The Stochastic Galerkin Method for Darcy Flow Problem with Log-Normal Random Field Coefficients

Michal BERES[1,2,3], *Simona DOMESOVA*[1,2,3]

[1]Department of Applied Mathematics, Faculty of Electrical Engineering and Computer Science, VSB–Technical University of Ostrava, 17. listopadu 15/2172, 708 33 Ostrava, Czech Republic
[2]IT4Innovations National Supercomputing Center, VSB–Technical University of Ostrava, 17. listopadu 15/2172, 708 33 Ostrava, Czech Republic
[3]Institute of Geonics of the Czech Academy of Sciences, Studentska 1768, 708 00 Ostrava, Czech Republic

michal.beres@vsb.cz, simona.domesova@vsb.cz

DOI: 10.15598/aeee.v15i2.2280

Abstract. *This article presents a study of the Stochastic Galerkin Method (SGM) applied to the Darcy flow problem with a log-normally distributed random material field given by a mean value and an autocovariance function. We divide the solution of the problem into two parts. The first one is the decomposition of a random field into a sum of products of a random vector and a function of spatial coordinates; this can be achieved using the Karhunen-Loeve expansion. The second part is the solution of the problem using SGM. SGM is a simple extension of the Galerkin method in which the random variables represent additional problem dimensions. For the discretization of the problem, we use a finite element basis for spatial variables and a polynomial chaos discretization for random variables. The results of SGM can be utilised for the analysis of the problem, such as the examination of the average flow, or as a tool for the Bayesian approach to inverse problems.*

Keywords

Darcy flow, Gaussian random field, Karhunen-Loeve decomposition, polynomial chaos, Stochastic Galerkin method.

1. Introduction and Problem Setting

The standard mathematical modelling approach usually works with deterministic parameters of solved problems, such as boundary conditions, material constants, etc. These simplifications neglect the inherent uncertainty in these parameters, which can lead to unreliable solutions. The fast development of computational resources allows computations of massive problems, e.g. solving the partial differential equations with uncertainty in parameters. Popular methods for calculation of these challenges are for example the Multilevel Monte Carlo method, stochastic collocation method (see [16]) or the stochastic Galerkin method which is presented in this article. The comparison between SGM and Monte Carlo method can be found in [17]. We focus on a model problem with a very high uncertainty in input parameters, for which we demonstrate the application of SGM.

Consider a 2D Darcy flow problem on the unit square according to Fig. 1. We assume that the pressure

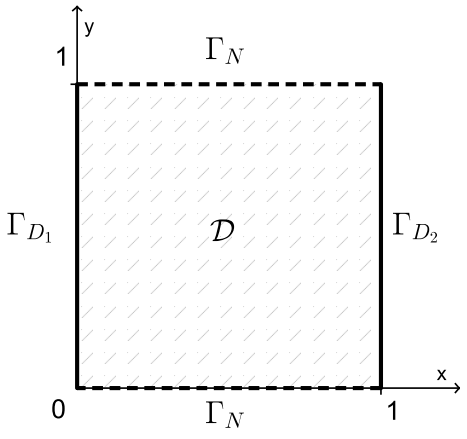


Fig. 1: Illustration of physical domain with boundary.

equals one on Γ_{D_1}, zero on Γ_{D_2}, and the flow equals zero on the rest of the boundary.

$$\begin{cases} \forall \omega \in \Omega : & \\ -\mathrm{div}\left(k\left(x;\omega\right)\nabla u\left(x;\omega\right)\right)=0, & \forall x \in \mathcal{D}, \\ u\left(x;\omega\right)=1, & \forall x \in \Gamma_{D_1}, \\ u\left(x;\omega\right)=0, & \forall x \in \Gamma_{D_2}, \\ \frac{\partial u(x;\omega)}{\partial n(x)}=0, & \forall x \in \Gamma_N, \end{cases} \tag{1}$$

where ω represents an event from the sample space Ω. In general the functions $k\left(x;\omega\right)$ and $u\left(x;\omega\right) : \mathcal{D}\times\Omega\rightarrow\mathbb{R}$ are understood as random fields. That means that the function $K_0 = k\left(x_0;\omega\right) : \Omega\rightarrow\mathbb{R}$ is a random variable for every fixed $x_0 \in \mathcal{D}$ and the function $k_0\left(x\right) = k\left(x;\omega_0\right) : \mathcal{D}\rightarrow\mathbb{R}$ is a function on the spatial domain for every fixed $\omega_0 \in \Omega$. A random field is a type of a random process, see for example [1]: chapter 5.

The problem Eq. (1) is very complex because the random field k represents an infinite number of random variables. To solve this kind of problems, we need to reduce the number of random variables to some finite, ideally small, number. We can achieve this using the Karhunen-Loeve (KL) decomposition, which will be discussed in the following section. For a more detailed insight into SGM and the KL decomposition, see [4].

We choose a log-normally distributed random material. This type of random field appears to be a reasonable choice for the description of permeability properties of natural materials like sandstone or gravel, see [6]. The log-normal random field is defined as an exponential of the Gaussian (normal) Random Field (GRF):

$$k\left(x;\omega\right)=\exp\left(g\left(x;\omega\right)\right). \tag{2}$$

The GRF is fully described by its mean value $\mu\left(x\right) = \mathbb{E}\left(g\left(x;\omega\right)\right)$ and its autocovariance function $c\left(x,y\right) = \mathrm{cov}\left(g\left(x;\omega\right),g\left(y;\omega\right)\right)$. In this paper, we consider the GRF specified by:

$$\mu\left(x\right)=0\,,c\left(x,y\right)=\sigma^2\exp\left(\frac{\left\|x-y\right\|}{\lambda}\right), \tag{3}$$

where σ is the scale parameter and λ is the correlation length parameter. The scale parameter affects the amplitude (equivalent samples with different values of σ differ only in scale) and the correlation length parameter changes the correlation/distance ratio between two points.

2. The Karhunen-Loeve Decomposition

In this section, we will discuss the approximation of the random field by a function of a random vector and a physical variable. This can be achieved using the truncated Karhunen-Loeve (KL) decomposition. The existence of the KL decomposition is stated in the following theorem, see [1]: theorem 7.52.

Theorem 1. *Let $\mathcal{D}\subset\mathbb{R}^d$. Consider a random field $\{k\left(x;\omega\right) : x\in\mathcal{D}\}$ and suppose that $k\in L^2\left(\Omega,L^2\left(\mathcal{D}\right)\right)$. Then*

$$k\left(x;\omega\right)=\mu\left(x\right)+\sum_{j=1}^{\infty}\sqrt{\lambda_j}\psi_j\left(x\right)\xi_j\left(\omega\right), \tag{4}$$

where the sum converges in $L^2\left(\Omega,L^2\left(\mathcal{D}\right)\right)$,

$$\xi_j\left(\omega\right):=\frac{1}{\sqrt{\lambda_j}}\int_{\mathcal{D}}\left(k\left(x;\omega\right)-\mu\left(x\right)\right)\psi_j\left(x\right)dx, \tag{5}$$

and $\{\lambda_j,\psi_j\}$ denotes the eigenvalues and the eigenfunctions of the autocovariance operator $\mathcal{C} : L^2\left(\mathcal{D}\right)\rightarrow L^2\left(\mathcal{D}\right)$

$$\left(\mathcal{C}f\right)\left(x\right):=\int_{\mathcal{D}}c\left(x,y\right)f\left(y\right)dy, \tag{6}$$

where $c\left(x,y\right) := cov\left(k\left(x;\omega\right),k\left(y;\omega\right)\right)$; $\lambda_1\geq\lambda_2\geq\cdots\geq 0$ and $\lim\limits_{k\rightarrow\infty}\lambda_k=0$. The random variables ξ_j have zero mean, unit variance and are pairwise uncorrelated.

In the case of the log-normal random material field, it is simpler to decompose the underlying GRF, which lies in space $L^2\left(\Omega,L^2\left(\mathcal{D}\right)\right)$, see for example [1]: corollary 4.41. Therefore its KL decomposition exists.

In the case of the KL decomposition of the GRF, the random variables ξ_i will also be Gaussian $\xi_i\sim N\left(0;1\right)$. Note that the random variables ξ_i are in general only uncorrelated, but for the Gaussian random variables, it implies independence.

The most difficult part of the KL decomposition is the spectral decomposition of the operator $\mathcal{C}$, which we can obtain by solving the eigenvalue problem

$$\int_{D}c\left(x,y\right)\psi_i\left(y\right)\mathrm{d}y=\lambda_i\psi_i\left(x\right),\forall i\in\mathbb{N}. \tag{7}$$

This is the Fredholm integral equation, which can be solved by the Galerkin method.

We can obtain the weak formulation of the Eq. (7) by multiplying it by a test function $v\in L^2\left(\mathcal{D}\right)$ and integrating over the domain $\mathcal{D}$.

$$\begin{cases} \text{Find } \psi_i\left(x\right)\in L^2\left(\mathcal{D}\right),\lambda_i\in\mathbb{R}^+ : \\ \int\limits_{\mathcal{D}}v\left(x\right)\int\limits_{\mathcal{D}}c\left(x,y\right)\psi_i\left(y\right)\mathrm{d}y\mathrm{d}x= \\ =\lambda_i\int\limits_{\mathcal{D}}v\left(x\right)\psi_i\left(x\right)\mathrm{d}x,\forall v\left(x\right)\in L^2\left(\mathcal{D}\right). \end{cases} \tag{8}$$

The next step is the discretization of the weak formulation. First consider a basis $\langle \phi_1(x), \ldots, \phi_n(x) \rangle = V_n \subset L^2(\mathcal{D})$, so the Galerkin formulation takes the form of:

$$\begin{cases} \text{Find } \psi_i(x) = \sum_{j=1}^{n} \overline{\psi_{ij}} \phi_j(x), \lambda_i \in \mathbb{R}^+ : \\ \int_{\mathcal{D}} \phi_j(x) \int_{\mathcal{D}} c(x,y) \psi_i(y) \, \mathrm{d}y \mathrm{d}x = \\ = \lambda_i \int_{\mathcal{D}} \phi_j(x) \psi_i(x) \, \mathrm{d}x, \forall \phi_j(x), \end{cases} \tag{9}$$

where $\overline{\psi_i}$ is a representation of the eigenvector ψ_i in the V_n basis. The solution of Eq. (9) can be rephrased into a generalized eigenvalue problem:

$$\mathbf{A}\overline{\psi_i^n} = \lambda_i^n \mathbf{W} \overline{\psi_i^n}, \tag{10}$$

where

$$\mathbf{A}_{ij} = \int_{\mathcal{D}} \int_{\mathcal{D}} c(x,y) \phi_i(y) \phi_j(x) \, \mathrm{d}y \mathrm{d}x, \tag{11}$$

$$\mathbf{W}_{ij} = \int_{\mathcal{D}} \phi_i(x) \phi_j(x) \, \mathrm{d}x. \tag{12}$$

2.1. The Choice of the Basis

We choose a discretization $\langle \phi_1(x), \ldots, \phi_n(x) \rangle = V_n$ of the space $L^2(\mathcal{D})$ ($\mathcal{D} = \langle a, b \rangle \times \langle c, d \rangle$) as a tensor product of 1D bases of the spaces $L^2(\langle a, b \rangle)$ and $L^2(\langle c, d \rangle)$. Therefore basis functions are in the form of

$$\phi_i(x) = \varphi_i^1(x_1) \varphi_i^2(x_2), \tag{13}$$

where $\varphi_i^1(x_1) \in V_n^1 := \langle \varphi_1^1(x_1), \ldots, \varphi_{m_1}^1(x_1) \rangle$ and $\varphi_i^2(x_2) \in V_n^2 := \langle \varphi_1^2(x_2), \ldots, \varphi_{m_2}^2(x_2) \rangle$ are 1D functions. The dimension of the space $V_n = V_n^1 \otimes V_n^1$ then equals $m_1 \cdot m_2$.

For the calculation of the problem Eq. (10), it is useful to consider the spaces V_n^1 and V_n^2 with specific properties. These properties will be demonstrated on 1D basis denoted by $\langle \varphi_1(x), \ldots, \varphi_m(x) \rangle$.

1) The Orthogonality of $\langle \varphi_1(x), \ldots, \varphi_m(x) \rangle$

The complexity of the problem Eq. (10) decreases if we transform the generalized eigenvalue problem into a standard eigenvalue problem ($\mathbf{W} = \mathbf{I}$). We can achieve this by assuring all of the 1D bases $\langle \varphi_1(x), \ldots, \varphi_m(x) \rangle$ to be orthonormal:

$$\forall i, j : \int_a^b \varphi_i(x) \varphi_j(x) \, \mathrm{d}x = \delta_{i,j}. \tag{14}$$

2) The (Anti)-Symmetricity of $\langle \varphi_1(x), \ldots, \varphi_m(x) \rangle$

The matrix $\mathbf{A}$ will generally be dense, but some unique properties of the chosen autocovariance function can be utilised to obtain partial sparsity of the matrix. First, we define the function

$$p(x,y) = \sigma^2 \exp\left(\frac{\sqrt{x^2 + y^2}}{\lambda} \right), \tag{15}$$

which is related to the autocovariance function by

$$p(|x_1 - y_1|, |x_2 - y_2|) = c((x_1, x_2), (y_1, y_2)). \tag{16}$$

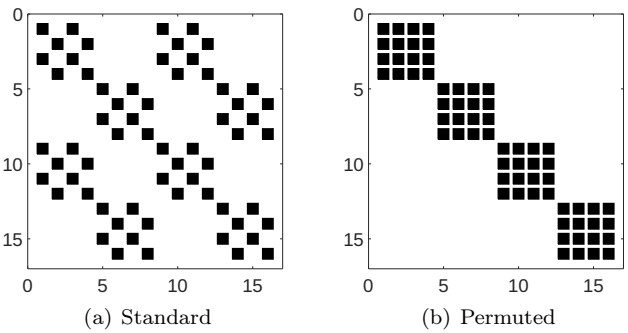
Next, consider the translation of the integral over the domain $\mathcal{D} = \langle a, b \rangle \times \langle c, d \rangle$ into the integral over the domain $\langle -\alpha, \alpha \rangle \times \langle -\beta, \beta \rangle$, where $\alpha = \frac{b-a}{2}$ and $\beta = \frac{d-c}{2}$ (translate the center of $\mathcal{D}$ into $(0,0)$). Note, that this will not affect the function $p(|x_1 - y_1|, |x_2 - y_2|)$. For simplicity we denote the basis functions $\varphi_i(x) = \varphi_i^1(x_1) \varphi_i^2(x_2)$ as before the translation of the integral. After these modifications, the formula for the elements of the matrix $\mathbf{A}$ is

$$\mathbf{A}_{i,j} = \int_{-\beta}^{\beta} \int_{-\beta}^{\beta} f(x_2, y_2) \varphi_i^2(x_2) \varphi_j^2(y_2) \, \mathrm{d}y_2 \mathrm{d}x_2, \tag{17}$$

where

$$f(x_2, y_2) = \int_{-\alpha}^{\alpha} \int_{-\alpha}^{\alpha} p(|x_1 - y_1|, |x_2 - y_2|) \cdot \cdot \varphi_i^1(x_1) \varphi_j^1(y_1) \, \mathrm{d}x_1 \mathrm{d}y_1. \tag{18}$$

If $\varphi_i^1(x_1)$ is symmetric and $\varphi_j^1(y_1)$ is anti-symmetric (or vice versa), the function $f(x_2, y_2)$ is constant zero and $\mathbf{A}_{i,j} = 0$. The same holds for the functions $\varphi_i^2(x_2)$ and $\varphi_j^2(y_2)$.



(a) Standard (b) Permuted

Fig. 2: Nonzero entries of matrix $\mathbf{A}$ (discretization 4×4).

The symmetricity/anti-symmetricity (or "parity") of the $\langle \varphi_1(x), \ldots, \varphi_m(x) \rangle$ basis results into a stronger property than just a partial sparsity of the matrix $\mathbf{A}$.

The nonzero pattern of the matrix $\mathbf{A}$ for the standard ordering according to the polynomial degree (see Fig. 2(a)) can be reordered into a block diagonal structure (see Fig. 2(b)). The permutation can be done by reordering the basis functions according to their parity in each dimension. The block diagonal structure leads to the solution of four smaller eigenvalue problems instead of one bigger eigenvalue problem. Note that in 3D we get eight diagonal blocks.

The next part of the Galerkin method is the particular choice of the discrete basis (orthonormal, symmetric/anti-symmetric). The most suitable bases for the Galerkin discretization Eq. (9) are e.g. piecewise constant functions, polynomials, or trigonometric functions. For the solution of similar problem using the wavelet basis see [18].

3) The Piecewise Constant Basis

First consider an interval (a, b) divided into an even number (say $2n$) of subintervals t_i of the same size $|t_i| = \frac{b-a}{2n}$. Than the symmetric basis functions are

$$\forall i = 1, \dots n : \varphi_i^S(x) = \begin{cases} \alpha, & x \in t_i, t_{2n-i+1}, \\ 0, & \text{otherwise}, \end{cases} \tag{19}$$

and the anti-symmetric basis functions are

$$\forall i = 1, \dots n : \varphi_i^A(x) = \begin{cases} \alpha, & x \in t_i, \\ -\alpha, & x \in t_{2n-i+1}, \\ 0, & \text{otherwise}, \end{cases} \tag{20}$$

where $\alpha = \sqrt{\frac{n}{b-a}}$. This construction also ensures the orthonormality of the basis.

4) The Polynomial Basis

In the case of polynomial basis, the choice is simple; the Legendre polynomials are both orthogonal and alternately symmetric or anti-symmetric. For general use, the Legendre polynomials need to be shifted into the desired interval and normalised. An explicit formula for the standard Legendre polynomials on $\langle -1, 1\rangle$ is

$$P_n(x) = 2^n \sum_{k=0}^{n} (-1)^k x^k \binom{n}{k} \binom{\frac{n+k-1}{2}}{n}. \tag{21}$$

The normalised Legendre polynomials on $\langle a, b\rangle$ are defined as

$$\varphi_n(x) = \sqrt{\frac{2n+1}{b-a}} P_n\left(\frac{2}{b-a}x - \frac{a+b}{b-a}\right). \tag{22}$$

5) The Trigonometric Basis

As the trigonometric basis, we choose the solution of the same problem in 1D (on $\langle 0, 1\rangle$). The 1D basis consists of eigenvectors of the operator

$$(\mathcal{C}f)(x) := \int_0^1 \sigma^2 \exp\left(\frac{\sqrt{(x-y)^2}}{\lambda}\right) f(y)\,\mathrm{d}y, \tag{23}$$

with the same parameters σ and λ. This problem has an analytic solution, see [5]. Explicit formulas for the basis functions are

$$\varphi_n'(x) = a_n \cos(w_n x) + b_n \sin(w_n x), \tag{24}$$

where

$$a_n = \lambda w_n b_n, b_n = \left(\frac{\lambda^2 w_n^2 + 1}{2} + \lambda\right)^{-\frac{1}{2}} \tag{25}$$

and w_n is the n-th smallest solution of

$$\left(\lambda^2 w^2 - 1\right) \sin(w) = 2\lambda w \cos(w). \tag{26}$$

Then we shift the functions into the interval $\langle a, b\rangle$:

$$\varphi_n(x) = \sqrt{\frac{1}{b-a}} \varphi_n'\left(\frac{1}{b-a}x - \frac{a}{b-a}\right). \tag{27}$$

2.2. Numerical Results

In this part, we compute the approximation of the KL decomposition and compare the bases mentioned above. To evaluate the integrals in the calculation of $\mathbf{A}$ (see Eq. (11)), we use the Gauss-Legendre quadrature with 100 points per dimension. Note, that this is a computationally expensive task, because we need to evaluate the integrand at 100^4 points for each nonzero entry of the matrix $\mathbf{A}$.

First, we illustrate the fast decay of the eigenvalues (see Fig. 3) and the approximation of the first and the thirteenth eigenfunction (see Fig. 4). The results are obtained using the trigonometric basis of 400 basis functions (20 in each dimension). Figure 3 shows that for a good approximation of the random field we need only the first few eigenpairs, i.e., the components of

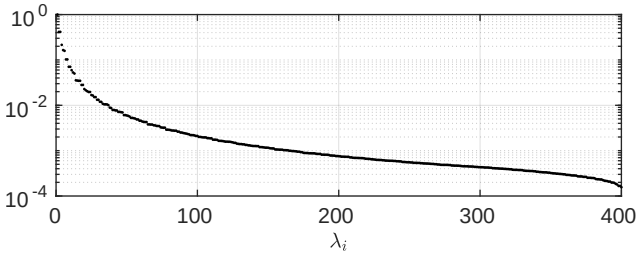


Fig. 3: Approximation of first 400 eigenvalues.

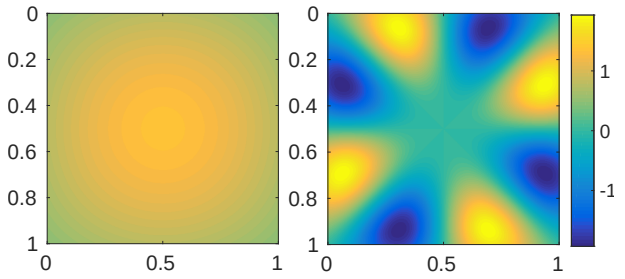


Fig. 4: 1st eigenvector (left) and 13th eigenvector (right).

the KL decomposition. This is an important property of the KL decomposition, that ensures a small approximation error while using the truncated KL decomposition. Some autocovariance functions provide a good estimate of the maximum approximation error of the truncated KL decomposition, see for example [7] and [1].

The next step is a comparison of the approximation error of the selected bases. The explicit eigenfunctions are unknown; therefore we use the approximation obtained by using the piecewise constant basis of $500 \times 500 = 250000$ functions as a reference solution. We measure the error as the $L^2(\mathcal{D})$ norm of the difference between the reference solution $\psi_i(\boldsymbol{x})$ and the Galerkin approximation $\psi_i^h(\boldsymbol{x})$ (both multiplied by the square root of the corresponding eigenvalue):

$$\mathrm{err}_i = \left\| \sqrt{\lambda_i}\psi_i(\boldsymbol{x}) - \sqrt{\lambda_i^h}\psi_i^h(\boldsymbol{x}) \right\|_{L^2(\mathcal{D})}. \tag{28}$$

The results in Fig. 5 and Fig. 6 show that the polynomial basis has the best approximation property. The stagnation of the convergence of the polynomial basis is caused by the choice of the reference solution. Note

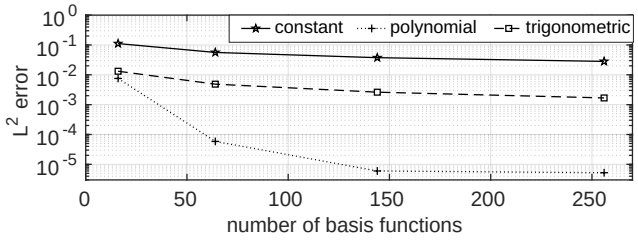


Fig. 5: Approximation of err_1 for different bases.

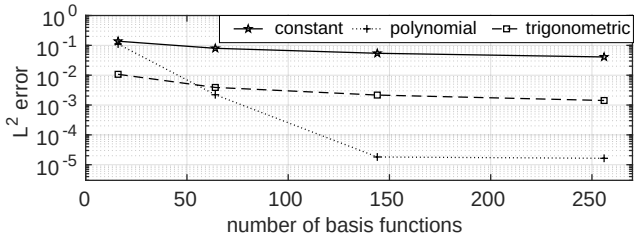


Fig. 6: Approximation of err_{13} for different bases.

that the convergence of the Galerkin method depends on the precision of the numerical integration, which is lower for fast oscillating functions like polynomials and trigonometric functions. On the other hand, the piecewise constant basis needs a lower number of integration points for a larger basis.

3. The Stochastic Galerkin Method

This section focuses on a solution of the model problem Eq. (1) with a random material properties in the form of the truncated KL decomposition of a random field. The fast decay of the eigenvalues (see Fig. 3) ensures a low Approximation Error (AE) of the truncated KL decomposition. We consider a log-normally distributed random material properties. Therefore our random material properties take the form

$$k(x;\boldsymbol{Z}) = \exp\left(\sum_{i=1}^{N} k_i(x) Z_i\right), \tag{29}$$

where $\boldsymbol{Z} = (Z_1, \ldots, Z_N)$ denotes a random vector of independent standard normal variables, and $k_i(x)$ are known functions.

The formulation of the model problem with the material field $k(x;\boldsymbol{Z})$ changes to the following:

$$\begin{cases} \forall \boldsymbol{Z} = (Z_1, \ldots, Z_N), Z_i \sim N(0;1): & \\ -\mathrm{div}\left(k(x;\boldsymbol{Z})\nabla u(x;\boldsymbol{Z})\right) = 0, & \forall x \in \mathcal{D}, \\ u(x;\boldsymbol{Z}) = 1, & \forall x \in \Gamma_{D_1}, \\ u(x;\boldsymbol{Z}) = 0, & \forall x \in \Gamma_{D_2}, \\ \dfrac{\partial u(x;\boldsymbol{Z})}{\partial n(x)} = 0, & \forall x \in \Gamma_N. \end{cases} \tag{30}$$

We can understand the formulation Eq. (30) as a parametric differential equation with additional properties of the parameters defined by their distribution. The solution of this problem is the function $u(x;\boldsymbol{Z}) : \mathcal{D}\times\mathbb{R}^N \to \mathbb{R}$, which is an approximation of the original solution $u(x;\omega) : \mathcal{D}\times\Omega \to \mathbb{R}$.

The next step is a weak formulation of Eq. (30), the details of which can also be found in [16]. A random field is understood as an operator from the space $L^2\left(\Omega, L^2(\mathcal{D})\right)$, as mentioned in Sec. 2. In the case of the solution $u(x;\omega)$ (also a random field), the homogeneous part $u_H(x;\omega)$ is in the space $H_0^1(\mathcal{D})$ for every fixed ω. Therefore the resulting random field is in the space $V := L^2\left(\Omega, H_0^1(\mathcal{D})\right) + u_0(x;\omega)$, where $u_0(x;\omega)$ is an arbitrary function from the space $L^2\left(\Omega, H^1(\mathcal{D})\right)$ that satisfies the Dirichlet boundary conditions. When using the truncated KL decomposition form of the random material properties, the solution $u(x;\boldsymbol{Z})$ is in

the space $u_0(x;\boldsymbol{Z}) + L^2_{\mathrm{d}F\boldsymbol{Z}}\left(\mathbb{R}^N, H^1_0(\mathcal{D})\right)$. The space $L^2_{\mathrm{d}F\boldsymbol{Z}}\left(\mathbb{R}^N, H^1_0(\mathcal{D})\right)$ is defined as:

$$L^2_{\mathrm{d}F\boldsymbol{Z}}\left(\mathbb{R}^N, H^1_0(\mathcal{D})\right) := \left\{f : \int_{\mathbb{R}^N} \|f(\boldsymbol{Z})\|^2_{H^1_0(\mathcal{D})}\,\mathrm{d}F\boldsymbol{Z} < \infty\right\}, \quad (31)$$

where $\mathrm{d}F\boldsymbol{Z}$ denotes "according to the distribution of the random vector $\boldsymbol{Z}$". In the case of continuous random variables, it is equal to the probability density of $\boldsymbol{Z}$. Similarly, as in the KL theorem, it can be shown that the space $L^2_{\mathrm{d}F\boldsymbol{Z}}\left(\mathbb{R}^N, H^1_0(\mathcal{D})\right)$ and tensor product of the spaces $H^1_0(\mathcal{D})$ and $L^2_{\mathrm{d}F\boldsymbol{Z}}\left(\mathbb{R}^N\right)$ are isometrically isomorphic (see e.g. [2] and [3]). The space $V_H := H^1_0(\mathcal{D}) \otimes L^2_{\mathrm{d}F\boldsymbol{Z}}\left(\mathbb{R}^N\right)$ is also a Hilbert space. Construction of the weak formulation is similar to the standard approach: multiply the equation by a function from the space V_H, integrate over the domain on which $u(x;\boldsymbol{Z})$ is defined with its natural measure (distribution of $\boldsymbol{Z}$ and L^2 measure), and use the Green formula to transfer the derivatives to the test function. If we consider the solution in the form $u(x;\boldsymbol{Z}) = u_H(x;\boldsymbol{Z}) + u_0(x;\boldsymbol{Z})$, where $u_0 \in V := H^1(\mathcal{D}) \otimes L^2_{\mathrm{d}F\boldsymbol{Z}}\left(\mathbb{R}^N\right)$ satisfies the Dirichlet boundary conditions and $u_H \in V_H$, we obtain the following weak formulation of the problem

$$\begin{cases} \text{Find } u_H(x;\boldsymbol{Z}) \in V_H,\ \forall v(x;\boldsymbol{Z}) \in V_H : \\ a(u_H; v) = -a(u_0; v), \end{cases} \quad (32)$$

where

$$a(u,v) := \\ = \int_{\mathbb{R}^N}\int_{\mathcal{D}} k(x;\boldsymbol{Z})\,\nabla_x u(x;\boldsymbol{Z})\,\nabla_x v(x;\boldsymbol{Z})\,\mathrm{d}x\,\mathrm{d}F\boldsymbol{Z}. \quad (33)$$

Although the material does not have positive lower or upper bound, the weak formulation is well posed, see [8].

The next step of the solution is the Galerkin discretization. We can construct the discretization using the tensor product of discrete bases of the spaces $H^1_0(\mathcal{D})$ and $L^2_{\mathrm{d}F\boldsymbol{Z}}\left(\mathbb{R}^N\right)$. The natural choice of the discrete basis of the space $H^1_0(\mathcal{D})$ (spatial variable) are finite elements. The discrete basis of the space $L^2_{\mathrm{d}F\boldsymbol{Z}}\left(\mathbb{R}^N\right)$ (stochastic variable) will be the polynomial chaos.

3.1. The Polynomial Chaos Expansion

This part serves as a brief introduction to the polynomial chaos (PC) (or the generalised polynomial chaos), a more thorough description of the PC can be found in [9], [1] and [4]. First, we focus on the PC of a single random variable and then we show the transition to the PC of a random vector of independent random variables.

Let Z be a random variable with all $2n$ moments finite, i.e. $\forall n \in \mathbb{N} : \mathbb{E}\left(Z^{2n}\right) < \infty$. The PC of such variable is a set of N polynomials $\psi_i(Z)$, $i \in 1,\ldots,N$ (i denotes the degree of the polynomial) satisfying $\mathbb{E}(\psi_i(Z)\,\psi_j(Z)) = \gamma_i\delta_{ij}$, where $\gamma_i = \mathbb{E}\left(\psi_i(Z)^2\right)$. In the following text, we consider a normalised PC ($\gamma_i = 1$). We can understand the PC of a random variable Z as an orthonormal basis of the space $L^2_{\mathrm{d}FZ}(\mathbb{R})$. Consequently, we can approximate every function $f(Z) \in L^2_{\mathrm{d}FZ}(\mathbb{R})$ by the PC of the random variable Z:

$$f(Z) \approx f_n(Z) = \sum_{i=0}^{n} \psi_i(Z)\,\mathbb{E}(\psi_i(Z)\,f(Z)), \quad (34)$$

where $f_n(Z)$ is the projection of $f(Z)$ into the space of polynomials up to the degree n, and

$$\lim_{n\to\infty} \|f_n(Z) - f(Z)\|_{L^2_{\mathrm{d}FZ}(\mathbb{R})} = 0, \quad (35)$$

see [2], [19] and [20].

In the case of a random vector $\boldsymbol{Z}$ of N independent random variables, the space $L^2_{\mathrm{d}F\boldsymbol{Z}}\left(\mathbb{R}^N\right)$ can be handled as a tensor product $L^2_{\mathrm{d}FZ_1}(\mathbb{R}) \otimes \ldots \otimes L^2_{\mathrm{d}FZ_N}(\mathbb{R})$. The PC of the random vector $\boldsymbol{Z}$ consists of products of the PC basis functions of the random variables Z_i:

$$\Psi_i(\boldsymbol{Z}) = \prod_{k=1}^{N} \psi_{i_k}(Z_k), \quad (36)$$

where i denotes a multi-index of size N.

The PC expansions are well known for most of the standard probability distributions, see the Wiener-Askey scheme in Tab. 1. According to this scheme, the polynomial chaos for a normal random variable consists of the Hermite polynomials (we will use the normalised Hermite polynomials).

Tab. 1: The Wiener-Askey scheme, see [4]: Table 5.1.

Distribution	Polynomial chaos
Gaussian	Hermite polynomials
Uniform	Legendre polynomials
Gamma	Laguerre polynomials
Beta	Jacobi polynomials
Poisson	Charlier polynomials
Negative Binomial	Miexner polynomials
Binomial	Krawtchouk polynomials
Hypogeometric	Hahn polynomials

3.2. Assembling the Stiffness Matrix

As mentioned before, discretization of the space V_H can be considered as a tensor product of discretization of the space $H^1_0(\mathcal{D})$ and discretization of the space $L^2_{\mathrm{d}F\boldsymbol{Z}}\left(\mathbb{R}^N\right)$. As discretization

of the space $H_0^1(\mathcal{D})$ we will use the finite elements basis $V_D^h := \langle\varphi_1(x),\ldots,\varphi_{N_D}(x)\rangle \subset \langle\varphi_1(x),\ldots,\varphi_{N_D}(x),\varphi_{N_D+1}(x),\ldots,\varphi_{N_{DD}}(x)\rangle \subset H^1(\mathcal{D})$, and as discretization of the space $L^2_{\mathrm{d}F_{\boldsymbol{Z}}}(\mathbb{R}^N)$ we will use the Hermite polynomials basis $V_P^h := \langle\psi_1(\boldsymbol{Z}),\ldots,\psi_{N_P}(\boldsymbol{Z})\rangle$ (we now simplify the aforementioned multi-index into the simple index). We denote the discretized test function space as $V_H^h := V_D^h \otimes V_P^h \subset V_H$. Then, the discretized solution and the discretized particular solution will take the form of

$$u_H^h(x;\boldsymbol{Z}) = \sum_{i=1}^{N_D}\sum_{j=1}^{N_P}(\overline{u_H})_{ij}\,\varphi_i(x)\,\psi_j(\boldsymbol{Z}), \tag{37}$$

and

$$u_0^h(x,\boldsymbol{Z}) = \sum_{i=N_D+1}^{N_{DD}}(\overline{u_0})_i\,\varphi_i(x)\,\psi_1(\boldsymbol{Z}), \tag{38}$$

where $\psi_1(\boldsymbol{Z})$ is the constant element of the PC basis. The Galerkin formulation of the problem:

$$\begin{cases}\text{Find } \overline{u_H}\in\mathbb{R}^{N_DN_P}, \forall\varphi\in V_D^h, \forall\psi\in V_P^h:\\ a\left(u_H^h,\varphi\psi\right) = -a\left(u_0^h,\varphi\psi\right),\end{cases} \tag{39}$$

is equivalent to the system of linear equations

$$\mathbf{A}\overline{u_H} = \overline{b}, \tag{40}$$

where individual elements of $\mathbf{A}$ and $\overline{b}$ are defined as:

$$(\mathbf{A})_{ij,kl} = \int_{\mathbb{R}^N}\int_{\mathcal{D}}\exp\left(\sum_{m=1}^{N}k_m(x)\,Z_m\right)\cdot \cdot\nabla\varphi_i(x)\,\psi_j(\boldsymbol{Z})\,\nabla\varphi_k(x)\,\psi_l(\boldsymbol{Z})\,\mathrm{d}x\mathrm{d}F\boldsymbol{Z}, \tag{41}$$

$$\left(\overline{b}\right)_{ij} = \int_{\mathbb{R}^N}\int_{\mathcal{D}}\exp\left(\sum_{m=1}^{N}k_m(x)\,Z_m\right)\nabla\varphi_i(x)\cdot \cdot\psi_j(\boldsymbol{Z})\sum_{k=N_D+1}^{N_{DD}}(\overline{u_0})_k\,\varphi_k(x)\,\psi_1(\boldsymbol{Z})\,\mathrm{d}x\mathrm{d}F\boldsymbol{Z}. \tag{42}$$

Calculation of elements of the matrix $\mathbf{A}$ and the right-hand side $\overline{b}$ seems complicated, but the integral over the stochastic domain $\mathbb{R}^N$ can be evaluated analytically. Using the commutativity of multiplication and Fubini's theorem, we can transform the expression of $(\mathbf{A})_{ij,kl}$ and $\left(\overline{b}\right)_{ij}$ to

$$(\mathbf{A})_{ij,kl} = \int_{\mathcal{D}}\nabla\varphi_i(x)\,\nabla\varphi_k(x)\,k_{j,l}(x)\,\mathrm{d}x, \tag{43}$$

$$\left(\overline{b}\right)_{ij} = \int_{\mathcal{D}}\nabla\varphi_i(x)\sum_{k=N_D+1}^{N_{DD}}(\overline{u_0})_k\,\varphi_k(x)\,k_{j,1}(x)\,\mathrm{d}x, \tag{44}$$

where

$$k_{i,j}(x) = \int_{\mathbb{R}^N}\exp\left(\sum_{m=1}^{N}k_m(x)\,Z_m\right)\cdot \cdot\psi_i(\boldsymbol{Z})\,\psi_j(\boldsymbol{Z})\,\mathrm{d}F\boldsymbol{Z}. \tag{45}$$

We obtain the values of $k_{i,j}(x)$ for some $x\in\mathcal{D}$ by rewriting the expression using the definition of the PC basis (see Eq. (36)):

$$k_{i,j}(x) = \prod_{m=1}^{N}\int_{\mathbb{R}}\exp\left(k_m(x)\,Z_m\right)\cdot \cdot\psi_{i_m}(Z_m)\,\psi_{j_m}(Z_m)\,\mathrm{d}FZ_m, \tag{46}$$

and merging the part of the random field $\exp(k_m(x)\,Z_m)$ corresponding to the random variable Z_m with its probability density function (currently denoted as $\mathrm{d}FZ_m$). The elements of the product take the form of

$$\int_{\mathbb{R}}\psi_{i_m}(Z_m)\,\psi_{j_m}(Z_m)\cdot \cdot\frac{1}{\sqrt{2\pi}}\exp\left(\frac{2k_m(x)\,Z_m - Z_m^2}{2}\right)\mathrm{d}Z_m, \tag{47}$$

which can be expressed by the shifted random variables $\widetilde{Z_m}\sim N(k_m(x),1)$. The probability density of $\widetilde{Z_m}$ is

$$\mathrm{d}F\widetilde{Z_m} := \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\left(Z_m^2-k_m(x)\right)^2}{2}\right)\mathrm{d}Z_m. \tag{48}$$

Therefore $k_{i,j}(x)$ can be expressed as:

$$k_{i,j}(x) = \prod_{m=1}^{N}\exp\left(k_m(x)^2\right)\cdot \cdot\int_{\mathbb{R}}\psi_{i_m}(Z_m)\,\psi_{j_m}(Z_m)\,\mathrm{d}F\widetilde{Z_m}. \tag{49}$$

Calculation of the stochastic part of the integral now reduces to the evaluation of

$$\int_{\mathbb{R}}\psi_{i_m}(Z_m)\,\psi_{j_m}(Z_m)\,\mathrm{d}F\widetilde{Z_m}, \forall m=1,\ldots,N. \tag{50}$$

The functions $\psi_{i_m}(Z_m)$ are polynomials; therefore the expression Eq. (50) can be written as a linear combination of the moments of $\widetilde{Z_m}$ (normal distribution with given mean), which are known.

Another approach to the construction of the matrix $\mathbf{A}$ is using the PC expansion of the truncated KL decomposition of the random material properties

$$\sum_{i=1}^{N_{PC}}\widetilde{k_i}(x)\,\psi_i(\boldsymbol{Z}) \approx \exp\left(\sum_{i=1}^{N}k_i(x)\,Z_i\right). \tag{51}$$

This approach allows us to separate the stochastic and the physical part of the integral in the matrix $\mathbf{A}$ calculation

$$(\mathbf{A})_{ij,kl} = \sum_{m=1}^{N_{PC}} \int_{\mathcal{D}} \widetilde{k_m}(x) \nabla\varphi_k(x) \nabla\varphi_i(x) \,\mathrm{d}x \cdot \cdot \int_{\mathbb{R}^N} \psi_m(\boldsymbol{Z}) \psi_j(\boldsymbol{Z}) \psi_l(\boldsymbol{Z}) \,\mathrm{d}F\boldsymbol{Z}. \quad (52)$$

The separation leads to the "tensor form" of the matrix $\mathbf{A}$:

$$\mathbf{A} = \sum_{m=1}^{N_{PC}} \mathbf{G}_{\mathrm{m}} \otimes \mathbf{K}_{\mathrm{m}}, \quad (53)$$

where

$$(\mathbf{K}_{\mathrm{m}})_{ij} = \int_{\mathcal{D}} \widetilde{k_m}(x) \nabla\varphi_i(x) \nabla\varphi_j(x) \,\mathrm{d}x, \quad (54)$$

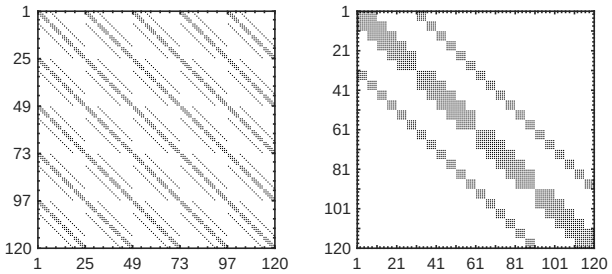
$$(\mathbf{G}_{\mathrm{m}})_{ij} = \int_{\mathbb{R}^N} \psi_m(\boldsymbol{Z}) \psi_i(\boldsymbol{Z}) \psi_j(\boldsymbol{Z}) \,\mathrm{d}F\boldsymbol{Z}. \quad (55)$$

This approach allows compressed storage of the matrix $\mathbf{A}$ in the form of pairs of the matrices $(\mathbf{K}_{\mathrm{m}}, \mathbf{G}_{\mathrm{m}})$; the matrix $\mathbf{A}$ does not need to be constructed. On the other hand, this approach suffers from an additional AE of the PC representation of the random material properties, see Eq. (51). The tensor form of the matrix can also be assembled by rephrasing the problem, see [10], but this approach leads to a non-symmetric system.

The matrix $\mathbf{A}$ can have two different types of structure, which depends on the ordering of the basis functions:

- $(\mathbf{A})_{ij,kl}$ denotes ordering $(\varphi_1\psi_1), (\varphi_2\psi_1), \ldots, (\varphi_{N_D}\psi_1), (\varphi_1\psi_2), \ldots$, or
- $(\mathbf{A})_{ji,lk}$ denotes ordering $(\varphi_1\psi_1), (\varphi_1\psi_2), \ldots, (\varphi_1\psi_{N_P}), (\varphi_2\psi_1), \ldots$.

We can see the nonzero pattern of the matrix $\mathbf{A}$ for both structures in Fig. 7. We can use this property for the construction of a preconditioner.



(a) $(\mathbf{A})_{ij,kl}$ (b) $(\mathbf{A})_{ji,lk}$

Fig. 7: Different structures of matrix $\mathbf{A}$ ($N_D = 24, N_P = 5$).

3.3. Numerical Results

This part reports on numerical results for the solved problem. For the numerical experiments, we choose a regular 50×50 grid for the physical domain. We solve the problem using the full matrix $\mathbf{A}$ construction to avoid the AE of the PC expansion of the random material properties.

The main aims of this section are the evaluation of the mean and the variance/standard deviation of the resulting random field and the comparison of the SGM solution with the standard FEM solution.

The system of equations (structure $(\mathbf{A})_{ij,kl}$) was solved using the flexible GMRES method and a double block diagonal preconditioner

$$\mathbf{M} = \mathbf{D}_2^{-1}\mathbf{P}^T\mathbf{D}_1^{-1}\mathbf{P}, \quad (56)$$

where $\mathbf{P}$ is a permutation matrix for reordering a vector from the $(\mathbf{A})_{ij,kl}$ ordering to the $(\mathbf{A})_{ji,lk}$ ordering, $\mathbf{D}_1$ is the block diagonal of the matrix $(\mathbf{A})_{ji,lk}$ and $\mathbf{D}_2$ is the block diagonal of the matrix $(\mathbf{A})_{ij,kl}$.

There exist other types of preconditioners for the SGM systems, for example: block-diagonal preconditioner, see [12]; kronecker proconditioner, see [15]; or hirearchical Schur preconditioner, see [11]. Some insight into effectivity of preconditioners can be obtained using the strengthened Cauchy-Bunyakowski-Schwarz constant, see [13] and [14].

1) Mean Value and Standard Deviation of the Solution

The SGM solution of the given problem offers an easy way to obtain the mean and the variance of the solution. We can express the mean value of the solution

$$\mu_{u^h}(x) : \mathcal{D} \to \mathbb{R} := \int_{\mathbb{R}^N} u^h(x, \boldsymbol{Z}) \,\mathrm{d}F\boldsymbol{Z}, \quad (57)$$

as

$$\mu_{u^h}(x) = \int_{\mathbb{R}^N} \sum_{j=1}^{N_P} u_j(x) \psi_j(\boldsymbol{Z}) \,\mathrm{d}F\boldsymbol{Z} = u_1(x), \quad (58)$$

where

$$u_j(x) = \sum_{i=1}^{N_{DD}} (\overline{u})_{ij} \varphi_i(x). \quad (59)$$

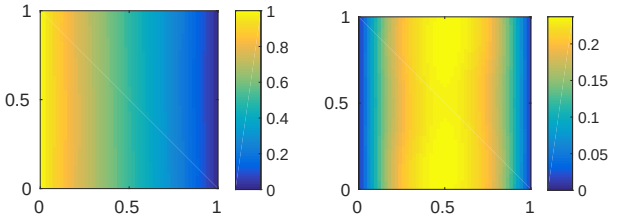
The variance of the solution

$$\sigma^2_{u^h}(x) : \mathcal{D} \to \mathbb{R} := \int_{\mathbb{R}^N} \left(u^h(x, \boldsymbol{Z}) - \mu_{u^h}(x)\right)^2 \mathrm{d}F\boldsymbol{Z}, \quad (60)$$

can be expressed as

$$\begin{aligned}\sigma^2_{u^h}(x) &= \int_{\mathbb{R}^N}\left(\sum_{j=2}^{N_P} u_j(x)\,\psi_j(\boldsymbol{Z})\right)^2 \mathrm{d}F\boldsymbol{Z}\\ &= \int_{\mathbb{R}^N}\sum_{j=2}^{N_P}\sum_{i=2}^{N_P} u_j(x)\,\psi_j(\boldsymbol{Z})\,u_i(x)\,\psi_i(\boldsymbol{Z})\,\mathrm{d}F\boldsymbol{Z}\\ &= \sum_{j=2}^{N_P} u_j^2(x)\,. \end{aligned} \quad (61)$$

The mean value and the standard deviation of solution are shown in Fig. 8. The results are obtained using $N = 5$ elements of the truncated KL decomposition and the PC up to the degree 10. The results agree with the intuition - the mean value corresponds to the deterministic solution with the homogeneous material and the standard deviation is higher for the points farther from the Dirichlet boundary.



(a) mean value $\mu_{u^h}(x)$ (b) standard deviation $\sigma_{u^h}(x)$

Fig. 8: Characteristics of the solution.

2) Comparing SGM and FEM Solutions

The main advantage of the SGM solution is that it gives us an approximation of $u(x, \boldsymbol{Z})$ for all possible values of $\boldsymbol{Z}$ (random material instances). Therefore we can use the SGM solution instead of the FEM calculation for a specific material sample. In the following text, $u(x, \boldsymbol{Z})$ denotes the FEM solution for a random material sample (using the truncated KL decomposition) and $u^h(x, \boldsymbol{Z})$ denotes the SGM solution. The AE of SGM will be measured using the $H^1(\mathcal{D})$ norm and the $L^2(\mathcal{D})$ norm of the difference between the FEM and the SGM solution; the precision of the SGM approximation over the whole domain of $\boldsymbol{Z}$ will be measured using

$$err_{H^1}\left(u^h\right) := \int_{\mathbb{R}^N} \left\|u(x,\boldsymbol{Z}) - u^h(x,\boldsymbol{Z})\right\|_{H^1} \mathrm{d}F\boldsymbol{Z}, \quad (62)$$

and

$$err_{L^2}\left(u^h\right) := \int_{\mathbb{R}^N} \left\|u(x,\boldsymbol{Z}) - u^h(x,\boldsymbol{Z})\right\|_{L^2} \mathrm{d}F\boldsymbol{Z}. \quad (63)$$

The integral in the calculation of $err_{H^1}\left(u^h\right)$ and $err_{L^2}\left(u^h\right)$ will be approximated by the Monte Carlo method with 10^6 samples.

The first numerical test describes the dependence of the AE on the iterative solution precision and on the maximum polynomial degree of the PC basis. We construct the PC basis with the maximum polynomial degree n as

$$\left\langle \Psi_a(\boldsymbol{Z}) := \prod_{i=1}^{N} \psi_{a_i}(Z_i) \right\rangle_{\sum a_i \le n}. \quad (64)$$

The results of the numerical tests are shown in Fig. 9. We can observe the following:

- The convergence in the L^2 norm is comparable with the convergence in the H^1 norm. In the rest of the text, we will use only the H^1 norm.
- The precision of the iterative solution affects the SGM AE only slightly, the difference between relative residual error 10^{-5} and 10^{-9} is negligible.
- The convergence is almost linear when increasing the maximum polynomial degree.

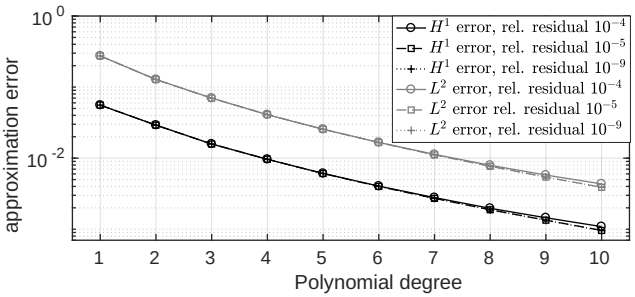


Fig. 9: Comparison of error for different iterative precision and polynomial degree.

The next numerical test will compare two basic approaches to the construction of the PC base. The first construction is given by Eq. (64), where the sum of degrees of polynomials over all dimensions is bounded by n. The second construction is

$$\left\langle \Psi_a(\boldsymbol{Z}) := \prod_{i=1}^{N} \psi_{a_i}(Z_i) \right\rangle_{\forall i: a_i \le n}, \quad (65)$$

where the degree of a polynomial is bounded by n in every dimension. The size of the PC basis is $\binom{n+N}{n}$ for the first construction and n^N for the second construction.

The error $err_{H^1}\left(u^h\right)$ for both PC bases is shown in Fig. 10 and Fig. 11. Figure 10 shows the dependence of

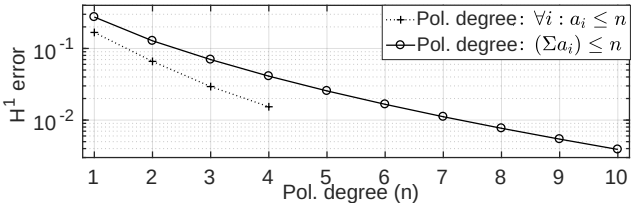


Fig. 10: Error dependence on the maximal degree of the PC basis.

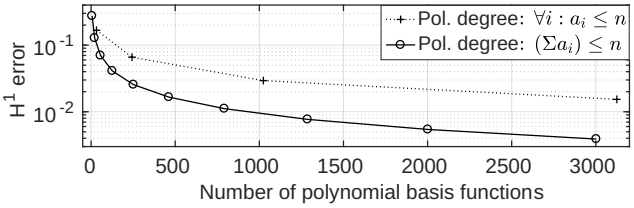


Fig. 11: Error dependence on the size of the PC basis.

$err_{H^1}\left(u^h\right)$ on the maximal degree of the polynomials, and Fig. 11 shows the dependence of error on the size of the basis. We can observe that the second approach has a linear convergence rate and if compared to the first approach, it has lower $err_{H^1}\left(u^h\right)$ for the same n. On the other hand, the first approach has lower AE for the same size of the basis.

The previous numerical tests demonstrated that there is a considerable difference between two approaches to the construction of the PC basis. This behaviour can motivate a search for an optimal construction of the PC basis. Here we will try to construct an "optimal" PC basis based on a different importance (or the impact on the AE) of each KL decomposition random variable (KLV).

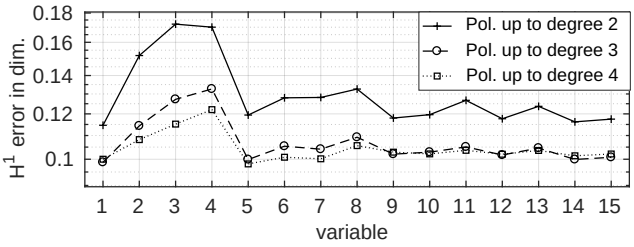
First, we investigate the AE in some specific subsets of the stochastic dimension. We want to observe the AE when changing only one random variable and the AE when fixing one random variable to zero. Therefore we propose two error indicators. First, the AE measured only across the Z_i part of the stochastic dimension

$$err^i_{H^1}\left(u^h\right) := \int_{\mathbb{R}^N} \left\|u\left(x, \boldsymbol{Z}_i\right) - u^h\left(x, \boldsymbol{Z}_i\right)\right\|_{H^1} \mathrm{d}F\boldsymbol{Z}, \tag{66}$$

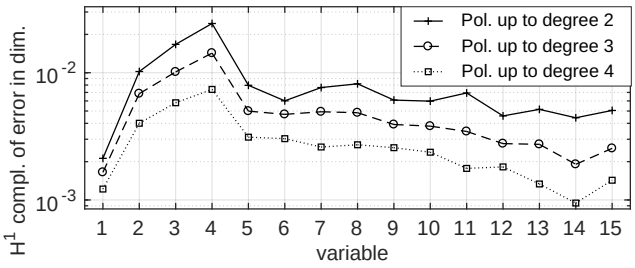
where $\boldsymbol{Z}_i = (0, \ldots, 0, Z_i, 0, \ldots, 0)$. Second, the AE measured as the complement to $err_{H^1}\left(u^h\right)$

$$err^{\bar{i}}_{H^1}\left(u^h\right) := err_{H^1}\left(u^h\right) - \\ - \int_{\mathbb{R}^N} \left\|u\left(x, \boldsymbol{Z}_{\bar{i}}\right) - u^h\left(x, \boldsymbol{Z}_{\bar{i}}\right)\right\|_{H^1} \mathrm{d}F\boldsymbol{Z}, \tag{67}$$

where $\boldsymbol{Z}_{\bar{i}} = (Z_1, \ldots, Z_{i-1}, 0, Z_{i+1}, \ldots, Z_N)$.



(a) $err^i_{H^1}\left(u^h\right)$



(b) $err^{\bar{i}}_{H^1}\left(u^h\right)$

Fig. 12: Indicators of the relative impact on the approximation error.

The results of the numerical test of the two proposed error indicators are shown in Fig. 12. The test was performed for $N = 15$ and $n = 2, 3, 4$ (bounded sum of degrees). Both $err^i_{H^1}\left(u^h\right)$ and $err^{\bar{i}}_{H^1}\left(u^h\right)$ exhibit a similar behaviour, the relative impact on the AE is the greatest for the 2$^{\text{nd}}$, 3$^{\text{rd}}$ and 4$^{\text{th}}$ KLV. The results of $err^{\bar{i}}_{H^1}\left(u^h\right)$ in Fig. 12 may be more credible, because they include a dependence between single KLVs.

The behaviour observed in the previous paragraph can be supported by the sensitivity analysis. The sensitivity analysis, in the context of this paper, is understood as a comparison of $err_{H^1}\left(u^h\right)$ for different bases. The tested bases will be constructed as

$$\Psi^{n,m}_i := \Psi^n_i \cup \Psi^m, \tag{68}$$

where

$$\Psi^m := \left\langle \Psi_a\left(\boldsymbol{Z}\right) := \prod_{j=1}^{N} \psi_{a_j}\left(Z_j\right) \right\rangle_{\sum a_j \leq m}, \tag{69}$$

$$\Psi^n_i := \left\langle \Psi_a\left(\boldsymbol{Z}\right) := \prod_{j=1, j\neq i}^{N} \psi_{a_j}\left(Z_j\right) \right\rangle_{\sum_{j\neq i} a_j \leq n}. \tag{70}$$

We performed numerical tests for $\Psi^{4,3}_i$ (denoted as "pol. degree -1"), $\Psi^{4,2}_i$ (denoted as "pol. degree -2") and $\Psi^{4,1}_i$ (denoted as "pol. degree -3"). The results (see Fig. 13) reflect the increase of $err_{H^1}\left(u^h\right)$ when removing the polynomials of the highest degree for a specific random variable. Note, that the sensitivity analysis calculation is extremely computationally expensive; therefore it cannot be used for the evaluation of the importance of a KLV.

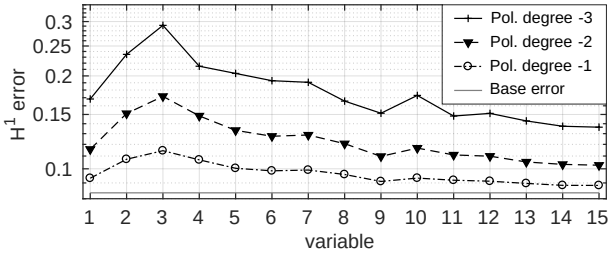


Fig. 13: Sensitivity analysis ($\Psi_i^{4,3}$, $\Psi_i^{4,2}$ and $\Psi_i^{4,1}$).

The performed numerical tests are computationally expensive and require the calculation of the approximations of the solution. Therefore we cannot use them for the construction of the PC basis. Another way to estimate the importance of the KLV is the examination of the functions $k_i(x)$ from the truncated KL decomposition form of the random material properties, see Eq. (29). The different importance of single KLVs is caused only by the disparity between the functions k_i. The objective is to find a function $f : L^2(\mathcal{D}) \to \mathbb{R}^+$ which for each $k_i(x)$ returns similar values as the relative impact indicators or sensitivity analysis for single KLVs. A straightforward approach to estimate the relative impact on the AE is to take the $L^2(\mathcal{D})$ norm of $k_i(x)$. However, a single $k_i(x)$ has the same relative impact on the AE if a constant $k_i(x) = k_i^c(x) + c_i$ is added to it, because

$$k(x;\boldsymbol{Z}) = c(\boldsymbol{Z})\, k^c(x;\boldsymbol{Z}), \tag{71}$$

where $c(\boldsymbol{Z}) = \exp\left(\sum_{i=1}^{N} c_i Z_i\right)$ and $k^c(x;\boldsymbol{Z}) = \exp\left(\sum_{i=1}^{N} k_i^c(x) Z_i\right)$, and

$$\begin{aligned} -\operatorname{div}\left(c(\boldsymbol{Z})\, k^c(x;\boldsymbol{Z})\, \nabla u(x;\boldsymbol{Z})\right) &= 0 \\ -\operatorname{div}\left(k^c(x;\boldsymbol{Z})\, \nabla u(x;\boldsymbol{Z})\right) &= 0. \end{aligned} \tag{72}$$

Therefore the following approximation is proposed:

$$f(k_i(x)) := \min_{c\in\mathbb{R}} \|k_i(x) - c\|_{L^2(\mathcal{D})}, \tag{73}$$

which equals to the

$$f(k_i(x)) = \left\| k_i(x) - \int_{\mathcal{D}} k_i(y)\,\mathrm{d}y \right\|_{L^2(\mathcal{D})}. \tag{74}$$

The values of the $f(k_i(x))$ for the KL decomposition elements $1,\ldots,15$ can be seen in Fig. 14. Compared to the previous relative impact indicators and sensitivity analysis, the function $f(k_i(x))$ has a similar behaviour.

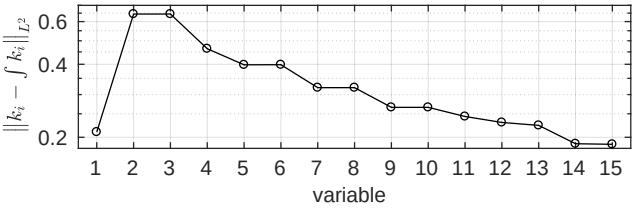


Fig. 14: Estimate of the relative impact on the AE.

The next step is the construction of the PC basis according to the given weights. We proceed as follows:

- we normalize the weights from Eq. (74) so that the maximum is 0.5: $w_i = f(k_i(x)) \Big/ \left(2 \max_{k\in 1,\ldots,N} f(k_i(x))\right)$, the vector $w = (w_1,\ldots,w_N)$ is obtained,
- we calculate the weight of each polynomial of the PC basis from its multi-index $a = (a_1,\ldots,a_N)$

$$w_a = \prod_{i=1}^{N} w_i^{a_i}, \tag{75}$$

- the PC basis of N_P polynomials is constructed from the first N_P polynomials with the highest weight w_a.

The numerical results of the AE dependence on the size of the PC basis for the weighted polynomial ordering compared to the previous construction approaches can be seen in Fig. 15. Note that in addition to the lower AE for the same size of the PC basis, the proposed construction also allows an arbitrary size of the basis.

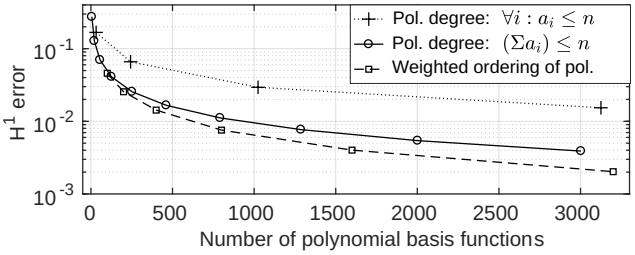


Fig. 15: Error dependence on the PC basis construction.

4. Conclusion

The paper focused on the complete solution of the Darcy flow problem with a random material. We outlined a path from the differential equation with a random field as an input parameter to an approximation of the solution, which is also a random field. We dealt with the spectral decomposition of the operator given

by the autocovariance function, and we used the outcoming decomposition for assembling the SGM system of equations.

In the section about the KL decomposition, we also provided an enhanced approach to the construction of the Galerkin solution, which leads to the block diagonal structure of the outcoming system, and numerical experiments concerning the choice of the basis for the Galerkin solution.

In the section about the SGM solution, we performed extensive numerical experiments regarding the convergence in the stochastic dimension and the AE of the different PC bases. Then we proposed the construction of the PC basis with a better AE.

Acknowledgment

This work was supported by The Ministry of Education, Youth and Sports from the National Programme of Sustainability (NPU II) project "IT4Innovations excellence in science - LQ1602". The work is partially supported by Grant of SGS No. SP2017/56, VSB–Technical University of Ostrava, Czech Republic.

References

[1] LORD, G. J., C. E. POWELL and T. SHARDLOW. *An introduction to computational stochastic PDEs*. New York: Cambridge University Press, 2014. ISBN 978-0-521-89990-1.

[2] LIGHT, W. A. and E. W. CHENEY. *Approximation theory in tensor product spaces*. New York: Springer, 1985. ISBN 3-540-16057-4.

[3] DA PRATO, G. and J. ZABCZYK. *Second order partial differential equations in Hilbert spaces*. Cambridge: Cambridge University Press, 2002. ISBN 978-0-521-77729-2.

[4] XIU, D. *Numerical methods for stochastic computations: a spectral method approach*. New Jersey: Princeton University Press, 2010. ISBN 978-0-691-14212-8.

[5] ZHANG, D. and Z. LU. An efficient, high-order perturbation approach for flow in random porous media via Karhunen–Loeve and polynomial expansions. *Journal of Computational Physics*. 2004, vol. 194, iss. 2, pp. 773–794. ISSN 0021-9991. DOI: 10.1016/j.jcp.2003.09.015.

[6] HOEKSEMA, R. J. and P. K. KITANIDIS. Analysis of the spatial structure of properties of selected aquifers. *Water Resources Research*. 1985, vol. 21, iss. 4, pp. 563–572. ISSN 0043-1397. DOI: 10.1029/WR021i004p00563.

[7] LI, C. F., Y. T. FENG, D. R. J. OWEN, D. F. LI and I. M. DAVIS. A Fourier–Karhunen–Loeve discretization scheme for stationary random material properties in SFEM. *International journal for numerical methods in engineering*. 2008, vol. 73, iss. 13, pp. 1942–1965. ISSN 0029-5981. DOI: 10.1002/nme.2160.

[8] GITTELSON, C. J. Stochastic Galerkin discretization of the log-normal isotropic diffusion problem. *Mathematical Models and Methods in Applied Sciences*. 2010, vol. 20, iss. 2, pp. 237–263. ISSN 0218-2025. DOI: 10.1142/S0218202510004210.

[9] XIU, D. and G. EM KARNIADAKIS. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*. 2003, vol. 24, iss. 2, pp. 619–644. ISSN 1064-8275. DOI: 10.1137/S1064827501387826.

[10] ULLMANN, E., H. C. ELMAN, and O. G. ERNST. Efficient Iterative Solvers for Stochastic Galerkin Discretizations of Log-Transformed Random Diffusion Problems. *SIAM Journal on Scientific Computing*. 2012, vol. 34, iss. 2, pp. A659–A682. ISSN 1064-8275. DOI: 10.1137/110836675.

[11] SOUSEDIK, B., R. G. GHANEM and E. T. PHIPPS. Hierarchical Schur Complement Preconditioner for the Stochastic Galerkin Finite Element Methods: Dedicated to Professor Ivo Marek on the Occasion of His 80th Birthday. *Numerical Linear Algebra with Applications*. 2014, vol. 21, iss. 1, pp. 136–151. ISSN 1070-5325. DOI: 10.1002/nla.1869.

[12] POWELL, C. E. and H. C. ELMAN. Block-Diagonal Preconditioning for Spectral Stochastic Finite-Element Systems. *IMA Journal of Numerical Analysis*. 2009, vol. 29, iss. 2, pp. 350–375. ISSN 0272-4979. DOI: 10.1093/imanum/drn014.

[13] PULTAROVA, I. Adaptive Algorithm for Stochastic Galerkin Method. *Applications of Mathematics*. 2015, vol. 60, iss. 5, pp. 551–571. ISSN 0862-7940. DOI: 10.1007/s10492-015-0111-9.

[14] PULTAROVA, I. Hierarchical Preconditioning for the Stochastic Galerkin Method: Upper Bounds to the Strengthened CBS Constants. *Computers and Mathematics with Applications*. 2016, vol. 71, iss. 4, pp. 949–964. ISSN 0898-1221. DOI: 10.1016/j.camwa.2016.01.006.

[15] ULLMANN, E. A Kronecker Product Preconditioner for Stochastic Galerkin Finite Eleme

Nt Discretizations. *SIAM Journal on Scientific Computing*. 2010, vol. 32, iss. 2, pp. 923–946. ISSN 1064-8275. DOI: 10.1137/080742853.

[16] BABUSKA, I., F. NOBILE and R. TEMPONE. A Stochastic Collocation Method for Elliptic Partial Differential Equations with Random Input Data. *SIAM Journal on Numerical Analysis*. 2007, vol. 45, iss. 3, pp. 1005–1034. ISSN 0036-1429. DOI: 10.1137/050645142.

[17] BABUSKA, I., R. TEMPONE and G. E. ZOURARIS. Galerkin Finite Element Approximations of Stochastic Elliptic Partial Differential Equations. *SIAM Journal on Numerical Analysis*. 2004, vol. 42, iss. 2, pp. 800–825. ISSN 0036-1429. DOI: 10.1137/S0036142902418680.

[18] OLIVEIRA, S. P., F. WISNIEWSKI and J. S. AZEVEDO. A wavelet Galerkin approximation of Fredholm integral eigenvalue problems with bidimensional Haar functions. *Proceeding Series of the Brazilian Society of Applied and Computational Mathematics*. 2014, vol. 2, iss. 1, pp. 1–6. ISSN 2359-0793. DOI: 10.5540/03.2014.002.01.0060.

[19] TIMAN, A. F. *Theory of Approximation of Functions of a Real Variable*. New York: Dover Publications, 1994. ISBN 978-0-486-67830-6.

[20] CHENEY, E. W. *Introduction to Approximation Theory*. New York: Chelsea Pub. Co., 1982. ISBN 978-0-821-81374-4.

About Authors

Michal BERES graduated from the Faculty of Electrical Engineering and Computer Science, VSB–Technical University of Ostrava in 2015. He is currently a Ph.D. student at the same university. His research interests include uncertainty in the numerical models.

Simona DOMESOVA graduated from VSB–Technical University of Ostrava in 2015, where she studied Computational Mathematics. Currently, she is a Ph.D. student at the same university. Her research interests include stochastic modeling.