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 \((\Omega, \mathcal{A}, P)\) 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. \((\Omega, \mathcal{A}, P)\).
### Notation

\[
\langle \xi \rangle := \int_{\Omega} \xi(\omega) \, dP(\omega)
\]

expected value of RV \( \xi : \Omega \rightarrow \mathbb{R} \)

\[
\bar{a}(x) := \langle a(x, \cdot) \rangle
\]

mean of RF \( a \) at \( x \in D \)

\[
\text{Cov}_a(x, y) := \langle (a(x, \cdot) - \bar{a}(x))(a(y, \cdot) - \bar{a}(y)) \rangle
\]

covariance of RF \( a \) at \( x, y \in D \)

\[
\text{Var}_a(x) := \text{Cov}_a(x, x)
\]

variance of RF \( a \) at \( x \in D \)

\[
\sigma_a(x) := \sqrt{\text{Var}_a(x)}
\]

standard deviation of RF \( a \) at \( x \in D \)

\[
L_P^2(\Omega) := \{ \xi : \langle \xi^2 \rangle < \infty \}
\]

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

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

\[
C : L^2(D) \to L^2(D), \quad (Cu)(x) = \int_D u(y) c(x, y) \, dy,
\]

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

\[
\langle \xi_m \rangle = 0 \forall m, \quad \langle \xi_m \xi_n \rangle = \delta_{m,n} \forall m, n
\]

such that the **Karhunen-Loève (KL) expansion**

\[
a(x, \omega) = \bar{a}(x) + \sum_{m=1}^{\infty} \sqrt{\lambda_m} a_m(x) \xi_m(\omega)
\]

(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

$$A : L^2(D) \to L^2_P, \quad f(x) \mapsto (Af)(\omega) := \int_D f(x)a(x, \omega) \, dx,$$

$$A^* : L^2_P \to L^2(D), \quad \xi(\omega) \mapsto (A^*\xi)(x) = \int_\Omega \xi(\omega)a(x, \omega) \, dP(\omega)$$

$$C = A^* A.$$
Common Covariance Models

\[ \text{Cov}_a(x, y) = c(x, y) = c(\rho), \quad \rho = \|x - y\| \]

**exponential**
\[ c(r) = \sigma^2 e^{-\rho/\ell} \]

**Bessel**
\[ c(r) = \sigma^2 \frac{\rho}{\ell} K_1 \left( \frac{\rho}{\ell} \right) \]

**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 $a_m(x)$,

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

$$\int_D \text{Var}_a(x) \, dx = \sum_{m=1}^{\infty} \lambda_m \langle a_m, a_m \rangle_D = \text{trace } C.$$  

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

$$\text{Var}_a \equiv \sigma^2 > 0, \quad \sum_m \lambda_m = |D| \sigma^2.$$
Truncated KL Expansion

For computational purposes, KL expansion truncated after $M$ terms:

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

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

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

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

\[ \mathcal{V}^h = \{ \phi : D \rightarrow \mathbb{R} : \phi|_T \in \mathcal{P}_k \ \forall T \in \mathcal{T} \} \subset L^2(D). \]

- Discrete eigenvalue problem: find pairs $(\lambda^h_m, a^h_m)$ such that

\[ (Ca^h_m, \phi) = \lambda^h_m (a^h_m, \phi) \quad \forall \phi \in \mathcal{V}^h, \quad m = 1, 2, \ldots \]

 corresponds to generalized matrix eigenvalue problem

\[ Cx = \lambda Mx, \quad [C]_{i,j} = (C\phi_j, \phi_i), [M]_{i,j} = (\phi_j, \phi_i), \]

\[ i, j = 1, 2, \ldots, N = \dim \mathcal{V}^h. \]

- $C$ large and dense, $M$ 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 \( \mathcal{V}^h \).

Discrete eigenpairs \( \{ (\lambda^h_m, a^h_m) \}_{m=1}^N \)

If covariance operator is piecewise smooth, then for any \( r > 0 \)

\[
0 \leq \lambda_m - \lambda^h_m \leq K_r \left( h^{2(k+1)} \lambda_{m-r}^1 + h^{4(k+1)} \lambda_{m-2r}^1 \right),
\]

\[
\| (I - P_h) a_m \|_{L^2(D)} \leq K_r \lambda_{m-r}^k h^{k+1}.
\]

[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 $\tilde{C}$, 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 \( \mathcal{K}_m(A, v) \)

\[
AQ_m = Q_m T_m + \beta_{m+1} q_{m+1} e_m^\top, \quad Q_m = [q_1, \ldots, q_m], \quad Q_m^\top Q_m = I_m,
\]

(2) compute eigenpairs \( T_m y_j = \vartheta_j y_j, j = 1, \ldots, m \),

(3) select \( k < m \) Ritz vectors to retain, \( Y_k := [y_1, \ldots, y_k] \),

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

\[
A \hat{Q}_k = \hat{Q}_k \hat{T}_k + \beta_{m+1} \hat{q}_{k+1} s^\top \quad \text{with} \quad \hat{q}_{k+1} = q_{m+1} \quad \text{and} \quad s := Y_k^\top e_m,
\]

(5) extend \( \text{span}\{\hat{q}_1, \ldots, \hat{q}_{m+1}\} \) to Krylov space of order \( m \) with Lanczos-type decomposition

\[
A \hat{Q}_m = \hat{Q}_m \hat{T}_m + \beta_{m+1} \hat{q}_{m+1} e_m^\top
\]
After restart cycle, projection $\hat{T}_m$ of $A$ on new Krylov space in 

$$A\hat{Q}_m = \hat{Q}_m\hat{T}_m + \hat{\beta}_{m+1}\hat{q}_{m+1}e_m^\top$$

has the form

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

**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 ($= \text{diag}(\vartheta_1, \ldots, \vartheta_k)$).
- 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
  \[
  \min\{\text{diam}(D_s), \text{diam}(D_t)\} \leq \eta \text{ dist}(D_s, D_t)
  \]
  satisfied by associated domains, $\eta$ is the admissibility parameter.
Karhunen-Loève Approximation of Random Fields Using Hierarchical Matrix Techniques Harrachov 2007
Remarks:

- “Algebraic variant” of fast multipole method
- Admissibility parameter $\eta$ 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 $a^h_m \approx a_m$, $\lambda^h_m \approx \lambda_m$
- Lanczos approximation of discrete eigenpairs
- hierarchical matrix approximation $\tilde{C} \approx C$
Numerical Example

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

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 \((\ell = 0.5)\)
## Performance of TRL

<table>
<thead>
<tr>
<th>$\ell$</th>
<th>$N$</th>
<th># evs</th>
<th>% variance</th>
<th>$m$</th>
<th>restarts</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>402</td>
<td>36</td>
<td>94.99</td>
<td>44</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>1608</td>
<td>36</td>
<td>95.66</td>
<td>44</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>6432</td>
<td>36</td>
<td>95.87</td>
<td>44</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>25728</td>
<td>36</td>
<td>95.88</td>
<td>44</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>102912</td>
<td>36</td>
<td>95.51</td>
<td>44</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>402</td>
<td>10</td>
<td>95.30</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>1608</td>
<td>10</td>
<td>95.46</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>6432</td>
<td>10</td>
<td>95.50</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>25728</td>
<td>10</td>
<td>95.51</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>102912</td>
<td>10</td>
<td>95.51</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>402</td>
<td>4</td>
<td>95.30</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>1608</td>
<td>4</td>
<td>96.06</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>6432</td>
<td>4</td>
<td>96.10</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>25728</td>
<td>4</td>
<td>96.10</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>102912</td>
<td>4</td>
<td>96.11</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>
Timings

generation of hierarchical matrix approximation

eigenvalue calculation

Karhunen-Loève Approximation of Random Fields Using Hierarchical Matrix Techniques
Harrachov 2007
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