Institut für Numerische Mathematik und Optimierung

Karhunen-Loève Approximation of Random Fields Using Hierarchical Matrix Techniques

Oliver Ernst

Computational Methods with Applications
Harrachov, CR, August 19–25, 2007

Collaborators

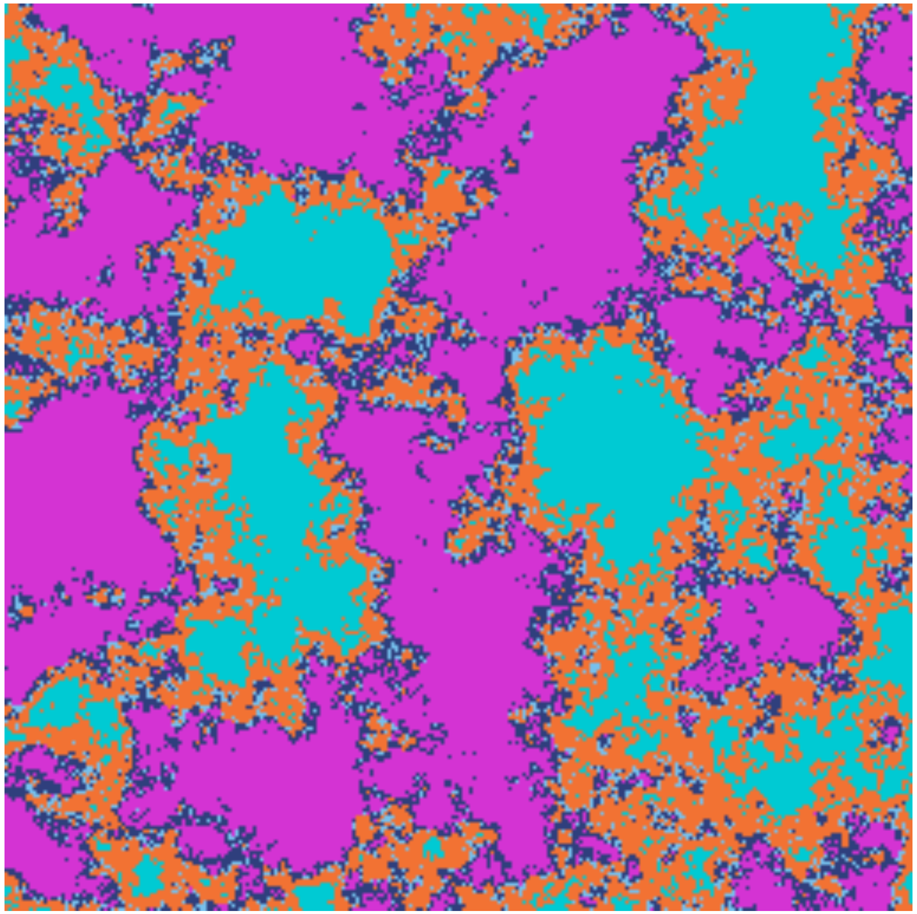
- Catherine Powell, David Silvester
  University of Manchester, School of Mathematics,
  Manchester, UK
- Ingolf Busch, Michael Eiermann, Elisabeth Ullmann,
  TU Bergakademie Freiberg,
  Freiberg, Germany

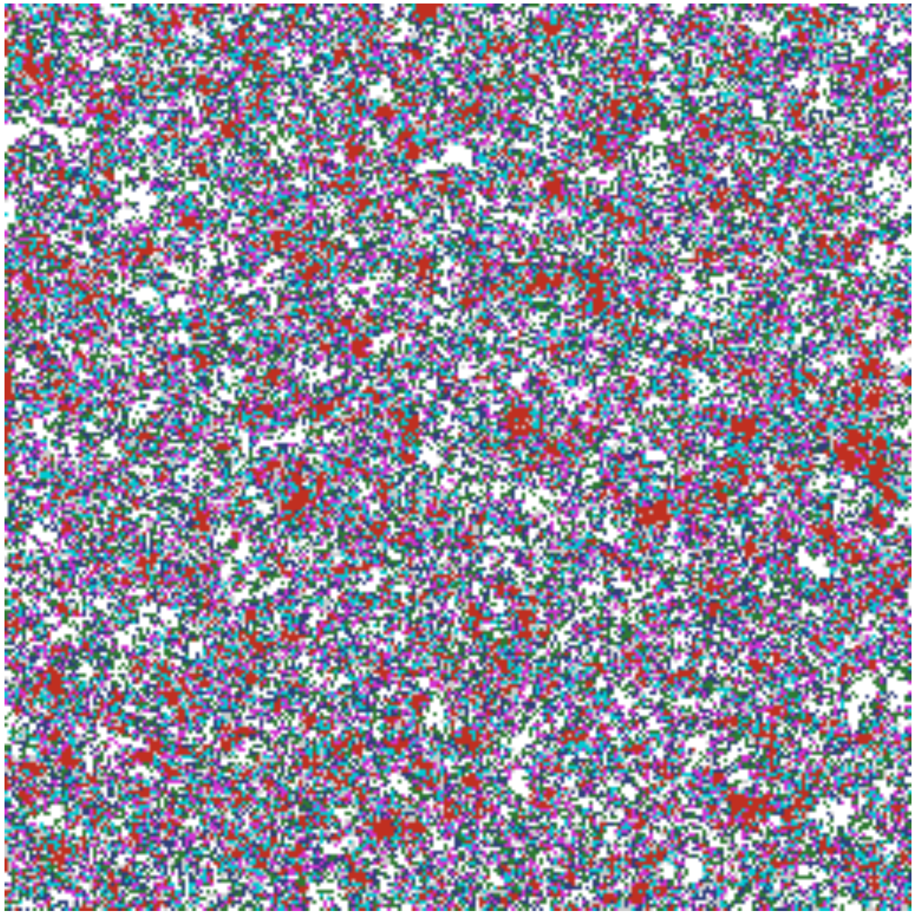
Support: DAAD/British Council

Outline

- Random fields and the Karhunen-Loève expansion
- Discretization of the covariance operator
- Solution of the discrete eigenvalue problem
- A numerical example

Random Fields





Formally

- stochastic process indexed by a spatial coordinate $x \in D \subset \mathbb{R}^d$, D bounded, i.e.,
- measurable function $a : D \times \Omega \to \mathbb{R}$, where $(\Omega, \mathscr{A}, P)$ is a given probability space
- For $\omega \in \Omega$ fixed, $a(\cdot, \omega)$ is a realization of the random field, i.e., a function $D \to \mathbb{R}$.
- For $x \in D$ fixed, $a(x, \cdot)$ is a random variable (RV) w.r.t. $(\Omega, \mathscr{A}, P)$.

Notation

$\langle \xi \rangle := \int_\Omega \xi(\omega)\, dP(\omega)$	expected value of RV $\xi : \Omega \to \mathbb{R}$
$\overline{a}(x) := \langle a(x,\cdot) \rangle$	mean of RF a at $x \in D$
$\mathrm{Cov}_a(x,y) := \langle (a(x,\cdot) - \overline{a}(x))(a(y,\cdot) - \overline{a}(y)) \rangle$	covariance of RF a at $x, y \in D$
$\mathrm{Var}_a(x) := \mathrm{Cov}_a(x,x)$	variance of RF a at $x \in D$
$\sigma_a(x) := \sqrt{\mathrm{Var}_a(x)}$	standard deviation of RF a at $x \in D$
$L^2_P(\Omega) := \{\xi : \langle \xi^2 \rangle < \infty\}$	RV of second order

A RF is of second order, if $a(x,\cdot) \in L^2_P(\Omega)$ for all $\boldsymbol{x} \in D$.

Theorem (Karhunen-Loève expansion). *Given a second order RF $a = a(x,\omega)$ with continuous covariance function $c(x,y) := \mathrm{Cov}_a(x,y)$, denote by $\{(\lambda_m, a_m(x))\}$ the eigenpairs of the (compact) integral operator*

$$C : L^2(D) \to L^2(D), \qquad (Cu)(x) = \int_D u(y)\, c(x,y)\, dy,$$

there exists a sequence $\{\xi_m\}_{m\in\mathbb{N}}$ of random variables with

$$\langle \xi_m \rangle = 0 \;\forall m, \qquad \langle \xi_m \xi_n \rangle = \delta_{m,n} \;\forall m,n$$

such that the Karhunen-Loève (KL) expansion

$$a(x,\omega) = \overline{a}(x) + \sum_{m=1}^{\infty} \sqrt{\lambda_m}\, a_m(x)\, \xi_m(\omega) \qquad \text{(KL)}$$

converges uniformly on D and in L^2_P.

Note:

- Covariance functions $c(x, y)$ are continuous on $\overline{D} \times \overline{D}$ as well as symmetric and of positive type.
- Therefore covariance operators C are compact, hence spectra $\Lambda(C)$ consist of countably many eigenvalues accumulating at most at zero.
- Covariance operators are selfadjoint and positive semidefinite.

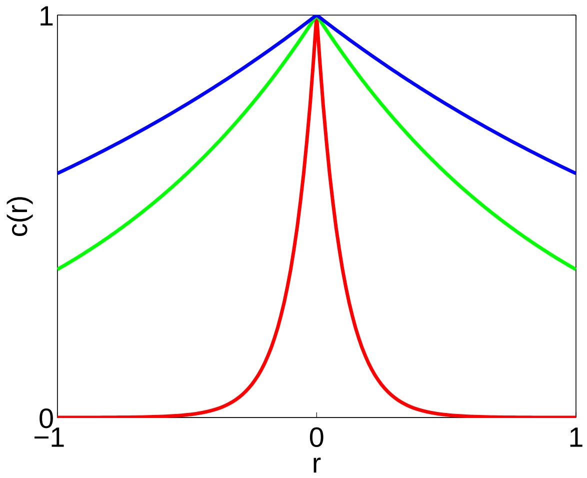
Analogy

Singular value expansion of integral operator

$$
\begin{aligned}
&A : L^2(D) \to L^2_P, && f(x) \mapsto (Af)(\omega) := \int_D f(x) a(x,\omega)\, dx,\\
&A^* : L^2_P \to L^2(D), && \xi(\omega) \mapsto (A^*\xi)(x) = \int_\Omega \xi(\omega) a(x,\omega)\, dP(\omega)\\
&C = A^* A.
\end{aligned}
$$

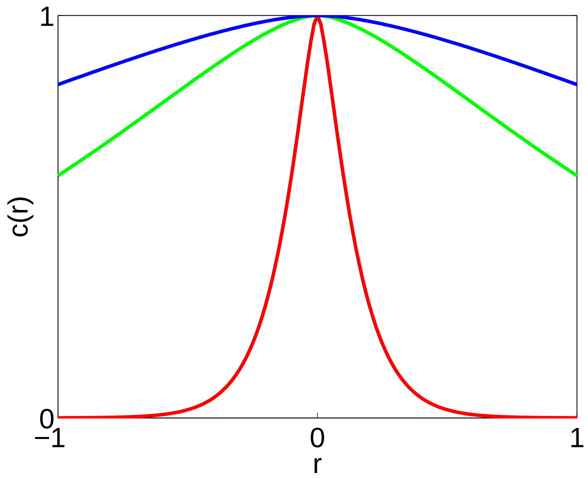
Common Covariance Models

$$\mathrm{Cov}_a(x, y) = c(x, y) = c(\rho), \quad \rho = \|x - y\|$$



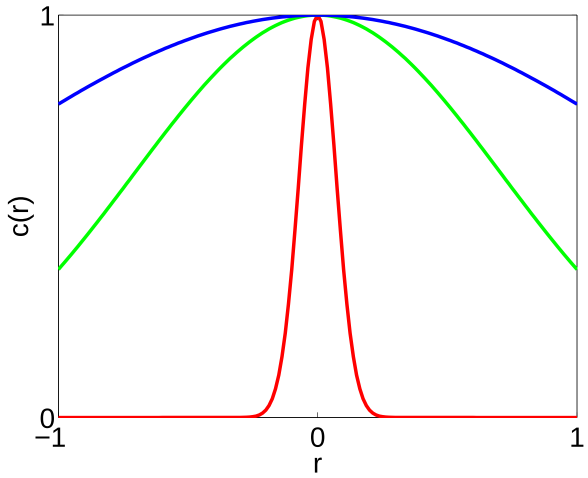
exponential

$$c(r) = \sigma^2 e^{-\rho/\ell}$$



Bessel

$$c(r) = \sigma^2 \frac{\rho}{\ell} K_1(\frac{\rho}{\ell})$$



Gaussian

$$c(r) = \sigma^2 e^{-\rho^2/\ell^2}$$

$\ell > 0$ is a measure of the “correlation length”, here $\ell = 0.1, 1, 2$.

Variance

For normalized eigenfunctions $a_m(x)$,

$$\mathrm{Var}_a(x) = c(x,x) = \sum_{m=1}^{\infty} \lambda_m a_m(x)^2,$$

$$\int_D \mathrm{Var}_a(x)\,dx = \sum_{m=1}^{\infty} \lambda_m \underbrace{(a_m, a_m)_D}_{=1} = \mathrm{trace}\, C.$$

For constant variance (e.g., stationary RF),

$$\mathrm{Var}_a \equiv \sigma^2 > 0, \qquad \sum_m \lambda_m = |D|\,\sigma^2.$$

Truncated KL Expansion

For computational purposes, KL expansion truncated after M terms:

$$a^{(M)}(x,\omega) = \overline{a}(x) + \sum_{m=1}^{M} \sqrt{\lambda_m}\, a_m(x)\, \xi_m(\omega).$$

Truncation error

$$\langle \|a - a^{(M)}\|_D^2 \rangle = \sum_{m=M+1}^{\infty} \lambda_m.$$

Choose M such that sufficient amount of total variance of RF is retained.

Eigenvalue Decay

Roughly: the smoother the kernel, the faster $\{\lambda_m\}_{m\in\mathbb{N}} \to 0$.

More precisely: if $D \subset \mathbb{R}^d$, then if the kernel function c is

piecewise H^r :	$\lambda_m \le c_1 m^{-r/d}$
piecewise smooth :	$\lambda_m \le c_2 m^{-r}$ for any $r > 0$
piecewise analytic :	$\lambda_m \le c_3 \exp(-c_4 m^{1/d})$

for suitable constants c_1, c_2, c_3, c_4.

Note: piecewise smoothness of kernel also leads to bounds on derivatives of eigenfunctions a_m in $L^\infty(D)$.

Proven e.g. in [Schwab & Todor (2006)], [Todor (2006)]

Galerkin Discretization

- $\mathscr{T}_h$ admissible finite element triangulation of D
- finite dimensional subspace of piecewise polynomials

$$\mathscr{V}^h = \{\phi : D \to \mathbb{R} \ : \ \phi|_T \in \mathscr{P}_k \ \forall T \in \mathscr{T}\} \subset L^2(D).$$

- Discrete eigenvalue problem: find pairs (λ_m^h, a_m^h) such that

$$(Ca_m^h, \phi) = \lambda_m^h (a_m^h, \phi) \qquad \forall \phi \in \mathscr{V}^h, \qquad m = 1, 2, \ldots$$

  corresponds to generalized matrix eigenvalue problem

$$\boldsymbol{Cx} = \lambda \boldsymbol{Mx}, \qquad [\boldsymbol{C}]_{i,j} = (C\phi_j, \phi_i), [\boldsymbol{M}]_{i,j} = (\phi_j, \phi_i),$$
$$i, j = 1, 2, \ldots, N = \dim \mathscr{V}^h.$$

- $\boldsymbol{C}$ large and dense, $\boldsymbol{M}$ can be made diagonal using suitable basis.

Discretization Error

Discrete operator given by $C_h = P_h C P_h$, P_h the $L^2(D)$ orthogonal projection to $\mathscr{V}^h$.

Discrete eigenpairs $\{(\lambda_m^h, a_m^h)\}_{m=1}^N$

If covariance operator is piecewise smooth, then for any $r > 0$

$$0 \leq \lambda_m - \lambda_m^h \leq K_r \left(h^{2(k+1)} \lambda_m^{1-r} + h^{4(k+1)} \lambda_m^{-2r} \right),$$

$$\|(I - P_h) a_m\|_{L^2(D)} \leq K_r \lambda_m^{-r} h^{k+1}.$$

[Todor (2006)]

Solution of Matrix Eigenvalue Problem

- Only fixed number of leading eigenpairs required, suggests restarted Krylov subspace technique.

  We use the Thick-Restart Lanczos (TRL) method [Simon & Wu (2000)].

  Idea: limit dimension of Krylov space to fixed m, save some desired approximate eigenpairs, generate new Krylov space which contains these retained approximations (restart).

- Krylov methods require inexpensive matrix-vector product.
  We obtain this by replacing C by a hierarchical matrix approximation $\widetilde{C}$, for which matrix vector products can be computed in $O(N \log N)$ operations [Hackbusch (1999)].

Thick-Restart Lanczos Cycle

(1) Given Lanczos decomposition of Krylov space $\mathscr{K}_m(A, \boldsymbol{v})$

$$AQ_m = Q_m T_m + \beta_{m+1} \boldsymbol{q}_{m+1} \boldsymbol{e}_m^\top, \qquad Q_m = [\boldsymbol{q}_1, \dots, \boldsymbol{q}_m],\ Q_m^\top Q_m = I_m,$$

(2) compute eigenpairs $T_m \boldsymbol{y}_j = \vartheta_j \boldsymbol{y}_j$, $j = 1, \dots, m$,

(3) select $k < m$ Ritz vectors to retain, $Y_k := [\boldsymbol{y}_1, \dots, \boldsymbol{y}_k]$,

(4) set $\widehat{Q}_k := Q_m Y_k$, $\widehat{T}_k := \widehat{Q}_k^\top T_m \widehat{Q}_k$ to obtain

$$A\widehat{Q}_k = \widehat{Q}_k \widehat{T}_k + \beta_{m+1} \widehat{\boldsymbol{q}}_{k+1} \boldsymbol{s}^\top \qquad \text{with } \widehat{\boldsymbol{q}}_{k+1} = \boldsymbol{q}_{m+1} \text{ and } \boldsymbol{s} := Y_k^\top \boldsymbol{e}_m,$$

(5) extend $\operatorname{span}\{\widehat{\boldsymbol{q}}_1, \dots, \widehat{\boldsymbol{q}}_{m+1}\}$ to Krylov space of order m with Lanczos-type decomposition

$$A\widehat{Q}_m = \widehat{Q}_m \widehat{T}_m + \widehat{\beta}_{m+1} \widehat{\boldsymbol{q}}_{m+1} \boldsymbol{e}_m^\top$$

After restart cycle, projection $\widehat{T}_m$ of A on new Krylov space in

$$A\widehat{Q}_m = \widehat{Q}_m\widehat{T}_m + \widehat{\beta}_{m+1}\widehat{\boldsymbol{q}}_{m+1}\boldsymbol{e}_m^\top$$

has the form

$$\widehat{T}_m = \begin{bmatrix} \widehat{T}_k & \widehat{\beta}_m \boldsymbol{s} & & & \\ \widehat{\beta}_m \boldsymbol{s}^\top & \widehat{\alpha}_{k+1} & \widehat{\beta}_{k+1} & & \\ & \widehat{\beta}_{k+1} & \ddots & \ddots & \\ & & \ddots & \ddots & \widehat{\beta}_m \\ & & & \widehat{\beta}_m & \widehat{\alpha}_m \end{bmatrix}.$$

Note: Leading $k \times k$ block is diagonal.

Remarks:

- Mathematically equivalent to implicitly restarted Lanczos method and other augmented Krylov techniques, but more efficient.
- Takes advantage of symmetry (ARPACK uses full recurrences).
- Projected matrix $\widehat{T}_k$ readily available ($= \mathrm{diag}(\vartheta_1, \ldots, \vartheta_k)$).
- Eigenvector residual norms from coordinate calculations (like in standard symmetric Lanczos).
- Well-known reorthogonalization techniques can be incorporated.
- For covariance problem: no shift-invert techniques required.
- Note: Need efficient matrix-vector product.

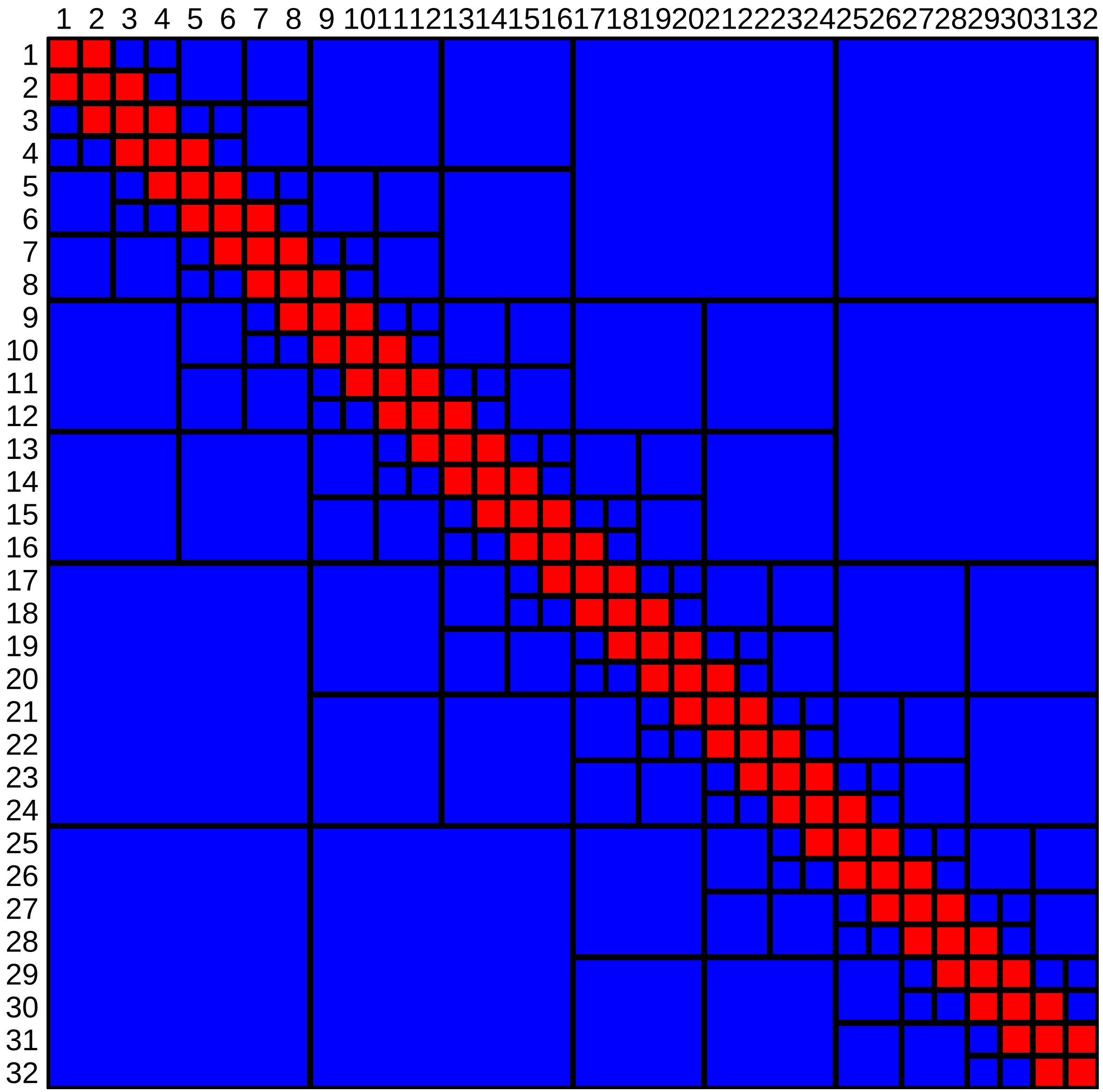
Hierarchical Matrix Approximation

Idea: (recall survey in Monday's plenary talk of W. Hackbusch)

- Partition dense matrix into square blocks of 2 types
  - near field blocks: computed and stored as usual
  - far field blocks: approximated by matrix of low rank $UV^\top$, computed by interpolation of kernel, store factors U, V.
- blocks correspond to clusters of degrees of freedom, i.e., clusters of supports of Galerkin basis functions
- block for pair of clusters s, t in near field if admissibility condition

$$\min\{\operatorname{diam}(D_s), \operatorname{diam}(D_t)\} \leq \eta \operatorname{dist}(D_s, D_t)$$

satisfied by associated domains, η is the admissibility parameter.


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

Remarks:

- “Algebraic variant” of fast multipole method
- Admissibility parameter η scales with correlation length.
- Necessary smoothness requirements satisfied for all common covariance kernels.
- Resulting data-sparse representation of discretized integral operator can be applied to a vector in $O(N \log N)$ operations (for N DOF).
- Need efficient quadrature for near field.

An optimal approximation must thus balance the errors due to

- truncation of the KL series,
- Galerkin error in approximation $a_m^h \approx a_m$, $\lambda_m^h \approx \lambda_m$
- Lanczos approximation of discrete eigenpairs
- hierarchical matrix approximation $\widetilde{\boldsymbol{C}} \approx \boldsymbol{C}$

Numerical Example

Bessel covariance kernel

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

Discretization: piecewise constant functions w.r.t. triangular mesh on D

Hierarchical matrix parameters:

interpolation polynomial degree : 4

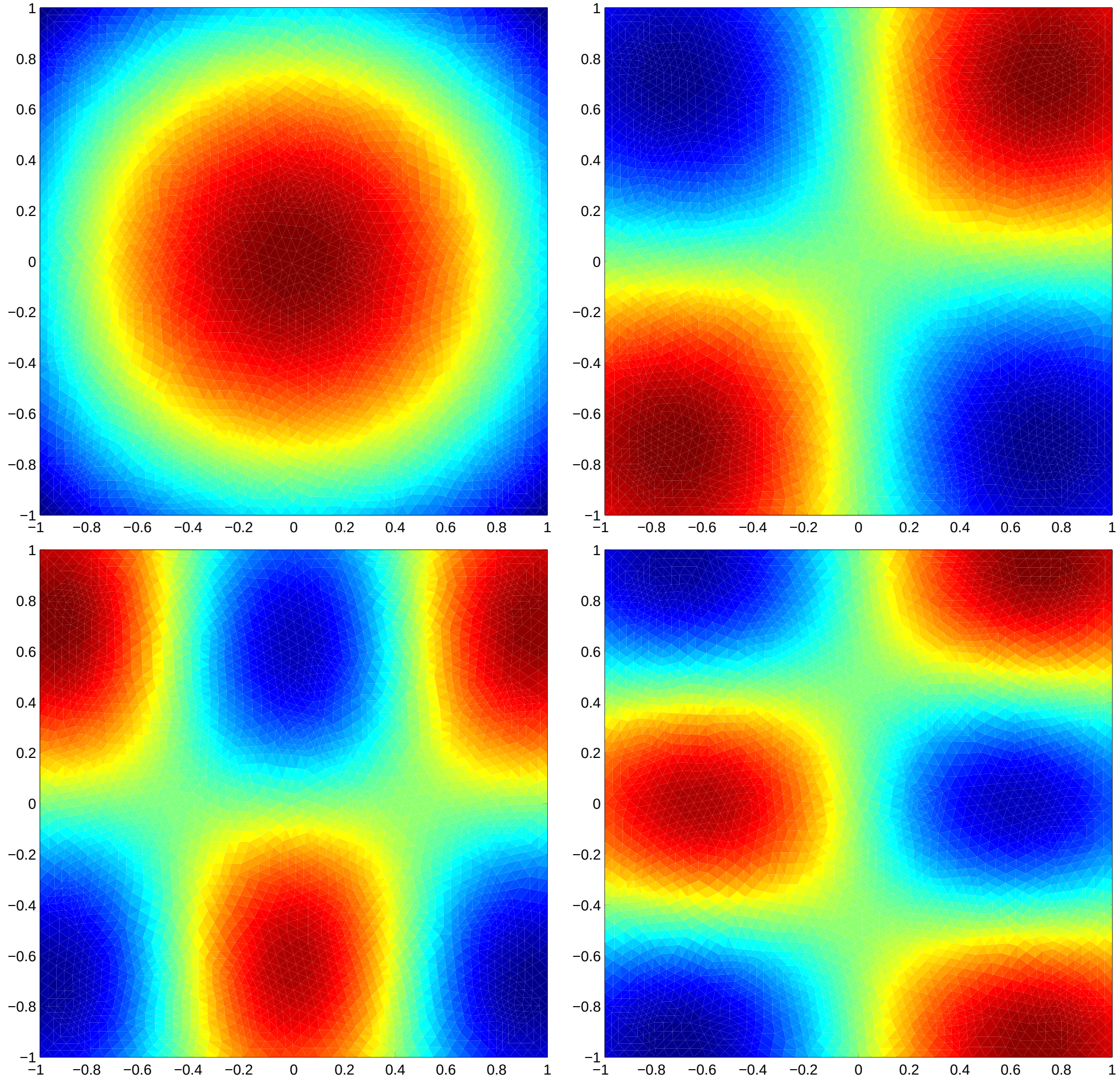
admissibility constant : $\eta = 1/\ell$

minimal block size : 62

Computations: MATLAB R2007a, Intel Xeon 5160, 3 GHz, 16 GB RAM

calls to HLib-1.3 library (MPI Leipzig) via MEX

Some modes ($\ell = 0.5$)



mode 1

mode 4

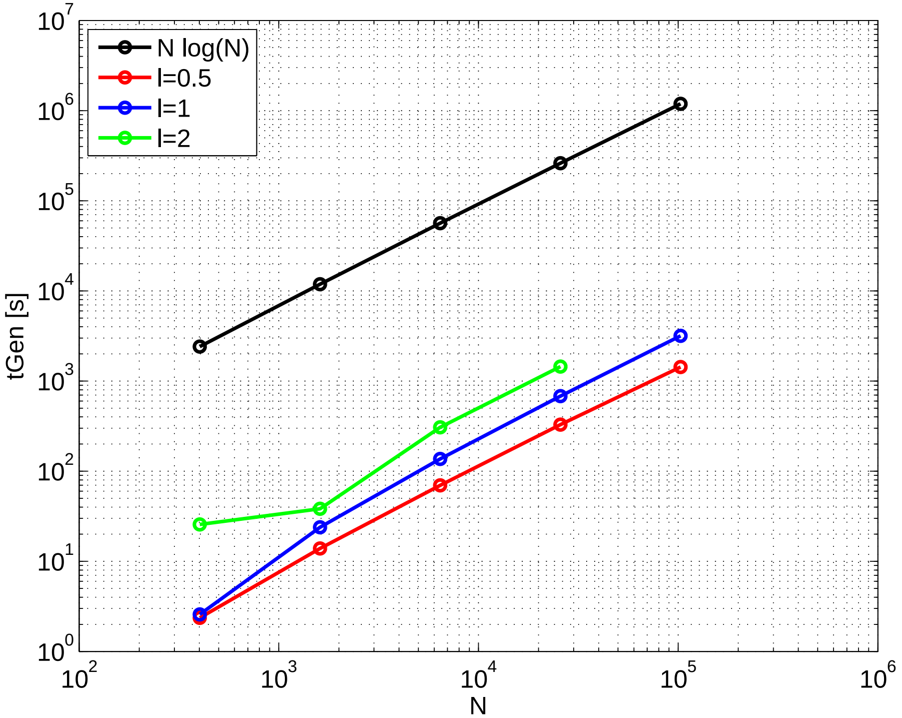
mode 7

mode 8

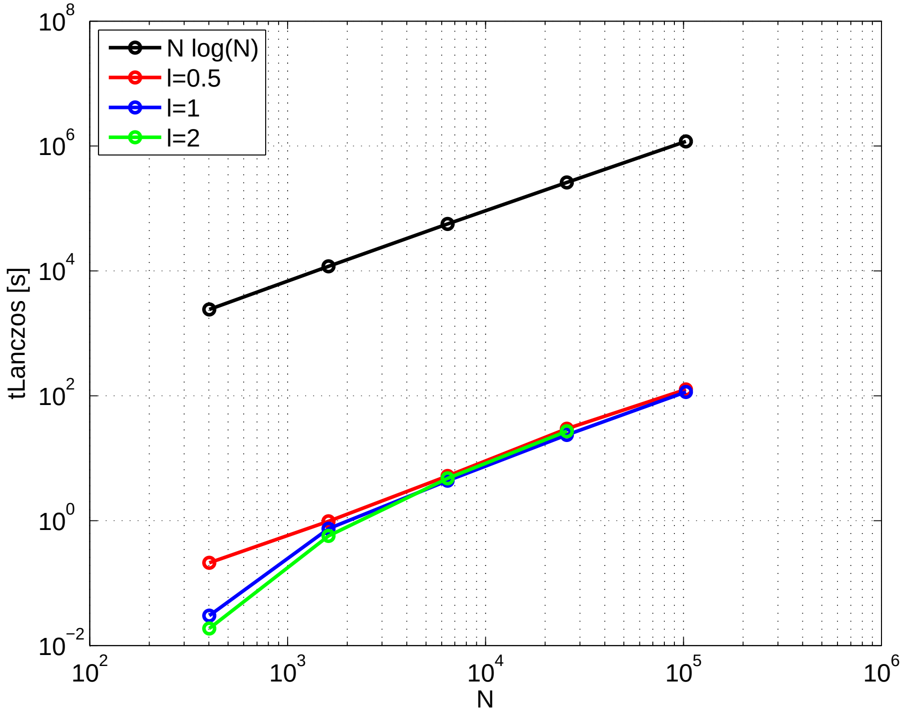
Performance of TRL

	N	# evs	% variance	m	restarts
$\ell = 0.5$	402	36	94.99	44	5
	1608	36	95.66	44	6
	6432	36	95.87	44	5
	25728	36	95.88	44	5
	102912	36	95.51	44	5
$\ell = 1$	402	10	95.30	14	8
	1608	10	95.46	14	8
	6432	10	95.50	14	8
	25728	10	95.51	14	9
	102912	10	95.51	14	9
$\ell = 2$	402	4	95.30	7	8
	1608	4	96.06	7	7
	6432	4	96.10	7	7
	25728	4	96.10	7	7
	102912	4	96.11	7	7

Timings



generation of hierarchical matrix approximation



eigenvalue calculation

Conclusions

- Covariance eigenvalue problem challenging due to its size
- Can exploit regularity of covariance kernels
- Lanczos combined with hierarchical matrix approximation promising
- Becomes intractable for very small correlation lengths (too many relevant modes)

Ongoing Work

- more careful tuning of hierarchical matrix approximation parameters
- multiple eigenvalues (symmetries in the domain)
- extend optimal quadrature techniques to 3D
- higher order FE approximation