SDE BASED REGRESSION FOR LINEAR RANDOM PDES

FELIX ANKER, CHRISTIAN BAYER, MARTIN EIGEL, MARCEL LADKAU, JOHANNES NEUMANN, AND JOHN SCHOENMAKERS

Abstract. A simulation based method for the numerical solution of PDEs with random coefficients is presented. By the Feynman-Kac formula, the solution can be represented as conditional expectation of a functional of a corresponding stochastic differential equation driven by independent noise. A time discretization of the SDE for a set of points in the domain and a subsequent Monte Carlo regression lead to an approximation of the global solution of the random PDE. We provide an initial error and complexity analysis of the proposed method along with numerical examples illustrating its behavior.

1. Introduction

Many applications in the applied sciences, e.g. in engineering and computational biology, involve uncertainties of model parameters. These can for instance be related to coefficients of media, i.e. material properties, the domains and boundary data. The uncertainties may result from heterogeneities of media and incomplete knowledge or inherent stochasticity of parameters. With steadily increasing computing power, the research field of uncertainty quantification has become a rapidly growing and vividly active area which covers many aspects of dealing with such uncertainties for problems of practical interest.

In this work, we are concerned with the description of a novel numerical approach for the solution of PDEs with stochastic data. More specifically, we consider the Darcy equation related to the modeling of groundwater flow given by

$$-\nabla \cdot (\kappa(x)\nabla u(x)) = f(x), \quad x \in D, \tag{1.1a}$$

$$u(x) = g(x), \quad x \in \partial D. \tag{1.1b}$$

Here, the solution u is the hydraulic head, κ denotes the conductivity coefficient describing the porosity of the medium, f is a source term and the Dirichlet boundary data is defined by g. The computational domain in d dimensions is $D \subset \mathbb{R}^d$. In what follows, we suppose that D is a convex polygon and all data is sufficiently smooth such that the problem always exhibits a unique solution which itself is smooth. A detailed regularity analysis is beyond the scope of this paper. In principle, although we restrict our investigations to a stochastic coefficient κ, any data of the PDE can be modeled as being stochastic. This model is quite popular for analytical and numerical examinations since it is one of the simplest models which reveals some major difficulties that also arise in more complex stochastic models. Moreover, it is of practical relevance, e.g. in the context of groundwater contamination, and the deterministic second order elliptic PDE is a well-studied model problem.

2010 *Mathematics Subject Classification.* 35R60, 47B80, 60H35, 65C20, 65N12, 65N22, 65J10,65C05 .

Key words and phrases. partial differential equations with random coefficients, random PDE, uncertainty quantification, Feynman-Kac, stochastic differential equations, stochastic simulation, stochastic regression, Monte-Carlo, Euler-Maruyama.

We would like to thank the anonymous referees for their helpful comments. We are also grateful to Anders Szepessy for clarifying discussions.

When stochastic data is assumed, an adequate description of the stochastic fields has to be chosen. This can for instance be based on actual measurements, expert-knowledge or simplifying assumptions regarding the statistics. For actual computations, a suitable representation amenable for the employed numerical method is required. It is a common assumption that the considered fields are transformed Gaussian fields and are thus completely specified by the first two moments. Another usual simplification is the finite dimensional noise assumption which states that a field only depends on a finite number of random variables.

In fact, any stochastic field $\kappa : \Omega \times D \to \mathbb{R}$ with finite variance can be represented in terms of

$$\kappa(\omega, x) = E[\kappa] + \sum_{m=1}^{\infty} a_m(x)\xi_m(\omega) \tag{1.2}$$

where the product of the sum separates the dependence on $\omega \in \Omega$ and $x \in D$. A typical method to obtain such a representation is the Karhunen-Loève expansion (KL) which will be used in the numerical examples with a finite number of terms (truncated KL). In this case, the basis a_m consists of eigenfunctions of the covariance integral operator weighted by the eigenvalues of this operator. The smoothness of the a_m is directly related to the covariance function used to model the respective stochastic field, see e.g. [9, 38, 26]. A common choice is the Whittle-Matèrn covariance which includes the smooth Gaussian covariance and the rather rough exponential covariance.

A variety of numerical methods is available to obtain approximate solutions of the model problem (1.1) with random data and we only refer to [26, 35, 22] for an overview in the context of uncertainty quantification (UQ). These methods often rely on the separation of the deterministic and the stochastic space and introduce separate discretizations [34]. Common methods are based on sampling of the stochastic space, the projection onto an appropriate stochastic basis or a perturbation analysis. The most well-known sampling approach is the Monte Carlo (MC) method which is very robust and easy to implement. Recent developments include the quite successful application of multilevel ideas for variance reduction and advances with structured point sequences (Quasi-MC), cf. [8, 11, 19, 21, 5]. (Pseudo-)Spectral methods represent a popular class of projection techniques which can e.g. be based on interpolation (Stochastic Collocation) [2, 31, 32] or orthogonal projections with respect to the energy norm induced by the differential operator of the random PDE (Stochastic Galerkin FEM) [16, 27, 4, 3, 15, 13]. These methods are more involved to analyze and implement but offer the benefit of possibly drastically improved convergence rates when compared to standard Monte Carlo sampling. The deterministic discretization often relies on the finite element method (FEM) which also holds for MC.

The aim of this paper is the description of a novel numerical approach which is based on the observation that the random PDE (1.1) is directly related to a stochastic differential equation driven by a stochastic process, namely

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \tag{1.3}$$

with appropriate coefficients b and σ, Brownian motion W and additional boundary conditions. For deterministic data κ, f, g, for any $x \in D$, the Feynman-Kac formula leads to a collection of random variables $\phi^x = \phi^x(\kappa, f, g)$ such that $u(x) = E[\phi^x]$, i.e. the deterministic solution at x is equivalent to the expectation of the random variable. When the data is stochastic, the solution $u(\omega, x)$ of the random PDE at $x \in D$ can be expressed as the *conditional expectation* $u(\omega, x) = E[\phi^x \,|\, \kappa, f, g]$ and the variance of $u(x)$ can be bounded by the variance of ϕ^x. To determine ϕ^x at points $x \in D$, a classical Euler method can be

employed. Given a finite set of sampling points in D, the approximate expectation of the solution $u(\cdot) = E[u(\omega, \cdot)]$ is determined by solving a regression problem with respect to a basis of orthogonal global polynomials. By this, we obtain a representation of the (expectation of the) solution field which is globally defined and smooth. One can regard the proposed method as a combination of sampling and reconstruction techniques, making use of classical stochastic solution techniques point-wise and a global polynomial projection in a least squares sense. When compared to MC which samples a stochastic space $(\Omega, \mathcal{F}, P)$ by (typically) determining a FEM solution at every point and subsequently averaging the solutions, our method determines realizations of stochastic solutions at points in the physical domain D and determines the expectation by a global minimization subject to a basis in the physical space. Thus, the method does not require any type of deterministic solver. Moreover, it can be parallelized extremely well.

The structure of the paper is as follows: In Section 2, we elaborate on the representation of deterministic and stochastic PDEs in terms of stochastic differential equations. Moreover, we recall the Euler method as a way to determine point-wise stochastic solutions numerically. Based on a set of stochastic solutions in the physical domain, the reconstruction of a global approximation by means of a regression approach is described in Section 3. In addition to the presentation of the method, we also provide a global convergence analysis. In Section 4, we comment on the expected overall convergence and complexity properties of the method. This should be considered as initial and not in-depth analysis which provides pointers to further required research. The paper is concluded with numerical examples in Section 5 where the performance of our method is demonstrated with a very smooth and a less smooth field.

2. Stochastic representations

In this section, we construct proper stochastic representations of solutions of *stochastic* PDEs in terms of solutions of *stochastic* differential equations. That is, our goal is to construct an SDE such that the solution $u(\omega, x)$ of the SPDE at some point $x \in D$ can be expressed as conditional expectation of some functional of the solution to the SDE.

We first give an extensive reminder of stochastic representations of certain deterministic PDEs. The material presented here is standard and we refer, for instance, to the comprehensive presentation by Milstein and Tretyakov [29].

2.1. Stochastic representations for deterministic PDEs.

We consider the following stochastic differential equation (SDE):

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \tag{2.1a}$$

$$X_0 = x, \tag{2.1b}$$

where $x \in \mathbb{R}^d$ is a deterministic point, W is a d-dimensional standard Brownian motion defined on a probability space $(\Omega, \mathcal{F}, P)$ and $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are (say) uniformly Lipschitz continuous functions. We could just as easily consider a Brownian motion with a different dimension. Sometimes we shall additionally consider the derived processes

$$Y_t \coloneqq \exp\left(\int_0^t c(X_s)ds\right), \tag{2.1c}$$

$$Z_t \coloneqq \int_0^t f(X_s)Y_s ds, \tag{2.1d}$$

for some functions $c, f : \mathbb{R}^d \to \mathbb{R}$. In particular, for every $T > 0$, (2.1) has a unique (strong) solution on $[0, T]$, i.e., there is a unique process X – adapted to the filtration generated by the Brownian motion W – which satisfies

$$\int_0^T E[|X_t|^2]dt < \infty$$

and

$$X_t = x + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s,$$

where the integral w.r.t. W is considered in the Itô-sense. If we want to stress the dependence on the initial value we write $X_t^x := X_t$. Of course, Y and Z depend on x as well, and we shall write Y_t^x and Z_t^x if we want to stress this dependence. (As usual, all equalities between random variables are supposed to hold a.s. unless otherwise noted.)

The *infinitesimal generator* of the SDE (2.1) is the differential operator L acting on test functions by

$$Lf(x) = \sum_{i=1}^n b_i(x)\frac{\partial}{\partial x_i}f(x) + \frac{1}{2}\sum_{i,j=1}^n a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j}f(x), \tag{2.2}$$

where $a(x) := \sigma(x)^\top \sigma(x)$. We employ the following version of Ito's formula: for any function $F : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ which is C^1 in time and C^2 in space it holds

$$F(t, X_t) = F(0, x) + \int_0^t \left(\frac{\partial}{\partial t}F(s, X_s) + LF(s, X_s)\right)ds + \int_0^t \sum_{i=1}^n \sum_{j=1}^d \frac{\partial}{\partial x_i}F(s, X_s)\sigma_{ij}(X_s)dW_s^j.$$

Moreover, recall that for any process $u(s)$ (adapted to the filtration generated by W), $t \mapsto \int_0^t u(s)dW_s$ is a martingale on $[0, T]$ provided that $\int_0^T E\left[|u(t)|^2\right]dt < \infty$. In particular, it holds that

$$E\left[\int_0^t u(s)dW_s\right] = 0.$$

After these preparations, we consider stochastic representations for solutions of PDEs posed in terms of L. For simplicity, we first consider the Cauchy problem for $(t, x) \in [0, \infty[\times\mathbb{R}^n$,

$$\frac{\partial}{\partial t}u(t, x) = Lu(t, x) + c(x)u(t, x) + f(x), \tag{2.3a}$$

$$u(0, x) = g(x). \tag{2.3b}$$

Assuming that the solution $u \in C^{1,2}([0, \infty[\times\mathbb{R}^n)$ and that the coefficients of L and c, f are Lipschitz continuous, we can use that

$$u(t, x) = E\left[g\left(X_t^x\right)Y_t^x + Z_t^x\right], \quad t \geq 0,\ x \in \mathbb{R}^d. \tag{2.4}$$

Indeed, fix some $T > 0$ and consider the functional $F(t, x) := u(T - t, x)$. We assume for simplicity that $c = f = 0$ which implies that $Y \equiv 1$ and $Z \equiv 0$. Ito's formula and (2.3a)

yield

$$\begin{aligned} u(0, X_T^x) - u(T, x) &= \int_0^T \left(-\frac{\partial}{\partial t} u(T-t, X_t^x) + Lu(T-t, X_t^x) \right) dt \\ &\quad + \int_0^T \sum_{i=1}^d \sum_{j=1}^d \frac{\partial}{\partial x_i} u(T-t, X_t^x) \sigma_{ij}(X_t^x) dW_t^j \\ &= \int_0^T \sum_{i=1}^d \sum_{j=1}^d \frac{\partial}{\partial x_i} u(T-t, X_t^x) \sigma_{ij}(X_t^x) dW_t^j. \end{aligned}$$

Inserting (2.3b) and taking expectations on both sides, we obtain

$$u(T, x) = E\left[g(X_T^x) \right].$$

Remark 2.1. Here and in what follows we will generally be rather liberal with conditions and assumptions, i.e., we will assume enough regularity of all data (σ, b, c, f, g and, where relevant, the domain D) such that the objects under consideration are well-defined, and the numerical approximations converge. A convenient assumption to work with will be the usual uniform ellipticity assumption, even though most of the facts will be true under weaker assumptions. Uniform ellipticity will generally also allow us to assume that the PDEs have classical solutions, which is important for the Ito formula. It is, however, very easy to generalize to the case of time-dependent coefficients, for instance.

Remark 2.2. Note that we have to compute a new solution of the SDE (2.1) for every position $x \in D$ for which we want to obtain the solution $u(t, x)$ of (2.3).

Remark 2.3. Notice that the numerically challenging part of (2.1) is the simulation of the paths of the process X. Given such paths, finding Y and Z is a standard one-dimensional numerical integration problem.

For computations, we need to solve two problems:

(i) Find an approximation $\overline{X}_N$ of X_t, which we can actually compute.
(ii) Given such an approximation, $E\left[g\left(\overline{X}_N \right) \right]$ is computed by a (quasi) Monte Carlo method.

As the second step is not different from other applications of the Monte Carlo method, we will mainly focus on the discretization of the SDE (2.1). Clearly the most popular approximation method for SDEs is a straight-forward generalization of the Euler method for ODEs. Indeed, let $0 = t_0 < \cdots < t_N = t$ be a time grid and denote $\Delta t_i := t_i - t_{i-1}$, $\Delta W_i := W_{t_i} - W_{t_{i-1}}$, $\Delta t_{\max} := \max_i \Delta t_i$. Set $\overline{X}_0 := x$ and iteratively define

$$\overline{X}_i := \overline{X}_{i-1} + b\left(\overline{X}_{i-1} \right) \Delta t_i + \sigma\left(\overline{X}_{i-1} \right) \Delta W_i, \quad i = 1, \ldots, N. \tag{2.5}$$

Under very weak assumptions we have *strong* convergence with rate $1/2$, i.e.,

$$E\left[\left| X_t - \overline{X}_N \right| \right] \leq C \sqrt{\Delta t_{\max}}$$

for some constant C independent of $\Delta t_{\max}$. More relevant in most applications (including our own application) is the concept of *weak* approximation. Fortunately, the Euler scheme typically (for instance, when the diffusion is hypo-elliptic) exhibits first order weak convergence, i.e., for any suitable test function $F : \mathbb{R}^d \to \mathbb{R}$ it holds

$$\left| E[F(X_t)] - E\left[F\left(\overline{X}_N \right) \right] \right| \leq C \Delta t_{\max} \tag{2.6}$$

with a constant C independent of $\Delta t_{\max}$.

Remark 2.4. Higher order weak schemes (i.e., schemes leading to higher order weak convergence) do exist and are often used for the Cauchy problem. For simplicity, and also because higher order methods are hard to find for boundary value problems, we focus on the Euler scheme. Higher-order strong schemes are only available in very special cases, we refer to Kloeden and Platen [20] for more information on high-order (strong or weak) schemes. Adaptive time-stepping schemes have been constructed by Szepessy et al. [36] based on the error representations of Talay and Tubaro [37]. Moreover, implicit Euler schemes may be preferable if the SDE solution exhibits instability.

Next, on a (smooth) domain D we consider the parabolic problem (2.3a) with Dirichlet boundary conditions

$$\frac{\partial}{\partial t}u(t,x) = Lu(t,x) + c(x)u(t,x) + f(x), \quad t \geq 0, \quad x \in D, \tag{2.7a}$$

$$u(t,x) = g(x), \quad x \in \partial D, \tag{2.7b}$$

$$u(0,x) = u_0(x), \quad x \in \overline{D}. \tag{2.7c}$$

Of course, we assume suitable compatibility conditions on u_0 and g. Moreover, let

$$\tau := \tau_x := \inf\{\, t \geq 0 \;\big|\; X_t^x \in D^c \,\} \tag{2.8}$$

denote the (random) first exit time of X from D. This leads to the stochastic representation

$$u(t,x) = E\left[\left(u_0(X_t^x)\mathbf{1}_{\tau_x \geq t} + g(X_{\tau_x}^x)\mathbf{1}_{\tau_x < t}\right) Y^x_{\min(t,\tau_x)} + Z^x_{\min(t,\tau_x)}\right]. \tag{2.9}$$

In this case, we use the very same Euler discretization (2.5) as before. In particular, τ is approximated by the first exit time $\overline{\tau}$ of the discrete time process $\overline{X}_i$, $i = 0, \ldots, N$. Note that there are two sources of errors in the approximation of τ by $\overline{\tau}$:

(i) the error in the approximation of X by $\overline{X}$;
(ii) the possibility that exit occurs between two grid points t_i and t_{i+1}.

Unfortunately, the second source of error reduces the weak error rate, i.e., the approximation error for $u(t,x)$, to the rate $1/2$. However, there are adaptive time-step refinements, which have empirically shown to be very successful for improving the order of convergence to an observed order 1 again, we refer to Dzougoutov et al. [12] for the stopped diffusion and to Bayer et al. [6] for an adaptive scheme for reflected and stopped diffusions. We also refer to Gobet and Menozzi [18] for an alternative scheme based on shifting the boundary.

For the elliptic problem

$$Lu(x) + c(x)u(x) + f(x) = 0, \quad x \in D, \tag{2.10a}$$

with the Dirichlet boundary condition

$$u(x) = g(x), \quad x \in \partial D, \tag{2.10b}$$

the stochastic representation is essentially obtained by the fact that the solution of the elliptic problem is obtained as limit for $t \to \infty$ from the solution of the parabolic problem. Hence, the stochastic representation of the Dirichlet problem (2.10b) is given by

$$u(x) = E\left[g(X_\tau^x)Y_\tau^x + Z_\tau^x\right], \quad x \in \overline{D}, \tag{2.11}$$

for the stopping time τ given in (2.8).

Remark 2.5. Similar stochastic representations exist for both parabolic and elliptic Neumann problems, based on reflected diffusion processes. Naturally, further extensions in the case of mixed boundary conditions are possible, too. We again refer to [29].

Example 2.6. We recognize Darcy's law as introduced in (1.1) as special case of (2.2), (2.10) with

$$b(x) \equiv \nabla\kappa(x), \quad \sigma(x) \equiv \sqrt{2\kappa(x)}I_d, \quad x \in D.$$

f and g are consistently used in (1.1) and (2.10). As $c \equiv 0$, the stochastic representation reads

$$dX_t^x = \nabla\kappa(X_t^x)dt + \sqrt{2\kappa(X_t^x)}dW_t, \quad X_0^x = x,$$
$$u(x) = E\left[g(X_\tau^x) + \int_0^\tau f(X_s^x)ds\right],$$

for $x \in D$.

Remark 2.7. The stochastic representations considered here are all for linear PDEs, more precisely for the PDEs (2.3), (2.7), (2.10). It is, however, possible to extend the method to some classes of non-linear PDEs. On the one hand, we refer to [30] for a layer method based on stochastic representations for a class of stochastic Navier Stokes equations. On the other hand, there are representations of more general non-linear PDEs by backward stochastic differential equations, more precisely systems of second order forward-backward SDEs. We refer to the review [39]. We note that all these representations for deterministic PDEs can be extended to the case of random coefficients in the same manner as the stochastic representations for linear PDEs covered in this article.

Remark 2.8. Similar representations hold for certain classes of linear non-local equations, where L is the infinitesimal generator of a Lévy process, such as fractional Laplacians. The SDE is the driven by the corresponding Lévy process instead of the Brownian motion.

2.2. **Stochastic representations for stochastic PDEs.** This section is concerned with the solutions of random PDEs. This means that the differential operator L and possible the initial and/or boundary values are random. For the sake of concreteness, let us concentrate on the case of a random elliptic Dirichlet problem in the sense of (2.10a) with (2.10b).

In order for the above constructions to make sense for *random* data b, σ, f, c, g, we need to choose a Brownian motion W *independent* of the other sources of randomness. This means we need to work on a probability space $(\Omega, \mathcal{F}, P)$ large enough such that the random fields and processes b, σ, f, c, g and W are all defined on the same probability space and W is independent of the data. Given constructions of a probability space carrying the fields b, σ, f, c, g and another probability space carrying the Brownian motion W, this simplifies to choosing the product space as the joint probability space.

More precisely, let us suppose that a Brownian motion W is defined on a probability space $(\Omega_1, \mathcal{F}_1, P_1)$ where we shall write $W_t(\omega_1)$ to stress the interpretation of W as a process on $(\Omega_1, \mathcal{F}_1, P_1)$.

Moreover, we are given a second probability space $(\Omega_2, \mathcal{F}_2, P_2)$ on which the random fields $b = b(\omega_2, x)$, $\sigma = \sigma(\omega_2, x)$, $c = c(\omega_2, x)$, $f = f(\omega_2, x)$, $x \in D$, and $g(\omega_2, x)$, $x \in \partial D$, $\omega_2 \in \Omega_2$ are defined. Additionally, we consider the product probability space

$$(\Omega, \mathcal{F}, P) := (\Omega_1, \mathcal{F}_1, P_1) \otimes (\Omega_2, \mathcal{F}_2, P_2) = (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, P_1 \otimes P_2).$$

Obviously, we can extend any random variable Y on $(\Omega_1, \mathcal{F}_1, P_1)$ to a random variable on $(\Omega, \mathcal{F}, P)$ by setting $Y(\omega_1, \omega_2) := Y(\omega_1)$, for any $\omega = (\omega_1, \omega_2) \in \Omega$, and likewise for random variables Z defined on $(\Omega_2, \mathcal{F}_2, P_2)$. Note that the construction implies that any such extended random variables Y and Z are independent. In this way, we obtain a Brownian motion W and random fields b, σ, g, c, f on $(\Omega, \mathcal{F}, P)$ with W independent of the data.

We interpret the random fields as random variables taking values in some function spaces.[1] More precisely, the random variables $b(\omega) \coloneqq b(\omega, \cdot)$, $\sigma(\omega) \coloneqq \sigma(\omega, \cdot)$ and $g(\omega) \coloneqq g(\omega, \cdot)$, $c(\omega) \coloneqq c(\omega, \cdot)$, $f(\omega) \coloneqq f(\omega, \cdot)$ should assume values in the space of Lipschitz continuous functions in order to avoid complications with regard to the related SDEs.

On the one hand, inserting these into (2.1), for $x \in \overline{D}$ we obtain systems of (random) SDEs of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \tag{2.12a}$$

$$X_0 = x. \tag{2.12b}$$

It is clear that the solution $X_t = X_t(\omega)$ can only be considered on the full probability space $(\Omega, \mathcal{F}, P)$. As in (2.1c) and (2.1d), we further define

$$Y_t := \exp\left(\int_0^t c(X_s)ds\right), \tag{2.12c}$$

$$Z_t := \int_0^t f(X_s)Y_s ds. \tag{2.12d}$$

On the other hand, we consider the random PDE given by the coefficients b, σ, g, c, f. More precisely, define the random matrix field $a \coloneqq \sigma^\top \sigma$ and the random differential operator L by

$$L(\omega_2)h(x) = \sum_{i=1}^n b_i(\omega_2)(x)\frac{\partial}{\partial x_i}h(x) + \frac{1}{2}\sum_{i,j=1}^n a_{ij}(\omega_2)(x)\frac{\partial^2}{\partial x_i \partial x_j}h(x). \tag{2.13}$$

Obviously, L is obtained by inserting the random coefficients into the deterministic formula for L as given in (2.2). Next, consider $u = u(\omega_2)(x) = u(\omega_2, x)$, the random solution of the random PDE

$$Lu + cu + f = 0, \tag{2.14a}$$

$$u(\omega_2, x) = g(\omega_2, x), \quad x \in \partial D. \tag{2.14b}$$

From (2.11), we immediately derive the following stochastic representation for u: define the stopping time $\tau = \tau_x$ as in (2.8) (on the full probability space $(\Omega, \mathcal{F}, P)$), i.e.,

$$\tau \coloneqq \tau_x \coloneqq \inf\{\, t \geq 0 \mid X_t \in D^c \,\}. \tag{2.15}$$

This leads to the solution

$$u(\omega_2, x) = \int_{\Omega_1} \left[g\left(\omega_2, X^x_{\tau_x(\omega_1,\omega_2)}(\omega_1, \omega_2)\right) Y^x_{\tau_x(\omega_1,\omega_2)}(\omega_1, \omega_2) + Z^x_{\tau_x(\omega_1,\omega_2)}(\omega_1, \omega_2)\right] P_1(d\omega_1), \tag{2.16}$$

by integrating out the randomness induced by the Brownian motion W. In a more probabilistic notation and using the joint probability space $(\Omega, \mathcal{F}, P)$, we can write

$$u(x) = E\left[g\left(X^x_{\tau_x}\right) Y^x_{\tau_x} + Z^x_{\tau_x} \middle| b, \sigma, f, c, g \right] = E\left[g\left(X^x_{\tau_x}\right) Y^x_{\tau_x} + Z^x_{\tau_x} \middle| \mathcal{F}_2 \right],$$

assuming that $\mathcal{F}_2 = \sigma(b, \sigma, f, c, g)$ *and* equating the σ-algebras $\mathcal{F}_2$ and $\{\Omega_1 \times A \mid A \in \mathcal{F}_2\} \subset \mathcal{F}$.

[1] This may not be true for the fields themselves, but it is certainly true for a suitable mollification of the random fields, e.g., by a finite truncation of the Karhunen-Loève expansion.

Remark 2.9. By independence between the coefficient fields and the added Brownian motion W, the analysis of the SDE with random coefficients and the stochastic representations remain basically unchanged compared to the deterministic coefficient case, mainly replacing unconditional with conditional expectations when required. In particular, the Euler scheme (2.5) can be extended in the obvious way to the above problem, retaining the usual rates of convergence as discussed above.

Assuming we are only interested in $E[u(x)]$, we can, of course, directly take the full expectation in (2.16), namely

$$E[u(x)] = E\left[g\left(X^x_{\tau_x}\right)Y^x_{\tau_x} + Z^x_{\tau_x}\right]. \tag{2.17}$$

Hence, we do not need to use a nested Monte Carlo simulation procedure in the end. In this context, note that

$$\operatorname{var} u(x) \le \operatorname{var}\left[g\left(X^x_{\tau_x}\right)Y^x_{\tau_x} + Z^x_{\tau_x}\right].$$

In fact, we have the elementary result that

$$\begin{aligned}\operatorname{var}\left[g\left(X^x_{\tau_x}\right)Y^x_{\tau_x} + Z^x_{\tau_x}\right] &= \operatorname{var} E\left[g\left(X^x_{\tau_x}\right)Y^x_{\tau_x} + Z^x_{\tau_x}\middle|\mathcal{F}_2\right] + E\operatorname{var}\left[g\left(X^x_{\tau_x}\right)Y^x_{\tau_x} + Z^x_{\tau_x}\middle|\mathcal{F}_2\right]\\ &= \operatorname{var} u(x) + E\operatorname{var}\left[g\left(X^x_{\tau_x}\right)Y^x_{\tau_x} + Z^x_{\tau_x}\middle|\mathcal{F}_2\right].\end{aligned} \tag{2.18}$$

Note that the second term in the variance decomposition (2.18) can be interpreted as the additional variance introduced by the Brownian motion W. Since it is non-negative by definition, it increases the total variance of the estimator.

Assumption 2.10. The problem data is sufficiently regular. In particular, for Darcy's law (see (1.1) and Example 2.6), the results of this paper hold if coefficients κ, $\nabla\kappa$, f, g are a.s. Lipschitz continuous on D, a convex polygon in $\mathbb{R}^d$, additionally $\kappa, \nabla\kappa, f, g \in L^\infty(\Omega\times D)$ and, $\kappa^{-1} \in L^\infty(\Omega \times D)$.

Discussion of the assumptions. We generally assume sufficient regularity such that the stochastic representation $g\left(X^x_{\tau_x}\right)Y^x_{\tau_x} + Z^x_{\tau_x}$ is square integrable and the (adaptive) Euler scheme converges with weak order one. In the spirit of this exploratory paper, we do not try to provide a refined error analysis leading to minimal conditions. (In particular, note that no proof of convergence rate one for the adaptive Euler scheme – even in the case of deterministic coefficients – has been given in the literature, to the best of our knowledge, cf. [12, 6].)

Regarding the step from deterministic to random coefficients, the representation (2.17) holds provided that the deterministic counterpart holds for P-almost all (random) coefficients, *and* the resulting random variable u – obtained by conditional expectations in (2.16) – is integrable. Similarly, error expansions for the case of deterministic coefficients (when available, such as for the non-adaptive Euler scheme) can be extended to the case of stochastic coefficients provided that the terms are integrable (w.r.t. the random coefficients).

However, it seems more tricky to obtain square integrability of $g\left(X^x_{\tau_x}\right)Y^x_{\tau_x} + Z^x_{\tau_x}$. In particular, square integrability of the stopping time τ is not straight-forward since there seems to be a lack of explicit bounds of moments of hitting times for SDEs with deterministic coefficients (explicit in terms of said coefficients) in the literature. Hence, one needs to directly analyze the stochastic problem. To this end, let us specifically consider the Darcy problem, i.e., the operator $L = \nabla\cdot\kappa\nabla$ for a random field κ as in Assumption 2.10.

Lemma 2.11. *Under Assumption 2.10, we have*

$$\operatorname{var}\left[g\left(X^x_{\tau_x}\right) + Z^x_{\tau_x}\right] < \infty$$

for almost every $x \in D$.

Proof. We only need to prove that $E[\tau^2] < \infty$. To this end, note that $w(t,x) \coloneqq P(\tau_x > t \mid \kappa)$, $t \geq 0$, satisfies the (random) PDE

$$\begin{aligned} \frac{\partial}{\partial t} w(t,x) &= Lw(t,x), \\ w(t,x) &= 0, \quad x \in \partial D, \\ w(0,x) &= 1, \quad x \in D, \end{aligned}$$

see, for instance, [29]. Moments of τ can now directly be related to moments of the solution of the above random PDE by

$$\begin{aligned} E[\tau_x] &= \int_0^\infty E[w(t,x)]dt, \\ E[\tau_x^2] &= \int_0^\infty tE[w(t,x)]dt. \end{aligned}$$

Note that from standard PDE theory [14, 17] the operator L is uniformly elliptic in the sense that for any $v \in H_0^1(D)$ we have

$$A_\kappa(v,v) \coloneqq \int_D \kappa(x) \left|\nabla v(x)\right|^2 dx \geq C_1 \left\|v\right\|_{H^1}^2$$

for some (random) constant $C_1 > 0$ only depending on $\inf_{x \in D} \kappa(x)$. Multiplying both sides of the PDE by $w(t,x)$ and integrating in x, we obtain

$$\int_D \partial_t w(t,x) w(t,x) dx = -\int_D \kappa(x) \left|\nabla w(t,x)\right|^2 dx.$$

By the energy estimate, we arrive at the bound

$$\partial_t \left\|w(t,\cdot)\right\|_{L^2}^2 \leq -C_1 \left\|w(t,\cdot)\right\|_{H^1}^2 \leq -C_1 \left\|w(t,\cdot)\right\|_{L^2}^2.$$

By a simple Grönwall argument, we obtain exponential decay

$$\left\|w(t,\cdot)\right\|_{L^2}^2 \leq \mathrm{vol}(D) e^{-2C_1 t},$$

implying that

$$\int_D E[\tau_x^2] dx \leq \mathrm{vol}(D)\, E\left[\frac{1}{4C_1^2}\right] < \infty$$

by Assumption 2.10, as κ is uniformly bounded away from 0 in x and ω. Hence, $E[\tau_x^2] < \infty$ for almost every $x \in D$. □

3. Regression based methods and their error analysis

The construction of global approximate solutions based on probabilistic representations for different random PDEs may generically be carried out via solving a regression problem connected with a probabilistic representation

$$v(x) = E\left[\Phi^x\right], \quad x \in D \subset \mathbb{R}^d. \tag{3.1}$$

In (3.1), $x \in D$ is a generic point in space and Φ^x is a real valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that the map

$$(x,\omega) \in D \times \Omega \to \Phi^x(\omega)$$

is $\mathcal{B}(D) \otimes \mathcal{F}$ measurable.

In the context of this paper, we consider $v(x) = E[u(x)]$, u being the solution of the random PDE (2.14). By (2.17), the above relation holds for

$$\Phi^x := g\left(X^x_{\tau_x}\right) Y^x_{\tau_x} + Z^x_{\tau_x}.$$

For each fixed $x \in D$, we consider the sequence of i.i.d. copies $(\Phi^x_m, m = 1, 2, \ldots)$. Let further μ be some probability measure on $(D, \mathcal{B}(D))$, U be a random variable on D with distribution μ, and $(U_n, n = 1, 2, \ldots)$ a sequence of i.i.d. copies of U that are independent of any Φ^x_m, $m = 1, 2, \ldots$, $x \in D$. One now may estimate $v(x)$ via regression procedures based on a Monte Carlo simulation of

$$\begin{aligned}
&\Phi^{U_1}_1, \ldots, \Phi^{U_1}_M,\\
&\Phi^{U_2}_{M+1}, \ldots, \Phi^{U_2}_{2M},\\
&\ldots\\
&\Phi^{U_N}_{(N-1)M+1}, \ldots, \Phi^{U_N}_{NM},
\end{aligned}$$

i.e. in condensed notation

$$\left(\Phi^{U_n}_{(n-1)M+1}, \ldots, \Phi^{U_n}_{nM}\right), \quad n = 1, \ldots, N. \tag{3.2}$$

Notice once more that all these random variables are, by construction, independent.

Generally one distinguishes between regression estimators of local and global nature. Below we give a concise recap, where we introduce for ease of notation the (point-wise) averages

$$\overline{\Phi}^x_n := \overline{\Phi}^x_{n,M} := \frac{1}{M}\sum_{m=1}^{M} \Phi^x_{(n-1)M+m}, \quad x \in D. \tag{3.3}$$

Remark 3.1. Although the boundary data is typically known, a regression approach as presented here does not enforce or impose boundary conditions explicitly. In the numerical experiments, we manually place a set of "exact solution points" on the entire boundary. There are several ways to get a more accurate matching of boundary conditions. However, this does not affect the presented error analysis.

3.1. **Recap of regression estimators.** In this section we recall some common regression estimators starting with some local ones.

Local regression. Let $\mathcal{K}(x) \geq 0$ be some kernel function on $\mathbb{R}^d$ satisfying $\int \mathcal{K}(x)dx = 1$ and $\int x^i \mathcal{K}(x)dx = 0$, $i = 1, \ldots, d$. For a band-width $\delta > 0$ with $\delta \downarrow 0$ in relation to $N \to \infty$, in a suitable way, one defines the kernel estimator

$$\widehat{v}(x) := \frac{\sum_{n=1}^{N} \mathcal{K}\left(\frac{x-U_n}{\delta}\right) \overline{\Phi}^{U_n}_{n,M}}{\sum_{n'=1}^{N} \mathcal{K}\left(\frac{x-U_{n'}}{\delta}\right)}.$$

Another related type of local regression is the so called *local polynomial regression,* where one solves the following weighted least squares problem for some polynomial degree p at a fixed point x_0 :

$$\underset{\{\beta_\eta;\, |\eta|\leq p\}}{\arg\min} \sum_{n=1}^{N} \mathcal{K}\left(\frac{x_0 - U_n}{\delta}\right) \cdot \left[\overline{\Phi}^{U_n}_{n,M} - \sum_{|\eta|\leq p} \beta_\eta (U_n - x_0)^\eta\right]^2$$

with $\eta \in \mathbb{N}^d_0$ being a generic multi-index, which gives the approximations

$$\widehat{v}(x_0) := \widehat{\beta}_0, \quad \partial_\eta \widehat{v}(x_0) := \eta! \widehat{\beta}_\eta.$$

It should be noted that local polynomial regression is particularly favorable if one needs to estimate derivatives at a fixed point. As some further local estimators we mention the *Nadaraya-Watson estimator,* and the *k-nearest neighbor estimators* (e.g. see [23] for more details).

An application of local regression methods to the random PDE problem at hand is postponed to a future paper. Here, we focus on *global* regression.

Fully stochastic global regression. Global Monte Carlo regression estimators are extremely popular in finance where they are used for the pricing of American options ([25] and [40]). The general procedure is as follows. One takes a system of (continuous basis) functions $\psi_k : D \to \mathbb{R}$, $k = 1, 2, \ldots$, and considers for a Monte Carlo sample (3.2) and a fixed number K, the solution of the regression problem

$$\widehat{\gamma} \coloneqq \operatorname*{arg\,min}_{\gamma \in \mathbb{R}^K} \frac{1}{N} \sum_{n=1}^{N} \left(\overline{\Phi}_{n,M}^{U_n} - \sum_{k=1}^{K} \gamma_k \psi_k(U_n) \right)^2 . \tag{3.4}$$

The solution of (3.4) is straightforward and given by

$$\widehat{\gamma} = \left(\mathcal{M}^\top \mathcal{M} \right)^{-1} \mathcal{M}^\top \mathcal{Y}, \tag{3.5}$$

where $\mathcal{M} \in \mathbb{R}^{N \times K}$ with entries given by

$$\mathcal{M}_{n,k} \coloneqq \psi_k(U_n), \quad n = 1, \ldots, N, \quad k = 1, \ldots, K,$$

and

$$\mathcal{Y} \coloneqq \left[\overline{\Phi}_{1,M}^{U_1}, \ldots, \overline{\Phi}_{N,M}^{U_N} \right]^\top .$$

One thus obtains the approximation

$$\widehat{v}(x) \coloneqq \sum_{k=1}^{K} \widehat{\gamma}_k \psi_k(x), \tag{3.6}$$

which, of course, depends on N, M, K and the choice of basis functions $\psi_1, \ldots, \psi_K$.

Semi stochastic global regression. In several applications, in particular in finance, the random variable U can be simulated (e.g. via an SDE) but its distribution is not explicitly known otherwise. Therefore the construction of $\widehat{\gamma}$ via (3.5) requires the inversion of a random matrix $\mathcal{M}^\top \mathcal{M}$. In some situations this inversion may be ill-conditioned and may be the source of suboptimal convergence properties unless some kind of regularization is included. In the present context of random PDEs we, however, have full control of the choice of the distribution of U. Note that

$$\frac{1}{N} \left[\mathcal{M}^\top \mathcal{M} \right]_{kl} = \frac{1}{N} \sum_{n=1}^{N} \psi_k(U_n) \psi_l(U_n) \xrightarrow{N \to \infty} \int_D \psi_k(x) \psi_l(x) \mu(dx) =: \mathcal{G}_{kl}. \tag{3.7}$$

For favorable choices of μ and/or ψ_k, the matrix $\mathcal{G}$ may be pre-computed by some quadrature method or even be known in closed form. In particular, if the basis functions ψ_k are chosen orthonormal w.r.t. μ, then $\mathcal{G}$ is the identity matrix. So it is natural to replace (3.5) with the estimate

$$\overline{\gamma} \coloneqq \frac{1}{N} \mathcal{G}^{-1} \mathcal{M}^\top \mathcal{Y} \tag{3.8}$$

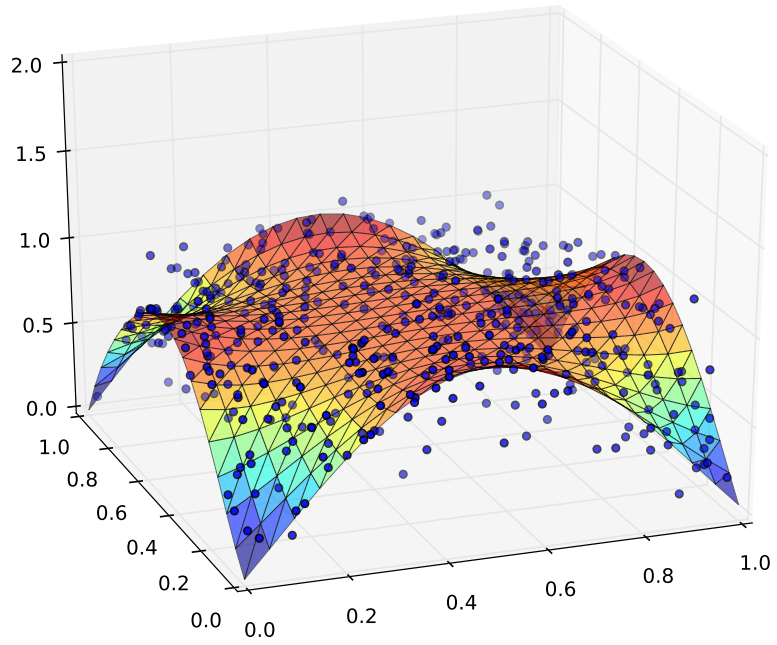

FIGURE 1. Subset of the regression points $\Phi_i^{U_i}$ and the solution of the regression procedure.

leading to the approximate solution

$$\overline{v}(x) \coloneqq \sum_{k=1}^{K} \overline{\gamma}_k \psi_k(x). \tag{3.9}$$

The (semi-)stochastic global regression method is visualized in Figure 1. The dots in the three-dimensional plot correspond to the samples $\left(U_i, \Phi_i^{U_i}\right)$ (or a subset thereof), and the surface is the outcome of the regression procedure, i.e., the graph of $\overline{v}$ on the domain D. Notice that Figure 1 is based on the actual calculations in Section 5.

Deterministic global regression. As an alternative to the semi stochastic procedure (3.8), (3.9), we may alternatively carry out spatial regression with respect to some deterministic set of, e.g. uniform, grid points x_n, $n = 1, \ldots, N$. This leads to the regression problem

$$\widetilde{\gamma} \coloneqq \operatorname*{arg\,inf}_{\gamma \in \mathbb{R}^K} \frac{1}{N} \sum_{n=1}^{N} \left(\overline{\Phi}_{n,M}^{x_n} - \sum_{k=1}^{K} \gamma_k \psi_k(x_n) \right)^2. \tag{3.10}$$

By the deterministic design matrix by $\mathcal{N} \in \mathbb{R}^{N \times K}$ with entries given by

$$\mathcal{N}_{n,k} \coloneqq \psi_k(x_n), \quad n = 1, \ldots, N, \quad k = 1, \ldots, K,$$

the solution is given by

$$\widetilde{\gamma} = \left(\mathcal{N}^\top \mathcal{N}\right)^{-1} \mathcal{N}^\top \widetilde{\mathcal{Y}}, \quad \widetilde{\mathcal{Y}} \coloneqq \left[\overline{\Phi}_{1,M}^{x_1}, \ldots, \overline{\Phi}_{N,M}^{x_N}\right]^\top, \tag{3.11}$$

yielding the approximation

$$\widetilde{v}(x) \coloneqq \sum_{k=1}^{K} \widetilde{\gamma}_k \psi_k(x). \tag{3.12}$$

3.2. **Convergence analysis of global regression estimators.** The convergence analysis of the full stochastic regression is highly nontrivial and relies on the theory of empirical processes. For instance, in the case of American options see [10], and in the case of Markovian control problems [7]. The convergence analysis for the semi stochastic regression and the deterministic regression approach is significantly easier. Since the latter procedures are more suitable for our purposes, we will restrict our analysis to them.

Remark 3.2. In what follows, we assume that the probability space $(\Omega, \mathcal{F}, P)$ is large enough to carry a sequence of independent samples of Φ^U. Given an $\mathcal{F}$-measurable random field $\alpha = \alpha(x)$, $x \in D$, which may depend on a sequence of random variables U_n, we denote

$$\|\alpha\|^2_{L^2(\Omega\times D; P\otimes\mu)} \coloneqq \int_D E[\alpha(x)^2]\mu(dx).$$

Hence, μ is used in a double role here: it denotes the distribution of the random variables U_n (defined on Ω) as well as a measure on $(D, \mathcal{B}(D))$.

Semi stochastic regression statistics. For the scalar product $f, g \in L^2_\mu(D)$ defined by

$$\langle f, g\rangle_{L^2_\mu} \coloneqq \langle f, g\rangle \coloneqq \int_D f(x)g(x)\mu(dx),$$

and the corresponding norm $\|\cdot\|_{L^2_\mu}$, let

$$\gamma^\circ \coloneqq \operatorname*{arg\,min}_{\gamma\in\mathbb{R}^K} \left\| v - \sum_{k=1}^{K} \gamma_k \psi_k \right\|^2 \tag{3.13}$$

and define

$$v_K(x) \coloneqq \sum_{k=1}^{K} \gamma^\circ_k \psi_k(x), \quad x \in D. \tag{3.14}$$

v_K is the (exact) L^2_μ-projection of v to span$\{\psi_1, \dots, \psi_k\}$. Hence, we have that

$$\langle v - v_K, \psi_k\rangle = 0, \quad k = 1, \dots, K,$$

i.e.,

$$\langle v, \psi_k\rangle - \sum_{l=1}^{K} \mathcal{G}_{kl}\gamma^\circ_l = 0.$$

By defining

$$[w]_k \coloneqq \langle v, \psi_k\rangle, \quad k = 1, \dots, K,$$

we thus have

$$\gamma^\circ = \mathcal{G}^{-1} w.$$

Note that the semi-stochastic regression estimator $\overline{v}$ is a standard Monte Carlo estimator of the projection v_K, as (for simplicity for $\mathcal{G} = I_K$)

$$\overline{\gamma}_k = \frac{1}{N}\sum_{n=1}^{N} \psi_k(U_n)\overline{\Phi}^{U_n}_{n,M} \approx \langle \psi_k, v\rangle.$$

Hence, the standard Monte Carlo error formula shows that

$$E\left[(\overline{\gamma}_k - \langle \psi_k\,, v\rangle)^2\right] = \frac{\operatorname{var}\left[\psi_k(U_n)\overline{\Phi}^{U_n}_{n,M}\right]}{N}.$$

On the other hand, we are really interested in $E\left[\|\overline{v} - v_K\|^2\right]$. An immediate application of the standard formula gives an error estimate of the order K^2/N, as the K estimates $\overline{\gamma}_1, \ldots, \overline{\gamma}_K$ are not independent. However, we shall see that a more careful analysis will give us an estimate of order K/N, see Proposition 3.3 below.

Let us write

$$\begin{aligned}\frac{1}{N}\left[\mathcal{M}^\top \mathcal{Y}\right]_k &= \frac{1}{N}\sum_{n=1}^N \psi_k(U_n)\mathcal{Y}_n = \frac{1}{N}\sum_{n=1}^N \psi_k(U_n)\overline{\Phi}^{U_n}_{n,M} \\ &=: \frac{1}{N}\sum_{n=1}^N \psi_k(U_n)\left(v(U_n) + \frac{\eta_n}{\sqrt{M}}\right) \qquad (3.15)\end{aligned}$$

with i.i.d. random variables

$$\eta_n \coloneqq \sqrt{M}\left(\overline{\Phi}^{U_n}_{n,M} - v(U_n)\right) = \frac{1}{\sqrt{M}}\sum_{m=1}^M \left(\Phi^{U_n}_{(n-1)M+m} - v(U_n)\right)$$

satisfying $E\left[\eta_1 | U_1\right] = 0$. For the variances we have

$$\begin{aligned}\operatorname{var}\left[\,\eta_1 |\, U_1\right] &= \operatorname{var}\left[\,\Phi^U \big|\, U\right] \\ \operatorname{var}\left[\eta_1\right] &= \operatorname{var}\left[\Phi^U - v(U)\right] = E\left[\left(\Phi^U - v(U)\right)^2\right] \\ &= \int \mu(dx) E\left[(\Phi^x - v(x))^2\right] = \int \mu(dx)\operatorname{var}\left[\Phi^x\right].\end{aligned}$$

From (3.15) we then define

$$\frac{1}{N}\left[\mathcal{M}^\top \mathcal{Y}\right]_k =: \langle v, \psi_k\rangle + \frac{\xi_k}{\sqrt{N}} + \frac{\vartheta_k}{\sqrt{NM}}$$

with

$$\begin{aligned}\xi_k &\coloneqq \frac{1}{\sqrt{N}}\sum_{n=1}^N \left(\psi_k(U_n)v(U_n) - \langle v, \psi_k\rangle\right), \\ \vartheta_k &\coloneqq \frac{1}{\sqrt{N}}\sum_{n=1}^N \psi_k(U_n)\eta_n = \frac{1}{\sqrt{NM}}\sum_{n=1}^N\sum_{m=1}^M \psi_k(U_n)\left(\Phi^{U_n}_{(n-1)M+m} - v(U_n)\right),\end{aligned}$$

where the ξ_k and ϑ_k have zero mean for each $k = 1, \ldots, K$, and variances

$$\begin{aligned}\operatorname{var}\left[\xi_k\right] &= \operatorname{var}\left[\psi_k(U)v(U)\right] = E\left[\psi_k^2(U)v^2(U)\right] - E\left[\psi_k(U)v(U)\right]^2 \\ &= \int \mu(dx)\psi_k^2(x)v^2(x) - \left(\int \mu(dx)\psi_k(x)v(x)\right)^2, \qquad (3.16)\end{aligned}$$

$$\operatorname{var}\left[\vartheta_k\right] = \operatorname{var}\left[\psi_k(U_1)\eta_1\right] = \int \mu(dx)\psi_k^2(x)\operatorname{var}\left[\Phi^x\right], \qquad (3.17)$$

respectively. From (3.8) we then have

$$\begin{aligned}\overline{\gamma} &= \mathcal{G}^{-1}\left(w + \frac{\xi}{\sqrt{N}} + \frac{\vartheta}{\sqrt{MN}}\right)\\ &= \gamma^{\circ} + \frac{\mathcal{G}^{-1}\xi}{\sqrt{N}} + \frac{\mathcal{G}^{-1}\vartheta}{\sqrt{MN}}\end{aligned}$$

with $\xi \coloneqq [\xi_1, \dots, \xi_K]^\top$ and $\vartheta \coloneqq [\vartheta_1, \dots, \vartheta_K]^\top$. We so obtain a point-wise approximation error,

$$\begin{aligned}\overline{v}(x) - v(x) &= \psi^\top(x)\,(\overline{\gamma} - \gamma^{\circ}) + v_K(x) - v(x)\\ &= \psi^\top(x)\left(\frac{\mathcal{G}^{-1}\xi}{\sqrt{N}} + \frac{\mathcal{G}^{-1}\vartheta}{\sqrt{MN}}\right) + v_K(x) - v(x).\end{aligned}$$

with $\psi \coloneqq [\psi_1, \dots, \psi_K]^\top$. Since ξ and ϑ have zero mean, the point-wise bias—i.e., the error from projecting v to the span of the basis function $\psi_1, \dots, \psi_K$—is equal to

$$\overline{\delta}(x) \coloneqq v_K(x) - v(x)$$

and the $L^2_\mu(D)$-norm of the bias equals,

$$\left\|\overline{\delta}\right\| = \|v - v_K\|_{L^2_\mu}. \tag{3.18}$$

The bias depends on properties of v and the basis functions ψ and shall not be analyzed in this section.

On the other hand, let $\overline{\varepsilon}(x)$ denote the regression error on top of the bias $\overline{\delta}$, i.e.,

$$\overline{\varepsilon}(x) \coloneqq \overline{v}(x) - v_K(x).$$

Clearly, $\overline{\varepsilon}(x)$ is related with the "statistical error" of a standard Monte Carlo procedure. Let us denote

$$\overline{\mathrm{var}}_\mu \coloneqq \|\overline{\varepsilon}\|^2_{L^2(\Omega\times D;P\otimes\mu)},$$

which can be seen as the average (w.r.t. μ) over the point-wise variance of $\overline{v}$. We have

$$\begin{aligned}\overline{\mathrm{var}}_\mu &= \int_D \mu(dx) E\left[\left(\frac{\mathcal{G}^{-1}\xi}{\sqrt{N}} + \frac{\mathcal{G}^{-1}\vartheta}{\sqrt{MN}}\right)^\top \psi(x)\psi^\top(x)\left(\frac{\mathcal{G}^{-1}\xi}{\sqrt{N}} + \frac{\mathcal{G}^{-1}\vartheta}{\sqrt{MN}}\right)\right]\\ &= E\left[\left(\frac{\mathcal{G}^{-1}\xi}{\sqrt{N}} + \frac{\mathcal{G}^{-1}\vartheta}{\sqrt{MN}}\right)^\top \left(\frac{\xi}{\sqrt{N}} + \frac{\vartheta}{\sqrt{MN}}\right)\right]\\ &= \frac{1}{N}E\left[\xi^\top\mathcal{G}^{-1}\xi\right] + \frac{2}{N\sqrt{M}}E\left[\vartheta^\top\mathcal{G}^{-1}\xi\right] + \frac{1}{MN}E\left[\vartheta^\top\mathcal{G}^{-1}\vartheta\right],\end{aligned} \tag{3.19}$$

by using (3.7) and the symmetry of $\mathcal{G}$.

Let $\overline{\lambda}^K_{\min} > 0$ be the smallest eigenvalue of $\mathcal{G} = \mathcal{G}^K$. We then have by (3.16), (3.17), and Cauchy-Schwartz, the estimate

$$\overline{\mathrm{var}}_\mu \le \frac{1}{\overline{\lambda}^K_{\min}N}\sum_{k=1}^K \mathrm{var}\left[\xi_k\right] + \frac{2}{\overline{\lambda}^K_{\min}N\sqrt{M}}\sum_{k=1}^K \sqrt{\mathrm{var}\left[\xi_k\right]}\sqrt{\mathrm{var}\left[\vartheta_k\right]} + \frac{1}{\overline{\lambda}^K_{\min}MN}\sum_{k=1}^K \mathrm{var}\left[\vartheta_k\right]. \tag{3.20}$$

The following proposition is now obvious.

Proposition 3.3. *Suppose that the set of basis functions is such that the variances (3.16) and (3.17) are bounded by $\mathcal{V}$ for all k if $K \to \infty$, and that $\overline{\lambda}_{\min}^K \geq \overline{\lambda}_{\min} > 0$ if $K \to \infty$. We then have the estimate*

$$\|\overline{\varepsilon}\|^2_{L^2(\Omega\times D;P\otimes\mu)} \leq \frac{\mathcal{V}}{\overline{\lambda}_{\min}}\left(1 + \frac{2}{\sqrt{M}} + \frac{1}{M}\right)\frac{K}{N}.$$

Remark 3.4. In view of the above proposition, it is inefficient to take M larger than one or two in the semi stochastic regression. For example, $M = 2$ doubles the computation time but reduces the variance bound approximately by a factor 0.73 only. Moreover, assuming a choice of μ-orthonormal basis functions $\psi_1, \ldots, \psi_K$, the matrix $\mathcal{G}$ is the identity matrix and, consequently, the parameter $\overline{\lambda}_{\min} = 1$.

Deterministic regression statistics. Similar to (3.7) we introduce the matrix F and the vector z defined by

$$[F]_{kl} \coloneqq \frac{1}{N}\sum_{n=1}^{N} \psi_k(x_n)\psi_l(x_n), \tag{3.21}$$

$$[z]_l \coloneqq \frac{1}{N}\sum_{n=1}^{N} v(x_n)\psi_l(x_n), \quad k, l = 1, \ldots, K.$$

We thus have that

$$\frac{1}{N}\left[\mathcal{N}^T\mathcal{N}\right]_{kl} = \frac{1}{N}\sum_{n=1}^{N}\psi_k(x_n)\psi_l(x_n) = [F]_{kl},$$

and

$$\begin{aligned}\frac{1}{N}\left[\mathcal{N}^T\widetilde{\mathcal{Y}}\right]_k &= \frac{1}{N}\sum_{n=1}^{N}\psi_k(x_n)\widetilde{\mathcal{Y}}_n \\ &= \frac{1}{N}\sum_{n=1}^{N}\psi_k(x_n)\overline{\Phi}^{x_n}_{n,M} \\ &= \frac{1}{N}\sum_{n=1}^{N}\psi_k(x_n)\left(v(x_n) + \frac{\widetilde{\eta}_n}{\sqrt{M}}\right)\end{aligned}$$

with independent random variables

$$\widetilde{\eta}_n \coloneqq \widetilde{\eta}(x_n) \coloneqq \sqrt{M}\left(\overline{\Phi}^{x_n}_{n,M} - v(x_n)\right) = \frac{1}{\sqrt{M}}\sum_{m=1}^{M}\left(\Phi^{x_n}_{(n-1)M+m} - v(x_n)\right),$$

satisfying $E[\widetilde{\eta}_n] = 0$ and

$$\operatorname{var}[\widetilde{\eta}_n] = \operatorname{var}[\Phi^{x_n}] \ = E\left[(\Phi^{x_n} - v(x_n))^2\right].$$

So we can write

$$\frac{1}{N}\left[\mathcal{N}^T\widetilde{\mathcal{Y}}\right]_k = z_k + \frac{\widetilde{\vartheta}^{N,M}_k}{\sqrt{NM}},$$

where for $k = 1, \ldots, K$,

$$\widetilde{\vartheta}_k \coloneqq \frac{1}{\sqrt{N}}\sum_{n=1}^{N}\psi_k(x_n)\widetilde{\eta}(x_n)$$

has zero mean and variance,

$$\operatorname{var}\left[\widetilde{\vartheta}_k\right] = \frac{1}{N}\sum_{n=1}^{N}\psi_k^2(x_n)\operatorname{var}\left[\Phi^{x_n}\right]. \tag{3.22}$$

Next we proceed with

$$\widetilde{\gamma} = \left(\frac{1}{N}\mathcal{N}^T\mathcal{N}\right)^{-1}\frac{1}{N}\mathcal{N}^T\widetilde{\mathcal{Y}} = F^{-1}z + \frac{1}{\sqrt{NM}}F^{-1}\widetilde{\vartheta}.$$

We obtain the random (point-wise) error

$$\begin{aligned}\widetilde{\varepsilon}(x) &:= \psi^\top(x)\widetilde{\gamma} - v(x)\\ &= \psi^\top(x)F^{-1}z - v(x) + \frac{1}{\sqrt{NM}}\psi^\top(x)F^{-1}\widetilde{\vartheta}.\end{aligned} \tag{3.23}$$

Since $\widetilde{\vartheta} = \left[\widetilde{\vartheta}_1, \dots, \widetilde{\vartheta}_k\right]$ has zero mean, the point-wise bias is thus equal to

$$\widetilde{\delta}(x) := \widetilde{\delta}^{N,K}(x) := \psi^\top(x)\left(F^{N,K}\right)^{-1}z^{N,K} - v(x). \tag{3.24}$$

Obviously, with

$$\mathcal{A}_D := \int_D dx$$

it holds that

$$F_{kl}^{N,K} \underset{N\to\infty}{\longrightarrow} F_{kl}^{\infty,K} := \frac{1}{\mathcal{A}_D}\int_D \psi_k(y)\psi_l(y)dy \quad \text{and} \quad z_k^{N,K} \underset{N\to\infty}{\longrightarrow} z_k^{\infty,K} := \frac{1}{\mathcal{A}_D}\int_D v(y)\psi_k(y)dy,$$

hence we may define

$$\begin{aligned}\widetilde{\delta}^{\infty,K}(x) &:= \psi^\top(x)\left(F^{\infty,K}\right)^{-1}z^{\infty,K} - v(x)\\ &=: \psi^\top(x)\gamma^{\dagger,K} - v(x),\end{aligned} \tag{3.25}$$

where we introduce

$$\gamma^{\dagger,K} := \left(F^{\infty,K}\right)^{-1}z^{\infty,K}.$$

In fact, $\psi^\top\gamma^{\dagger,K}$ is the projection of the true solution v on the span of the basis functions, with respect to the scalar product

$$\langle f, g\rangle_{L_D^2} := \int_D f(y)g(y)dy.$$

Thus, the (global) projection error with respect to the corresponding norm satisfies

$$\left\|\widetilde{\delta}^{\infty,K}\right\|_{L_D^2}^2 := \left\|\psi^\top\gamma^{\dagger,K} - v\right\|_{L_D^2}^2 = \int_D\left(\psi^\top(y)\gamma^{\dagger,K} - v(y)\right)^2 dy.$$

It is natural to measure the global bias (3.24) with respect to this norm also, that is,

$$\left\|\widetilde{\delta}^{N,K}\right\|_{L_D^2}^2 = \left\|\widetilde{\delta}\right\|_{L_D^2}^2 = \int_D\left(\psi^\top(y)\left(F^{N,K}\right)^{-1}z^{N,K} - v(y)\right)^2 dy.$$

By writing

$$\begin{aligned}\widetilde{\delta}^{N,K}(x) &= \widetilde{\delta}^{\infty,K}(x) + \psi^\top(x)\left(\left(F^{\infty,K}\right)^{-1}\left(z^{N,K} - z^{\infty,K}\right) + \left(\left(F^{N,K}\right)^{-1} - \left(F^{\infty,K}\right)^{-1}\right)z^{N,K}\right)\\ &=: \widetilde{\delta}^{\infty,K}(x) + \mathcal{R}_{N,K}(x),\end{aligned}$$

we obtain the global estimate

$$\left\|\widetilde{\delta}^{N,K}\right\|_{L_D^2} = \left\|\widetilde{\delta}^{\infty,K} + \mathcal{R}_{N,K}\right\|_{L_D^2} \le \left\|\widetilde{\delta}^{\infty,K}\right\|_{L_D^2} + \left\|\mathcal{R}_{N,K}\right\|_{L_D^2},$$

where the residual term can be bounded from above by

$$\left\|\mathcal{R}_{N,K}\right\|_{L^2_D} \leq \|\psi\|_{L^2_D}\left(\left\|\left(F^{\infty,K}\right)^{-1}\left(z^{N,K}-z^{\infty,K}\right)\right\|_K + \left\|\left(\left(F^{N,K}\right)^{-1}-\left(F^{\infty,K}\right)^{-1}\right)z^{N,K}\right\|_K\right), \tag{3.26}$$

and where

$$z^{N,K}-z^{\infty,K} = \text{const}\frac{d}{N^{1/d}} \quad \text{and} \quad \left[\left(F^{N,K}\right)^{-1}-\left(F^{\infty,K}\right)^{-1}\right]_{kl} = \text{const}\frac{d}{N^{1/d}},$$

i.e.,

$$\left\|\widetilde{\delta}^{N,K}\right\|_{L^2_D} \approx \left\|\widetilde{\delta}^{\infty,K}\right\|_{L^2_D} + \text{const}\frac{d}{N^{1/d}}. \tag{3.27}$$

So, unlike in the semi stochastic case where the bias was only depending on K, the bias (3.24) depends on N and K.

Remark 3.5. Similarly to the semi-stochastic regression analyzed above, we could now also switch to a "semi-deterministic" regression by replacing F_{kl} with $\mathcal{A}_D^{-1}\int_D \psi_k(y)\psi_l(y)dy$. This would help avoid any stability issues in the above linear system when $N < K$. Apart from this, the second term in (3.26) would disappear, but the subsequent complexity analysis would remain unchanged.

We now proceed with the estimation of the variance in (3.23). Similar to (3.19) we obtain for the L^2_D-norm of the point-wise variance, $\frac{1}{\sqrt{NM}}\psi^\top(x)F^{-1}\widetilde{\vartheta}$

$$\begin{aligned}\widetilde{\text{Var}}_{L^2_D} &:= \frac{1}{NM}\int_G E\left[\left(F^{-1}\widetilde{\vartheta}\right)^\top \psi(y)\psi^\top(y)F^{-1}\widetilde{\vartheta}\right]dy\\ &= \frac{1}{NM}E\left[\widetilde{\vartheta}^\top F^{-1}F^{\infty,K}F^{-1}\widetilde{\vartheta}\right]\end{aligned}$$

by using the symmetry of F. Let $\lambda^K_{\min}$ be the smallest eigenvalue of $F^{\infty,K}$. Since for fixed K,

$$F^{-1}F^{\infty,K}F^{-1} = \left(F^{N,K}\right)^{-1}F^{\infty,K}\left(F^{N,K}\right)^{-1} \to F^{\infty,K}, \quad \text{if } N\to\infty,$$

we may assume that N is large enough, such that the smallest eigenvalue of $F^{-1}F^{\infty,K}F^{-1}$, $\lambda^{N,K}_{\min}$ say, satisfies $\lambda^{N,K}_{\min} > \lambda^K_{\min}/2$. Then, analogue to (3.20), we obtain the variance bound,

$$\widetilde{\text{Var}}_{L^2_D} \leq \frac{2}{\lambda^K_{\min}NM}\sum_{k=1}^K \text{Var}\left[\widetilde{\vartheta}_k\right].$$

We so have the following proposition.

Proposition 3.6. *Suppose that the set of basis functions is such that the variances (3.22) are bounded by $\widetilde{\mathcal{V}}$ for all N and k if $N, K \to \infty$. Suppose that $\lambda^K_{\min} > \widetilde{\lambda}_{\min}$ for each K. Let us take $N > K$ such that $\lambda^{N,K}_{\min} \geq \widetilde{\lambda}^K_{\min}/2 > \widetilde{\lambda}_{\min}/2$. We then have the estimate*

$$\widetilde{\text{Var}}_{L^2_D} \leq \frac{2\widetilde{\mathcal{V}}}{\widetilde{\lambda}_{\min}}\frac{K}{NM}.$$

Let us have a look at the bias of the deterministic regression procedure again. It is clear that, even if K is such that $\left\|\widetilde{\delta}^{\infty,K}\right\|_{L^2_D}$ is small, a too small N causes extra bias due to the deterministic integration error, cf. (3.27). This in contrast to the semi stochastic method where the bias (3.18) is independent of the number N of Monte Carlo samples of the random variable $\mathcal{U}$. Therefore, for larger dimensions d the semi stochastic method outperforms the deterministic one.

For a closer look at the bias of the deterministic regression, observe that

$$\left\|\widetilde{\delta}^{N,K}\right\|_{L^2_D} \longrightarrow \left\|\widetilde{\delta}^{\infty,K}\right\|_{L^2_D}.$$

It is clear that, even if K is such that $\left\|\widetilde{\delta}^{\infty,K}\right\|_{L^2_D}$ is small, a too small N causes extra bias due to the deterministic integration error. This is in contrast to the semi stochastic method where the bias (3.18) is independent of the number N of Monte Carlo samples of the random variable U.

The size of the deterministic integration error as a function of the number N of grid points depends on the nature of the point set $x_1, \dots, x_N$. For instance, in the case of a uniform tensor grid $x_1, \dots, x_N$ in dimension d, the integration error is, in principle, of order $\frac{1}{N^{1/d}}$, and, hence, the bias gives

$$\left\|\widetilde{\delta}^{N,K}\right\|_{L^2_D} \approx \left\|\widetilde{\delta}^{\infty,K}\right\|_{L^2_D} + \text{const}\,\frac{d}{N^{1/d}}, \tag{3.28a}$$

where the second term amounts to the deterministic integration error of the present integration method based on ρ_N. On the other hand, if we choose $x_1, \dots, x_N$ as the first N points of a d-dimensional low-discrepancy sequence, then the integration error can be reduced to $\frac{\log^d N}{N}$, leading to

$$\left\|\widetilde{\delta}^{N,K}\right\|_{L^2_D} \approx \left\|\widetilde{\delta}^{\infty,K}\right\|_{L^2_D} + \text{const}\,\frac{\log^d(N)}{N}. \tag{3.28b}$$

We refer to [28] for a detailed exposition of the low-discrepancy case and to [41] for an analysis based on collocation grids.

4. Convergence and complexity analysis

This section is concerned with a first analysis of the convergence and complexity of the proposed numerical method. We point out that these are only first results and a more thorough analysis should be carried out in subsequent work.

In order to clarify the proposed algorithm, we sketch a possible implementation in pseudo code. First, the Euler scheme for the simulation of Brownian motion is depicted in Algorithm 4.1. For the sake of simplicity, only uniform timestepping is included. Second,

In : Field realization κ, initial value $x \in D$, time-increment Δt.
Out: Trajectory $\left(\overline{X}_{t_0}, \dots, \overline{X}_{\overline{\tau}}\right)$ of the approximate path and approximate hitting time $\overline{\tau}$.

$\overline{X}_{t_0} := x$, $i := 0$
while $\overline{X}_{t_i} \in D$ **do**
  Simulate the increment of the Brownian motion
  $\Delta W_i \sim \mathcal{N}(0, \Delta t)$
  $\overline{X}_{t_{i+1}} := \overline{X}_{t_i} + \nabla\kappa(\overline{X}_{t_i})\Delta t + \sqrt{2\kappa(\overline{X}_{t_i})}\Delta W_i$
  $i := i + 1$
return $\left(\overline{X}_{t_0}, \dots, \overline{X}_{\overline{\tau}}\right)$

Algorithm 4.1: Trajectory simulation of Brownian motion by Euler scheme.

a sketch of the semi-stochastic regression algorithm is shown in Algorithm 4.2 based on the Euler scheme. The deterministic regression algorithm has a very similar structure.

In : Number N of samples to be computed, number K of basis functions and basis functions $\psi_1, \dots, \psi_K : D \to \mathbb{R}$, a probability measure μ on D, the matrix $\mathcal{G}$ defined in (3.7), time increment Δt for use by the Euler method.

Out: Solution approximation $\overline{v}(x) \coloneqq \sum_{k=1}^{K} \overline{\gamma}_k \psi_k(x)$ for any $x \in D$.

Allocate an N-dimensional array $\mathcal{Y}$ and $N \times K$ dimensional array $\mathcal{M}$

for $i = 1, \dots, N$ **do**

Sample $x_i \sim \mu$, sample realizations κ_i, f_i, g_i from the distribution of the conductivity field, the source field and the boundary field, respectively.

Call `Euler`$(\kappa_i, x_i, \Delta t)$ Algorithm 4.1, returning a pair of $\left(\overline{X}_{t_0}, \dots, \overline{X}_{\overline{\tau}}\right)$ and $\overline{\tau}$.

$\Phi_i \coloneqq g_i\left(\overline{X}_{\overline{\tau}}\right) + \int_0^{\overline{\tau}} f_i\left(\overline{X}_s\right) \mathrm{d}s$ (with suitable quadrature)

$\mathcal{Y}_i \coloneqq \Phi_i$, $\mathcal{M}_{i,k} \coloneqq \psi_k(x_i)$, $k = 1, \dots, K$

$i \coloneqq i + 1$

$\overline{\gamma} \coloneqq \frac{1}{N} \mathcal{G}^{-1} \mathcal{M}^\top \mathcal{Y}$

return $\overline{v}(\cdot) \coloneqq \sum_{k=1}^{K} \overline{\gamma}_k \psi_k(\cdot)$

Algorithm 4.2: Global regression algorithm to determine solution approximation based on point-wise samples.

Coming back to the analysis of the error, let us first consider the case of a fixed point $x \in D \subset \mathbb{R}^d$, i.e., we want to approximate the solution $u(x)$ or $E[u(x)]$ for this particular single point x. Recall the elliptic model problem (1.1) with stochastic coefficient κ, $\omega \in \Omega$. Then we can naturally decompose the error into four parts:

(i) The error from approximating the stochastic fields κ, g and f;
(ii) the error from the discretization of the SDE (2.1a) and the functional (2.1d);
(iii) the truncation error in the regression, i.e., the error introduced by choosing a finite, non-dense set of basis functions;
(iv) the integration error in the regression, i.e., the error introduced by *computing* an approximate projection of the true solution to the basis functions only.

In this paper, we are mainly concerned with the second and fourth sources of errors of the method. We believe this can be justified for pragmatic reasons: while a quite general and convincing error and convergence analysis can be given for all mentioned sources of errors, the influence of the first and third error component is more difficult to describe without imposing very specific assumptions on both the coefficients and the solution of the random PDE. We just make the following remarks:

- We only consider stochastic fields with finite variance which can be represented by a Karhunen-Loève expansion. These fields are described by a covariance function the regularity of which directly determines the smoothness of the realizations of the stochastic field. Note that for the considered application, there always is some correlation between sufficiently close points in the domain (i.e. coloured noise). The number of expansion terms required for an adequate approximation again directly depends on the regularity of the covariance. If we assume an expansion (1.2)

of the form

$$\kappa_\iota(x,\omega) = E[\kappa] + \sum_{m=1}^{\iota} \sqrt{\nu_m}\varphi_m(x)\xi_m(\omega)$$

which is exact for $\iota = \infty$, then the truncation error for M terms is determined by

$$\|\kappa_\infty - \kappa_M\|^2_{L^2(\Omega,L^2(D))} \leq \sum_{m=M+1}^{\infty} \nu_m.$$

The ν_m are the eigenvalues of the covariance integral operator and their decay behavior thus determines the truncation error with respect to the number of KL terms M. In some cases, the decay behavior is known a priori for which we refer to [9, 38, 26]. Note that alternate techniques to generate field realizations can be employed equally well with our method. For instance, turning band methods and circulant embedding are frequently used approaches, see [26].

- For a given set of basis functions $\psi_1, \dots, \psi_K$, the truncation error of the regression method, denoted as *bias* in Section 3, is given by

$$v(x) - v_K(x),$$

where v_K denotes the projection of v to $\operatorname{span}\{\psi_1, \dots, \psi_K\}$, cf. (3.13) and (3.14). Clearly, if $v \in L^2(D;\mu)$, then the error will converge to 0 as $K \to \infty$, for any reasonable sequence ψ_k of basis functions, for instance an orthonormal basis of $L^2(D;\mu)$. The *speed* of convergence, however, depends on the regularity of the solution as well as the choice of the basis functions. For instance, if the solution is actually analytic, then the error decays exponentially, for good choices of basis functions. Algebraic convergence (with rate depending on the dimension) can be obtained when the solution is differentiable with square integrable derivatives up to a certain order. For more information we refer to Pinsker [33].

In this paper, we assume that the true solution v can be approximated by $v_K \in \operatorname{span}\{\psi_1, \dots, \psi_K\}$ with an error

$$\|v - v_K\|_{L^2_\mu} \leq e(K).$$

We remark that a similar analysis can be done for other norms, say $\|\cdot\|_{H^1(D)}$. This would require a similar error analysis of the regression method as shown for the L^2 norm. We only assume that $\lim_{K\to\infty} e(K) = 0$ and that e is invertible with inverse e^{-1}.

We first discuss the time-discretization error (ii). By the empirically well-established first order weak convergence of the adaptive Euler scheme, see [12, 6], an error tolerance ϵ_{disc} will yield approximate solutions $\overline{X}$, $\overline{Y}$, $\overline{Z}$ and a corresponding stopping time $\overline{\tau}$ such that the corresponding expected value $v_{\text{disc}}(x) \coloneqq E\left[g\left(\overline{X}_{\overline{\tau}}\right)\overline{Y}_{\overline{\tau}} + \overline{Z}_{\overline{\tau}}\right]$ satisfies $\|v - v_{\text{disc}}\|_{L^2_\mu} \leq \epsilon_{\text{disc}}$ at a computational cost of order $\epsilon_{\text{disc}}^{-1}$ *on average* for computing one realization. Notice that the computational cost is a random variable as the hitting time at the boundary is random. However, uniform ellipticity and boundedness of the domain D (cf. Assumption 2.10 and Lemma 2.11) guarantee that the hitting time is square integrable, both for the approximate process $\overline{X}$ and the true solution X.

We next consider the integration error in the regression, term (iv), first concentrating on the semi-stochastic case. Recall that the outcome of the semi-stochastic global regression $\overline{v} = \overline{v}_K = \sum_{k=1}^K \overline{\gamma}_k\psi_k$ is obtained from $\overline{\gamma} = \frac{1}{N}\mathcal{G}^{-1}\mathcal{M}^\top\mathcal{Y}$, see (3.9) and (3.8). By Proposition 3.3, for instance assuming that $v = v_{\text{disc}}$ and $\operatorname{var}\Phi^x$ are uniformly bounded in x on D (as guaranteed by Assumption 2.10) and, for simplicity, that $(\psi_k)_{k\geq 1}$ are orthonormal

w.r.t. μ, the error is bounded by

$$\|v_K - \overline{v}_K\|^2_{L^2(\Omega\times D;P\otimes\mu)} \leq \text{const}\left(1 + \frac{2}{\sqrt{M}} + \frac{1}{M}\right)\frac{K}{N}. \tag{4.1}$$

Clearly, $M = 1$ is the optimal choice, so we disregard other possible choices henceforth.

Remark 4.1. For this discussion, we note that the regression actually approximates v_{disc}, not the true solution v. Hence, the assumptions needed in Section 3 have to hold for the discretized solution v_{disc} uniformly in ϵ_{disc}.

In total, we obtain the error decomposition

$$\|v - \overline{v}_K\|_{L^2\left(\Omega\times\widetilde{D};P\otimes\widetilde{\mu}\right)} \leq \|v - v_{\text{disc}}\|_{L^2_\mu} + \|v_{\text{disc}} - v_K\|_{L^2_\mu} + \|v_K - \overline{v}_K\|_{L^2(\Omega\times D;P\otimes\mu)} =: \epsilon \tag{4.2}$$

with overall error bound ϵ.

For fixed N and K, the computational cost of the regression part of the algorithm is proportional to NK, if the cost of computing the N realizations $\Phi_n^{U_n}$, $n = 1,\dots,N$, is neglected. For fixed K, we need to choose N proportional to $K\epsilon^{-2}$, so the computational cost of the regression is proportional to $K^2\epsilon^{-2}$. On the other hand, we need to sample N realizations of Φ^U at cost proportional to ϵ^{-1} each, which amounts to cost proportional to $K\epsilon^{-3}$. Finally, by assumption, we need to choose K proportional to $e^{-1}(\epsilon)$ to achieve a truncation error bounded by ϵ. To summarize, we obtain

Proposition 4.2. *Given Assumption 2.10 and a discretization error tolerance ϵ in $\|\cdot\|_{L^2(\Omega\times D;P\otimes\mu)}$. Then the average computational cost C of the semi-stochastic global regression algorithm with adaptive time-discretization is bounded by*

$$C \leq C_1 e^{-1}(\epsilon)\epsilon^{-3} + C_2 e^{-2}(\epsilon)\epsilon^{-2}.$$

Remark 4.3. Note that the computational cost is, superficially, independent of the dimension d. As discussed earlier, $e^{-1}(\epsilon)$ could be anything between, say, $\log\epsilon^{-1}$ and ϵ^{-d} (or even worse). Hence, it is not clear which of the two terms is dominant. One might expect the constant C_1 to be typically much bigger than C_2, as C_2 essentially just entails a floating point multiplication, whereas C_1 is the entire computational cost of one step of the adaptive Euler scheme.

Remark 4.4. If we are only interested in the point-wise error, i.e., if we only want to compute $v(x)$ for one specific value $x \in D$, then we can replace the regression analysis by a simple Monte Carlo analysis. The computational cost of our method for computing $v(x)$ at tolerance ϵ in RMSE sense is then proportional to ϵ^{-3}, independent of the dimension d.

For the deterministic regression procedure analyzed in Proposition 3.6 (variance) and (3.28) (bias), we need to replace the estimate (4.1) by

$$\|v_K - \overline{v}_K\|^2_{L^2(\Omega\times D;P\otimes dx)} \leq \text{const}\left(\frac{K}{NM} + \frac{d^2}{N^{2/d}}\right), \tag{4.3a}$$

for a uniform, tensorized grid $x_1,\dots,x_N$ and by

$$\|v_K - \overline{v}_K\|^2_{L^2(\Omega\times D;P\otimes dx)} \leq \text{const}\left(\frac{K}{NM} + \frac{d^2}{N^2}\right), \tag{4.3b}$$

(ignoring logarithmic terms) for the low-discrepancy case, where we use the special choice $\mu = dx|_D$. The cost-optimal choice for M will now depend on K — treated as fixed at

this stage —, the error tolerance ϵ and the dimension d. The computational cost C of performing the algorithm is

$$C \leq \text{const}\left(MN\epsilon^{-1} + KN + K^2\right)$$

on average, corresponding to the average cost of computing MN samples at accuracy ϵ by the adaptive Euler-Maruyama algorithm, the cost of multiplying a $K \times N$-matrix with an N-dimensional vector and the cost of solving a $K \times K$-linear system, cf. (3.11). Hence, one has to minimize the cost given that the error (4.3) is bounded by ϵ^2.

Let us first consider the case of a tensorized uniform grid. We may assume that both error contributions in (4.3a) are of order ϵ^2, implying that

$$N = \text{const}\, d^d \epsilon^{-d} \quad \text{and} \quad M = \text{const}\max\left(Kd^{-d}\epsilon^{d-2}, 1\right).$$

Hence, for the computational cost it holds

$$C \leq \text{const}\left(\max\left(K\epsilon^{-3}, d^d\epsilon^{-(d+1)}\right) + d^d K\epsilon^{-d} + K^2\right).$$

No further calculation is needed for the case of $x_1, \dots, x_N$ being based on a low-discrepancy sequence, as this case essentially (up to logarithmic terms) corresponds to the case $d = 1$ in the uniform case.

Proposition 4.5. *Assume the conditions of Proposition 4.2 and an error tolerance ϵ in the sense of $\|\cdot\|_{L^2(\Omega\times D;P\otimes dx)}$.*
a) If the grid $x_1, \dots, x_N$ is a uniform, tensorized grid in dimension d, then the average computational cost C of the deterministic global regression algorithm with adaptive time-discretization is bounded by

$$C \leq C_1 \max\left(e^{-1}(\epsilon)\epsilon^{-3}, d^d\epsilon^{-(d+1)}\right) + C_2 d^d e^{-1}(\epsilon)\epsilon^{-d} + C_3 e^{-2}(\epsilon)$$

with constants $C_1, C_2, C_3 > 0$.
b) Up to logarithmic terms, the above bound holds with $d = 1$ regardless of the dimension of the space if the point set $x_1, \dots, x_N$ has low discrepancy.

Remark 4.6. Even in the analytic case ($e(K) \sim e^K$) we already see the curse of dimensionality in the deterministic regression case. It appears because of the inherent numerical approximation of integrals w.r.t. μ based on the grid, i.e., the approximation error $\mu \approx \rho_N$, in the notation of Section 3. Further, note that we have not considered stability restrictions ($N \gg K$) on the choice of N and K induced by the design matrix $\mathcal{N}^\top\mathcal{N}$. Similarly to Remark 4.3, we note that typically $C_1 \gg C_2, C_3$.

$e(K)$	Semi-stochastic reg.	Det. reg. (tensor)	Det. reg. (low disc.)
e^{-K}	ϵ^{-3}	$\max(\epsilon^{-(d+1)}, \epsilon^{-3})$	ϵ^{-3}
$K^{-1/d}$	$\epsilon^{-(3+d)}$	$\epsilon^{-(3+d)}$	$\epsilon^{-(3+d)}$

TABLE 1. Asymptotic computational costs for deterministic (tensor grid or low discrepancy sequence) and semi-stochastic regression for error tolerance ϵ. Logarithmic terms are ignored, C_2 and C_3 are set to 0.

We end these theoretical complexity considerations by a sketchy asymptotic comparison of deterministic and semi-stochastic regression techniques, see Table 1 and Table 2. We compare the case of a very fast decay of the error in terms of the number of basis functions K with the case of a slow small decay, i.e., the case of an analytic solution v with a merely

$e(K)$	Semi-stochastic reg.	Det. reg. (tensor)	Det. reg. (low disc.)
e^{-K}	ϵ^{-3}	$\max(\epsilon^{-(d+1)}, \epsilon^{-3})$	ϵ^{-3}
$K^{-1/d}$	$\epsilon^{-(2+2d)}$	$\epsilon^{-(3+d)} + \epsilon^{-2d}$	$\epsilon^{-(3+d)} + \epsilon^{-2d}$

TABLE 2. Asymptotic computational costs for deterministic (tensor grid or low discrepancy sequence) and semi-stochastic regression for error tolerance ϵ. Logarithmic terms are ignored, C_2 and C_3 are kept.

square-integrable solution v. In Table 1, we only consider the cost of generating samples from the solution of the SDE. This is usually the dominant contribution to the overall computational costs, as the cost of computing one step in the stochastic Euler scheme is much larger than the cost of simple floating point multiplications. In Table 2, we treat all the contributions as equivalent, which is adequate for a true asymptotic analysis. We see that the semi-stochastic regression is clearly superior in the highly regular case, at least for $d > 2$ and it is never worse than deterministic regression based on a uniform, tensorized grid in the realistic scenario with C_2 and C_3 ignored. This is a consequence of the curse of dimensionality. On the other hand, the deterministic regression algorithm based on a low discrepancy point set $x_1, \dots, x_N$ seems comparable to the semi-stochastic algorithm. We should note, however, that this very simple analysis does not take the constants into account. Recall that all the presented algorithms are, overall, stochastic in nature. A pure QMC version of the algorithm—i.e., an approach where the random coefficients and the Brownian motion are replaced by their deterministic counterpart—seems difficult due to the very high dimensionality.

5. Examples

For the benchmark problem at hand we consider a constant right-hand side $f \equiv 1$ on the unit square domain $D = [0,1]^2 \subset \mathbb{R}^2$. The Dirichlet boundary data $g = \sin(\pi x_1) + \sin(\pi x_2)$ is enforced on the whole boundary ∂D of the domain. Here, x_i denotes component i of the coordinate vector x. The permeability tensor κ is constructed in a form resembling a Karhunen-Loève expansion. It exhibits the characteristics of a separable covariance function on the unit square and is easily controlled with respect to amplitude and frequency of the field. More precisely, in (1.2) we set $E[\kappa] = 1$ and consider the coefficients a_m with $m = 1, 2, \dots$ and

$$\begin{aligned} a_m(x) &= \alpha_m \cos(2\pi\beta_1(m)x_1)\cos(2\pi\beta_2(m)x_2), \quad \alpha_m = A_\alpha m^{-\sigma_\alpha}, \\ \beta_1(m) &= m - k(m)(k(m)+1)/2, \quad \beta_2(m) = k(m) - \beta_1(m), \\ k(m) &= \lfloor -1/2 + \sqrt{1/4 + 2m} \rfloor. \end{aligned} \tag{5.1}$$

The parameters must verify $\sigma_\alpha > 1$ and $0 < A_\alpha < 1/\zeta(\sigma_\alpha)$ with the Riemann zeta function ζ. For uniformly distributed random variables $\varphi_m \sim U(-1,1)$, the parameters $c_a, \varepsilon_a > 0$ and the truncation length $t_a \in \mathbb{N}$ determine the (computational) random field κ by

$$\kappa(x) = \frac{c_a}{\alpha_{\min}} \left(\sum_{m=1}^{t_a} a_m(x)\varphi_m + \alpha_{\min} \right) + \varepsilon_a. \tag{5.2}$$

Remark 5.1. We suppose the model (5.2) of the field κ in an t_a-term expansion to be the exact representation. This, of course, usually is not the case and one has to consider how the truncation affects the solution accuracy. Here, the representation merely serves the purpose of an easy to compute miscellaneous random field which fulfills Assumption 2.10.

However, an explicit representation of κ, based on a countable set of random variables, is neither necessary nor used in the presented method.

The constant $\alpha_{\min}$ is set to the absolute value of the minimum for the sum over the a_m, i.e. $\alpha_{\min} = \sum_{m=1}^{t_a} \alpha_m = \left|\min_{x\in D} \sum_{m=1}^{t_a} a_m(x)\varphi_m\right| = \sum_{m=1}^{t_a} \alpha_m$.

The basis functions ψ_k for $k = 1, \dots, K$ in the regression methods from Subsection 3.1 are chosen as the Legendre polynomials. We choose the polynomial degree 4 for each spatial direction which results in $K = 25$ basis functions. In case that locations x are drawn uniformly in D, this choice admits the advantage of $G = I$ with the identity matrix I in (3.7) such that no inverse of G needs to be computed in (3.8).

A reference solution for this problem is obtained with a simple Monte Carlo approach utilizing a standard finite element solver. A set of $N = 10^6$ samples κ^i for $i = 1, \dots, N$ is drawn from the random variable κ and the finite element method gives a discrete solution u_h^i on some very fine mesh with approximately $3 \cdot 10^6$ degrees of freedom for each sample. Subsequently, the stochastic estimator for the expected value is given by $\hat{u}_h = N^{-1} \sum_{i=1}^{N} u_h^i$ and we assume that $E[u] \approx \hat{u}_h$ is a sufficiently close approximation.

As described in Section 4, we note that we only consider the error from the time-discretization and the regression steps of our algorithm. In particular, we work with an exact stochastic field given by a truncated expansion ("finite dimensional noise assumption") and also fix the number of basis functions. This implies that the error will converge to a positive value given by the bias, i.e., by the error induced by projection to the fixed set of basis functions.

In the following subsections, we first demonstrate point-wise convergence of the scheme at an arbitrary location (the center) in the domain. Then, global convergence of the regression approximation is examined with two different coefficient fields. First, a "smooth" field with few expansion terms and thus only low frequencies is considered. Second, a "rough" field with many expansion terms is employed. This also contains high frequency components and thus exhibits large gradients. With both settings, the convergence for a structured (deterministic) and a random (stochastic) selection of regression points in the domain D is depicted.

Remark 5.2 (Notes regarding the implementation)**.** We implemented the Euler-Maruyama Monte-Carlo solver in C++ as a Python module. Using `OpenMP`, the solver calculates the trajectories in parallel on an arbitrary amount of processors. To sample the discrete Brownian motion we use a 64-bit Mersenne-Twister random number generator (RNG) from the C++ standard library. To ensure independently sampled random variables, each thread has his own distinctly seeded RNG. Being a Python module, we can easily compare our solution with the reference solution computed with the finite element solver `FEniCS` [24, 1] and references therein. Moreover, the fully/semi stochastic and deterministic global regression methods from Subsection 3.1 are also implemented in Python using the `NumPy` package.

5.1. **Convergence in one point.** To visualize the outcome of the simulation algorithm detailed in the preceding subsection, we set $\sigma_\alpha = 2$ and $A_\alpha = 0.6$ in (5.1) and $t_a = 5$, $c_a = 1$ and $\varepsilon_a = 5 \cdot 10^{-4}$ in (5.2). This yields a rather smooth coefficient field. This first test compares the results of the implemented solver with the reference solution at a single point $x \in D$. We examine convergence in two ways in Figure 2. On the left, we observe a convergence rate of 1 in the Euler-Maruyama scheme by decreasing the initial time step Δt_0. Note that an adaptive time step calculation is applied which reduces the time

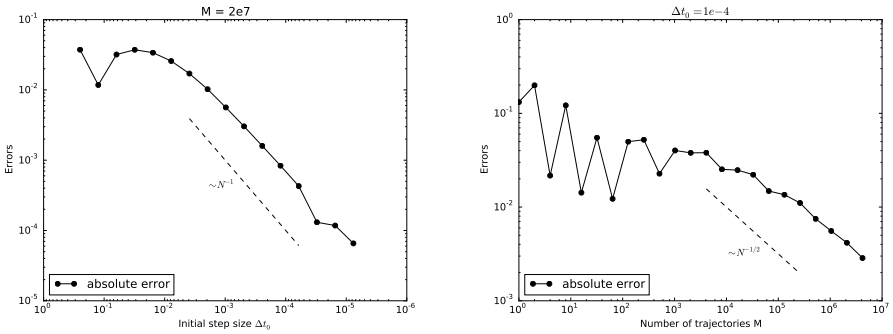

FIGURE 2. Convergence of the solution in $x = [0.5, 0.5]$.

steps close to the boundary ∂D. On the right, we can see a convergence rate of $1/2$ due to increasing the number of trajectories M in the Monte Carlo simulation.

5.2. **Experiment - smooth benchmark field.** We again choose $\sigma_\alpha = 2$ and $A_\alpha = 0.6$ in (5.1) and $t_a = 5$, $c_a = 1$ and $\varepsilon_a = 5 \cdot 10^{-4}$ in (5.2). This time we are interested in convergence in the whole domain. We thus measure the errors in the $L^2(D)$ and $H^1(D)$ norms.

Deterministic global regression. The first part of this experiment applies the deterministic global regression from (3.10) to solutions with the above example problem data. On a given uniform triangulation of D, a single trajectory is computed for each vertex. The convergence of the errors in the $L^2(D)$ and $H^1(D)$ norms is depicted in Figure 3 with square markers. Here, N is the total number of grid points and $M = 1$ is the number of trajectories starting from each grid point. We also set an initial time step Δt_0 small enough such that the first term on the right-hand side in (4.2) is smaller than the other error contributions. Hence, Figure 3 shows the error resulting from the global regression. From Proposition 3.6 we expect to see a RMSE of order $\frac{1}{\sqrt{MN}}$ in terms of the total number of trajectories MN, i.e., a convergence rate of $1/2$. Hence, the numerical results are in line with the theory.

Semi stochastic global regression. The second part of this experiment uses uniformly distributed points in the domain D together with the semi stochastic global regression from (3.7). Once again, for each sample point a single trajectory is computed. The results are depicted in Figure 3 with circular markers. They show the same behavior of the L^2 error as in the first experiment with a slightly improved performance in the H^1 norm. The convergence rate is the same as in the previous part of the experiment, as predicted by Proposition 3.3.

5.3. **Experiment - rough benchmark field.** We now choose $\sigma_\alpha = 1.1$ and $A_\alpha = 0.0009$ in (5.1). Furthermore we drop the first 1000 terms of the sum in (5.2) and take $t_a = 1020$, $c_a = 1$ and $\varepsilon_a = 5 \cdot 10^{-4}$. This results in a coefficient field which only includes higher frequencies in the expansion (1.2). We thus call this field "rough" when compared to the "smooth" first field. A sample realization is depicted in Figure 4. With this, the experiment from Section 5.2 are repeated in the following.

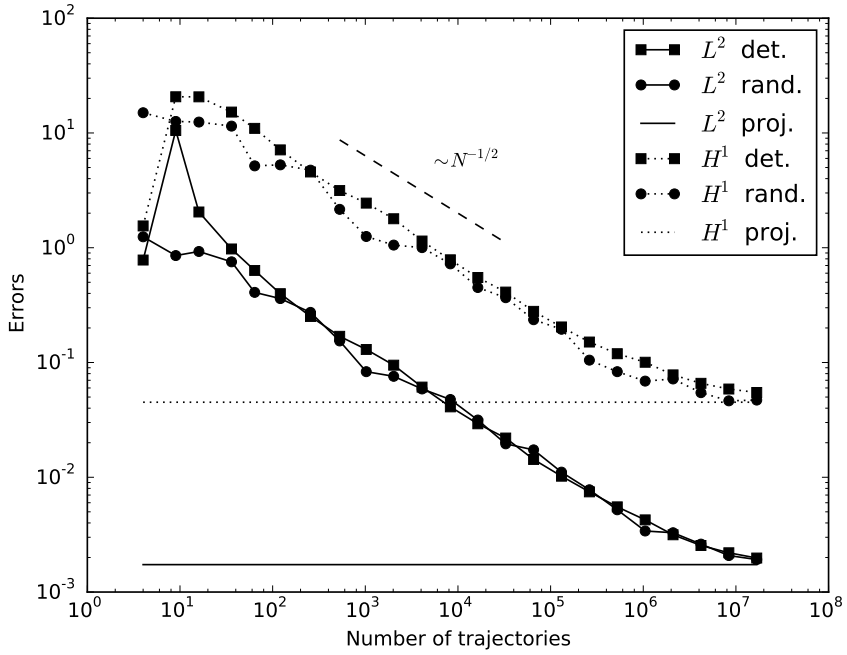

FIGURE 3. The L^2 and H^1 errors from the experiments with deterministic global regression (square) and semi stochastic global regression (circle) on a smooth benchmark field as well as the L^2 and H^1 error of the projection into the polynomial regression space (i.e., the bias, no marker).

Deterministic global regression. As in the former experiment, we launch one trajectory from each vertex of a uniform grid of the domain D. The errors in the $L^2(D)$ and $H^1(D)$ norms are again depicted with square markers in Figure 5. Although the stochastic field now exhibits high oscillations, we can still observe the anticipated convergence rate of $1/2$ in both error norms.

Semi stochastic global regression. To conclude this second experiment, as before, we sample on uniformly distributed points in D and compute one trajectory from each point. Then again, the semi stochastic global regression is applied and we observe a convergence rate of $1/2$ (circular markers in Figure 5) and a slight performance improvement compared to the deterministic global regression.

5.4. **Comparison.** Comparing the two experiments in Sections 5.2 and 5.3, we can see that the projection error in the $H^1(D)$ norm is reached at around 10^7 and the $L^2(D)$ error around $2 \cdot 10^{-3}$ sampled points in both cases. However, the $L^2(D)$ error of the projection is approximately one order of magnitude lower in the case of the rough field. A possible explanation for this last observation could be that realizations of the rough field exhibit much smaller global $L^2(D)$ norms than the smooth fields. This directly leads to smaller absolute values of the solutions (high frequencies of the coefficient are smoothed by the differential operator) and hence smaller L^2 errors.

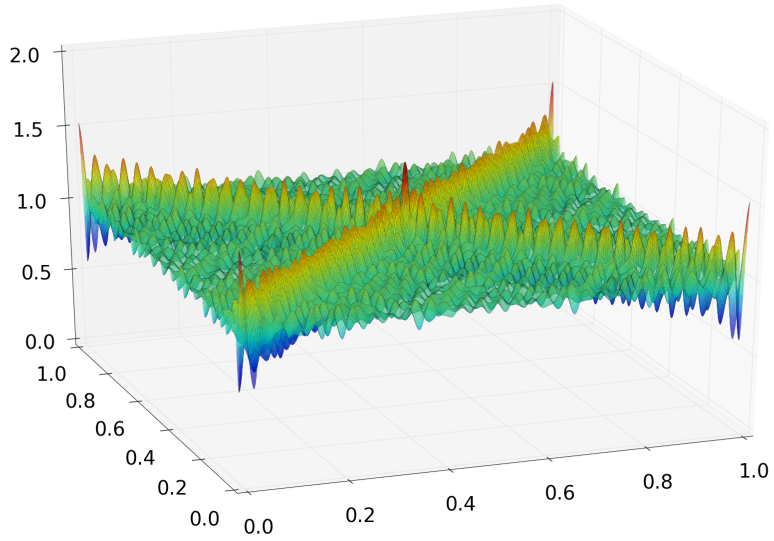

FIGURE 4. Example realization of the rough benchmark field

In summary, the experiments of this section illustrate that the described algorithm exhibits the predicted convergence behaviour for stochastic fields of different smoothness.

Certainly, further experiments and a more detailed analysis will have to be carried out to assess the possibilities and the limitations for the application of this numerical method.

REFERENCES

[1] The FEniCS Project. `http://fenicsproject.org`.

[2] Ivo Babuška, Fabio Nobile, and Raúl Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.*, 45(3):1005–1034, 2007.

[3] Ivo Babuška, Raúl Tempone, and Georgios E. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numer. Anal.*, 42(2):800–825, 2004.

[4] Ivo Babuška, Raúl Tempone, and Georgios E. Zouraris. Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation. *Comput. Methods Appl. Mech. Engrg.*, 194(12-16):1251–1294, 2005.

[5] Andrea Barth, Annika Lang, and Christoph Schwab. Multilevel Monte Carlo method for parabolic stochastic partial differential equations. *BIT*, 53(1):3–27, 2013.

[6] Christian Bayer, Anders Szepessy, and Raúl Tempone. Adaptive weak approximation of reflected and stopped diffusions. *Monte Carlo Methods Appl.*, 16(1):1–67, 2010.

[7] Denis Belomestny, Anastasia Kolodko, and John Schoenmakers. Regression methods for stochastic control problems and their convergence analysis. *SIAM J. Control Optim.*, 48(5):3562–3588, 2009/10.

[8] J. Charrier, R. Scheichl, and A. L. Teckentrup. Finite element error analysis of elliptic PDEs with random coefficients and its application to multilevel Monte Carlo methods. *SIAM J. Numer. Anal.*, 51(1):322–352, 2013.

[9] George Christakos. *Random field models in earth sciences*. Courier Corporation, 2012.

[10] Emmanuelle Clément, Damien Lamberton, and Philip Protter. An analysis of a least squares regression method for American option pricing. *Finance Stoch.*, 6(4):449–471, 2002.

[11] K. A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Comput. Vis. Sci.*, 14(1):3–15, 2011.

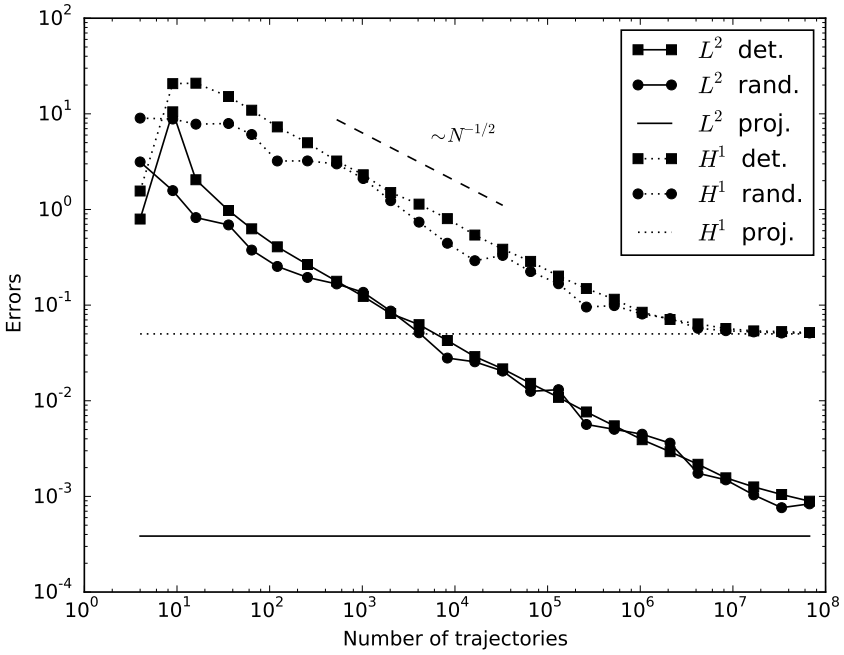

FIGURE 5. The L^2 and H^1 errors from the experiments with deterministic global regression (square) and semi stochastic global regression (circle) on a rough benchmark field as well as the L^2 and H^1 error of the projection into the polynomial regression space (no marker).

[12] Anna Dzougoutov, Kyoung-Sook Moon, Erik von Schwerin, Anders Szepessy, and Raúl Tempone. Adaptive Monte Carlo algorithms for stopped diffusion. In *Multiscale methods in science and engineering*, volume 44 of *Lect. Notes Comput. Sci. Eng.*, pages 59–88. Springer, Berlin, 2005.

[13] Martin Eigel, Claude Jeffrey Gittelson, Christoph Schwab, and Elmar Zander. Adaptive stochastic Galerkin FEM. *Comput. Methods Appl. Mech. Engrg.*, 270:247–269, 2014.

[14] Lawrence C. Evans. *Partial differential equations*. Graduate studies in mathematics. American Mathematical Society, 1998.

[15] Philipp Frauenfelder, Christoph Schwab, and Radu Alexandru Todor. Finite elements for elliptic problems with stochastic coefficients. *Comput. Methods Appl. Mech. Engrg.*, 194(2-5):205–228, 2005.

[16] Roger G Ghanem and Pol D Spanos. *Stochastic finite elements: a spectral approach*. Courier Corporation, 2003.

[17] David Gilbarg and Neil S Trudinger. *Elliptic partial differential equations of second order*. springer, 2015.

[18] Emmanuel Gobet and Stéphane Menozzi. Stopped diffusion processes: Boundary corrections and overshoot. *Stochastic Processes and their Applications*, 120(2):130 – 162, 2010.

[19] I. G. Graham, F. Y. Kuo, D. Nuyens, R. Scheichl, and I. H. Sloan. Quasi-Monte Carlo methods for elliptic PDEs with random coefficients and applications. *J. Comput. Phys.*, 230(10):3668–3694, 2011.

[20] Peter E. Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[21] Frances Y. Kuo, Christoph Schwab, and Ian H. Sloan. Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients. *SIAM J. Numer. Anal.*, 50(6):3351–3374, 2012.

[22] Olivier P Le Maître and Omar M Knio. *Introduction: Uncertainty Quantification and Propagation*. Springer, 2010.

[23] Hannelore Liero. Strong uniform consistency of nonparametric regression function estimates. *Probab. Theory Related Fields*, 82(4):587–614, 1989.

[24] Anders Logg, Kent-Andre Mardal, and Garth Wells. *Automated solution of differential equations by the finite element method: The FEniCS book*, volume 84. Springer Science & Business Media, 2012.

[25] F.A. Longstaff and E.S. Schwartz. Valuing American options by simulation: a simple least-squares approach. *Review of Financial Studies*, 14(1):113–147, 2001.

[26] Gabriel J Lord, Catherine E Powell, and Tony Shardlow. *An Introduction to Computational Stochastic PDEs*. Number 50. Cambridge University Press, 2014.

[27] Hermann G. Matthies and Andreas Keese. Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Comput. Methods Appl. Mech. Engrg.*, 194(12-16):1295–1331, 2005.

[28] Giovanni Migliorati and Fabio Nobile. Analysis of discrete least squares on multivariate polynomial spaces with evaluations at low-discrepancy point sets. *J. Complexity*, 31(4):517–542, 2015.

[29] G. N. Milstein and M. V. Tretyakov. *Stochastic numerics for mathematical physics*. Scientific Computation. Springer-Verlag, Berlin, 2004.

[30] G.N. Milstein and M.V. Tretyakov. Layer methods for stochastic navierstokes equations using simplest characteristics. *Journal of Computational and Applied Mathematics*, 302:1 – 23, 2016.

[31] Fabio Nobile, Raul Tempone, and Clayton G. Webster. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2411–2442, 2008.

[32] Fabio Nobile, Raul Tempone, and Clayton G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2309–2345, 2008.

[33] M.S. Pinsker. Optimal filtering of square-integrable signals in Gaussian noise. *Probl. Inf. Transm.*, 16:120–133, 1980.

[34] Christoph Schwab and Claude Jeffrey Gittelson. Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numer.*, 20:291–467, 2011.

[35] Ralph C Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*, volume 12. SIAM, 2013.

[36] Anders Szepessy, Raúl Tempone, and Georgios E. Zouraris. Adaptive weak approximation of stochastic differential equations. *Comm. Pure Appl. Math.*, 54(10):1169–1214, 2001.

[37] Denis Talay and Luciano Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4):483–509 (1991), 1990.

[38] Radu Alexandru Todor and Christoph Schwab. Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients. *IMA J. Numer. Anal.*, 27(2):232–261, 2007.

[39] Nizar Touzi. Second order backward SDEs, fully nonlinear PDEs, and applications in finance. In *Proceedings of the International Congress of Mathematicians*, volume 4, pages 3132–3150, 2010.

[40] J. Tsitsiklis and B. Van Roy. Regression methods for pricing complex American style options. *IEEE Trans. Neural. Net.*, 12(14):694–703, 2001.

[41] Tao Zhou, Akil Narayan, and Zhiqiang Xu. Multivariate discrete least-squares approximations with a new type of collocation grid. *SIAM J. Sci. Comput.*, 36(5):A2401–A2422, 2014.