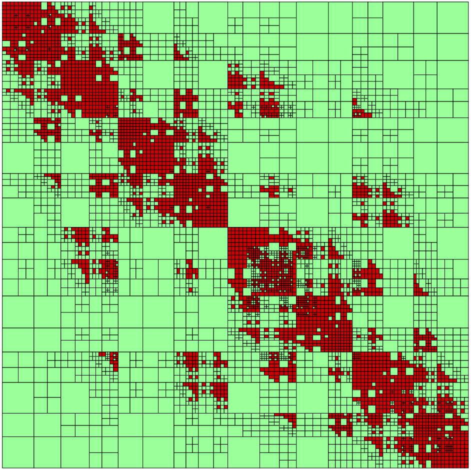
$$\boldsymbol{A}\hat{\boldsymbol{Q}}_{k+1} = \hat{\boldsymbol{Q}}_{k+1}\hat{\boldsymbol{T}}_{k+1} + \beta_{k+1}\hat{\boldsymbol{q}}_{k+2}\boldsymbol{e}_{k+1}^T$$

- Not a proper Lanczos decomposition ($\hat{\boldsymbol{T}}_{k+1}$ not tridiagonal), but can now continue with 3-term recurrence.

$$\hat{\boldsymbol{T}}_{k+1} = \begin{pmatrix} \hat{\boldsymbol{T}}_k & \beta_m\boldsymbol{s} \\ \beta_m\boldsymbol{s}^T & \hat{\alpha}_{k+1}\end{pmatrix}$$

Hierarchical Matrix Approximation

- Algebraic variant of fast multipole method, [Hackbusch et al., 2000]
- Partition dense matrix into rectangular blocks of 2 types
  - full near-field blocks,
  - low-rank far field blocks
- blocks correspond to clusters of degrees of freedom, i.e., clusters of supports of Galerkin basis functions
- yields data-sparse representation of matrix, construction $O(N \log N)$, matrix-vector product in $O(N)$.



Hierarchical Matrix Approximation
Far-field case

- If Δ_i and Δ_j are well separated, the covariance function can be appproximated by a low degree interpolation:

$$c(\boldsymbol{x}, \boldsymbol{y}) \approx \tilde{c}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{k=1}^{r} c(\boldsymbol{x}_k, \boldsymbol{y}) \ell_k(\boldsymbol{x})$$

- ℓ_k are Lagrange polynomials to the interpolation points $\{\boldsymbol{x}_k\}_{k=1}^r$
- This is also true for two well-separated clusters σ and τ of elements of the triangulation $\mathscr{T}_h$

- For $[\boldsymbol{C}]_{ij}$ for triangles $\Delta_i \in \sigma$ and $\Delta_j \in \tau$:

$$
\begin{aligned}
[\boldsymbol{C}]_{ij} &= \int_{\Delta_i}\int_{\Delta_j} \phi_i(\boldsymbol{y}) c(\boldsymbol{x},\boldsymbol{y}) \phi_j(\boldsymbol{x})\, d\boldsymbol{y}\, d\boldsymbol{x} \\
&\approx \int_{\Delta_i}\int_{\Delta_j} \phi_i(\boldsymbol{x}) \left(\sum_{k=1}^{r} c(\boldsymbol{x}_k,\boldsymbol{y}) \ell_k(\boldsymbol{x}) \right) \phi_j(\boldsymbol{y})\, d\boldsymbol{y}\, d\boldsymbol{x} \\
&= \sum_{k=1}^{r} \left(\int_{\Delta_i} c(\boldsymbol{x}_k,\boldsymbol{y}) \phi_i(\boldsymbol{y})\, d\boldsymbol{y} \right) \left(\int_{\Delta_j} \ell_k(\boldsymbol{x}) \phi_j(\boldsymbol{x})\, d\boldsymbol{x} \right) \\
&= [A_{\sigma,\tau} B_\tau]_{ij}
\end{aligned}
$$

$\Rightarrow$ whole blocks (maybe after permutation of the indices) of $\boldsymbol{C}$ can be approximated by a rank-r matrix in factored form

- Assembly of hierarchical matrix approximation of $\boldsymbol{C}$:
  (1) Divide triangulation into cluster (cluster tree, clustering strategy, minimal cluster size)
  (2) Determine for each pair of clusters whether corresponding matrix block can be approximated by a low rank matrix (block cluster tree)
  (3) admissibility condition for cluster pair (σ, τ):

$$\min(\operatorname{diam}(\sigma), \operatorname{diam}(\tau)) \leq \eta \operatorname{dist}(\sigma, \tau)$$

  (4) compute for each matrix block the low rank approximation (admissible) or the full block (inadmissible)

$\Rightarrow$ assembly of hierarchical matrix and the matrix-vector-product have complexity of $O(N \log N)$.

$\Rightarrow$ Lanczos solver for integral operator becomes scalable.

- If clusters well-separated covariance function smooth in $\boldsymbol{x}$ and $\boldsymbol{y}$

$\Rightarrow$ interpolate the covariance function in both variables:

$$\begin{aligned}[\boldsymbol{C}]_{ij} &\approx \sum_{l=1}^{r}\sum_{k=1}^{r}\left(\int_{\Delta_i}\ell_l(\boldsymbol{y})\phi_i(\boldsymbol{y})\,d\boldsymbol{y}\right)c(\boldsymbol{x}_k,\boldsymbol{y}_l)\left(\int_{\Delta_j}\ell_k(\boldsymbol{x})\phi_j(\boldsymbol{x})\,d\boldsymbol{x}\right)\\ &= [V_\sigma S_{\sigma,\tau}W_\tau]_{ij}\end{aligned}$$

- another admissibility condition:

$$\max(\operatorname{diam}(\sigma),\operatorname{diam}(\tau)) \le \eta\,\operatorname{dist}(\sigma,\tau)$$

- Together with certain other techniques (nested cluster basis) matrix-vector product complexity reduced to $O(N)$

Next...

- Bessel covariance (Matérn family, $\nu = 1$):

$$k_1(\boldsymbol{x}, \boldsymbol{y}) = \frac{\|\boldsymbol{x} - \boldsymbol{y}\|}{c} K_1\left(\frac{\|\boldsymbol{x} - \boldsymbol{y}\|}{c}\right), \qquad \boldsymbol{x}, \boldsymbol{y} \in D = [-1, 1]^2.$$

- $\mathscr{U}_n$: piecewise constants on triangulation of D
- hierarchical matrix parameters

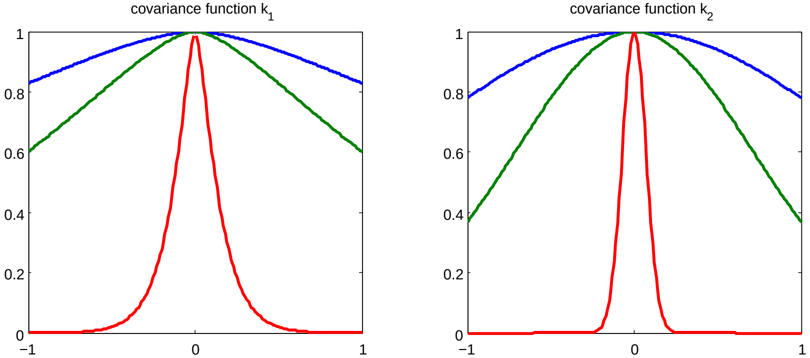
degree of interpolation	:	4
admissibility parameter	:	$1/c$
minimal cluster size	:	62

- 5 largest eigenvalues with restart length 10
- **Environment:**
  MATLAB 2012a on single node machine,
  Opteron 6136 (2.4 GHz), 256 GB RAM,
  Calls to HLib 1.4 / HLib Pro libraries (MPI Leipzig) via MEX

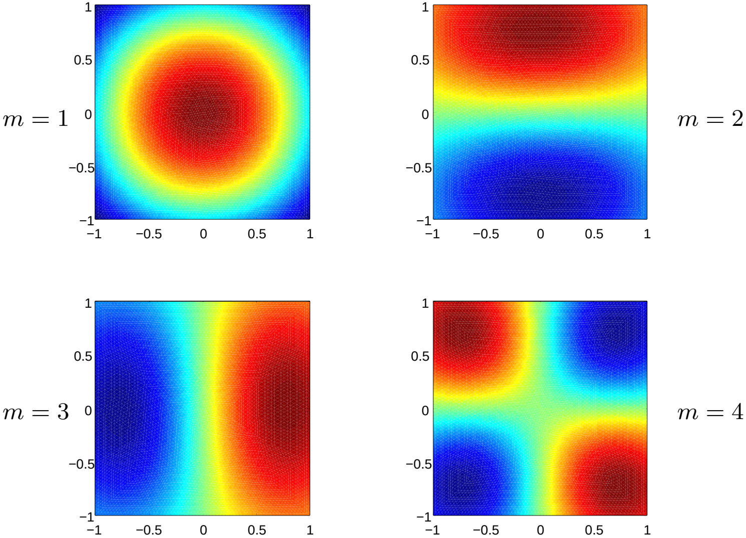
- Gaussian covariance kernel:

$$k_2(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{y}\|^2}{c^2}\right)$$

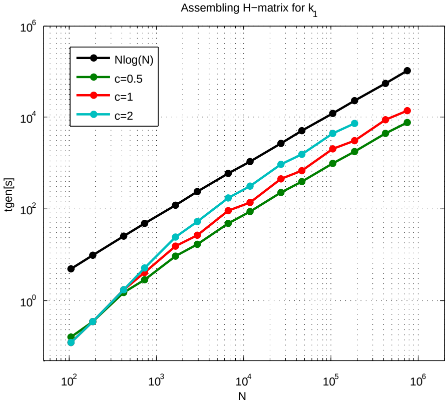
- correlation lengths $c = 0.1, 1, 2$

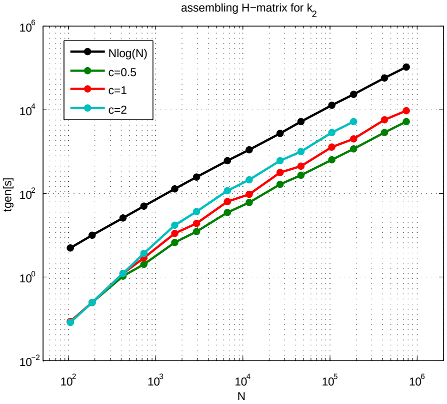


Numerical examples
First four eigenmodes $c = 0.5$

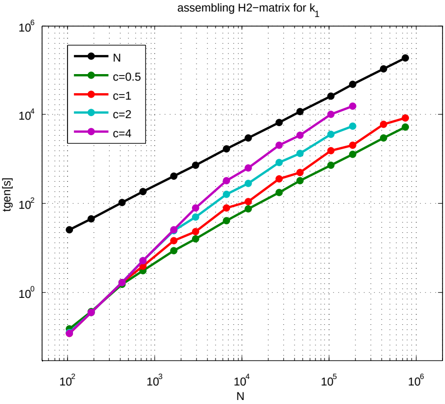


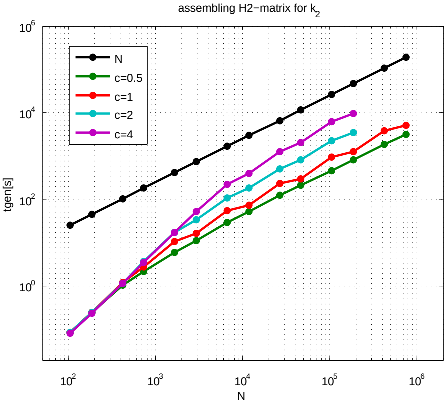
Numerical examples
Assembly timings, $\mathscr{H}$ matrices



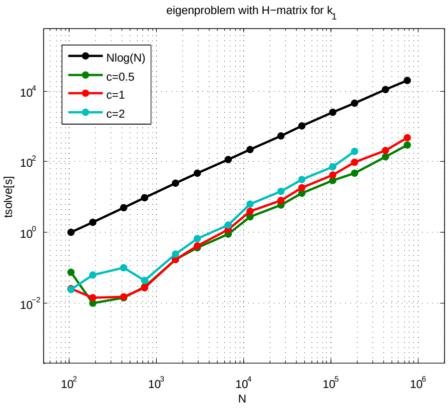


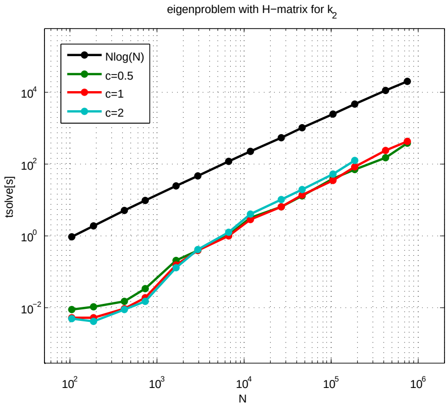
Numerical examples
Assemby timings, $\mathscr{H}^2$ matrices



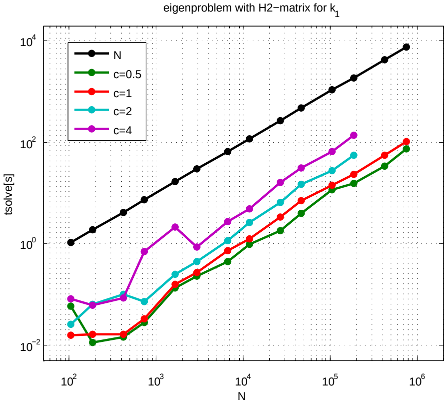


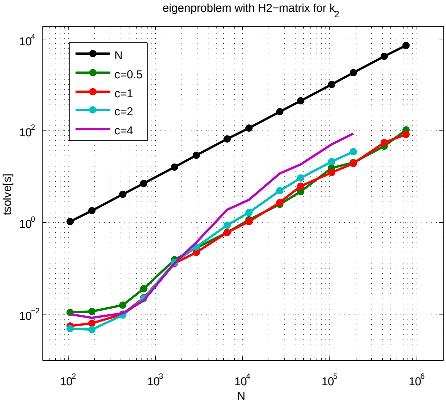
Numerical examples
Solution timings, $\mathscr{H}$ matrices



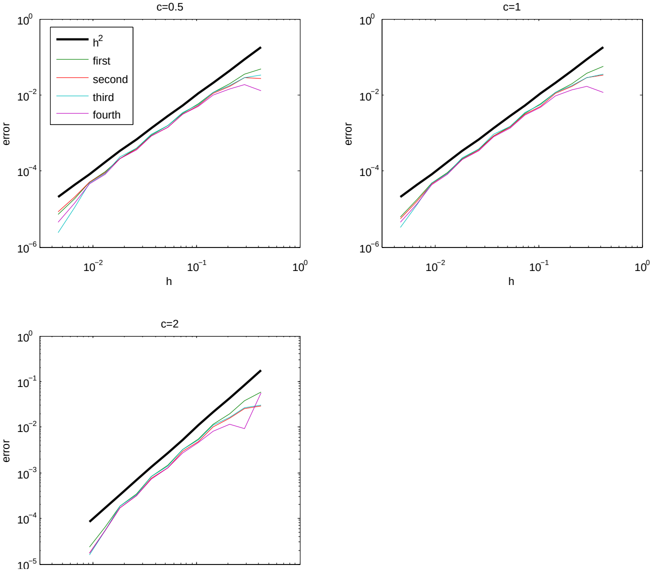


Numerical examples
Solution timings, $\mathscr{H}^2$ matrices

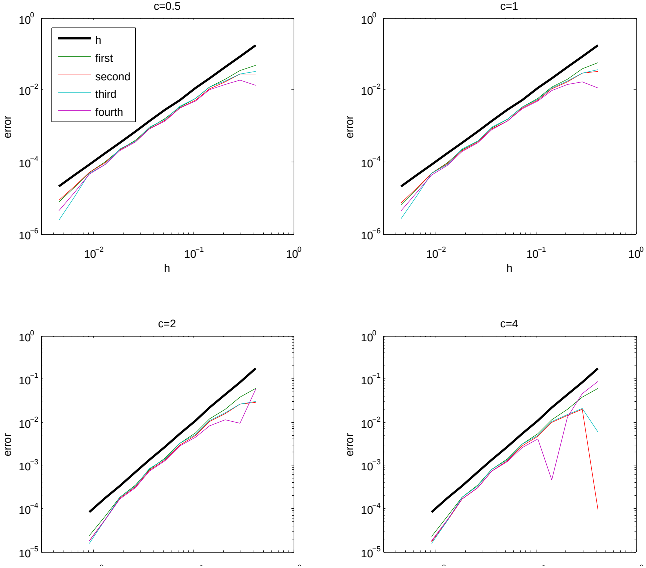




Numerical examples
Convergence for k_1, $\mathscr{H}$ matrices

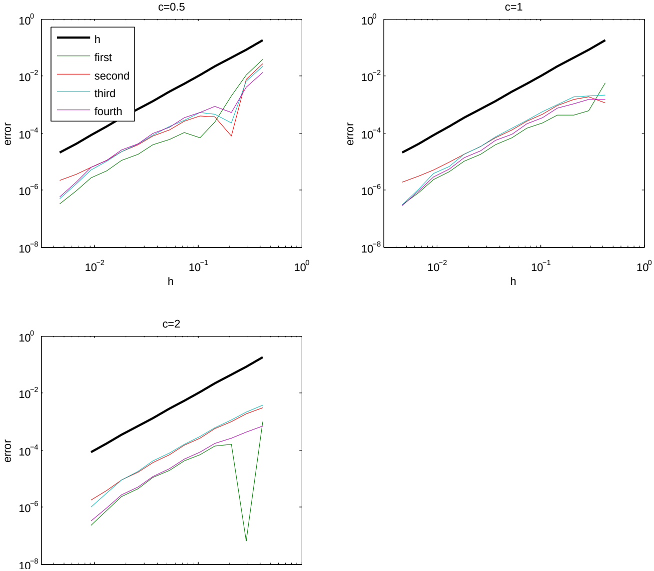


Numerical examples
Convergence for k_1, $\mathscr{H}^2$ matrices



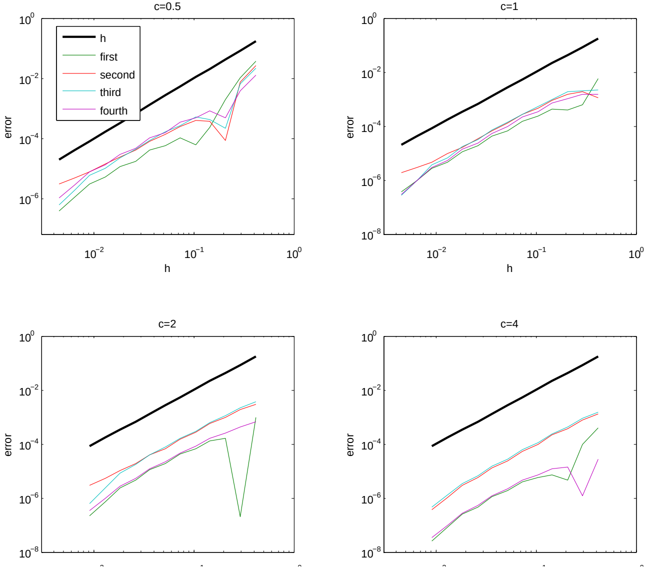
Numerical examples

Convergence for k_2, $\mathscr{H}$ matrices

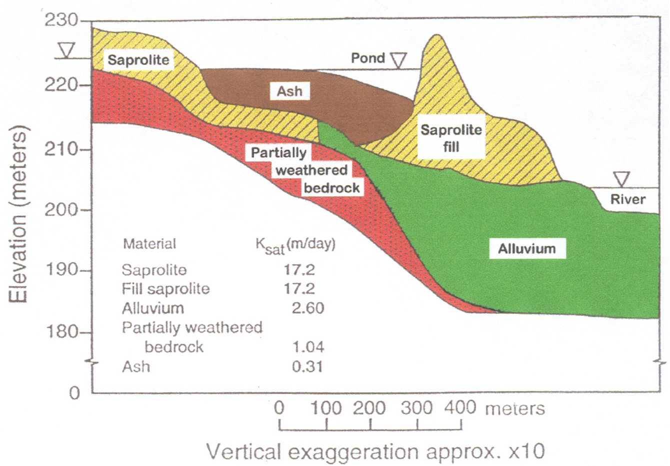


Numerical examples

Convergence for k_2, $\mathscr{H}^2$ matrices

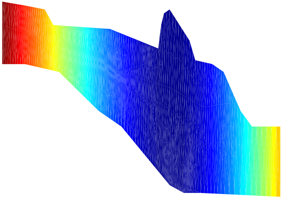


[Rivière & Wheeler, 1999]

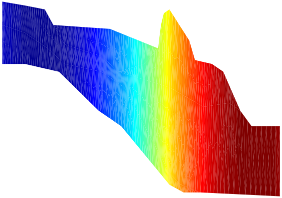


Numerical examples

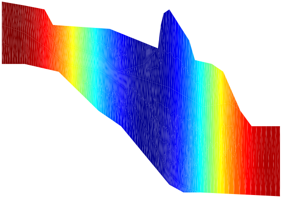
L-site: covariance modes, Bessel correlation, $\ell = 400m$



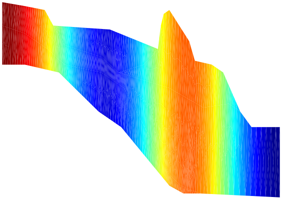
mode 1



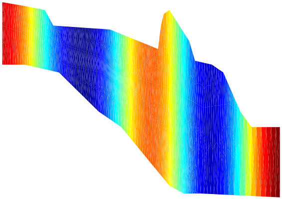
mode 2



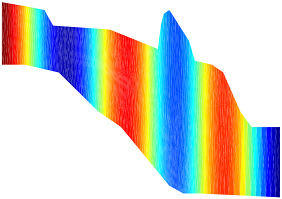
mode 3



mode 4



mode 5



mode 6

Concluding Remarks

Summary

- Scalable covariance eigenvalue solver based on restarted (block) Lanczos and hierarchical matrix approximation
- Adapted quadrature
- Can incorporate conditioning on measured data, Kriging, REML estimates of mean.
- Flexible w.r.t. kernel, geometry.

In progress

- 3D (essentially just the quadrature).
- Pcw. linears/quadratics.
- Localized alternatives to KL (joint with Raul Tempone).

References

- Robert J. Adler and Jonathan E. Taylor.
  Random Fields and Geometry.
  Springer Monographs in Mathematics. Springer-Verlag, New York, 2007.
- Alain Berlinet and Christine Thomas-Agnan.
  Reproducing Kernel Hilbert Spaces in Probability and Statistics.
  Kluwer Academic, Norwell, MA, 2004.
- Claude Gittelson.
  Representation of Gaussian fields in series with independent coefficients.
  IMA Journal of Numerical Analysis, 32:294–319, 2012.
- Wolfgang Hackbusch.
  Hierarchische Matrizen: Algorithmen und Analysis.
  Springer-Verlag, 2009.
- Reuven Meidan.
  Reproducing kernel Hilbert spaces of distributions and generalized stochastic processes.
  SIAM J. Math. Anal., 10(1):62–70, 1979.