Expansion of random field gradients using hierarchical matrices

Ingolf Busch[1], **Oliver G. Ernst**[1,*], and **Elisabeth Ullmann**[1]

[1] Institut für Numerische Mathematik und Optimierung, Technische Universität Bergakademie Freiberg, D-09596 Freiberg, Germany

We present two expansions for the gradient of a random field. In the first approach, we differentiate its truncated Karhunen-Loève expansion. In the second approach, the Karhunen-Loève expansion of the random field gradient is computed directly. Both strategies require the solution of dense, symmetric matrix eigenvalue problems which can be handled efficiently by combining hierachical matrix techniques with a thick-restart Lanczos method.

1 Introduction

Many techniques in stochastic uncertainty quantification, e.g., stochastic Galerkin finite element methods rely heavily on computations with random fields $a : D \times \Omega \to \mathbb{R}$, where $(\Omega, \mathfrak{A}, P)$ is a probability space and $D \subset \mathbb{R}^d$ is a bounded spatial domain. Practical representations of random fields separate the stochastic and deterministic dependence in the form

$$a(\boldsymbol{x}, \omega) = \sum_m \varphi_m(\boldsymbol{x}) \xi_m(\omega) \tag{1}$$

with deterministic functions φ_m and random variables ξ_m. For random fields with finite variance, the most popular option is the Karhunen-Loève (KL) expansion [1]. Here, we assume that a is a random field with mean value $\bar{a}(\boldsymbol{x}) := \langle a(\boldsymbol{x}, \cdot)\rangle$ and covariance function $c(\boldsymbol{x}, \boldsymbol{y}) = \langle (a(\boldsymbol{x}, \cdot) - \bar{a}(\boldsymbol{x}))(a(\boldsymbol{y}, \cdot) - \bar{a}(\boldsymbol{y}))\rangle$, $\boldsymbol{x}, \boldsymbol{y} \in D$, where $\langle \cdot \rangle$ denotes the expectation with respect to the probability measure P. The KL expansion of a is

$$a(\boldsymbol{x}, \omega) = \bar{a}(\boldsymbol{x}) + \sum_{m=1}^{\infty} \sqrt{\lambda_m} a_m(\boldsymbol{x}) \xi_m(\omega), \tag{2}$$

with uncorrelated, centered random variables $\{\xi_m\}_{m=1}^{\infty}$ and eigenpairs $(\lambda_m, a_m)_{m=1}^{\infty}$ of the self-adjoint, compact integral operator

$$C : L^2(D) \to L^2(D) \qquad (Cu)(\boldsymbol{x}) = \int_D c(\boldsymbol{x}, \boldsymbol{y}) u(\boldsymbol{y}) \, d\boldsymbol{y}. \tag{3}$$

In actual computations, the sum in (2) is truncated after the M leading terms, yielding an approximation of a of the form

$$a^{(M)}(\boldsymbol{x}, \omega) = \bar{a}(\boldsymbol{x}) + \sum_{m=1}^{M} \sqrt{\lambda_m} a_m(\boldsymbol{x}) \xi_m(\omega). \tag{4}$$

Galerkin discretization of the KL eigenproblem with piecewise constant shape functions results in a generalized matrix eigenvalue problem $\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{B}\boldsymbol{x}$, where $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ can be chosen diagonal but $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is a dense, symmetric positive definite matrix. An effective way to solve this matrix eigenvalue problem is outlined in [2] where hierarchical matrix techniques [3,4] are combined with a thick-restart Lanczos method [5]; this approach costs $O(n \log n)$ operations.

Assume that a is continuously differentiable in the mean-square (m.-s.) sense[2] (see, e.g., [6, Chapter 2]). We are interested in a representation of the gradient random field ∇a of the form (1) with deterministic vector-valued functions $\boldsymbol{\varphi}_m : D \to \mathbb{R}^d$. Here we present two expansion strategies for ∇a. In the first approach, we truncate the KL expansion of a and then compute the gradient $\nabla a^{(M)}$ which requires the gradient of KL eigenfunctions. In the second approach, the KL expansion of ∇a is computed directly. We build upon the work in [2] and achieve a total cost of $O(n \log n)$ operations for both strategies.

2 Differentiation of Karhunen-Loève eigenfunctions

Partial derivatives of a KL eigenfunction a_m in (4) (where we may assume that $\lambda_m > 0$) can be computed using the integral eigenproblem equation

$$\int_D c(\boldsymbol{x}, \boldsymbol{y}) a_m(\boldsymbol{y}) \, d\boldsymbol{y} = \lambda_m a_m(\boldsymbol{x}) \tag{5}$$

*Corresponding author: Email ernst@math.tu-freiberg.de, phone +00 49 3731 39-3246, fax +00 49 3731 39-3595

[2] Necessary and sufficient for this is the existence and continuity of $\nabla \bar{a}$ and the partial derivatives $(\partial^2 / \partial x_j \partial y_k) c(\boldsymbol{x}, \boldsymbol{y})$, $\boldsymbol{x}, \boldsymbol{y} \in D$, $j, k = 1, \ldots, d$.

derived from (3). Since a is assumed m.-s. continuously differentiable[2], the requirements of Lebesgue's Dominated Convergence Theorem are satisfied (see, for example, [7, Chapter 2]). It follows that we can interchange the order of differentiation and integration in (5) and obtain

$$\frac{\partial a_m}{\partial x_i}(\boldsymbol{x}) = \lambda_m^{-1} \int_D \frac{\partial c}{\partial x_i}(\boldsymbol{x}, \boldsymbol{y}) a_m(\boldsymbol{y})\, d\boldsymbol{y}, \qquad i = 1, \ldots, d. \tag{6}$$

In stochastic finite element methods the assembly of the stiffness matrix requires the evaluation of (6) at certain quadrature nodes $\{\boldsymbol{x}_\ell\} \in D$. We discretize the action of the integral operator on the right-hand side of (6) by a Galerkin projection onto the subspace Z_h of piecewise constant functions on a given triangulation $\mathscr{T}_h$ of D. Let $Z_h = \operatorname{span}\{\phi_1(\boldsymbol{x}), \ldots, \phi_n(\boldsymbol{x})\} \subset L^2(D)$. With $\partial a_m(\boldsymbol{x})/\partial x_i \approx \sum_j a_{m,j}^{(i)} \phi_j(\boldsymbol{x})$, $i = 1, \ldots, d$, and $a_m(\boldsymbol{x}) \approx \sum_j a_{m,j}^{(0)} \phi_j(\boldsymbol{x})$, it follows that the n coefficients that determine the partial derivative of an (approximate) eigenfunction satisfy the Galerkin equations

$$\sum_{j=1}^{n} a_{m,j}^{(i)} \int_D \phi_j(\boldsymbol{x}) \phi_k(\boldsymbol{x})\, d\boldsymbol{x} = \lambda_m^{-1} \sum_{j=1}^{n} a_{m,j}^{(0)} \int_D \int_D \frac{\partial c}{\partial x_i}(\boldsymbol{x}, \boldsymbol{y}) \phi_j(\boldsymbol{y}) \phi_k(\boldsymbol{x})\, d\boldsymbol{y}\, d\boldsymbol{x}.$$

for $k = 1, \ldots, n$ and $i = 1, \ldots, d$. The matrix formulation of these equations reads

$$\boldsymbol{Q}\, \boldsymbol{a}_m^{(i)} = \lambda_m^{-1} \boldsymbol{K}_i\, \boldsymbol{a}_m^{(0)}, \tag{7}$$

where $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$ denotes the Gramian matrix of $\{\phi_1, \ldots, \phi_n\}$ with respect to the $L^2(D)$ inner product, $\boldsymbol{K}_i \in \mathbb{R}^{n \times n}$ is a discrete integral operator with kernel function $\partial c/\partial x_i$, the vector $\boldsymbol{a}_m^{(0)} \in \mathbb{R}^n$ contains the coefficients of the KL eigenfunction a_m, and the vectors $\boldsymbol{a}_m^{(i)} \in \mathbb{R}^n$ contain the coefficients of $\partial a_m/\partial x_i$, $i = 1, \ldots, d$, respectively. The action of $\boldsymbol{Q}^{-1}$ can be computed in $O(n)$ operations since $\boldsymbol{Q}$ is a diagonal matrix in this setting. However, each matrix $\boldsymbol{K}_i$ in (7) is in general a dense, non-symmetric $n \times n$ matrix. Therefore we approximate $\boldsymbol{K}_i$ by a hierarchical $\mathscr{H}^2$-matrix (see [3, 4] and references therein), which allows us to perform matrix-vector products with $\boldsymbol{K}_i$ in $O(n)$ operations. The assembly of an $\mathscr{H}^2$-matrix costs $O(n \log n)$ operations. The vector $\boldsymbol{a}_m^{(0)}$ solves a discretized version of the KL integral eigenproblem (5) and can be computed with $O(n \log n)$ complexity as outlined in [2]. In summary, then, approximations to each component of the gradient ∇a_m of KL eigenfunctions can be obtained with $O(n \log n)$ complexity; thus computing $\nabla a^{(M)}$ costs $O(n \log n)$ operations.

3 Karhunen-Loève expansion of ∇a

Since a is m.-s. continuously differentiable[2] by assumption, it follows that the KL expansion of ∇a can be computed directly (cf. [8, Chapter 10]). By analogy with (2) we obtain

$$\nabla a(\boldsymbol{x}) = \nabla \bar{a}(\boldsymbol{x}) + \sum_{m=1}^{\infty} \sqrt{\gamma_m}\, \boldsymbol{g}_m(\boldsymbol{x}) \xi_m(\omega),$$

with uncorrelated, centered random variables $\{\xi_m\}_{m=1}^{\infty}$ and eigenpairs $(\gamma_m, \boldsymbol{g}_m)_{m=1}^{\infty}$ of the integral operator

$$C_\nabla : [L^2(D)]^d \to [L^2(D)]^d \qquad (C_\nabla \boldsymbol{u})(\boldsymbol{x}) = \int_D c_\nabla(\boldsymbol{x}, \boldsymbol{y}) \boldsymbol{u}(\boldsymbol{y})\, d\boldsymbol{y}. \tag{8}$$

Above, the kernel function is the matrix-valued covariance function of ∇a which reads (see, e.g., [6, Chapter 2])

$$c_\nabla : D \times D \to \mathbb{R}^{d \times d} \qquad [c_\nabla]_{j,k}(\boldsymbol{x}, \boldsymbol{y}) = (\partial^2/\partial x_j \partial y_k) c(\boldsymbol{x}, \boldsymbol{y}), \quad j, k = 1, \ldots, d,$$

where c is the covariance function of a. For example, if $\bar{a} = \langle a \rangle$ is differentiable and a has a Gaussian covariance function

$$c(\boldsymbol{x}, \boldsymbol{y}) = \exp(-(r/\rho)^2), \quad r = \|\boldsymbol{x} - \boldsymbol{y}\|_2, \quad \boldsymbol{x}, \boldsymbol{y} \in D, \quad \rho > 0, \tag{9}$$

then the gradient ∇a is well-defined with mean value $\langle \nabla a \rangle = \nabla \bar{a}$ and covariance function

$$c_\nabla(\boldsymbol{x}, \boldsymbol{y}) = (2/\rho^2) \exp(-(r/\rho)^2) \boldsymbol{I}_{d \times d} - (4/\rho^4) \exp(-(r/\rho)^2)(\boldsymbol{x} - \boldsymbol{y})(\boldsymbol{x} - \boldsymbol{y})^\top, \quad r = \|\boldsymbol{x} - \boldsymbol{y}\|_2, \quad \boldsymbol{x}, \boldsymbol{y} \in D. \tag{10}$$

Galerkin discretization of the vector KL eigenproblem derived from (8) results in a generalized matrix eigenvalue problem $\boldsymbol{C}\boldsymbol{x} = \lambda \boldsymbol{M}\boldsymbol{x}$ with $d \times d$ block matrices $\boldsymbol{C}, \boldsymbol{M} \in \mathbb{R}^{nd \times nd}$. The matrix $\boldsymbol{M}$ is block-diagonal and if piecewise constant Galerkin shape functions are used, then $\boldsymbol{M}$ can be chosen diagonal, but $\boldsymbol{C}$ is a dense, yet symmetric positive definite matrix. The diagonal blocks of $\boldsymbol{C}$ correspond to the covariance function of each component of ∇a, and the off-diagonal blocks of $\boldsymbol{C}$ are associated with the respective cross-covariances of the components of ∇a. We approximate each block of $\boldsymbol{C}$ by a hierarchical $\mathscr{H}^2$-matrix [3, 4]. Since the assembly of an $\mathscr{H}^2$-matrix costs $O(n \log n)$ operations, the block matrix $\boldsymbol{C}$ can be assembled in $O(d^2\, n \log n)$ operations. We use the thick-restart Lanczos method from [5] to compute leading eigenpairs satisfying $\boldsymbol{C}\boldsymbol{x} = \lambda \boldsymbol{M}\boldsymbol{x}$; this requires only matrix-vector multiplications with $\boldsymbol{C}$ which cost $O(d\, n)$ operations and the action of $\boldsymbol{M}^{-1}$ which can be computed with $O(d\, n)$ operations in this setting. In summary, then, since d is fixed for a given problem, approximate KL eigenpairs of ∇a can be obtained with $O(n \log n)$ complexity.

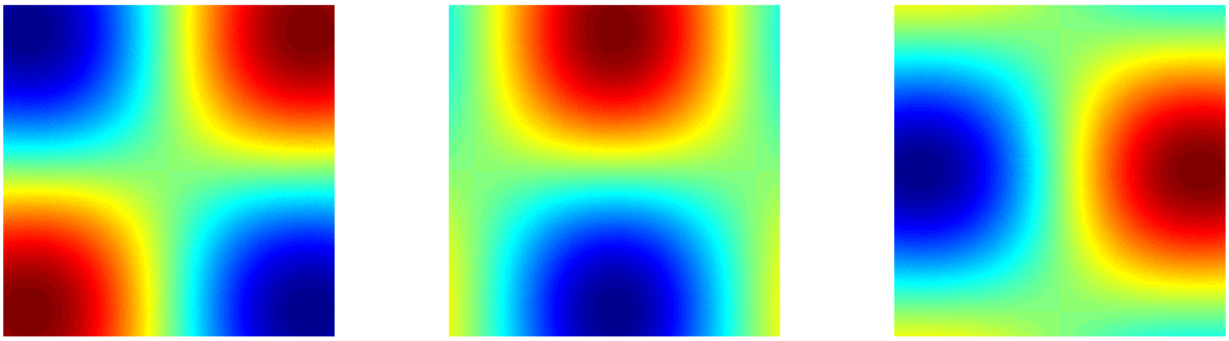

Fig. 1 KL eigenfunction a_4 (left), $\partial a_4/\partial x$ (middle), and $\partial a_4/\partial y$ (right) computed on a mesh containing 47,104 triangles. In this example, $D = [-1,1] \times [-1,1]$ and a has Gaussian covariance (9) with $\rho = 1$.

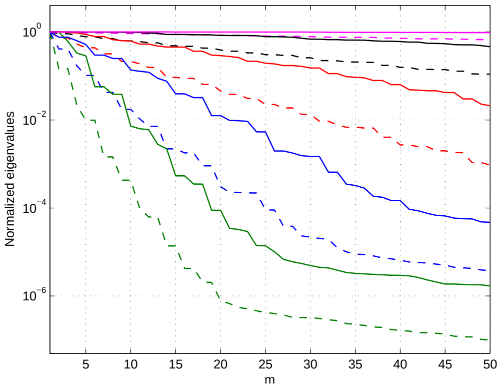

Fig. 2 Normalized KL eigenvalues λ_m/λ_1 of a (dashed line) and γ_m/γ_1 of ∇a (solid line) for a centered random field a with Gaussian covariance (9) on $D = [-1,1] \times [-1,1]$. In this example, $\rho = 2$ (green), $\rho = 1$ (blue), $\rho = 0.5$ (red), $\rho = 0.25$ (black), $\rho = 0.1$ (magenta). The KL eigenvalues have been computed on a mesh containing 188,416 triangles.

4 Numerical experiments

We consider a random field a with zero mean and Gaussian covariance function (9) on $D = [-1,1] \times [-1,1]$, hence ∇a is well-defined. The hierarchical matrix computations are carried out with software based upon the HLib package [9].

KL eigenfunctions and eigenvalues As proof of concept we plot the fourth leading eigenfunction in the KL expansion of a together with the two components of its gradient in Figure 1.

The smoothness of the covariance kernel determines the asymptotic KL eigenvalue decay rate (see, e.g., [10]). Since the Gaussian covariance function c in (9) and the covariance function c_∇ in (10) have the same smoothness we expect a similar asymptotic eigenvalue decay for both a and ∇a. This observation is consistent with Figure 2 where we plot the normalized KL eigenvalues of a and ∇a for various correlation lengths ρ.

Computational costs We compute the 50 leading terms of the gradient of the KL expansion of a (see Section 2) and the KL expansion of ∇a (see Section 3), respectively, on a sequence of uniformly refined triangular grids, resulting in $n = 104$ to $n = 753{,}664$. We use a Krylov space of dimension 100 in the thick-restart Lanczos method. The total computing times reported in Figure 3 agree well with our complexity estimates from Sections 2 and 3, that is we achieve $O(n \log n)$ complexity for both expansion strategies.

Convergence of expansions Truncation of the sum in the KL expansion of ∇a after the M leading terms yields an approximation $\nabla^{(M)} a$ to the random field gradient ∇a. We assess the quality of the expansions $\nabla a^{(M)}$ and $\nabla^{(M)} a$ by computing the quantities $\langle \|\nabla a^{(M)}\|^2 \rangle = \sum_{m=1}^{M} \lambda_m \|\nabla a_m\|^2$ and $\langle \|\nabla^{(M)} a\|^2 \rangle = \sum_{m=1}^{M} \gamma_m$, and comparing these with the total variance of ∇a, $\langle \|\nabla a\|^2 \rangle = \sum_{i=1}^{d} \int_D [c_\nabla]_{i,i}(\boldsymbol{x},\boldsymbol{x})\, d\boldsymbol{x} = \sum_{i=1}^{d} (2/\rho^2)|D| = (2/\rho^2)\, d\, |D|$; $\|\cdot\|$ denotes the canonical norm on $[L^2(D)]^d$.

If a is m.-s. continuously differentiable[2], then $\nabla a^{(M)}$ converges to ∇a in the m.-s. sense uniformly on D for $M \to \infty$.[3] Thus, for sufficiently smooth covariance functions $\nabla a^{(M)}$ converges to ∇a in the same sense as the truncated KL expansion $\nabla^{(M)} a$ does (see, e.g., [8, Chapter 10]). In particular, this holds for random fields with Gaussian covariance function (9).

[3] This result is proved in [11, Theorem A.2] for random functions $a : D \times \Omega \to \mathbb{R}$ defined on $D \subset \mathbb{R}$ but it carries over to random fields defined on $D \subset \mathbb{R}^d$ with $d > 1$. (The differentiability of a almost surely is not needed.)

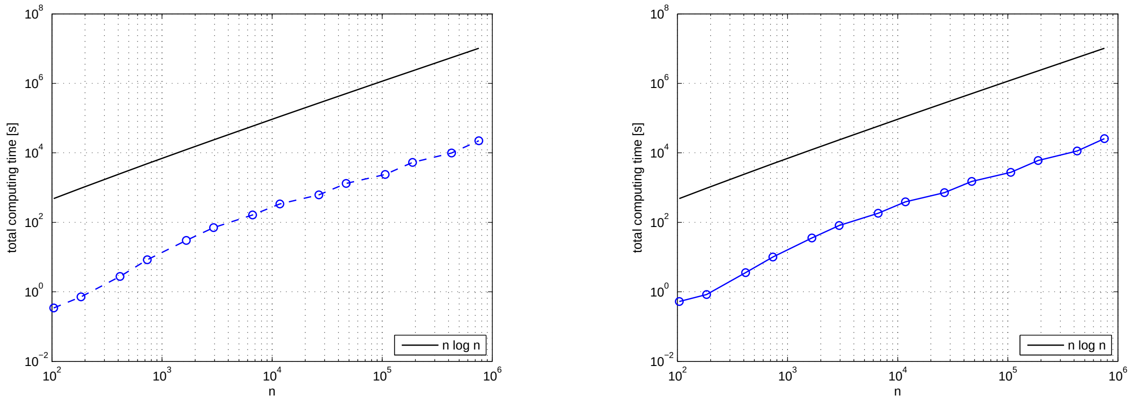

Fig. 3 Total time (in seconds) for computing the 50 leading terms of the gradient of the KL expansion of a (left) and the KL expansion of ∇a (right) for a centered random field a with Gaussian covariance (9) on $D = [-1, 1] \times [-1, 1]$. In this example, $\rho = 1$.

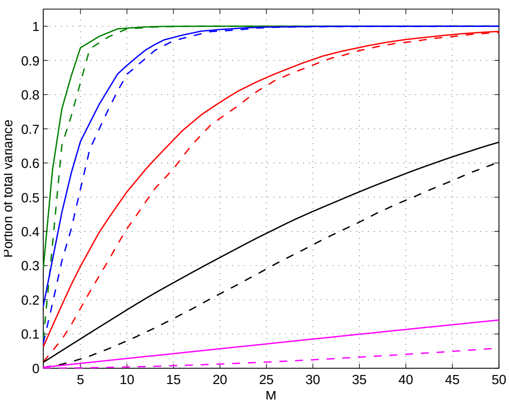

Fig. 4 Portion of total variance $\langle \|\nabla a\|^2 \rangle$ captured by $\langle \|\nabla a^{(M)}\|^2 \rangle$ (dashed line) and $\langle \|\nabla^{(M)} a\|^2 \rangle$ (solid line) for a centered random field a with Gaussian covariance (9) on $D = [-1, 1] \times [-1, 1]$. In this example, $\rho = 2$ (green), $\rho = 1$ (blue), $\rho = 0.5$ (red), $\rho = 0.25$ (black), $\rho = 0.1$ (magenta). The expansions $\nabla a^{(M)}$ and $\nabla^{(M)} a$ have been computed on a mesh containing 26,624 triangles.

The expansion $\nabla a^{(M)}$ is suboptimal, since $\nabla^{(M)} a$ is the best M-term approximation to ∇a with respect to the canonical norm on $[L^2(D)]^d \otimes L^2_P(\Omega)$. This observation is consistent with Figure 4, where we see that the portion of total variance captured by the best approximation $\nabla^{(M)} a$ is always larger than the portion captured by $\nabla a^{(M)}$.

5 Conclusions

We have presented expansions for the gradient of a random field obtained by differentiation of its KL expansion and by computing the KL expansion of the gradient directly. These strategies cost $O(n \log n)$ operations, where n is the number of elements in a triangulation of the spatial domain. For sufficiently smooth covariance functions both expansions converge to the random field gradient; in this case, use of the direct KL expansion of the gradient is clever since it is a best approximation.

Acknowledgements The second and third author gratefully acknowledge the support by grant ER 275/5-1 within the German Research Foundation (DFG) Priority Programme 1324.

References

[1] M. Loève, Probability Theory, 4th edition (Springer-Verlag, New York - Heidelberg - Berlin, 1978).
[2] M. Eiermann, O. G. Ernst, and E. Ullmann, Comput. Vis. Sci. **10**(1), 3–15 (2007).
[3] S. Börm, L. Grasedyck, and W. Hackbusch, Eng. Anal. Bound. Elem. **27**(5), 405–422 (2003).
[4] W. Hackbusch, Hierarchische Matrizen. Algorithmen und Analysis (Springer-Verlag, Berlin Heidelberg, 2009).
[5] K. Wu and H. Simon, SIAM J. Matrix Anal. Appl. **22**(2), 602–616 (2000).
[6] G. Christakos, Random Field Models in Earth Sciences (Dover, Mineola N.Y., 2005).
[7] A. N. Shiryaev, Probability, 2nd edition (Springer-Verlag, New York, 1996).
[8] V. S. Pugachev, Theory of Random Functions and its Application in Control Problems, Int. Series of Monographs on Automation and Automatic Control, Vol. 5 (Pergamon Press, Oxf.-Lond.-Edinb.-N.Y.-Paris-Frankfurt, 1965).
[9] S. Börm and L. Grasedyck, Hierachical Matrices Library (HLib) version 1.3, `http://www.hlib.org/`.
[10] C. Schwab and R. A. Todor, J. Comput. Phys. **217**, 100–122 (2006).
[11] T. T. Kadota, IEEE Trans. Inf. Theory **13**(2), 255–260 (1967).