Kernel-based Collocation Methods versus Galerkin Finite Element Methods for Approximating Elliptic Stochastic Partial Differential Equations

Gregory E. Fasshauer and Qi Ye*

Department of Applied Mathematics, Illinois Institute of Technology, Chicago, Illinois 60616 USA, fasshauer@iit.edu and qye3@iit.edu

Summary. We compare a kernel-based collocation method (meshfree approximation method) with a Galerkin finite element method for solving elliptic stochastic partial differential equations driven by Gaussian noise. The kernel-based collocation solution is a linear combination of reproducing kernels obtained from related differential and boundary operators centered at chosen collocation points. Its random coefficients are obtained by solving a system of linear equations with multiple random right-hand sides. The finite element solution is given as a tensor product of triangular finite elements and Lagrange polynomials defined on a finite-dimensional probability space. Its coefficients are obtained by solving several deterministic finite element problems. For the kernel-based collocation method, we directly simulate the (infinite-dimensional) Gaussian noise at the collocation points. For the finite element method, however, we need to truncate the Gaussian noise into finite-dimensional noises. According to our numerical experiments, the finite element method has the same convergence rate as the kernel-based collocation method provided the Gaussian noise is truncated using a suitable number terms.

Key words: Kernel-based collocation, meshfree approximation, Galerkin finite element, elliptic stochastic partial differential equations, Gaussian fields, reproducing kernel.

1 Introduction

Stochastic partial differential equations (SPDEs) form the basis of a recent, fast growing research area with many applications in physics, engineering and finance. However, it is often difficult to obtain an explicit form of the solution. Moreover, current numerical methods usually show limited success for high-dimensional problems and in complex domains. In our recent papers [4, 11],

* Corresponding Author

we use a kernel-based collocation method to approximate the solution of high-dimensional SPDE problems. Since parabolic SPDEs can be transformed into elliptic SPDEs using, e.g., an implicit time stepping scheme, solution of the latter represents a particularly important aspect of SPDE problems.

In this paper, we compare the use of a *kernel-based collocation method* [4, 11] (meshfree approximation method) and a *Galerkin finite element method* [1, 2] to approximate the solution of elliptic SPDEs. For kernel-based collocation, we directly simulate the Gaussian noise at a set of collocation points. For the Galerkin finite element method, on the other hand, we use a truncated Karhunen-Loéve expansion of the Gaussian noise in order to satisfy a finite-dimensional noise condition. For kernel-based collocation the same collocation locations are used to construct the deterministic basis and the random part. For the Galerkin finite element method one needs to separately set up the finite element basis on the spatial domain and the polynomials on the probability space. For a given kernel function, the convergence rate of the collocation solution depends only on the fill distance of the collocation points. The convergence rate of the finite element solution depends on the maximum mesh spacing parameter and the degrees of the polynomials defined on the finite-dimensional probability space. According to our numerical experiments, the truncation length of the Gaussian noise also affects the convergence results of the finite element method.

1.1 Problem Setting

Assume that $\mathcal{D}$ is a regular open bounded domain in $\mathbb{R}^d$. Let the stochastic process $\xi : \mathcal{D} \times \Omega_\xi \to \mathbb{R}$ be Gaussian with mean zero and covariance kernel $\Phi : \mathcal{D}\times\mathcal{D} \to \mathbb{R}$ defined on a probability space $(\Omega_\xi, \mathcal{F}_\xi, \mathbb{P}_\xi)$ (see Definition A.1). We consider an elliptic SPDE driven by the Gaussian noise ξ with Dirichlet boundary conditions

$$\begin{cases} \Delta u = f + \xi, & \text{in } \mathcal{D}, \\ u = 0, & \text{on } \partial\mathcal{D}, \end{cases} \tag{1}$$

where Δ is the Laplacian and $f : \mathcal{D} \to \mathbb{R}$ is a deterministic function. We can solve the SPDE (1) by either of the following two numerical methods.

Kernel-based collocation method (KC): We simulate the Gaussian noise ξ with covariance structure $\Phi(\boldsymbol{x}, \boldsymbol{y})$ at a finite collection of predetermined *collocation points*

$$X_{\mathcal{D}} := \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\} \subset \mathcal{D}, \quad X_{\partial\mathcal{D}} := \{\boldsymbol{x}_{N+1}, \cdots, \boldsymbol{x}_{N+M}\} \subset \partial\mathcal{D}$$

and approximate u using a kernel-based collocation method written as

$$u(\boldsymbol{x}) \approx \hat{u}(\boldsymbol{x}) := \sum_{k=1}^{N} c_k \Delta_2 \overset{*}{K}(\boldsymbol{x}, \boldsymbol{x}_k) + \sum_{k=1}^{M} c_{N+k} \overset{*}{K}(\boldsymbol{x}, \boldsymbol{x}_{N+k}), \quad \boldsymbol{x} \in \mathcal{D},$$

where $\overset{*}{K}$ is an integral-type kernel associated with a reproducing kernel K (see Equation (5) in Appendix Appendix A.). Here Δ_2 means that we differentiate with respect to the second argument, i.e., $\Delta_2 \overset{*}{K}(\boldsymbol{x}, \boldsymbol{x}_k) = \Delta_{\boldsymbol{y}} \overset{*}{K}(\boldsymbol{x}, \boldsymbol{y})|_{\boldsymbol{y}=\boldsymbol{x}_k}$. The unknown random coefficients $\boldsymbol{c} := (c_1, \cdots, c_{N+M})^T$ are obtained by solving a random system of linear equations (with constant deterministic system matrix and random right-hand side that varies with each realization of the noise). Details are provided in Section 2.

Galerkin finite element method (FE): Since the Galerkin finite element method is based on a finite-dimensional noise assumption (see [1, 2]), assuming $\Phi \in \mathrm{L}_2(\mathcal{D} \times \mathcal{D})$, we truncate the Gaussian noise ξ by a Karhunen-Loéve expansion, i.e.,

$$\xi \approx \xi^n := \sum_{k=1}^{n} \zeta_k \sqrt{q_k} \phi_k, \quad \text{and} \quad \zeta_k \sim i.i.d.\ \mathcal{N}(0,1), \quad k = 1, \cdots, n,$$

where q_k and ϕ_k are eigenvalues and eigenfunctions of the covariance kernel Φ, i.e., $\int_{\mathcal{D}} \Phi(\boldsymbol{x}, \boldsymbol{y}) \phi_k(\boldsymbol{y}) \mathrm{d}\boldsymbol{y} = q_k \phi_k(\boldsymbol{x})$. We approximate the original SPDE (1) by another elliptic SPDE driven by the truncated Gaussian noise ξ^n

$$\begin{cases} \Delta u^n = f + \xi^n, & \text{in } \mathcal{D}, \\ u^n = 0, & \text{on } \partial\mathcal{D}. \end{cases} \tag{2}$$

Next we combine the finite element method in spatial domain $\overline{\mathcal{D}} := \mathcal{D} \cup \partial\mathcal{D}$ and the collocation in the zeros of suitable tensor product orthogonal polynomials (Gaussian points) in the finite-dimensional probability space. We obtain the approximation as a tensor product of the finite element solutions defined on the spatial domain and the Lagrange polynomials defined on the finite-dimensional probability space, i.e., $u_{h,\boldsymbol{p}} \approx u^n \approx u$, where h is the maximum mesh spacing parameter and $\boldsymbol{p} = (p_1, \cdots, p_n)$ is the degree of the Lagrange polynomials. Details are provided in Section 3.

Remark 1.1. Because Φ is always positive semi-definite Mercer's theorem ensures that its eigenvalues $q_1 \geq q_2 \geq \cdots \geq 0$ and its eigenfunctions $\{\phi_k\}_{k=1}^{\infty}$ form an orthonormal base of $\mathrm{L}_2(\mathcal{D})$ so that $\Phi(\boldsymbol{x}, \boldsymbol{y}) = \sum_{k=1}^{\infty} q_k \phi_k(\boldsymbol{x}) \phi_k(\boldsymbol{y})$. Therefore $E \|\xi - \xi^n\|^2_{\mathrm{L}_2(\mathcal{D})} = \sum_{k=n+1}^{\infty} q_k \to 0$ when $n \to \infty$, and we can accurately represent the infinite-dimensional noise ξ by a (potentially long) truncated Karhunen-Loéve expansion.

2 Kernel-based Collocation Method

In the papers [4, 11] we use the Gaussian fields ΔS, S with means $\Delta\mu$, μ and covariance kernels $\Delta_1\Delta_2\overset{*}{K}$, $\overset{*}{K}$ (see Theorem A.1), respectively, to construct the collocation approximation $\hat{u}$ of the solution u of SPDE (1). Here $\Delta_1\Delta_2\overset{*}{K}(\boldsymbol{x}_j, \boldsymbol{x}_k) = \Delta_{\boldsymbol{x}}\Delta_{\boldsymbol{y}}\overset{*}{K}(\boldsymbol{x}, \boldsymbol{y})|_{\boldsymbol{x}=\boldsymbol{x}_j, \boldsymbol{y}=\boldsymbol{x}_k}$.

Because of the order $\mathcal{O}(\Delta) = 2$, we suppose that the reproducing kernel Hilbert space $\mathrm{H}_K(\mathcal{D})$ is embedded into the L_2-based Sobolev space $\mathcal{H}^m(\mathcal{D})$ where $m > 2 + d/2$.

Remark 2.1. Since we want to interpolate the values of the differential equation at the collocation points, $\Delta\omega(\boldsymbol{x})$ needs to be well-defined pointwise for each available solution $\omega \in \mathrm{H}_K(\mathcal{D}) \subseteq \mathcal{H}^m(\mathcal{D}) \subset \mathrm{C}^2(\overline{\mathcal{D}})$. This requires the Sobolev space $\mathcal{H}^m(\mathcal{D})$ to be smooth enough. If we just need a weak solution as for the finite element method, then the order needs to satisfy $m \geq 2$ only.

Since ξ is Gaussian with a known correlation structure, we can simulate the values of ξ at the collocation points $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$, i.e.,

$$\boldsymbol{\xi} := (\xi_{\boldsymbol{x}_1}, \cdots, \xi_{\boldsymbol{x}_N})^T \sim \mathcal{N}(\mathbf{0}, \Phi), \quad \text{where } \Phi := (\Phi(\boldsymbol{x}_j, \boldsymbol{x}_k))_{j,k=1}^{N,N}.$$

Consequently, we assume that the values $\{y_j\}_{j=1}^N$ and $\{y_{N+j}\}_{j=1}^M$ defined by

$$y_j := f(\boldsymbol{x}_j) + \xi_{\boldsymbol{x}_j}, \quad j = 1, \cdots, N, \quad y_{N+j} := 0, \quad j = 1, \cdots, M,$$

are known. Moreover we can also obtain the joint probability density function $p_{\boldsymbol{y}}$ of the random vector $\boldsymbol{y}_\xi := (y_1, \cdots, y_{N+M})^T$.

We define the product space

$$\Omega_{K\xi} := \Omega_K \times \Omega_\xi, \quad \mathcal{F}_{K\xi} := \mathcal{F}_K \otimes \mathcal{F}_\xi, \quad \mathbb{P}_\xi^\mu := \mathbb{P}^\mu \otimes \mathbb{P}_\xi,$$

where the probability measure $\mathbb{P}^\mu$ is defined on $(\mathrm{H}_K(\mathcal{D}), \mathcal{B}(\mathrm{H}_K(\mathcal{D}))) = (\Omega_K, \mathcal{F}_K)$ as in Theorem A.1, and the probability space $(\Omega_\xi, \mathcal{F}_\xi, \mathbb{P}_\xi)$ is given in the SPDE (1). We assume that the random variables defined on the original probability spaces are extended to random variables on the new probability space in the natural way: if random variables $V_1 : \Omega_K \to \mathbb{R}$ and $V_2 : \Omega_\xi \to \mathbb{R}$ are defined on $(\Omega_K, \mathcal{F}_K, \mathbb{P}^\mu)$ and $(\Omega_\xi, \mathcal{F}_\xi, \mathbb{P}_\xi)$, respectively, then

$$V_1(\omega_1, \omega_2) := V_1(\omega_1),\ V_2(\omega_1, \omega_2) := V_2(\omega_2), \quad \text{for each } \omega_1 \in \Omega_K \text{ and } \omega_2 \in \Omega_\xi.$$

Note that in this case the random variables have the same probability distributional properties, and they are independent on $(\Omega_{K\xi}, \mathcal{F}_{K\xi}, \mathbb{P}_\xi^\mu)$. This implies that the stochastic processes ΔS, S and ξ can be extended to the product space $(\Omega_{K\xi}, \mathcal{F}_{K\xi}, \mathbb{P}_\xi^\mu)$ while preserving the original probability distributional properties, and that $(\Delta S, S)$ and ξ are independent.

2.1 Approximation of SPDEs

Fix any $\boldsymbol{x} \in \mathcal{D}$. Let $\mathcal{A}_{\boldsymbol{x}}(v) := \{\omega_1 \times \omega_2 \in \Omega_{K\xi} : \omega_1(\boldsymbol{x}) = v\}$ for each $v \in \mathbb{R}$, and $\mathcal{A}_{PB}^{\boldsymbol{y}_\xi} := \{\omega_1 \times \omega_2 \in \Omega_{K\xi} : \Delta\omega_1(\boldsymbol{x}_1) = y_1(\omega_2), \ldots, \omega_1(\boldsymbol{x}_{N+M}) = y_{N+M}(\omega_2)\}$. Using the methods in [4] and Theorem A.1, we obtain

$$\mathbb{P}_\xi^\mu(\mathcal{A}_{\boldsymbol{x}}(v) | \mathcal{A}_{PB}^{\boldsymbol{y}_\xi}) = \mathbb{P}_\xi^\mu(S_{\boldsymbol{x}} = v | \boldsymbol{S}_{PB} = \boldsymbol{y}_\xi) = p_{\boldsymbol{x}}^\mu(v | \boldsymbol{y}_\xi),$$

where $p^{\mu}_{\boldsymbol{x}}(\cdot|\cdot)$ is the conditional probability density function of the random variable $S_{\boldsymbol{x}}$ given the random vector $\boldsymbol{S}_{PB} := (\Delta S_{\boldsymbol{x}_1}, \cdots, \Delta S_{\boldsymbol{x}_N}, S_{\boldsymbol{x}_{N+1}}, \cdots, S_{\boldsymbol{x}_{N+M}})^T$. (Here $\boldsymbol{y}_\xi$ is viewed as given values.) According to the natural extension rule, $p^{\mu}_{\boldsymbol{x}}$ is consistent with the formula (6). Then the approximation $\hat{u}(\boldsymbol{x})$ is solved by the maximization problem

$$\hat{u}(\boldsymbol{x}) = \underset{v\in\mathbb{R}}{\operatorname{argmax}} \sup_{\mu\in \mathrm{H}_K(\mathcal{D})} p^{\mu}_{\boldsymbol{x}}(v|\boldsymbol{y}_\xi).$$

If the covariance matrix

$$\overset{*}{\mathsf{K}}_{PB} := \begin{pmatrix} (\Delta_1\Delta_2\overset{*}{K}(\boldsymbol{x}_j,\boldsymbol{x}_k))_{j,k=1}^{N,N}, & (\Delta_1\overset{*}{K}(\boldsymbol{x}_j,\boldsymbol{x}_{N+k}))_{j,k=1}^{N,M} \\ (\Delta_2\overset{*}{K}(\boldsymbol{x}_{N+j},\boldsymbol{x}_k))_{j,k=1}^{M,N}, & (\overset{*}{K}(\boldsymbol{x}_{N+j},\boldsymbol{x}_{N+k}))_{j,k=1}^{M,M} \end{pmatrix} \in \mathbb{R}^{(N+M)\times(N+M)}$$

is nonsingular, then one solution of the above maximum problem has the form

$$\hat{u}(\boldsymbol{x}) := \sum_{k=1}^{N} c_k \Delta_2\overset{*}{K}(\boldsymbol{x},\boldsymbol{x}_k) + \sum_{k=1}^{M} c_{N+k}\overset{*}{K}(\boldsymbol{x},\boldsymbol{x}_{N+k}) = \boldsymbol{k}_{PB}(\boldsymbol{x})^T \overset{*}{\mathsf{K}}_{PB}{}^{-1}\boldsymbol{y}_\xi, \quad (3)$$

where $\boldsymbol{k}_{PB}(\boldsymbol{x}) := (\Delta_2\overset{*}{K}(\boldsymbol{x},\boldsymbol{x}_1), \cdots, \Delta_2\overset{*}{K}(\boldsymbol{x},\boldsymbol{x}_N), \overset{*}{K}(\boldsymbol{x},\boldsymbol{x}_{N+1}), \cdots, \overset{*}{K}(\boldsymbol{x},\boldsymbol{x}_{N+M}))^T$. This means that its random coefficients are obtained from the linear equation system $\overset{*}{\mathsf{K}}_{PB}\boldsymbol{c} = \boldsymbol{y}_\xi$.

The estimator $\hat{u}$ also satisfies the interpolation condition, i.e., $\Delta\hat{u}(\boldsymbol{x}_1) = y_1, \ldots, \Delta\hat{u}(\boldsymbol{x}_N) = y_N$ and $\hat{u}(\boldsymbol{x}_{N+1}) = y_{N+1}, \ldots, \hat{u}(\boldsymbol{x}_{N+M}) = y_{N+M}$. It is obvious that $\hat{u}(\cdot,\omega_2) \in \mathrm{H}_K(\mathcal{D})$ for each $\omega_2 \in \Omega_\xi$. Since the random part of $\hat{u}(\boldsymbol{x})$ is only related to $\boldsymbol{y}_\xi$, we can formally rewrite $\hat{u}(\boldsymbol{x},\omega_2)$ as $\hat{u}(\boldsymbol{x},\boldsymbol{y}_\xi)$ and $\hat{u}(\boldsymbol{x})$ can be transferred to a random variable defined on the finite-dimensional probability space $(\mathbb{R}^{N+M}, \mathcal{B}(\mathbb{R}^{N+M}), \mu_{\boldsymbol{y}})$, where the probability measure $\mu_{\boldsymbol{y}}$ is defined by $\mu_{\boldsymbol{y}}(\mathrm{d}\boldsymbol{v}) := p_{\boldsymbol{y}}(\boldsymbol{v})\mathrm{d}\boldsymbol{v}$. Moreover, the probability distributional properties of $\hat{u}(\boldsymbol{x})$ do not change when $(\Omega_\xi, \mathcal{F}_\xi, \mathbb{P}_\xi)$ is replaced by $(\mathbb{R}^{N+M}, \mathcal{B}(\mathbb{R}^{N+M}), \mu_{\boldsymbol{y}})$.

Remark 2.2. The random coefficients are obtained solving by system of linear equations that is slightly different from [4]. However the main ideas and techniques are the same as in [4]. For this estimator it is easier to derive error bounds and compare with Galerkin finite element method. A lot more details of the relationship between the two different estimators are provided in [11, Chapter 7].

2.2 Convergence Analysis

We assume that $u(\cdot,\omega_2)$ belongs to $\mathrm{H}_K(\mathcal{D})$ almost surely for $\omega_2 \in \Omega_\xi$. Therefore u can be seen as a map from Ω_ξ into $\mathrm{H}_K(\mathcal{D})$. So we have $u \in \Omega_{K\xi} = \Omega_K \times \Omega_\xi$.

We fix any $\boldsymbol{x} \in \mathcal{D}$ and any $\epsilon > 0$. Let the subset

$$\mathcal{E}_{\boldsymbol{x}}^{\epsilon} := \Big\{ \omega_1 \times \omega_2 \in \Omega_{K\xi} : |\omega_1(\boldsymbol{x}) - \hat{u}(\boldsymbol{x}, \omega_2)| \geq \epsilon, \\ \text{such that } \Delta\omega_1(\boldsymbol{x}_1) = y_1(\omega_2), \ldots, \omega_1(\boldsymbol{x}_{N+M}) = y_{N+M}(\omega_2) \Big\}.$$

Because $\Delta S_{\boldsymbol{x}}(\omega_1, \omega_2) = \Delta S_{\boldsymbol{x}}(\omega_1) = \Delta\omega_1(\boldsymbol{x})$, $S_{\boldsymbol{x}}(\omega_1, \omega_2) = S_{\boldsymbol{x}}(\omega_1) = \omega_1(\boldsymbol{x})$ and $\boldsymbol{y}_{\boldsymbol{\xi}}(\omega_1, \omega_2) = \boldsymbol{y}_{\boldsymbol{\xi}}(\omega_2)$ for each $\omega_1 \in \Omega_K$ and $\omega_2 \in \Omega_\xi$ (see Theorem A.1) we can deduce that

$$\begin{aligned} \mathbb{P}_{\xi}^{\mu}\left(\mathcal{E}_{\boldsymbol{x}}^{\epsilon}\right) &= \mathbb{P}_{\xi}^{\mu}\left(|S_{\boldsymbol{x}} - \hat{u}(\boldsymbol{x})| \geq \epsilon \text{ such that } \boldsymbol{S}_{PB} = \boldsymbol{y}_{\xi}\right) \\ &= \int_{\mathbb{R}^{N+M}} \int_{|v - \hat{u}(\boldsymbol{x}, \boldsymbol{v})| \geq \epsilon} p_{\boldsymbol{x}}^{\mu}(v|\boldsymbol{v}) p_{\boldsymbol{y}}(\boldsymbol{v}) \mathrm{d}v \mathrm{d}\boldsymbol{v} \\ &= \int_{\mathbb{R}^{N+M}} \operatorname{erfc}\left(\frac{\epsilon}{\sqrt{2}\sigma(\boldsymbol{x})}\right) p_{\boldsymbol{y}}(\boldsymbol{v}) \mathrm{d}\boldsymbol{v} = \operatorname{erfc}\left(\frac{\epsilon}{\sqrt{2}\sigma(\boldsymbol{x})}\right), \end{aligned}$$

where the variance of $p_{\boldsymbol{x}}^{\mu}$ is $\sigma(\boldsymbol{x})^2 = \overset{*}{K}(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}_{PB}(\boldsymbol{x})^T \overset{*}{\mathsf{K}}_{PB}{}^{-1} \boldsymbol{k}_{PB}(\boldsymbol{x})$ (see Equation (6) given in Appendix Appendix A.).

The reader may note that the form of the expression for the variance $\sigma(\boldsymbol{x})^2$ is analogous to that of the *power function* [5, 10], and we can therefore use the same techniques as in the proofs from [4, 5, 10, 11] to obtain a formula for the order of $\sigma(\boldsymbol{x})$, i.e.,

$$\sigma(\boldsymbol{x}) = \mathcal{O}(h_X^{m-2-d/2}),$$

where $h_X = \sup_{\boldsymbol{x} \in \mathcal{D}} \min_{\boldsymbol{x}_j \in X_{\mathcal{D}} \cup X_{\partial\mathcal{D}}} \|\boldsymbol{x} - \boldsymbol{x}_j\|_2$ is the *fill distance* of $X := X_{\mathcal{D}} \cup X_{\partial\mathcal{D}}$. This implies that

$$\sup_{\mu \in \mathrm{H}_K(\mathcal{D})} \mathbb{P}_{\xi}^{\mu}(\mathcal{E}_{\boldsymbol{x}}^{\epsilon}) = \mathcal{O}\left(\frac{h_X^{m-2-d/2}}{\epsilon}\right).$$

Because $|u(\boldsymbol{x}, \omega_2) - \hat{u}(\boldsymbol{x}, \omega_2)| \geq \epsilon$ if and only if $u \in \mathcal{E}_{\boldsymbol{x}}^{\epsilon}$ we conclude that

$$\sup_{\mu \in \mathrm{H}_K(\mathcal{D})} \mathbb{P}_{\xi}^{\mu}\left(\|u - \hat{u}\|_{\mathrm{L}_\infty(\mathcal{D})} \geq \epsilon\right) \leq \sup_{\mu \in \mathrm{H}_K(\mathcal{D}), \boldsymbol{x} \in \mathcal{D}} \mathbb{P}_{\xi}^{\mu}\left(\mathcal{E}_{\boldsymbol{x}}^{\epsilon}\right) \to 0, \text{ when } h_X \to 0.$$

Therefore we say that the estimator $\hat{u}$ converges to the exact solution u of the SPDE (1) in all probabilities $\mathbb{P}_{\xi}^{\mu}$ when h_X goes to 0.

Sometimes we know only that the solution $u \in \mathcal{H}^m(\mathcal{D})$. In this case, as long as the reproducing kernel Hilbert space is dense in the Sobolev space $\mathcal{H}^m(\mathcal{D})$ with respect to its Sobolev norm, we can still say that $\hat{u}$ converges to u in probability.

3 Galerkin Finite Element Method

The right hand side of the SPDE (2)

$$f_{\xi^n}(\boldsymbol{x},\boldsymbol{\zeta}) := f(\boldsymbol{x}) + \xi^n_{\boldsymbol{x}} = f(\boldsymbol{x}) + \sum_{k=1}^{n} \zeta_k \sqrt{q_k}\phi_k(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathcal{D},$$

and the random vector $\boldsymbol{\zeta} := (\zeta_1, \cdots, \zeta_n)^T$ has the joint standard normal density function

$$\rho_n(\boldsymbol{z}) := \prod_{k=1}^{n} \rho(z_k), \quad \boldsymbol{z} \in \mathbb{R}^n, \quad \text{where } \rho(z) := \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

Therefore we can replace the probability space $(\Omega_\xi, \mathcal{F}_\xi, \mathbb{P}_\xi)$ by a finite-dimensional probability space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu_{\boldsymbol{\zeta}})$ such that u^n and ξ^n have the same probability distributional properties on both probability spaces, where the probability measure $\mu_{\boldsymbol{\zeta}}$ is defined by $\mu_{\boldsymbol{\zeta}}(\mathrm{d}\boldsymbol{z}) := \rho_n(\boldsymbol{z})\mathrm{d}\boldsymbol{z}$.

In the paper [1] the numerical approximation $u_{h,\boldsymbol{p}}$ of the solution u^n of the SPDE (2) is sought in a finite-dimensional subspace $V_{h,\boldsymbol{p}}$ based on a tensor product, $V_{h,\boldsymbol{p}} := \mathrm{H}_h(\mathcal{D}) \otimes \mathcal{P}_{\boldsymbol{p}}(\mathbb{R}^n)$, where the following hold:

C1) $\mathrm{H}_h(\mathcal{D}) \subset \mathcal{H}^1_0(\mathcal{D})$ is a standard finite element space, which contains continuous piecewise polynomials defined on regular triangulations with a maximum mesh spacing parameter h.

C2) $\mathcal{P}_{\boldsymbol{p}}(\mathbb{R}^n) := \otimes_{k=1}^n \mathcal{P}_{p_k}(\mathbb{R}) \subset \mathrm{L}_{2,\rho_n}(\mathbb{R}^n)$ is the span of the tensor product of polynomials with degree at most $\boldsymbol{p} = (p_1, \cdots, p_n)$, where $\mathcal{P}_{p_k}(\mathbb{R})$ is a space of univariate polynomials of degree p_k for each $k = 1, \ldots, n$.

Thus the approximation $u_{h,\boldsymbol{p}} \in V_{h,\boldsymbol{p}}$ and $u_{h,\boldsymbol{p}}(\boldsymbol{x})$ is a random variable defined on the finite-dimensional probability space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu_{\boldsymbol{\zeta}})$.

Next we construct the Gaussian points

$$\boldsymbol{z}_{\boldsymbol{j}} := (z_{1,j_1}, \cdots, z_{n,j_n})^T, \quad \boldsymbol{j} \in \mathcal{N}_{\boldsymbol{p}} := \{\boldsymbol{j} \in \mathbb{N}^n : 1 \le j_k \le p_k + 1,\ k = 1, \ldots, n\},$$

where $z_{k,1}, \ldots, z_{k,p_k+1}$ are the roots of the Hermite polynomials η_{p_k+1} of degree $p_k + 1$ for each dimension $k = 1, \ldots, n$. The η_{p_k+1} are also orthogonal polynomials on the space $\mathcal{P}_{p_k}(\mathbb{R})$ with respect to a standard normal weight ρ. Here these Hermite polynomials are only used to set up the Gaussian points for approximating the Gaussian fields.

Let a polynomial base of $\mathcal{P}_{\boldsymbol{p}}(\mathbb{R}^n)$ be

$$l_{\boldsymbol{j}}(\boldsymbol{z}) := \prod_{k=1}^{n} l_{k,j_k}(z_k), \quad \boldsymbol{j} \in \mathcal{N}_{\boldsymbol{p}},$$

where $\{l_{k,j}\}_{j=1}^{p_k+1}$ is the Lagrange basis of $\mathcal{P}_{p_k}(\mathbb{R})$ for each $k = 1, \ldots, n$, i.e.,

$$l_{k,j} \in \mathcal{P}_{p_k}(\mathbb{R}), \quad l_{k,j}(z_{k,i}) = \delta_{ij}, \quad i, j = 1, \ldots, p_k + 1,$$

and δ_{ij} is the Kronecker symbol.

For each Gaussian point $\boldsymbol{z}_{\boldsymbol{j}} \in \mathcal{N}_{\boldsymbol{p}}$, we compute the finite element solution $u_h(\cdot, \boldsymbol{z}_{\boldsymbol{j}}) \in \mathrm{H}_h(\mathcal{D})$ of the equation

$$-\int_{\mathcal{D}} \nabla u_h(\boldsymbol{x}, \boldsymbol{z_j}) \nabla \gamma(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \int_{\mathcal{D}} f_{\xi^n}(\boldsymbol{x}, \boldsymbol{z_j}) \gamma(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \quad \text{for any } \gamma \in \mathrm{H}_h(\mathcal{D}).$$

The approximation $u_{h,\boldsymbol{p}}$ is the tensor product of the finite element solutions and the Lagrange polynomials, i.e.,

$$u_{h,\boldsymbol{p}}(\boldsymbol{x}, \boldsymbol{\zeta}) := \sum_{\boldsymbol{j} \in \mathcal{N}_{\boldsymbol{p}}} u_h(\boldsymbol{x}, \boldsymbol{z_j}) l_{\boldsymbol{j}}(\boldsymbol{\zeta}), \quad \boldsymbol{x} \in \mathcal{D}. \tag{4}$$

This indicates that $u_{h,\boldsymbol{p}}$ is interpolating at all Gaussian points $\boldsymbol{z_j} \in \mathcal{N}_{\boldsymbol{p}}$.

We assume that u^n belongs to $\mathrm{L}_{2,\rho_n}(\mathbb{R}^n) \otimes \mathcal{H}_0^1(\mathcal{D})$. According to [1, Theorem 4.1] and [1, Lemma 4.7] we get the error bound

$$\begin{aligned}
&\|E_{\rho_n}(u^n - u_{h,\boldsymbol{p}})\|_{\mathcal{H}_0^1(\mathcal{D})} \\
\leq & C_1 \inf_{w \in \mathrm{L}_{2,\rho_n}(\mathbb{R}^n) \otimes \mathrm{H}_h(\mathcal{D})} \left(\int_{\mathcal{D}} E_{\rho_n} \left| \nabla u^n(\boldsymbol{x}) - \nabla w(\boldsymbol{x}) \right|^2 \mathrm{d}\boldsymbol{x} \right)^{1/2} + C_2 \sum_{k=1}^{n} p_k^{3/2} e^{-r_k p_k^{1/2}}
\end{aligned}$$

with positive constants $r_1, \ldots, r_n$ and C_1, C_2 independent of h and $\boldsymbol{p}$.

4 Side-by-Side Comparison of Both Methods

4.1 Differences

- **Probability spaces:** For kernel-based collocation (KC) we transfer the probability space $(\Omega_\xi, \mathcal{F}_\xi, \mathbb{P}_\xi)$ to the tensor product probability space $(\Omega_{K\xi}, \mathcal{F}_{K\xi}, \mathbb{P}_\xi^\mu)$ such that the Gaussian noise ξ has the same probability distributional properties defined on both probability spaces, while for Galerkin finite elements (FE) we approximate the Gaussian noise ξ by the truncated Gaussian noise ξ^n such that $\lim_{n\to\infty} E\left\|\xi - \xi^n\right\|_{\mathrm{L}_2(\mathcal{D})} = 0$ and ξ^n has the same probability distributional properties on the probability space $(\Omega_\xi, \mathcal{F}_\xi, \mathbb{P}_\xi)$ and the finite-dimensional probability space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu_{\boldsymbol{\zeta}})$
- **Basis functions:** The bases of the KC solution $\hat{u}$ are the kernel functions $\Delta_2 \overset{*}{K}$ and $\overset{*}{K}$ centered at the collocation points $X_{\mathcal{D}} \subset \mathcal{D}$ and $X_{\partial\mathcal{D}} \subset \partial\mathcal{D}$, while the bases of the FE solution $u_{h,\boldsymbol{p}}$ are the tensor products of the triangular finite element bases defined on $\mathcal{D}$ and the Lagrange polynomials defined on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu_{\boldsymbol{\zeta}})$.
- **Simulation:** For KC we can simulate the Gaussian noise ξ at the collocation points $X_{\mathcal{D}}$ because we know its covariance kernel Φ, i.e., $\boldsymbol{\xi} = (\xi_{\boldsymbol{x}_1}, \cdots, \xi_{\boldsymbol{x}_N})^T \sim \mathcal{N}(\mathbf{0}, \Phi)$ and $\Phi = (\Phi(\boldsymbol{x}_j, \boldsymbol{x}_k))_{j,k=1}^{N,N}$. For FE we can simulate the random vector $\boldsymbol{\zeta} = (\zeta_1, \cdots, \zeta_2)^T \sim \mathcal{N}(\mathbf{0}, I_n)$ in order to introduce random variables on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu_{\boldsymbol{\zeta}})$.
- **Interpolation:** In KC $\Delta\hat{u}$ and $\hat{u}$ are interpolating at the collocation points $X_{\mathcal{D}} \cup X_{\partial\mathcal{D}} \subset \mathcal{D}$ in the domain space, respectively, while in FE $u_{h,\boldsymbol{p}}$ is interpolating at the Gaussian points $\mathcal{N}_{\boldsymbol{p}} \subset \mathbb{R}^n$ in the probability space.

- **Function spaces:** For KC, $\hat{u} \in \operatorname{span}\{\overset{*}{\Delta_2 K}(\cdot, \boldsymbol{x}_j), \overset{*}{K}(\cdot, \boldsymbol{x}_{N+k})\}_{j,k=1}^{N,M} \otimes \mathcal{P}_1(\mathbb{R}^{N+M}) \subset \mathrm{H}_K(\mathcal{D}) \otimes \mathrm{L}_{2,p_{\boldsymbol{y}}}(\mathbb{R}^{N+M})$, while for FE we have $u_{h,\boldsymbol{p}} \in \mathrm{H}_h(\mathcal{D}) \otimes \mathcal{P}_{\boldsymbol{p}}(\mathbb{R}^n) \subset \mathcal{H}_0^1(\mathcal{D}) \otimes \mathrm{L}_{2,\rho_n}(\mathbb{R}^n)$.
- **Approximation properties:** The KC result $\hat{u}$ approximates the solution u of the SPDE (1) and its convergence rate depends on the fill distance h_X of the collocation points, while the FE result $u_{h,\boldsymbol{p}}$ approximates the truncated solution u^n of the SPDE (2) with a convergence rate that depends on the maximum mesh spacing parameter h of the triangulation and the degree $\boldsymbol{p}$ of the Lagrange polynomials.

4.2 Relationship Between the Two Methods

Roughly speaking, the random parts of $\hat{u}$ and $u_{h,\boldsymbol{p}}$ are simulated by the normal random vectors $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$, respectively.

For the following we assume that Φ is positive definite on $\mathcal{D}$ and the dimensions of $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ are the same, i.e., $N = n$.

We firstly show the relationship between $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$. Since Φ is positive definite, we have the decomposition $\Phi = \mathsf{V}\mathsf{D}\mathsf{V}^T$, where D and V are the eigenvalues and eigenvector matrices of Φ, respectively. Therefore

$$\boldsymbol{\xi} \sim \mathsf{V}\mathsf{D}^{1/2}\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \Phi), \quad \boldsymbol{\zeta} \sim \mathsf{D}^{-1/2}\mathsf{V}^T\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, I_N).$$

We can also use $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ to approximate the Gaussian noise $\xi_{\boldsymbol{x}}$ for any fixed $\boldsymbol{x} \in \mathcal{D}$. Using simple kriging, we let

$$\hat{\xi}_{\boldsymbol{x}} := \boldsymbol{c}(\boldsymbol{x})^T\boldsymbol{\xi} = \boldsymbol{b}(\boldsymbol{x})^T\Phi^{-1}\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{b}(\boldsymbol{x})^T\Phi^{-1}\boldsymbol{b}(\boldsymbol{x})),$$

where $\boldsymbol{b}(\boldsymbol{x}) := (\Phi(\boldsymbol{x}, \boldsymbol{x}_1), \cdots, \Phi(\boldsymbol{x}, \boldsymbol{x}_N))^T$. According to [9],

$$E\left|\xi_{\boldsymbol{x}} - \hat{\xi}_{\boldsymbol{x}}\right|^2 = \Phi(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{b}(\boldsymbol{x})^T\Phi^{-1}\boldsymbol{b}(\boldsymbol{x}) = \mathcal{O}(h_{X_{\mathcal{D}}}^k),$$

when $\Phi \in \mathrm{C}^{2k}(\mathcal{D} \times \mathcal{D})$ and $h_{X_{\mathcal{D}}}$ is the fill distance of $X_{\mathcal{D}}$. For the Karhunen-Loéve expansion,

$$\xi_{\boldsymbol{x}}^N = \sum_{j=1}^{N} \sqrt{q_j}\phi_j(\boldsymbol{x})\zeta_j = \boldsymbol{\phi}(\boldsymbol{x})^T\mathsf{Q}^{1/2}\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\phi}(\boldsymbol{x})^T\mathsf{Q}\boldsymbol{\phi}(\boldsymbol{x})),$$

where $\boldsymbol{\phi}(\boldsymbol{x}) := (\phi_1(\boldsymbol{x}), \cdots, \phi_N(\boldsymbol{x}))^T$ and $\mathsf{Q} = \operatorname{diag}(q_1, \cdots, q_N)$. We also have

$$E\int_{\mathcal{D}} \left|\xi_{\boldsymbol{x}} - \xi_{\boldsymbol{x}}^N\right|^2 \mathrm{d}\boldsymbol{x} = \sum_{j=N+1}^{\infty} q_j.$$

This shows that the kernel-based method and the finite element method, respectively, use a kernel basis and a spectral basis to approximate Gaussian

fields. It also shows that we should suitably choose collocation points such that

$$\lim_{N\to\infty} h_{X_{\mathcal{D}}} = 0 \quad \Longrightarrow \quad \lim_{N\to\infty} \hat{\xi} = \lim_{N\to\infty} \xi^N = \xi.$$

Usually, the smoothness of Φ is related to its eigenvalues in the sense that the order k of continuous differentiability becomes large when the eigenvalues q_j decrease fast, e.g., $\Phi(x,y) = \sum_{j=1}^{\infty}(2\pi j)^{-2k}\sin(2\pi jx)\sin(2\pi jy)$.

Following these discussions, when the eigenvalues of the covariance kernel Φ decrease fast, then the Galerkin finite element method seems to be preferable to the kernel-based collocation method because we can truncate the Gaussian noise ξ at a low dimension. However, when the eigenvalues change slowly, then we may use the kernel-based collocation method because we are able to directly simulate ξ by its covariance structure.

4.3 Competitive Advantages

- The kernel-based collocation method is a meshfree approximation method. It does not require an underlying triangular mesh as the Galerkin finite element method does. For both methods, the (collocation) points can be placed at rather arbitrarily scattered locations which allows for the use of either deterministic or random designs, e.g., Halton or Sobol' points.
- The kernel-based collocation method can be applied to a high-dimensional domain $\mathcal{D}$ with complex boundary $\partial\mathcal{D}$. We can also generalize it to solve a system of elliptic SPDEs derived by vector Gaussian noises.
- The collocation method requires the SPDE solution to be smooth enough such that interpolation is well-behaved at each collocation point while the finite element method can solve non-smooth problems.
- The interpolation matrix $\overset{*}{K}_{PB}$ for the collocation method is usually a dense (and sometimes ill-conditioned) matrix. The finite element method, on the other hand, usually gets the solutions by a sparse linear system because its basis consists of local elements.
- To obtain the truncated Gaussian noise ξ^n for the finite element method we need to compute the eigenvalues and eigenfunctions of the covariance kernel Φ. This, however, is usually difficult to do, so one must estimate them. For the collocation method we need not worry about this issue.
- If the truncated dimension n for the finite element solutions is large, then the degree $\boldsymbol{p}$ of the polynomials has to become correspondingly large in order to satisfy a given error tolerance (see [2]). Once the kernel functions are fixed, the error of the collocation solution only depends on the collocation points.
- In the finite element method, the dimension of its polynomial space defined on the finite-dimensional probability space is equal to $n_{\boldsymbol{p}} = \prod_{k=1}^{n}(p_k+1)$. So we need to compute $n_{\boldsymbol{p}}$ deterministic finite element solutions. In the collocation method, we need to simulate the N-dimensional nonstandard

normal vector. Therefore, when $n \ll N$ we may choose the finite element method, while vice versa the collocation method may be preferable.

- For various covariance kernels Φ for the Gaussian noise ξ, the choice of reproducing kernels can affect the kernel-based collocation solution. How to choose the "best" kernels is still an open question. Polynomials are used to construct the approximations for the finite element method, and we only need to determine the appropriate polynomial degree.
- The paper [1] also discusses any other tensor-product finite dimensional noise. In the papers [4, 11] we only consider Gaussian noises. However, one may generalize the kernel-based collocation idea to other problems with colored noise.
- The finite element method works for any elliptic SPDE whose differential operator contains stochastic coefficients defined on a finite dimensional probability space. For the collocation method this idea requires further study.

5 Numerical Examples

In this section we present a few simple numerical experiments comparing the kernel-based collocation method to the Galerkin finite element method.

Let the domain $\mathcal{D} := (0,1)^2 \subset \mathbb{R}^2$ and the covariance kernel of the (finite-dimensional) noise be

$$\begin{aligned}\Phi(\boldsymbol{x},\boldsymbol{y}) :=& 4\pi^4 \sin(\pi x_1)\sin(\pi x_2)\sin(\pi y_1)\sin(\pi y_2)\\ &+ 16\pi^4 \sin(2\pi x_1)\sin(2\pi x_2)\sin(2\pi y_1)\sin(2\pi y_2)\end{aligned}$$

so that we are able to demonstrate the effects of a "correct" and "incorrect" truncation dimension for the finite element method. We use the deterministic function

$$f(\boldsymbol{x}) := -2\pi^2 \sin(\pi x_1)\sin(\pi x_2) - 8\pi^2 \sin(2\pi x_1)\sin(2\pi x_2)$$

and the Gaussian noise ξ with the covariance kernel Φ to set up the right hand side of the stochastic Poisson equation with Dirichlet boundary condition as in SPDE (1). Then its solution has the form

$$\begin{aligned}u(\boldsymbol{x}) :=& \sin(\pi x_1)\sin(\pi x_2) + \sin(2\pi x_1)\sin(2\pi x_2) + \zeta_1 \sin(\pi x_1)\sin(\pi x_2)\\ &+ \frac{\zeta_2}{2}\sin(2\pi x_1)\sin(2\pi x_2), \quad \boldsymbol{x} = (x_1, x_2) \in \mathcal{D},\end{aligned}$$

where ζ_1, ζ_2 are independent standard normal random variables defined on $(\Omega_\xi, \mathcal{F}_\xi, \mathbb{P}_\xi)$, i.e., $\zeta_1, \zeta_2 \sim$ i.i.d. $\mathcal{N}(0,1)$.

For the collocation methods, we use the C^4-Matérn function with shape parameter $\theta > 0$

$$g_\theta(r) := (3 + 3\theta r + \theta^2 r^2)e^{-\theta r}, \quad r > 0,$$

to construct the reproducing kernel (Sobolev-spline kernel)

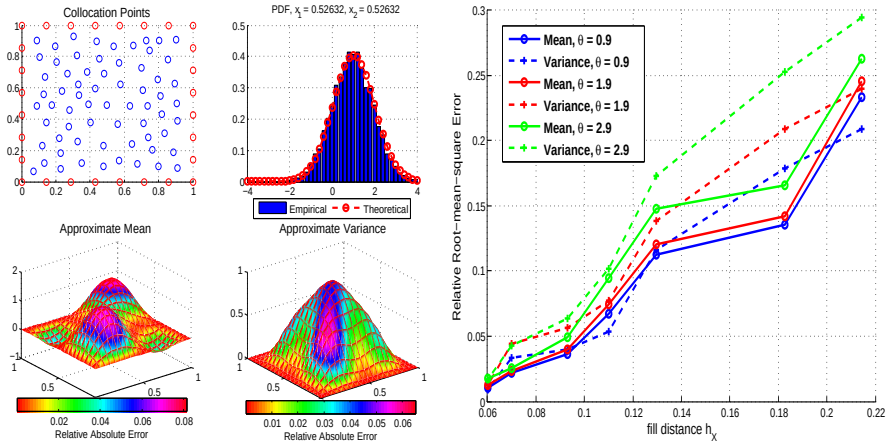
$$K_\theta(\boldsymbol{x},\boldsymbol{y}) := g_\theta(\|\boldsymbol{x}-\boldsymbol{y}\|_2), \quad \boldsymbol{x},\boldsymbol{y}\in\mathcal{D}.$$

According to [6], we can deduce that its reproducing kernel Hilbert space $\mathrm{H}_K(\mathcal{D})$ is equivalent to the L_2-based Sobolev space $\mathcal{H}^{3+1/2}(\mathcal{D}) \subset \mathrm{C}^2(\overline{\mathcal{D}})$. Then we can compute the integral-type $\overset{*}{K}_\theta(\boldsymbol{x},\boldsymbol{y}) = \int_0^1\int_0^1 K_\theta(\boldsymbol{x},\boldsymbol{z})K_\theta(\boldsymbol{y},\boldsymbol{z})\mathrm{d}z_1\mathrm{d}z_2$. Next we choose Halton points in $\mathcal{D}$ and uniform grid points on $\partial\mathcal{D}$ as collocation points. Using the kernel-based collocation method, we can set up the approximation $\hat{u}$ via formula (3).

Since we chose $\xi_{\boldsymbol{x}} = \zeta_1 2\pi^2\sin(\pi x_1)\sin(\pi x_2) + \zeta_2 4\pi^2\sin(2\pi x_1)\sin(2\pi x_2)$, we can let the dimension of the probability space be either $n=1$ or $n=2$ for the finite element method. The Gaussian points $z_{j,1},\cdots,z_{j,p_j+1}\in\mathbb{R}$ are computed as the roots of the Hermite polynomial of degree p_j+1, $j=1,2$. For $n=1$ we have $\xi^1 = \zeta_1 2\pi^2\sin(\pi x_1)\sin(\pi x_2)$, while $n=2$ gives $\xi^2=\xi$.

We approximate the mean and variance of the arbitrary random variable U by its sample mean and sample variance based on $s := 10000$ simulated sample paths using the above algorithm, i.e.,

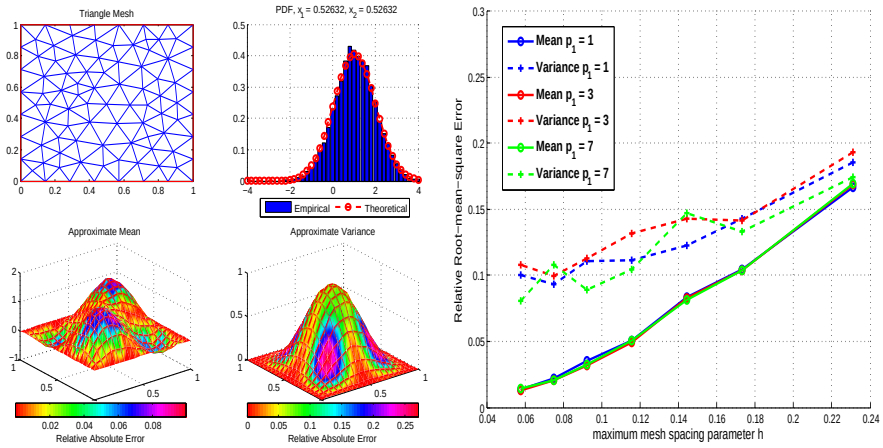
$$E(U) \approx \frac{1}{s}\sum_{k=1}^{s} U(\omega_k), \quad \mathrm{Var}(U) \approx \frac{1}{s}\sum_{k=1}^{s}\left(U(\omega_k) - \frac{1}{s}\sum_{j=1}^{s}U(\omega_j)\right)^2.$$



$N = 65$, $M = 28$ and $\theta = 0.9$ Convergence of Mean and Variance

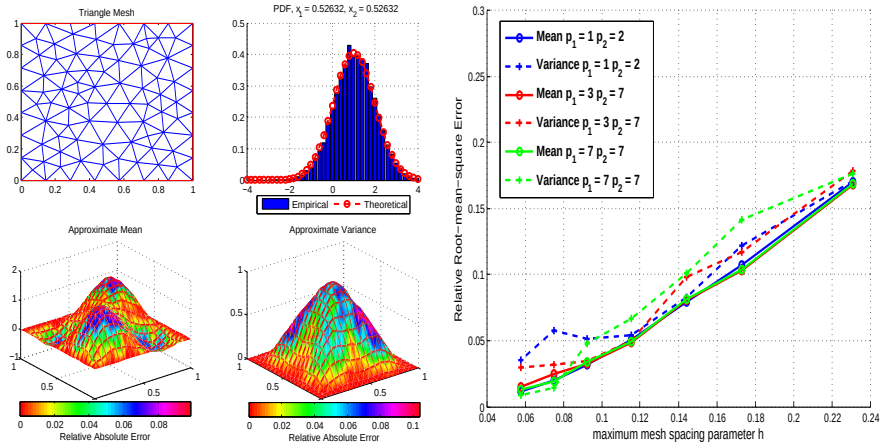
Figure 1. Kernel-based Collocation Method.

According to the numerical results (see Figures 1, 2 and 3), the approximate probability density functions are well-behaved for both numerical methods. The means and variances of the kernel-based collocation solutions are



$h = 0.15$ and $p_1 = 7$ Convergence of Mean and Variance

Figure 2. Galerkin Finite Element Methods, $n = 1$.



$h = 0.15$, $p_1 = 7$ and $p_2 = 3$ Convergence of Mean and Variance

Figure 3. Galerkin Finite Element Methods, $n = 2$.

smooth estimators, while the means and variances of the finite element solutions are piecewise smooth estimators. If we suitably truncate the Gaussian noise, then the finite element method has the same convergence rate as the kernel-based collocation method. If not, the kernel-based collocation method seems to do better than the finite element method for the variance.

Appendix A. Reproducing Kernels and Gaussian Fields

Let $K : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ be a reproducing kernel and $\mathrm{H}_K(\mathcal{D})$ be its reproducing-kernel Hilbert space (see [10, Definition 10.1]). If $K \in \mathrm{C}(\overline{\mathcal{D}} \times \overline{\mathcal{D}})$ then its *integral-type* kernel $\overset{*}{K}$ is defined by

$$\overset{*}{K}(\boldsymbol{x}, \boldsymbol{y}) := \int_{\mathcal{D}} K(\boldsymbol{x}, \boldsymbol{z}) K(\boldsymbol{y}, \boldsymbol{z}) \mathrm{d}\boldsymbol{z}, \quad \boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}. \tag{5}$$

Remark A.1. As in the proof of [4, Lemma 2.2], in order to "match the spaces", any other kernel that "dominates" K (in the sense of [8]) could play the role of the integral-type kernel $\overset{*}{K}$.

Let $\mathcal{H}^m(\mathcal{D})$ be the *classical* L_2*-based Sobolev space* of order $m \in \mathbb{N}$. The *differential operator* $P : \mathcal{H}^m(\mathcal{D}) \to \mathrm{L}_2(\mathcal{D})$ and the *boundary operator* $B : \mathcal{H}^m(\mathcal{D}) \to \mathrm{L}_2(\partial\mathcal{D})$ are linear combinations of the derivatives D^α, $\alpha \in \mathbb{N}_0^d$, with nonconstant coefficients defined on $\mathcal{D}$ and $\partial\mathcal{D}$ respectively, i.e., $P = \sum_{|\alpha| \le m} c_\alpha D^\alpha$ and $B = \sum_{|\alpha| \le m-1} b_\alpha D^\alpha|_{\partial\mathcal{D}}$ where $c_\alpha \in \mathrm{C}(\mathcal{D})$ and $b_\alpha \in \mathrm{C}(\partial\mathcal{D})$. Their orders are given by $\mathcal{O}(P) := m$ and $\mathcal{O}(B) := m - 1$.

Definition A.1 ([3, Definition 3.28]). *A stochastic process $S : \mathcal{D} \times \Omega \to \mathbb{R}$ is said to be* Gaussian *with mean $\mu : \mathcal{D} \to \mathbb{R}$ and covariance kernel $\Phi : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ if, for any pairwise distinct points $X_\mathcal{D} := \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\} \subset \mathcal{D}$, the random vector $\boldsymbol{S} := (S_{\boldsymbol{x}_1}, \cdots, S_{\boldsymbol{x}_N})^T$ is a multi-normal random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Phi}$, i.e., $\boldsymbol{S} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Phi})$, where $\boldsymbol{\mu} := (\mu(\boldsymbol{x}_1), \cdots, \mu(\boldsymbol{x}_N))^T$ and $\boldsymbol{\Phi} := (\Phi(\boldsymbol{x}_j, \boldsymbol{x}_k))_{j,k=1}^{N,N}$.*

We view the reproducing-kernel Hilbert space $\mathrm{H}_K(\mathcal{D})$ as a *sample space* and its Borel σ-field $\mathcal{B}(\mathrm{H}_K(\mathcal{D}))$ as a σ*-algebra* to set up the probability spaces.

Theorem A.1 ([4, Theorem 3.1]). *Suppose that the reproducing kernel Hilbert space $\mathrm{H}_K(\mathcal{D})$ is embedded into the Sobolev space $\mathcal{H}^m(\mathcal{D})$ with $m > d/2$. Further assume that the differential operator P and the boundary operator B have the orders $\mathcal{O}(P) < m - d/2$ and $\mathcal{O}(B) < m - d/2$. Given a function $\mu \in \mathrm{H}_K(\mathcal{D})$ there exists a probability measure $\mathbb{P}^\mu$ defined on $(\Omega_K, \mathcal{F}_K) = (\mathrm{H}_K(\mathcal{D}), \mathcal{B}(\mathrm{H}_K(\mathcal{D})))$ such that the stochastic processes PS, BS given by*

$$\begin{aligned} PS_{\boldsymbol{x}}(\omega) = PS(\boldsymbol{x}, \omega) &:= (P\omega)(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathcal{D} \subset \mathbb{R}^d, \quad \omega \in \Omega_K = \mathrm{H}_K(\mathcal{D}), \\ BS_{\boldsymbol{x}}(\omega) = BS(\boldsymbol{x}, \omega) &:= (B\omega)(\boldsymbol{x}), \quad \boldsymbol{x} \in \partial\mathcal{D}, \quad \omega \in \Omega_K = \mathrm{H}_K(\mathcal{D}), \end{aligned}$$

are Gaussian fields with means $P\mu$, $B\mu$ and covariance kernels $P_1 P_2 \overset{*}{K}$, $B_1 B_2 \overset{*}{K}$ defined on $(\Omega_K, \mathcal{F}_K, \mathbb{P}^\mu)$, respectively, where

$$P_1P_2\overset{*}{K}(\boldsymbol{x},\boldsymbol{y}) := P_{\boldsymbol{z}_1}P_{\boldsymbol{z}_2}\overset{*}{K}(\boldsymbol{z}_1,\boldsymbol{z}_2)|_{\boldsymbol{z}_1=\boldsymbol{x},\boldsymbol{z}_2=\boldsymbol{y}}, \quad \boldsymbol{x},\boldsymbol{y}\in\mathcal{D},$$
$$B_1B_2\overset{*}{K}(\boldsymbol{x},\boldsymbol{y}) := B_{\boldsymbol{z}_1}B_{\boldsymbol{z}_2}\overset{*}{K}(\boldsymbol{z}_1,\boldsymbol{z}_2)|_{\boldsymbol{z}_1=\boldsymbol{x},\boldsymbol{z}_2=\boldsymbol{y}}, \quad \boldsymbol{x},\boldsymbol{y}\in\partial\mathcal{D}.$$

When $P := I$ then we rewrite $PS = S$ which indicates that $S_{\boldsymbol{x}}(\omega) = \omega(\boldsymbol{x})$.

Remark A.2. The probability measure $\mathbb{P}^\mu$ defined in Theorem A.1 can be verified to be *Gaussian* (see [3, 4, 11]). The papers [6, 7, 11] show what kinds of reproducing kernels satisfy the conditions that their reproducing-kernel Hilbert spaces are embedded into the Sobolev spaces. One typical example is the Sobolev spline kernel computed by the Matérn function.

Given $X_{\mathcal{D}} := \{\boldsymbol{x}_j\}_{j=1}^N \subset \mathcal{D}$ and $X_{\partial\mathcal{D}} := \{\boldsymbol{x}_{N+j}\}_{j=1}^M \subset \partial\mathcal{D}$, [4, Corollary 3.2] shows that the random vector $\boldsymbol{S}_{PB} := (PS_{\boldsymbol{x}_1},\cdots,BS_{\boldsymbol{x}_{N+M}})^T$ defined on $(\Omega_K,\mathcal{F}_K,\mathbb{P}^\mu)$ has a *multi-normal* distribution with mean $\boldsymbol{m}_{PB}^\mu$ and covariance matrix $\overset{*}{\mathsf{K}}_{PB}$, i.e.,

$$\boldsymbol{S}_{PB} \sim \mathcal{N}(\boldsymbol{m}_{PB}^\mu, \overset{*}{\mathsf{K}}_{PB}),$$

where $\boldsymbol{m}_{PB}^\mu := (P\mu(\boldsymbol{x}_1),\cdots,P\mu(\boldsymbol{x}_N),B\mu(\boldsymbol{x}_{N+1}),\cdots,B\mu(\boldsymbol{x}_{N+M}))^T$ and

$$\overset{*}{\mathsf{K}}_{PB} := \begin{pmatrix} (P_1P_2\overset{*}{K}(\boldsymbol{x}_j,\boldsymbol{x}_k))_{j,k=1}^{N,N}, & (P_1B_2\overset{*}{K}(\boldsymbol{x}_j,\boldsymbol{x}_{N+k}))_{j,k=1}^{N,M} \\ (B_1P_2\overset{*}{K}(\boldsymbol{x}_{N+j},\boldsymbol{x}_k))_{j,k=1}^{M,N}, & (B_1B_2\overset{*}{K}(\boldsymbol{x}_{N+j},\boldsymbol{x}_{N+k}))_{j,k=1}^{M,M} \end{pmatrix}.$$

Fix any $\boldsymbol{x}\in\mathcal{D}$. We can compute the joint probability density function p_X^μ and p_J^μ of $\boldsymbol{S}_{PB}$ and $(S_{\boldsymbol{x}},\boldsymbol{S}_{PB})$ defined on $(\Omega_K,\mathcal{F}_K,\mathbb{P}^\mu)$ respectively. By Bayes' rule, we can obtain the *conditional probability density function* $p_{\boldsymbol{x}}^\mu$ of $S_{\boldsymbol{x}}$ given $\boldsymbol{S}_{PB}$ defined on $(\Omega_K,\mathcal{F}_K,\mathbb{P}^\mu)$, i.e., for each $\boldsymbol{v}\in\mathbb{R}^{N+M}$,

$$p_{\boldsymbol{x}}^\mu(v|\boldsymbol{v}) := \frac{p_J^\mu(v,\boldsymbol{v})}{p_X^\mu(\boldsymbol{v})} = \frac{1}{\sigma(\boldsymbol{x})\sqrt{2\pi}}\exp\left(-\frac{(v-m_{\boldsymbol{x}}^\mu(\boldsymbol{v}))^2}{2\sigma(\boldsymbol{x})^2}\right), \quad v\in\mathbb{R}, \tag{6}$$

where $m_{\boldsymbol{x}}^\mu(\boldsymbol{v}) := \mu(\boldsymbol{x}) + \boldsymbol{k}_{PB}(\boldsymbol{x})^T\overset{*}{\mathsf{K}}_{PB}{}^\dagger(\boldsymbol{v}-\boldsymbol{m}_{PB}^\mu)$, $\sigma(\boldsymbol{x})^2 := \overset{*}{K}(\boldsymbol{x},\boldsymbol{x}) - \boldsymbol{k}_{PB}(\boldsymbol{x})^T\overset{*}{\mathsf{K}}_{PB}{}^\dagger\boldsymbol{k}_{PB}(\boldsymbol{x})$ and $\boldsymbol{k}_{PB}(\boldsymbol{x}) := (P_2\overset{*}{K}(\boldsymbol{x},\boldsymbol{x}_1),\cdots,B_2\overset{*}{K}(\boldsymbol{x},\boldsymbol{x}_{N+M}))^T$. In particular, given the real observation $\boldsymbol{y} := (y_1,\cdots,y_{N+M})^T$, $S_{\boldsymbol{x}}$ conditioned on $\boldsymbol{S}_{PB} = \boldsymbol{y}$ has the probability density $p_{\boldsymbol{x}}^\mu(\cdot|\boldsymbol{y})$.

References

1. I. Babuška, F. Nobile and R. Tempone, *A Stochastic Collocation Method for Elliptic Partial Differential Equations with Random Input Data*, SIAM Rev., vol. 52, 2010, pp. 317–355.
2. I. Babuška, R. Tempone and G. E. Zouraris, *Galerkin Finite Element Approximations of Stochastic Elliptic Partial Differential Equations*, SIAM J. Numer. Anal., vol. 42, 2004, pp. 800–825.

3. A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Kluwer Academic Publishers, 2004.
4. I. Cialenco, G. E. Fasshauer and Q. Ye, *Approximation of Stochastic Partial Differential Equations by a Kernel-based Collocation Method*, Int. J. Comput. Math., Special Issue: Recent Advances on the Numerical Solutions of Stochastic Partial Differential Equations, 2012, to appear.
5. G. E. Fasshauer, *Meshfree Approximation Methods with* MATLAB, World Scientific Publishing Co. Pte. Ltd., 2007.
6. G. E. Fasshauer and Q. Ye, *Reproducing Kernels of Generalized Sobolev Spaces via a Green Function Approach with Distributional Operators*, Numer. Math., vol. 119, 2011, pp. 585–611.
7. G. E. Fasshauer and Q. Ye, *Reproducing Kernels of Sobolev Spaces via a Green Kernel Approach with Differential Operators and Boundary Operators*, Adv. Comput. Math., DOI: 10.1007/s10444-011-9264-6.
8. M. N. Lukić and J. H. Beder, *Stochastic Processes with Sample Paths in Reproducing Kernel Hilbert Spaces*, Trans. Amer. Math. Soc., vol. 353, 2001, pp. 3945–3969.
9. M. Scheuerer, R. Schaback and M. Schlather, *Interpolation of Spatial Data – A Stochastic or A Deterministic Problem?*, Data Page of R. Schaback's Research Group, 2010.
10. H. Wendland, *Scattered Data Approximation*, Cambridge University Press, 2005.
11. Q. Ye, *Analyzing Reproducing Kernel Approximation Methods via A Green Function Approach*, Ph.D. thesis, Illinois Institute of Technology, 2012.