Weak convergence of Galerkin approximations for fractional elliptic stochastic PDEs with spatial white noise

BIT Numerical Mathematics (2018) 58:881–906
https://doi.org/10.1007/s10543-018-0719-8

BIT

Weak convergence of Galerkin approximations for fractional elliptic stochastic PDEs with spatial white noise

David Bolin[1] · Kristin Kirchner[1] · Mihály Kovács[1]

Received: 12 January 2018 / Accepted: 1 August 2018 / Published online: 6 August 2018

Abstract

The numerical approximation of the solution to a stochastic partial differential equation with additive spatial white noise on a bounded domain is considered. The differential operator is assumed to be a fractional power of an integer order elliptic differential operator. The solution is approximated by means of a finite element discretization in space and a quadrature approximation of an integral representation of the fractional inverse from the Dunford–Taylor calculus. For the resulting approximation, a concise analysis of the weak error is performed. Specifically, for the class of twice continuously Fréchet differentiable functionals with second derivatives of polynomial growth, an explicit rate of weak convergence is derived, and it is shown that the component of the convergence rate stemming from the stochasticity is doubled compared to the corresponding strong rate. Numerical experiments for different functionals validate the theoretical results.

Keywords Stochastic partial differential equations · Weak convergence · Gaussian white noise · Fractional operators · Finite element methods · Galerkin methods · Matérn covariances · Spatial statistics

Mathematics Subject Classification 35S15 · 65C30 · 65C60 · 65N12 · 65N30

Communicated by Ragnar Winther.

This work was supported in part by the Swedish Research Council (Grant Nos. 2016-04187, 2017-04274), and the Knut and Alice Wallenberg Foundation (KAW 20012.0067).

✉ Kristin Kirchner
kristin.kirchner@chalmers.se

David Bolin
david.bolin@chalmers.se

Mihály Kovács
mihaly@chalmers.se

[1] Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, 412 96 Göteborg, Sweden

1 Introduction

The representation of Gaussian random fields as solutions to stochastic partial differential equations (SPDEs) has become a popular approach in spatial statistics in recent years. It was observed already in [21] and [22] that a Gaussian random field u on $\mathbb{R}^d$ with a covariance function of Matérn type [13] solves an SPDE of the form $(\kappa^2 - \Delta)^\beta u = \mathscr{W}$. Here, $\mathscr{W}$ is Gaussian white noise, $\kappa > 0$ is a parameter determining the practical correlation range of the field, and $\beta > d/4$ controls the smoothness parameter ν of the Gaussian Matérn field via the equality $\nu = 2\beta - d/2$.

Later, this relation was the incentive to consider the SPDE

$$(\kappa^2 - \Delta)^\beta u = \mathscr{W} \quad \text{in } \mathscr{D} \tag{1.1}$$

for Gaussian random field approximations of Matérn fields on bounded domains $\mathscr{D} \subsetneq \mathbb{R}^d$. On the boundary $\partial\mathscr{D}$, the operator $\kappa^2 - \Delta$ is augmented with, e.g., homogeneous Dirichlet or Neumann boundary conditions. In [12] it was shown that by restricting the value of β to $2\beta \in \mathbb{N}$ and by solving the stochastic problem (1.1) by means of a finite element method, the computational costs of many operations, which are needed for statistical inference, such as sampling and likelihood evaluations can be significantly reduced. This decrease in computing time is one of the main reasons for the popularity of the SPDE approach in spatial statistics. In addition, it facilitates various extensions of the Matérn model which are difficult to formulate using a covariance-based approach, see, for instance [2,5,10,12,20].

However, the constraint $2\beta \in \mathbb{N}$ imposed by [12] restricts the value of the smoothness parameter ν, which is the most important parameter when the model is used for prediction [17]. In [4] we showed that this restriction can be avoided by combining a finite element discretization in space with a quadrature approximation based on an integral representation of the inverse fractional power operator from the Dunford–Taylor calculus. We furthermore derived an explicit rate of convergence for the strong mean-square error of the proposed approximation for a class of fractional elliptic stochastic equations including (1.1).

In practice, it is often not only necessary to sample from the solution u to (1.1), but also to estimate the expected value $\mathbb{E}[\varphi(u)]$ of a certain real-valued quantity of interest $\varphi(u)$. The aim of this work is to provide a concise analysis of the weak error $|\mathbb{E}[\varphi(u)] - \mathbb{E}[\varphi(u_{h,k}^Q)]|$ for the approximation $u_{h,k}^Q$ proposed in [4]. This analysis includes the derivation of an explicit weak convergence rate for twice continuously Fréchet differentiable real-valued functions φ, whose second derivatives are of polynomial growth. Functions of this form occur in many applications, e.g., when integral means of the solution with respect to a certain subdomain of $\mathscr{D}$ are of interest, or when a transformation of the model is used as a component in a hierarchical model. An example of the latter situation is to consider logit or probit transformed Gaussian random fields for binary regression models, see, e.g., [16, §4.3.3].

We prove that, compared to the convergence rate of the strong error formulated in [4], the component of the weak convergence rate stemming from the stochasticity of the problem is doubled. To this end, two time-dependent stochastic processes are introduced, which at time $t = 1$ have the same probability distribution as the exact

solution u and the approximation $u_{h,k}^{Q}$, respectively. The weak error is then bounded by introducing an associated Kolmogorov backward equation on the interval $[0, 1]$ and applying Itô calculus.

The structure of this article is as follows: in Sect. 2 we formulate the equation of interest in a Hilbert space setting similarly to [4] and state our main result on weak convergence of the approximation in Theorem 2.1. A detailed proof of Theorem 2.1 is given in Sect. 3. For validating the theoretical result in practice, we describe the outcomes of several numerical experiments in Sect. 4. Finally, Sect. 5 concludes the article with a discussion.

2 Weak approximations

The subject of our investigations is the fractional order equation considered in [4],

$$L^{\beta} u = g + \mathscr{W}, \tag{2.1}$$

for $\beta \in (0, 1)$, where $\mathscr{W}$ denotes Gaussian white noise defined on a complete probability space $(\Omega, \mathscr{A}, \mathbb{P})$ with values in a separable Hilbert space H. Here and below, (in-)equalities involving random terms are meant to hold $\mathbb{P}$-almost surely, if not specified otherwise. Furthermore, we use the notation $X \overset{d}{=} Y$ to indicate that two random variables X and Y have the same probability distribution.

Similarly to [4], we make the following assumptions: $L \colon \mathscr{D}(L) \subset H \to H$ is a densely defined, self-adjoint, positive definite operator and has a compact inverse $L^{-1} \colon H \to H$. In this case, $-L$ generates an analytic strongly continuous semigroup $(S(t))_{t \geq 0}$ on H. The H-orthonormal eigenvectors of L are denoted by $\{e_j\}_{j \in \mathbb{N}}$ and the corresponding eigenvalues by $\{\lambda_j\}_{j \in \mathbb{N}}$. These values are listed in nondecreasing order and we assume that there exist constants $\alpha, c_\lambda, C_\lambda > 0$ such that

$$c_\lambda \, j^{\alpha} \leq \lambda_j \leq C_\lambda \, j^{\alpha} \qquad \forall j \in \mathbb{N}. \tag{2.2}$$

The action of the fractional power operator L^{β} in (2.1) is well-defined on

$$\dot{H}^{2\beta} := \mathscr{D}(L^{\beta}) = \left\{ \psi \in H : \|\psi\|_{2\beta}^2 := \|L^{\beta}\psi\|_H^2 = \sum_{j \in \mathbb{N}} \lambda_j^{2\beta} (\psi, e_j)_H^2 < \infty \right\},$$

which is itself a Hilbert space with inner product $(\phi, \psi)_{2\beta} := (L^{\beta}\phi, L^{\beta}\psi)_H$. Furthermore, there exists a unique continuous extension of L^{β} to an isometric isomorphism $L^{\beta} \colon \dot{H}^{r} \to \dot{H}^{r-2\beta}$ for all $r \in \mathbb{R}$, see [4, Lem. 2.1]. Here, for $s > 0$, the negative-indexed space $\dot{H}^{-s}$ is defined as the dual space of $\dot{H}^{s}$. After identifying the dual space H^* of $\dot{H}^0 := H$ via the Riesz map, we obtain the Gelfand triple $\dot{H}^{s} \hookrightarrow H \cong H^* \hookrightarrow \dot{H}^{-s}$ with continuous and dense embeddings. The norm on the dual space $\dot{H}^{-s}$ can be expressed by

$$\|g\|_{-s} = \sup_{\phi \in \dot{H}^s \setminus \{0\}} \frac{\langle g, \phi \rangle}{\|\phi\|_s} = \left(\sum_{j \in \mathbb{N}} \lambda_j^{-s} \langle g, e_j \rangle^2 \right)^{\frac{1}{2}},$$

where $\langle \cdot, \cdot \rangle$ denotes the duality pairing between $\dot{H}^{-s}$ and $\dot{H}^s$, [19, Proof of Lem. 5.1]. With this representation of the dual norm and the growth (2.2) of the eigenvalues λ_j at hand, it is an immediate consequence of a Karhunen–Loève expansion of the white noise $\mathscr{W}$ with respect to the H-orthonormal eigenvectors $\{e_j\}_{j \in \mathbb{N}}$ that $\mathscr{W}$ has mean-square regularity in $\dot{H}^{-s}$ for every $s > \alpha^{-1}$, see [4, Prop. 2.3]. Consequently, (2.1) has a solution $u \in L_2(\Omega; \dot{H}^{2\beta - s})$ for $s > \alpha^{-1}$ if $g \in \dot{H}^{-s}$.

2.1 The Galerkin approximation

In the following, let $(V_h)_{h \in (0,1)}$ be a family of subspaces of $\dot{H}^1 = \mathscr{D}(L^{1/2})$ with finite dimensions $N_h := \dim(V_h)$ and let $\Pi_h \colon H \to V_h$ be the H-orthogonal projection onto V_h. For $g \in H$, we define the finite element approximation of $v = L^{-1} g$ by $v_h = L_h^{-1} \Pi_h g$, where L_h denotes the Galerkin discretization of the operator L with respect to V_h, i.e.,

$$L_h \colon V_h \to V_h, \qquad (L_h \psi_h, \phi_h)_H = \langle L \psi_h, \phi_h \rangle \quad \forall \psi_h, \phi_h \in V_h.$$

We then consider the following numerical approximation of the solution u to (2.1)

$$u_{h,k}^Q := Q_{h,k}^\beta (\Pi_h g + \mathscr{W}_h^\Phi) \tag{2.3}$$

proposed in [4, Eq. (2.18)]. It is based on the following two components:

(a) The operator $Q_{h,k}^\beta$ is the quadrature approximation for $L_h^{-\beta}$ of [6]:

$$Q_{h,k}^\beta := \frac{2k \sin(\pi\beta)}{\pi} \sum_{\ell=-K^-}^{K^+} e^{2\beta y_\ell} \left(\mathrm{Id}_{V_h} + e^{2y_\ell} L_h \right)^{-1}. \tag{2.4}$$

The quadrature nodes $\{y_\ell = \ell k : \ell \in \mathbb{Z}, -K^- \le \ell \le K^+\}$ are equidistant with distance $k > 0$ and we set $K^- := \lceil \frac{\pi^2}{4\beta k^2} \rceil$ and $K^+ := \lceil \frac{\pi^2}{4(1-\beta)k^2} \rceil$.

(b) The white noise $\mathscr{W}$ in H is approximated by the square-integrable V_h-valued random variable $\mathscr{W}_h^\Phi$ given by $\mathscr{W}_h^\Phi := \sum_{j=1}^{N_h} \xi_j \phi_{j,h}$, where $\Phi := \{\phi_{j,h}\}_{j=1}^{N_h}$ is any basis of the finite element space V_h. The vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_{N_h})^T$ is multivariate Gaussian distributed with mean zero and covariance matrix $\mathbf{M}^{-1}$, where $\mathbf{M}$ denotes the mass matrix with respect to the basis Φ, i.e., $M_{ij} = (\phi_{i,h}, \phi_{j,h})_H$.

The main outcome of [4] is strong convergence of the approximation $u_{h,k}^Q$ in (2.3) to the solution u of (2.1) at an explicit rate. Subsequently, this work focusses on weak

approximations based on $u_{h,k}^Q$, i.e., we investigate the error

$$\left|\mathbb{E}[\varphi(u)] - \mathbb{E}[\varphi(u_{h,k}^Q)]\right| \tag{2.5}$$

for continuous functions $\varphi\colon H \to \mathbb{R}$.

Remark 2.1 In practice, the expected value $\mathbb{E}[\varphi(u_{h,k}^Q)]$ is approximated, e.g., by a Monte Carlo method. For this, usually a large number of realizations of $\varphi(u_{h,k}^Q)$ and, thus, of the approximation $u_{h,k}^Q$ in (2.3) is needed. Each of them requires a sample of the load vector $\mathbf{b}$ with entries $b_j := (\Pi_h g + \mathscr{W}_h^\Phi, \phi_{j,h})_H$. As pointed out in [4, Rem. 2.9], this is computationally feasible if the mass matrix $\mathbf{M}$ with respect to the finite element basis Φ is sparse, since the distribution of $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{M}^{-1})$ implies that

$$\mathbf{b} \sim \mathcal{N}(\mathbf{g}, \mathbf{M}), \qquad \mathbf{b} \overset{d}{=} \mathbf{g} + \mathbf{G}\mathbf{z},$$

where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\mathbf{G}$ is the Cholesky factor of $\mathbf{M} = \mathbf{G}\mathbf{G}^T$, and the vector $\mathbf{g}$ has entries $g_j := (g, \phi_{j,h})_H$.

2.2 Weak convergence

For bounding the error in (2.5), we start by introducing some more notation and assumptions. Let $\mathscr{E} := \{e_{j,h}\}_{j=1}^{N_h} \subset V_h$ be the H-orthonormal eigenvectors of the discrete operator L_h with corresponding eigenvalues $\{\lambda_{j,h}\}_{j=1}^{N_h}$ listed in nondecreasing order. In addition, the strongly continuous semigroup on V_h generated by $-L_h$ is denoted by $(S_h(t))_{t\geq 0}$.

We define the space $C^2(H;\mathbb{R})$ of twice continuously Fréchet differentiable functions $\varphi\colon H \to \mathbb{R}$, i.e., $\varphi \in C^2(H;\mathbb{R})$ if and only if

$$\varphi \in C(H;\mathbb{R}), \quad D\varphi \in C(H;H), \quad \text{and} \quad D^2\varphi \in C(H;\mathscr{L}(H)).$$

Here and below, using the Riesz representation theorem, we identify the first two Fréchet derivatives $D\varphi$ and $D^2\varphi$ of φ with functions taking values in H and in $\mathscr{L}(H)$, respectively. Furthermore, we say that the second derivative has polynomial growth of degree $p \in \mathbb{N}$, if there exists a constant $K > 0$ such that

$$\|D^2\varphi(\psi)\|_{\mathscr{L}(H)} \leq K\left(1 + \|\psi\|_H^p\right) \quad \forall \psi \in H. \tag{2.6}$$

All the properties of the finite element discretization, of the operator L, and of the function φ, which are of importance for our analysis of the weak error (2.5), are summarized in the assumption below.

Assumption 2.1 The finite element spaces $(V_h)_{h\in(0,1)} \subset \dot{H}^1$, the operator L in (2.1), and the function $\varphi\colon H \to \mathbb{R}$ in (2.5) satisfy the following:

(i) there exists $d \in \mathbb{N}$ such that $N_h = \dim(V_h) \propto h^{-d}$ for all $h > 0$;

(ii) there exist constants $C_1, C_2 > 0$, $h_0 \in (0, 1)$, as well as exponents $r, s > 0$ and $q > 1$ such that

$$\lambda_j \leq \lambda_{j,h} \leq \lambda_j + C_1 h^r \lambda_j^q,$$
$$\|e_j - e_{j,h}\|_H^2 \leq C_2 h^{2s} \lambda_j^q,$$

for all $h \in (0, h_0)$ and $j \in \{1, \ldots, N_h\}$;

(iii) the eigenvalues of L satisfy (2.2) for an exponent α with

$$\tfrac{1}{2\beta} < \alpha \leq \min\left\{ \tfrac{r}{(q-1)d}, \tfrac{2s}{qd} \right\},$$

where the values of $d \in \mathbb{N}$, $r, s > 0$, and $q > 1$ are the same as in (i)–(ii);

(iv) $s > 2\beta$ and for $0 \leq \theta \leq \sigma \leq s$ there exists a constant $C_3 > 0$ such that

$$\|(S(t) - S_h(t)\Pi_h)g\|_H \leq C_3 h^\sigma t^{\frac{\theta-\sigma}{2}} \|g\|_\theta \quad \forall t > 0,$$

for every $g \in \dot{H}^\theta$ and $h \in (0, h_0)$. Here, h_0 and s are as in (ii);

(v) $\varphi \in C^2(H; \mathbb{R})$ and $D^2\varphi$ has polynomial growth (2.6) of degree $p \geq 2$.

The following example shows that Assumptions 2.1(i)–(iv) are satisfied, e.g., for the motivating problem (1.1) related to approximations of Matérn fields, if $\beta > d/4$, when using continuous piecewise linear finite element bases.

Example 2.1 For $\kappa \geq 0$ and a bounded, convex, polygonal domain $\mathscr{D} \subset \mathbb{R}^d$, consider the stochastic model problem (1.1), i.e., the fractional order equation (2.1) for $g = 0$ and $L = \kappa^2 - \Delta$ on $H = L_2(\mathscr{D})$. Furthermore, we assume that the differential operator L is augmented with homogeneous Dirichlet boundary conditions on $\partial\mathscr{D}$. In this case, the eigenvalues $\{\lambda_j\}_{j\in\mathbb{N}}$ of L satisfy (2.2) for $\alpha = 2/d$ (see [8, Ch. VI.4] for $\mathscr{D} = (0, 1)^d$, the result for more general domains as above follows from the min–max principle). Consequently, the first inequality of Assumption 2.1(iii) holds if $\beta > d/4$.

In addition, if $(V_h)_{h\in(0,1)} \subset \dot{H}^1 = H_0^1(\mathscr{D})$ are finite element spaces with continuous piecewise linear basis functions defined with respect to a quasi-uniform family of triangulations, Assumption 2.1(i) holds and Assumptions 2.1(ii), (iv) are satisfied for $r = s = q = 2$, see [18, Thm. 6.1, Thm. 6.2] and [19, Thm. 3.5]. Thus,

$$s = 2 > 2\beta, \qquad \alpha = \tfrac{2}{d} = \min\left\{ \tfrac{r}{(q-1)d}, \tfrac{2s}{qd} \right\},$$

and Assumptions 2.1(i)–(iv) hold for all $\beta \in (d/4, 1)$.

We remark that Assumptions 2.1(i)–(iii) coincide with those of [4]. The strong $L_2(\Omega; H)$-convergence rate

$$\min\{d(\alpha\beta - 1/2), r, s\} \tag{2.7}$$

was derived in [4, Thm. 2.10] for the approximation $u_{h,k}^Q$ in (2.3) under a suitable calibration of the distance of the quadrature nodes k with the finite element mesh size h. Furthermore, a bound for the weak-type error

$$\left| \|u\|_{L_2(\Omega;H)}^2 - \|u_{h,k}^Q\|_{L_2(\Omega;H)}^2 \right|$$

was provided, showing convergence to zero with the rate $\min\{d(2\alpha\beta - 1), r, s\}$, see [4, Cor. 3.4]. In particular, the term $d(2\alpha\beta - 1)$ stemming from the stochasticity is doubled compared to the strong rate in (2.7).

In the following, we generalize this result to weak errors of the form (2.5) for functions $\varphi\colon H \to \mathbb{R}$, which are twice continuously Fréchet differentiable and have a second derivative of polynomial growth. The bound of the weak error in Theorem 2.1 is our main result.

Theorem 2.1 *Let Assumption 2.1 be satisfied. Let $\theta > \min\{d(2\alpha\beta - 1), s\} - 2\beta$, if $d(2\alpha\beta - 1) \geq 2\beta$, and set $\theta = 0$ otherwise. Then, for $g \in \dot{H}^\theta$ and for sufficiently small $h \in (0, h_0)$ and $k \in (0, k_0)$, the weak error in (2.5) admits the bound*

$$\begin{aligned} \left|\mathbb{E}[\varphi(u)] - \mathbb{E}[\varphi(u_{h,k}^Q)]\right| \leq C &\left(h^{\min\{d(2\alpha\beta-1),r,s\}} + e^{-\frac{\pi^2}{k}} h^{-d} + e^{-\frac{\pi^2}{2k}} + e^{-\frac{\pi^2}{2k}} f_{\alpha,\beta}(h) \right) \\ &\times \left(1 + e^{-\frac{p\pi^2}{2k}} h^{-\frac{pd}{2}} + \|g\|_H^{p+1} \right). \end{aligned} \tag{2.8}$$

Here, we set $f_{\alpha,\beta}(h) := h^{d(\alpha\beta-1)}$, if $\alpha\beta \neq 1$, and $f_{\alpha,\beta}(h) := |\ln(h)|$, if $\alpha\beta = 1$. The constant $C > 0$ is independent of h and k and the values of $\alpha, r, s > 0$, $d \in \mathbb{N}$, and $p \in \{2, 3, \ldots\}$ are those of Assumption 2.1.

Remark 2.2 In the derivation of the strong convergence rate (2.7), we balanced the error terms caused by the quadrature and by the finite element method by choosing the quadrature step size k sufficiently small with respect to the finite element mesh width h, namely $e^{-\pi^2/(2k)} \propto h^{d\alpha\beta}$, see [4, Table 1].

For calibrating the terms in the weak error estimate (2.8), we distinguish the cases $\alpha\beta < 1$, $\alpha\beta = 1$, and $\alpha\beta > 1$. If $\alpha\beta < 1$, then $d\alpha\beta > d(2\alpha\beta - 1)$ and we let $k > 0$ be such that $e^{-\pi^2/(2k)} \propto h^{d\alpha\beta}$. With this choice, the error estimate (2.8) simplifies to

$$\left|\mathbb{E}[\varphi(u)] - \mathbb{E}[\varphi(u_{h,k}^Q)]\right| \leq C h^{\min\{d(2\alpha\beta-1),r,s\}} \left(1 + \|g\|_H^{p+1}\right) (1 + \|g\|_\theta).$$

For $\alpha\beta > 1$ ($\alpha\beta = 1$) we achieve the same bound if k and h are calibrated such that $e^{-\pi^2/(2k)} \propto h^{d(2\alpha\beta-1)}$ ($e^{-\pi^2/(2k)} \max\{1, |\ln(h)|\} \propto h^d$). Note that the calibration for $\alpha\beta < 1$ coincides with the one for the strong error and that the term $d(2\alpha\beta - 1)$ in the derived weak convergence rate $\min\{d(2\alpha\beta - 1), r, s\}$ is doubled compared to the first term of the strong convergence rate (2.7).

Remark 2.3 We emphasize that (under the same assumptions) both the strong and weak convergence rates remain the same when approximating the solution u to

$$L^\beta u = \sigma(g + \mathscr{W})$$

by $u_{h,k}^{Q} := \sigma\, Q_{h,k}^{\beta}(\Pi_h g + \mathscr{W}_h^{\Phi})$, where $\sigma > 0$ is a constant parameter which scales the variance of u. This can be seen from the equality $\sigma^{-1} L^{\beta} = L_{\sigma}^{\beta}$ for $L_{\sigma} := \sigma^{-1/\beta} L$, combined with the fact that the eigenvalues of the operator L_{σ} satisfy the growth assumption (2.2) with the same exponent $\alpha > 0$ as the eigenvalues of L.

However, the constants $c_{\lambda}, C_{\lambda} > 0$ in (2.2) and the constants in the error estimates change. For instance, if $\varphi(u) := \|u\|_H^{p_*}$ for $p_* \in \mathbb{N}$, then the constant $C > 0$ in (2.8) will depend linearly on σ^{p_*}.

Note that one has to consider a problem of the form

$$(\kappa^2 - \Delta)^{\beta} u = \sigma \mathscr{W} \quad \text{for} \quad \sigma := \sigma_* (4\pi)^{\frac{d}{4}} \kappa^{2\beta - \frac{d}{2}} \sqrt{\frac{\Gamma(2\beta)}{\Gamma(2\beta - d/2)}}$$

when approximating a Matérn field with variance σ_*^2. Here and in what follows, $\Gamma(\cdot)$ denotes the Gamma function.

Remark 2.4 We also comment on how the error bound in (2.8) changes if instead of the family $(Q_{h,k}^{\beta})_{k>0}$ a different sequence of approximations $\{R_{h,n}^{\beta}\}_{n\in\mathbb{N}}$ of $L_h^{-\beta}$ is used. If there exists a function $E\colon \mathbb{N} \to \mathbb{R}_{\geq 0}$ such that $\lim_{n\to\infty} E(n) = 0$ as well as a constant $C > 0$, independent of h and n, such that

$$\|(L_h^{-\beta} - R_{h,n}^{\beta})\phi_h\|_H \leq C E(n) \|\phi_h\|_H \quad \forall \phi_h \in V_h,$$

it is an immediate consequence of the arguments in our proof that a bound of the weak error for the approximation $u_{h,n}^{R} := R_{h,n}^{\beta}(\Pi_h g + \mathscr{W}_h^{\Phi})$ is given by

$$\begin{aligned} \left|\mathbb{E}[\varphi(u)] - \mathbb{E}[\varphi(u_{h,k}^{Q})]\right| \leq C \Big(& h^{\min\{d(2\alpha\beta-1), r, s\}} + E(n)^2 h^{-d} + E(n) + E(n) f_{\alpha,\beta}(h) \Big) \\ & \times \Big(1 + E(n)^p h^{-\frac{pd}{2}} + \|g\|_H^{p+1}\Big)\,(1 + \|g\|_{\theta}). \end{aligned}$$

An example of such a family $\{R_{h,n}^{\beta}\}_{n\in\mathbb{N}}$ are the approximations of $L_h^{-\beta}$ proposed in [3], which are based on rational approximations of the function $x^{-\beta}$ of different degrees $n \in \mathbb{N}$.

3 The derivation of Theorem 2.1

The main idea in our derivation of the weak error estimate (2.8) is to introduce two time-dependent stochastic processes with the property that their (random) values at time $t = 1$ have the same distribution as the solution u to (2.1) and its approximation $u_{h,k}^{Q}$ in (2.3), respectively. We then use an associated Kolmogorov backward equation and Itô calculus to estimate the difference between these values.

3.1 The extension to time-dependent processes

Recall the eigenvalue-eigenvector pairs $\{(\lambda_j, e_j)\}_{j\in\mathbb{N}}$ of L as well as the parameter $\alpha > 0$ determining the growth of the eigenvalues via (2.2). In what follows, we assume

that $g \in H$ and $2\alpha\beta > 1$ so that the solution u to (2.1) satisfies $u \in L_2(\Omega; H)$. With the aim of introducing the time-dependent processes mentioned above, we start by defining

$$W^\beta(t) := \sum_{j\in\mathbb{N}} \lambda_j^{-\beta} B_j(t)\, e_j, \quad t \geq 0,$$

where $\{B_j\}_{j\in\mathbb{N}}$ is a sequence of independent real-valued Brownian motions adapted to a filtration $\mathscr{F} := (\mathscr{F}_t,\ t \geq 0)$. Owing to this construction, $(W^\beta(t),\ t \geq 0)$ is an $\mathscr{F}$-adapted H-valued Wiener process with covariance operator $L^{-2\beta}$, which is of trace-class if $2\alpha\beta > 1$. Since the random variables $\{B_j(1)\}_{j\in\mathbb{N}}$ are independent and identically $\mathscr{N}(0, 1)$-distributed, the spatial white noise $\mathscr{W}$ satisfies

$$\mathscr{W} \overset{d}{=} \sum_{j\in\mathbb{N}} B_j(1)\, e_j \quad \text{in } H.$$

The stochastic process $Y := (Y(t),\ t \in [0, 1])$ defined as the (strong) solution to the stochastic partial differential equation

$$\mathrm{d}Y(t) = \mathrm{d}W^\beta(t), \quad t \in [0, 1], \qquad Y(0) = L^{-\beta} g, \tag{3.1}$$

therefore takes the following random value in H at time $t = 1$,

$$Y(1) = Y(0) + \int_0^1 \mathrm{d}W^\beta(t) = L^{-\beta} g + W^\beta(1) \overset{d}{=} L^{-\beta}(g + \mathscr{W}) = u. \tag{3.2}$$

Its Gaussian distribution implies the existence of all moments, as shown in the following lemma.

Lemma 3.1 *Let $p \in \mathbb{N}$, $t \in [0, 1]$, and Y be the strong solution of (3.1). Then the p-th moment of $Y(t)$ exists and, for $p \geq 2$, it admits the following bound:*

$$\mathbb{E}\left[\|Y(t)\|_H^p\right] \leq 2^{p-1} \left(\|g\|_{-2\beta}^p + t^{\frac{p}{2}} \mu_p \operatorname{tr}(L^{-2\beta})^{\frac{p}{2}} \right). \tag{3.3}$$

Here, $\mu_p := \mathbb{E}[|Z|^p] = \sqrt{\frac{2^p}{\pi}}\, \Gamma\left(\frac{p+1}{2}\right)$ is the p-th absolute moment of $Z \sim \mathscr{N}(0, 1)$.

Proof For $p = 2$, the bound in (3.3) follows from the Itô isometry [15, Thm. 8.7(i)]:

$$\mathbb{E}\left[\|Y(t)\|_H^2\right] = \|L^{-\beta} g\|_H^2 + \int_0^t \operatorname{tr}(L^{-2\beta})\, \mathrm{d}s = \|g\|_{-2\beta}^2 + t\mu_2 \operatorname{tr}(L^{-2\beta}).$$

If $p \geq 3$, we estimate $\mathbb{E}[\|Y(t)\|_H^p] \leq 2^{p-1}(\|L^{-\beta}g\|_H^p + \mathbb{E}[\|W^\beta(t)\|_H^p])$. By Jensen's inequality we have

$$\mathbb{E}\Big|\sum_{j\in\mathbb{N}} \lambda_j^{-2\beta}|B_j(t)|^2\Big|^{\frac{p}{2}} \leq \mathbb{E}\Bigg[\Big|\sum_{j\in\mathbb{N}} \lambda_j^{-2\beta}\Big|^{\frac{p}{2}-1} \sum_{j\in\mathbb{N}} \lambda_j^{-2\beta}|B_j(t)|^p\Bigg].$$

Thus, the distribution of $\{B_j(t)\}_{j\in\mathbb{N}}$ implies that $\mathbb{E}[\|W^\beta(t)\|_H^p] \leq t^{p/2}\mu_p \operatorname{tr}(L^{-2\beta})^{p/2}$, and assertion (3.3) follows. □

In order to define a another stochastic process $\widetilde{Y} := (\widetilde{Y}(t),\ t \in [0,1])$ with the property $\widetilde{Y}(1) \stackrel{d}{=} u_{h,k}^Q$ in H, we recall the orthonormal eigenbasis $\mathscr{E} = \{e_{j,h}\}_{j=1}^{N_h} \subset V_h$ of L_h and define $P_h^\beta \colon H \to V_h$ for $\beta \in (0,1)$ by

$$P_h^\beta g := \sum_{j=1}^{N_h} \lambda_j^\beta (g, e_j)_H\, e_{j,h}. \tag{3.4}$$

Since V_h is finite-dimensional, the operator $Q_{h,k}^\beta \colon V_h \to V_h$ in (2.4) is bounded, $Q_{h,k}^\beta \in \mathscr{L}(V_h)$ for short, with norm

$$\|Q_{h,k}^\beta\|_{\mathscr{L}(V_h)} := \sup_{\psi_h \in V_h\setminus\{0\}} \frac{\|Q_{h,k}^\beta \psi_h\|_H}{\|\psi_h\|_H} < \infty.$$

We now consider the following stochastic partial differential equation

$$\mathrm{d}\widetilde{Y}(t) = Q_{h,k}^\beta P_h^\beta\, \mathrm{d}W^\beta(t), \quad t \in [0,1], \qquad \widetilde{Y}(0) = Q_{h,k}^\beta \Pi_h g. \tag{3.5}$$

Note that the reproducing kernel Hilbert space of W^β is $\dot{H}^{2\beta}$. The finite rank of the operator $Q_{h,k}^\beta P_h^\beta \colon H \to V_h$ implies that it is a Hilbert–Schmidt operator from $\dot{H}^{2\beta}$ to H. For this reason, existence and uniqueness of a (strong) solution $\widetilde{Y}$ to (3.5) is evident. Furthermore, the solution process $\widetilde{Y}$ satisfies

$$\widetilde{Y}(1) = \widetilde{Y}(0) + \int_0^1 Q_{h,k}^\beta P_h^\beta\, \mathrm{d}W^\beta(t) = Q_{h,k}^\beta(\Pi_h g + \mathscr{W}_h^{\mathscr{E}}),$$

where $\mathscr{W}_h^{\mathscr{E}} := \sum_{j=1}^{N_h} B_j(1)\, e_{j,h}$. To see that also $\widetilde{Y}(1) \stackrel{d}{=} u_{h,k}^Q$ holds in H, define the deterministic matrix $\mathbf{R}$ and the random vector $\mathbf{B}_1$ by

$$R_{ij} := (e_{i,h}, \phi_{j,h})_H, \quad 1 \leq i, j \leq N_h, \qquad \mathbf{B}_1 := (B_1(1), \ldots, B_{N_h}(1))^T,$$

i.e., $\mathbf{B}_1$ is the vector of the first N_h Brownian motions at time $t = 1$. Due to

$$(\mathbf{R}^T\mathbf{R})_{ij} = (\phi_{i,h}, \phi_{j,h})_H = M_{ij},$$

the vector $\boldsymbol{\xi} := \mathbf{R}^{-1}\mathbf{B}_1$ is $\mathcal{N}(\mathbf{0}, \mathbf{M}^{-1})$-distributed. In addition, by [4, Lem. 2.8] the V_h-valued random variables

$$\mathscr{W}_h^{\mathscr{E}} = \sum_{j=1}^{N_h} B_j(1)\, e_{j,h} \quad \text{and} \quad \mathscr{W}_h^{\Phi} := \sum_{j=1}^{N_h} \xi_j\, \phi_{j,h}$$

are equal in $L_2(\Omega; H)$. In particular, their first and second moments coincide. Since $\mathscr{W}_h^{\mathscr{E}}$ and $\mathscr{W}_h^{\Phi}$ are Gaussian random variables, their distributions are uniquely characterized by their first two moments and we conclude that

$$\widetilde{Y}(1) = Q_{h,k}^{\beta}(\Pi_h g + \mathscr{W}_h^{\mathscr{E}}) \overset{d}{=} Q_{h,k}^{\beta}(\Pi_h g + \mathscr{W}_h^{\Phi}) = u_{h,k}^{Q}. \tag{3.6}$$

3.2 The Kolmogorov backward equation and partition of the error

With the aim of bounding the weak error in (2.5) by means of Itô calculus, we introduce the following Kolmogorov backward equation associated with the stochastic partial differential equation (3.1) for Y and the function φ by

$$w_t(t,x) + \frac{1}{2}\operatorname{tr}\left(w_{xx}(t,x)L^{-2\beta}\right) = 0, \quad t \in [0,1],\ x \in H, \qquad w(1,x) = \varphi(x). \tag{3.7}$$

Here, $w_x := D_x w$ and $w_{xx} := D_x^2 w$ denote the first and second order Fréchet derivative of w with respect to $x \in H$. It is well-known [9, Rem. 3.2.1, Thm. 3.2.3] that the solution $w\colon [0,1] \times H \to \mathbb{R}$ to (3.7) is given in terms of the stochastic process Y in (3.1) by the following expectation

$$w(t,x) = \mathbb{E}[\varphi(x + Y(1) - Y(t))]. \tag{3.8}$$

Since $\varphi\colon H \to \mathbb{R}$ is twice continuously Fréchet differentiable, we can furthermore express the first two derivatives of w with respect to x in terms of φ and Y by

$$w_x(t,x) = \mathbb{E}[D\varphi(x + Y(1) - Y(t))], \tag{3.9}$$

$$w_{xx}(t,x) = \mathbb{E}[D^2\varphi(x + Y(1) - Y(t))]. \tag{3.10}$$

Let $\widetilde{Y}$ be the solution to (3.5). The application of Itô's lemma [7] to the stochastic process $(w(t, \widetilde{Y}(t)),\ t \in [0,1])$ yields

$$\begin{aligned} \mathrm{d}w(t,\widetilde{Y}(t)) = &\left(w_t(t,\widetilde{Y}(t)) + \frac{1}{2}\operatorname{tr}\left(w_{xx}(t,\widetilde{Y}(t))Q_{h,k}^{\beta}P_h^{\beta}L^{-2\beta}\left(Q_{h,k}^{\beta}P_h^{\beta}\right)^*\right)\right)\mathrm{d}t \\ &+ w_x(t,\widetilde{Y}(t))Q_{h,k}^{\beta}P_h^{\beta}\,\mathrm{d}W^{\beta}(t), \qquad t \in [0,1], \end{aligned} \tag{3.11}$$

where, for $T \in \mathscr{L}(H)$, the H-adjoint operator is denoted by T^*. To simplify the second term in (3.11), we define the operator $\widetilde{\Pi}_h \colon H \to V_h$ by

$$\widetilde{\Pi}_h g := \sum_{j=1}^{N_h} (g, e_j)_H \, e_{j,h}. \tag{3.12}$$

Note that in contrast to the H-orthogonal projection Π_h, the operator $\widetilde{\Pi}_h$ is neither self-adjoint ($\widetilde{\Pi}_h^* \neq \widetilde{\Pi}_h$) nor a projection ($\widetilde{\Pi}_h^2 \neq \widetilde{\Pi}_h$). We then use the following relation between $\widetilde{\Pi}_h$ and P_h^β from (3.4),

$$P_h^\beta L^{-\beta} g = \widetilde{\Pi}_h g \qquad \forall g \in H,$$

and express (3.11) as an integral equation for $t = 1$. Taking the expectation on both sides of this equation yields

$$\begin{aligned} \mathbb{E}[w(1, \widetilde{Y}(1))] &= w(0, Q_{h,k}^\beta \Pi_h g) \\ &\quad + \frac{1}{2} \mathbb{E} \int_0^1 \operatorname{tr} \left(w_{xx}(t, \widetilde{Y}(t)) \left(Q_{h,k}^\beta \widetilde{\Pi}_h \widetilde{\Pi}_h^* Q_{h,k}^{\beta *} - L^{-2\beta} \right) \right) \mathrm{d}t \end{aligned} \tag{3.13}$$

since $\widetilde{Y}(0) = Q_{h,k}^\beta \Pi_h g$ by (3.5) and $w_t(t, \widetilde{Y}(t)) = -\frac{1}{2} \operatorname{tr}\big(w_{xx}(t, \widetilde{Y}(t)) L^{-2\beta}\big)$ by (3.7).

As a final step in this subsection, we relate the quantity of interest $\mathbb{E}[\varphi(u)]$ with the expected value of $w(1, Y(1))$ and similarly for the approximation $\mathbb{E}[\varphi(u_{h,k}^Q)]$ and $w(1, \widetilde{Y}(1))$. For this purpose, we extend the equalities in (3.8)–(3.10) to the case that $x = \xi$ is a an H-valued random variable in the following lemma.

Lemma 3.2 *Let Assumption 2.1 (v) be satisfied. Then, for every $t \in [0, 1]$ and any $\mathscr{F}_t$-measurable random variable $\xi \in L_{p+2}(\Omega; H)$, it holds*

$$D_x^k w(t, \xi) = \mathbb{E}[D^k \varphi(\xi + Y(1) - Y(t)) \,|\, \mathscr{F}_t], \quad k \in \{0, 1, 2\}.$$

Proof For $k = 0$, this identity follows from [11, Lem. 4.1] with $N = p + 2$, $\xi_1 = \xi$ and $\xi_2 = Y(1) - Y(t)$, since $Y(t) \in L_{p+2}(\Omega; H)$ for all $t \in [0, 1]$ by Lemma 3.1 and $|\varphi(x)| \lesssim 1 + \|x\|_H^{p+2}$ as a consequence of (2.6).

Furthermore, for $y, z \in H$, we define $\varphi_y, \varphi_{y,z} \colon H \to \mathbb{R}$ by

$$\varphi_y(x) := (D\varphi(x), y)_H, \qquad \varphi_{y,z}(x) := (D^2\varphi(x)z, y)_H.$$

Since the inner product is bilinear and continuous with respect to both components, we find with (3.9)–(3.10) that

$$\begin{aligned} (w_x(t, x), y)_H &= \mathbb{E}[\varphi_y(x + Y(1) - Y(t))], \\ (w_{xx}(t, x)z, y)_H &= \mathbb{E}[\varphi_{y,z}(x + Y(1) - Y(t))]. \end{aligned}$$

Thus, again applying [11, Lem. 4.1] for $\xi_1 = \xi$ and $\xi_2 = Y(1) - Y(t)$ as well as $N = p + 1$ and $N = p$, respectively, yields

$$(w_x(t,\xi), y)_H = \mathbb{E}[\varphi_y(\xi_1 + \xi_2) \,|\, \mathscr{F}_t] = (\mathbb{E}[D\varphi(\xi + Y(1) - Y(t)) \,|\, \mathscr{F}_t], y)_H,$$
$$(w_{xx}(t,\xi)z, y)_H = \mathbb{E}[\varphi_{y,z}(\xi_1 + \xi_2) \,|\, \mathscr{F}_t] = (\mathbb{E}[D^2\varphi(\xi + Y(1) - Y(t)) \,|\, \mathscr{F}_t]z, y)_H$$

by bilinearity and continuity of the inner product. The separability of H and the arbitrary choice of $y,\ z \in H$ complete the proof of the assertion for $k \in \{1, 2\}$. □

Owing to Lemma 3.2 and the tower property for conditional expectations, the stochastic process $(w(t, Y(t)),\ t \in [0, 1])$ has no drift, i.e.,

$$\mathbb{E}[w(1, Y(1))] = \mathbb{E}[\varphi(Y(1))] = \mathbb{E}[w(0, Y(0))] = w(0, L^{-\beta} g). \tag{3.14}$$

Furthermore, it follows with (3.2) and (3.6) that

$$\mathbb{E}[w(1, Y(1))] = \mathbb{E}[\varphi(Y(1))] = \mathbb{E}[\varphi(u)], \tag{3.15}$$
$$\mathbb{E}[w(1, \widetilde{Y}(1))] = \mathbb{E}[\varphi(\widetilde{Y}(1))] = \mathbb{E}[\varphi(u_{h,k}^Q)]. \tag{3.16}$$

Summing up the observations in (3.13)–(3.16), we find that the difference between the quantity of interest $\mathbb{E}[\varphi(u)]$ and the expected value of the approximation $\varphi(u_{h,k}^Q)$ can be expressed by

$$\begin{aligned}\mathbb{E}[\varphi(u)] - \mathbb{E}[\varphi(u_{h,k}^Q)] &= w(0, L^{-\beta} g) - w(0, Q_{h,k}^\beta \Pi_h g) \\ &\quad - \frac{1}{2}\,\mathbb{E}\int_0^1 \operatorname{tr}\left(w_{xx}(t, \widetilde{Y}(t))\left(Q_{h,k}^\beta \widetilde{\Pi}_h \widetilde{\Pi}_h^* Q_{h,k}^{\beta *} - L^{-2\beta}\right)\right) \mathrm{d}t.\end{aligned}$$

This equality implies that the weak error (2.5) admits the following upper bound

$$\begin{aligned}\left|\mathbb{E}[\varphi(u)] - \mathbb{E}[\varphi(u_{h,k}^Q)]\right| &\le \left|w(0, L^{-\beta} g) - w(0, L_h^{-\beta} \Pi_h g)\right| \\ &\quad + \left|w(0, L_h^{-\beta} \Pi_h g) - w(0, Q_{h,k}^\beta \Pi_h g)\right| \\ &\quad + \frac{1}{2}\left|\mathbb{E}\int_0^1 \operatorname{tr}\left(w_{xx}(t, \widetilde{Y}(t))\left(\widetilde{Q}_{h,k}^\beta \widetilde{Q}_{h,k}^{\beta *} - \widetilde{L}_h^{-\beta} \widetilde{L}_h^{-\beta *}\right)\right) \mathrm{d}t\right| \\ &\quad + \frac{1}{2}\left|\mathbb{E}\int_0^1 \operatorname{tr}\left(w_{xx}(t, \widetilde{Y}(t))\left(\widetilde{L}_h^{-\beta} \widetilde{L}_h^{-\beta *} - L^{-2\beta}\right)\right) \mathrm{d}t\right| \\ &=: \mathrm{(I)} + \mathrm{(II)} + \mathrm{(III)} + \mathrm{(IV)},\end{aligned} \tag{3.17}$$

where we set $\widetilde{Q}_{h,k}^\beta := Q_{h,k}^\beta \widetilde{\Pi}_h$ and $\widetilde{L}_h^{-\beta} := L_h^{-\beta} \widetilde{\Pi}_h$.

The following subsections are structured as follows: In Sect. 3.3 we bound the deterministic error $\|(L^{-\beta} - L_h^{-\beta} \Pi_h) g\|_H$ caused by the finite element discretization. This result is essential for estimating the first error term (I) in (3.17). Secondly, we investigate the terms (II) and (III) stemming from applying the quadrature operator

$Q_{h,k}^{\beta}$ instead of the discrete fractional inverse $L_h^{-\beta}$ in Sect. 3.4. Finally, in Sect. 3.5 we estimate the trace in (IV) and combine all our results to prove Theorem 2.1.

3.3 The deterministic finite element error

In this subsection we focus on the deterministic error $\|(L^{-\beta} - L_h^{-\beta}\Pi_h)g\|_H$ caused by the inhomogeneity g. More precisely, we derive an explicit rate of convergence depending on the $\dot{H}^\theta$-regularity of g in Lemma 3.3 below. Subsequently, in Lemma 3.4, we apply this result to bound the first term of (3.17).

Lemma 3.3 *Suppose Assumption 2.1(iv) is satisfied. Set $\theta_* := d(2\alpha\beta - 1) - 2\beta$ and let $\theta > \min\{\theta_*, s - 2\beta\}$ if $\theta_* \geq 0$, and set $\theta = 0$ otherwise. Then there exists a constant $C > 0$, independent of h, such that*

$$\|(L^{-\beta} - L_h^{-\beta}\Pi_h)g\|_H \leq Ch^{\min\{d(2\alpha\beta-1),s\}}\|g\|_\theta \tag{3.18}$$

for all $g \in \dot{H}^\theta$ and sufficiently small $h \in (0, h_0)$.

Proof By applying [14, Ch. 2, Eq. (6.9)] to the negative fractional powers of L and L_h, we find

$$L^{-\beta} - L_h^{-\beta}\Pi_h = \frac{1}{\Gamma(\beta)}\int_0^\infty t^{\beta-1}(S(t) - S_h(t)\Pi_h)\,\mathrm{d}t.$$

Thus, Assumption 2.1(iv) yields for $0 \leq \theta_j \leq \sigma_j \leq s$ $(j = 1, 2)$ the estimate

$$\|(L^{-\beta} - L_h^{-\beta}\Pi_h)g\|_H \lesssim h^{\sigma_1}\|g\|_{\theta_1}\int_0^1 t^{\beta-1+\frac{\theta_1-\sigma_1}{2}}\,\mathrm{d}t + h^{\sigma_2}\|g\|_{\theta_2}\int_1^\infty t^{\beta-1+\frac{\theta_2-\sigma_2}{2}}\,\mathrm{d}t.$$

If $\theta_* \geq 0$, we let $\varepsilon > 0$ be such that $\theta = \min\{\theta_*, s - 2\beta\} + \varepsilon$ and we choose $\sigma_1 := \min\{d(2\alpha\beta - 1), s\}$, $\sigma_2 := s$, $\theta_1 := \min\{\theta, \sigma_1\}$, and $\theta_2 := 0$. We then obtain $\theta_1 - \sigma_1 = \min\{-2\beta + \varepsilon, 0\}$ and

$$\|(L^{-\beta} - L_h^{-\beta}\Pi_h)g\|_H \lesssim h^{\min\{d(2\alpha\beta-1),s\}}\Big(\tfrac{2}{\min\{\varepsilon,2\beta\}}\|g\|_{\theta_1} + \tfrac{2}{s-2\beta}\|g\|_H\Big).$$

For $\theta_* < 0$, we instead set $\sigma_1 := d(2\alpha\beta - 1)$, $\sigma_2 := s$, $\theta_1 := 0$, $\theta_2 := 0$, and we conclude in a similar way that

$$\|(L^{-\beta} - L_h^{-\beta}\Pi_h)g\|_H \lesssim h^{\min\{d(2\alpha\beta-1),s\}}\|g\|_H(-2\theta_*^{-1} + 2(s - 2\beta)^{-1}).$$

Since in both cases $\max\{\|g\|_{\theta_1}, \|g\|_{\theta_2}\} \leq \|g\|_\theta$ with θ defined as in the statement of the lemma, the bound (3.18) follows. □

Remark 3.1 We note that by letting $\sigma_1 = \sigma_2 := s$, $\theta_1 := s - 2\beta + \varepsilon$, and $\theta_2 := 0$ in the proof of Lemma 3.3 the optimal convergence rate for the deterministic error,

$$\|(L^{-\beta} - L_h^{-\beta}\Pi_h)g\|_H \leq Ch^s\|g\|_{s-2\beta+\varepsilon},$$

can be derived. The error estimate (3.18) is formulated in such a way that the smoothness $\theta \geq 0$ of $g \in \dot{H}^\theta$ is minimal for convergence with the rate $\min\{d(2\alpha\beta - 1), s\}$, which will dominate the overall weak error, stemming from the term (IV) in the partition (3.17), see Lemma 3.8.

We furthermore remark that the convergence result of Lemma 3.3 is in accordance with the result of [6, Thm. 4.3]. There the self-adjoint positive definite operator L is induced by an $H_0^1(\mathscr{D})$-coercive, symmetric bilinear form A:

$$\langle Lv, w\rangle := A(v, w) = \int_{\mathscr{D}} a(\mathbf{x})\nabla v(\mathbf{x}) \cdot \nabla w(\mathbf{x})\, \mathrm{d}\mathbf{x} \quad \forall v, w \in \dot{H}^1,$$

where $0 < a_0 \leq a(\mathbf{x}) \leq a_1$, $H := L_2(\mathscr{D})$, $\dot{H}^1 := H_0^1(\mathscr{D})$ and $\mathscr{D}$ is a bounded polygonal domain in $\mathbb{R}^d$, $d \in \{1, 2, 3\}$, with Lipschitz boundary. The discrete spaces $(V_h)_h$ considered in [6] are the finite element spaces with continuous piecewise linear basis functions defined with respect to a quasi-uniform family of triangulations. The convergence rate for the error $\|(L^{-\beta} - L_h^{-\beta}\Pi_h)g\|_H$ derived in [6, Thm. 4.3] is 2τ, if $g \in \dot{H}^\theta$ for $\theta > 2(\tau - \beta)$, if $\tau \geq \beta$, and $\theta = 0$ otherwise. Here, $\tau \in (0, 1]$ is such that the operators

$$L^{-1} \colon \widetilde{H}^{-1+\tau}(\mathscr{D}) \to \widetilde{H}^{1+\tau}(\mathscr{D}) \quad \text{and} \quad L \colon \widetilde{H}^{1+\tau}(\mathscr{D}) \to \widetilde{H}^{-1+\tau}(\mathscr{D})$$

are bounded with respect to the intermediate Sobolev spaces

$$\widetilde{H}^\varrho(\mathscr{D}) := \begin{cases} H_0^1(\mathscr{D}) \cap H^\varrho(\mathscr{D}), & \varrho \in [1, 2], \\ [L_2(\mathscr{D}), H_0^1(\mathscr{D})]_{\varrho,2}, & \varrho \in [0, 1], \\ [H^{-1}(\mathscr{D}), L_2(\mathscr{D})]_{1+\varrho,2}, & \varrho \in [-1, 0], \end{cases}$$

where $H^{-1}(\mathscr{D}) = \dot{H}^{-1}$ is the dual space of $H_0^1(\mathscr{D}) = \dot{H}^1$ and $[\cdot, \cdot]_{\varrho,q}$ denotes the real K-interpolation method.

According to this result of [6], the convergence rate $2\min\{d(\alpha\beta - 1/2), 1\}$ can be achieved if g is $\dot{H}^\theta$-regular for $\theta > \theta_*$ if $\theta_* := 2(\min\{d(\alpha\beta - 1/2), 1\} - \beta) \geq 0$ and $\theta = 0$ if $\theta_* < 0$. A comparison with (3.18) in Lemma 3.3 shows that the error estimates and regularity assumptions coincide for this particular case, since $s = 2$ for the choice of finite-dimensional subspaces $(V_h)_h$ in [6] specified above.

Having bounded the error between $L^{-\beta}g$ and $L_h^{-\beta}\Pi_h g$, an estimate of the first error term (I) in (3.17) is an immediate consequence of the fundamental theorem of calculus and the chain rule for Fréchet derivatives. This bound is formulated in the next lemma.

Lemma 3.4 *Let Assumptions 2.1 (iv)–(v) be satisfied and $2\alpha\beta > 1$. Define $\theta \geq 0$ as in Lemma 3.3. Then there exists a constant $C > 0$, independent of h, such that*

$$\left|w(0, L^{-\beta}g) - w(0, L_h^{-\beta}\Pi_h g)\right| \leq Ch^{\min\{d(2\alpha\beta-1), s\}}\|g\|_\theta(1 + \|g\|_H^{p+1})$$

for all $g \in \dot{H}^\theta$ and sufficiently small $h \in (0, h_0)$.

Proof Since the mapping $x \mapsto w(0, x)$ is Fréchet differentiable, we obtain by the fundamental theorem of calculus and the Cauchy–Schwarz inequality

$$\begin{aligned}
&\left|w(0, L_h^{-\beta}\Pi_h g) - w(0, L^{-\beta} g)\right| \\
&\quad = \left|\int_0^1 (w_x(0, L^{-\beta} g + t(L_h^{-\beta}\Pi_h - L^{-\beta})g), (L_h^{-\beta}\Pi_h - L^{-\beta})g)_H \,\mathrm{d}t\right| \\
&\quad \leq \|(L_h^{-\beta}\Pi_h - L^{-\beta})g\|_H \sup_{t\in[0,1]} \|w_x(0, L^{-\beta} g + t(L_h^{-\beta}\Pi_h - L^{-\beta})g)\|_H.
\end{aligned}$$

A bound for the first term is given by (3.18) in Lemma 3.3. For the second term, we use (3.9), $Y(0) = L^{-\beta} g$, and the polynomial growth (2.6) of $D^2\varphi$ to estimate

$$\begin{aligned}
\|w_x(0, L^{-\beta} g + t(L_h^{-\beta}\Pi_h - L^{-\beta})g)\|_H &\leq \mathbb{E}[\|D\varphi(Y(1) + t(L_h^{-\beta}\Pi_h - L^{-\beta})g)\|_H] \\
&\lesssim \left(1 + \mathbb{E}[\|Y(1)\|_H^{p+1}] + \|g\|_H^{p+1}\right)
\end{aligned}$$

for all $t \in [0, 1]$. The boundedness (3.3) of the $(p + 1)$-th moment of $Y(1)$ completes the proof, since the trace of $L^{-2\beta}$ is finite if $2\alpha\beta > 1$. □

3.4 The quadrature approximation

In this subsection we address the error terms (II) and (III) in (3.17), which are induced by the quadrature approximation $Q_{h,k}^{\beta}$ of $L_h^{-\beta}$. To this end, we start by stating the following result of [6, Lem. 3.4, Thm. 3.5] that bounds the error between the two operators on V_h.

Lemma 3.5 *The approximation $Q_{h,k}^{\beta}\colon V_h \to V_h$ of $L_h^{-\beta}$ in (2.4) admits the bound*

$$\|(Q_{h,k}^{\beta} - L_h^{-\beta})\phi_h\|_H \leq Ce^{-\frac{\pi^2}{2k}}\|\phi_h\|_H \quad \forall \phi_h \in V_h,$$

and it is bounded, $\|Q_{h,k}\|_{\mathscr{L}(V_h)} \leq C'$, for sufficiently small $h \in (0, h_0)$, $k \in (0, k_0)$, where the constants $C, C' > 0$ depend only on β and the smallest eigenvalue of L.

In the following, we use this error estimate of the quadrature approximation $Q_{h,k}^{\beta}$ for bounding the second term of (3.17) in Lemma 3.6 as well as the trace occurring in the third term of (3.17) in Lemma 3.7.

Lemma 3.6 *Suppose that Assumption 2.1(v) is satisfied and that $2\alpha\beta > 1$. Then there exists a constant $C > 0$, independent of h and k, such that*

$$\left|w(0, L_h^{-\beta}\Pi_h g) - w(0, Q_{h,k}^{\beta}\Pi_h g)\right| \leq Ce^{-\frac{\pi^2}{2k}}\|g\|_H\left(1 + \|g\|_H^{p+1}\right)$$

for all $g \in H$ and sufficiently small $h \in (0, h_0)$ and $k \in (0, k_0)$.

Proof As in the proof of Lemma 3.4, we apply the fundamental theorem of calculus and the chain rule for Fréchet derivatives. By (3.9) and Lemma 3.5 we then find

$$\begin{aligned}\left|w(0, Q_{h,k}^{\beta}\Pi_h g) - w(0, L_h^{-\beta}\Pi_h g)\right| &\le \|(Q_{h,k}^{\beta} - L_h^{-\beta})\Pi_h g\|_H \\ &\quad \times \sup_{t\in[0,1]} \mathbb{E}[\|D\varphi(L_h^{-\beta}\Pi_h g + t(Q_{h,k}^{\beta} - L_h^{-\beta})\Pi_h g + Y(1) - L^{-\beta} g)\|_H] \\ &\lesssim e^{-\frac{\pi^2}{2k}}\|g\|_H \left(1 + \mathbb{E}[\|Y(1)\|_H^{p+1}] + \|g\|_H^{p+1}\right).\end{aligned}$$

Again, the proof is completed by (3.3) and the fact that $\operatorname{tr}(L^{-2\beta}) < \infty$. □

Lemma 3.7 *Let Assumptions 2.1(i)–(iii) be satisfied. Then there exists a constant $C > 0$, independent of h and k, such that*

$$\left|\operatorname{tr}(T(\widetilde{Q}_{h,k}^{\beta}\widetilde{Q}_{h,k}^{\beta *} - \widetilde{L}_h^{-\beta}\widetilde{L}_h^{-\beta *}))\right| \le C\left(e^{-\frac{\pi^2}{k}}h^{-d} + e^{-\frac{\pi^2}{2k}} + e^{-\frac{\pi^2}{2k}} f_{\alpha,\beta}(h)\right)\|T\|_{\mathscr{L}(H)}$$

for every self-adjoint $T \in \mathscr{L}(H)$ and sufficiently small $h \in (0, h_0)$ and $k \in (0, k_0)$. Here, the function $f_{\alpha,\beta}$ is defined as in Theorem 2.1.

Proof By the definition of $\widetilde{\Pi}_h$ in (3.12) we have

$$\widetilde{\Pi}_h e_j = e_{j,h}, \quad j \in \{1, \dots, N_h\}, \qquad \widetilde{\Pi}_h e_j = 0, \quad j > N_h. \tag{3.19}$$

Therefore, the trace of interest simplifies to a finite sum,

$$\begin{aligned}\operatorname{tr}(T(\widetilde{Q}_{h,k}^{\beta}\widetilde{Q}_{h,k}^{\beta *} - \widetilde{L}_h^{-\beta}\widetilde{L}_h^{-\beta *})) &= \sum_{j=1}^{N_h}\left[(T Q_{h,k}^{\beta} e_{j,h}, Q_{h,k}^{\beta} e_{j,h})_H - (T L_h^{-\beta} e_{j,h}, L_h^{-\beta} e_{j,h})_H\right] \\ &= \sum_{j=1}^{N_h}(T(Q_{h,k}^{\beta} - L_h^{-\beta})e_{j,h}, (Q_{h,k}^{\beta} - L_h^{-\beta})e_{j,h})_H \\ &\quad + 2\sum_{j=1}^{N_h}(T(Q_{h,k}^{\beta} - L_h^{-\beta})e_{j,h}, L_h^{-\beta} e_{j,h})_H \\ &=: S_1 + 2S_2,\end{aligned} \tag{3.20}$$

where the second equality follows from the self-adjointness of $T \in \mathscr{L}(H)$.

The application of the Cauchy–Schwarz inequality and of Lemma 3.5 to the first sum yield the following upper bound

$$|S_1| \le \|T\|_{\mathscr{L}(H)} \sum_{j=1}^{N_h} \|(Q_{h,k}^{\beta} - L_h^{-\beta})e_{j,h}\|_H^2 \le C e^{-\frac{\pi^2}{k}} N_h \|T\|_{\mathscr{L}(H)}.$$

By Assumption 2.1(i) we thus have $|S_1| \lesssim e^{-\frac{\pi^2}{k}} h^{-d}\|T\|_{\mathscr{L}(H)}$.

The second sum can be bounded by

$$|S_2| \leq \|T\|_{\mathscr{L}(H)} \max_{1\leq j\leq N_h} \|(Q_{h,k}^{\beta} - L_h^{-\beta})e_{j,h}\|_H \sum_{j=1}^{N_h} \lambda_{j,h}^{-\beta}.$$

Finally, due to the approximation property of the discrete eigenvalues $\lambda_{j,h}$ in Assumption 2.1(ii) as well as the growth (2.2) of the exact eigenvalues λ_j we obtain $\lambda_{j,h}^{-\beta} \leq \lambda_j^{-\beta} \leq c_\lambda^{-\beta} j^{-\alpha\beta}$ and, for $\alpha\beta \neq 1$, we find

$$|S_2| \lesssim e^{-\frac{\pi^2}{2k}} \left(1 + N_h^{1-\alpha\beta}\right) \|T\|_{\mathscr{L}(H)} \lesssim e^{-\frac{\pi^2}{2k}} \left(1 + h^{d(\alpha\beta-1)}\right) \|T\|_{\mathscr{L}(H)},$$

where we have used Lemma 3.5 and Assumption 2.1(i). If $\alpha\beta = 1$, we instead estimate $|S_2| \lesssim e^{-\pi^2/(2k)}(1 + |\ln(h)|)\, \|T\|_{\mathscr{L}(H)}$. This completes the proof. □

3.5 Proof of Theorem 2.1

After having bounded the terms (I), (II), and (III) in the partition (3.17) of the weak error in the previous subsections, we now turn to estimating the final error term (IV). Furthermore, we bound the p-th moment of $\widetilde{Y}(t)$, where $\widetilde{Y}$ is the solution process of (3.5). We then combine all our results and prove Theorem 2.1.

Lemma 3.8 *Let Assumptions 2.1(i)–(iii) be satisfied. Then there exists a constant $C > 0$, independent of h, such that*

$$\left|\mathrm{tr}\left(T\left(\widetilde{L}_h^{-\beta}\widetilde{L}_h^{-\beta *} - L^{-2\beta}\right)\right)\right| \leq C h^{\min\{d(2\alpha\beta-1),r,s\}} \|T\|_{\mathscr{L}(H)}$$

for every self-adjoint $T \in \mathscr{L}(H)$ and sufficiently small $h \in (0, h_0)$.

Proof Similarly to (3.20) we use the self-adjointness of T and rewrite the trace as $\mathrm{tr}(T(\widetilde{L}_h^{-\beta}\widetilde{L}_h^{-\beta *} - L^{-2\beta})) = S_1 + S_2$, where

$$S_1 := \sum_{j\in\mathbb{N}} (T(\widetilde{L}_h^{-\beta} - L^{-\beta})e_j, \widetilde{L}_h^{-\beta} e_j)_H, \qquad S_2 := \sum_{j\in\mathbb{N}} (T(\widetilde{L}_h^{-\beta} - L^{-\beta})e_j, L^{-\beta} e_j)_H.$$

In order to estimate the terms S_1 and S_2, we note that for $j \in \{1, \ldots, N_h\}$

$$\|(\widetilde{L}_h^{-\beta} - L^{-\beta})e_j\|_H = \|\lambda_{j,h}^{-\beta} e_{j,h} - \lambda_j^{-\beta} e_j\|_H \leq |\lambda_{j,h}^{-\beta} - \lambda_j^{-\beta}| + \lambda_j^{-\beta} \|e_{j,h} - e_j\|_H.$$

By the mean value theorem, the existence of $\widetilde{\lambda}_j \in (\lambda_j, \lambda_{j,h})$ satisfying $\lambda_j^{-\beta} - \lambda_{j,h}^{-\beta} = \beta\widetilde{\lambda}_j^{-\beta-1}(\lambda_{j,h} - \lambda_j)$ is ensured. By Assumption 2.1(ii) we thus have

$$\|(\widetilde{L}_h^{-\beta} - L^{-\beta})e_j\|_H \leq \max\left\{\beta C_1, \sqrt{C_2}\right\} \left(h^r \lambda_j^{q-\beta-1} + h^s \lambda_j^{\frac{q}{2}-\beta}\right). \tag{3.21}$$

Owing to (3.19) the series S_1 simplifies to the finite sum

$$S_1 = \sum_{j=1}^{N_h} \lambda_{j,h}^{-\beta} (T(\widetilde{L}_h^{-\beta} - L^{-\beta})e_j, e_{j,h})_H.$$

Using (3.21) as well as Assumptions 2.1(i)–(iii), this sum can be bounded by

$$|S_1| \lesssim \|T\|_{\mathscr{L}(H)} \sum_{j=1}^{N_h} \left(h^r \lambda_j^{q-2\beta-1} + h^s \lambda_j^{\frac{q}{2}-2\beta} \right) \lesssim h^{\min\{d(2\alpha\beta-1),r,s\}} \|T\|_{\mathscr{L}(H)},$$

since $d\alpha(q-1) \le r$ and $d\alpha q/2 \le s$ by Assumption 2.1(iii).

For the second term we find

$$S_2 = \sum_{j=1}^{N_h} \lambda_j^{-\beta} (T(\widetilde{L}_h^{-\beta} - L^{-\beta})e_j, e_j)_H - \sum_{j>N_h} \lambda_j^{-2\beta} (Te_j, e_j)_H,$$

since $\widetilde{L}_h^{-\beta} e_j = 0$ for $j > N_h$ by (3.19). Therefore, the application of (3.21) yields

$$|S_2| \lesssim \|T\|_{\mathscr{L}(H)} \left(\sum_{j=1}^{N_h} \left(h^r \lambda_j^{q-2\beta-1} + h^s \lambda_j^{\frac{q}{2}-2\beta} \right) + \sum_{j>N_h} \lambda_j^{-2\beta} \right)$$

and $|S_2| \lesssim h^{\min\{d(2\alpha\beta-1),r,s\}} \|T\|_{\mathscr{L}(H)}$ follows from Assumptions 2.1(i), (iii). □

Lemma 3.9 *Suppose that Assumptions 2.1(i)–(iii) are satisfied. Let $p \in \mathbb{N}$, $t \in [0,1]$, and $\widetilde{Y}$ be the strong solution of (3.5). Then the p-th moment of $\widetilde{Y}(t)$ exists and, for $p \ge 2$, it admits the following bound:*

$$\mathbb{E}\big[\|\widetilde{Y}(t)\|_H^p\big] \le C\Big(1 + e^{-\frac{p\pi^2}{2k}} h^{-\frac{pd}{2}} + \|g\|_H^p\Big),$$

where the constant $C > 0$ is independent of h and k.

Proof Since $P_h^\beta W^\beta(t) = \sum_{j=1}^{N_h} B_j(t)\, e_{j,h}$, we obtain by Lemma 3.5, for any $p \ge 2$, that

$$\mathbb{E}\big[\|(Q_{h,k}^\beta - L_h^{-\beta}) P_h^\beta W^\beta(t)\|_H^p\big] \le C^p e^{-\frac{p\pi^2}{2k}}\, \mathbb{E}\Big|\sum_{j=1}^{N_h} B_j(t)^2\Big|^{\frac{p}{2}} \le C^p e^{-\frac{p\pi^2}{2k}} N_h^{\frac{p}{2}} t^{\frac{p}{2}} \mu_p,$$

where, again, $\mu_p := \mathbb{E}[|Z|^p]$ denotes the p-th absolute moment of $Z \sim \mathscr{N}(0,1)$ and the constant $C > 0$ is independent of h, k, and p. Furthermore, using $0 < \lambda_j \le \lambda_{j,h}$

of Assumption 2.1(ii) and applying the Hölder inequality gives

$$\mathbb{E}\big[\|L_h^{-\beta} P_h^\beta W^\beta(t)\|_H^p\big] = \mathbb{E}\Big|\sum_{j=1}^{N_h} \lambda_{j,h}^{-2\beta} B_j(t)^2\Big|^{\frac{p}{2}} \leq \operatorname{tr}(L^{-2\beta})^{\frac{p}{2}} t^{\frac{p}{2}} \mu_p,$$

where $\operatorname{tr}(L^{-2\beta}) < \infty$ by Assumption 2.1(iii). Thus, we obtain for the solution $\widetilde{Y}$ of (3.5) that for any $t \in [0,1]$ the bound

$$\begin{aligned}\mathbb{E}\big[\|\widetilde{Y}(t)\|_H^p\big] &= \mathbb{E}\big[\|Q_{h,k}^\beta \Pi_h g + (Q_{h,k}^\beta - L_h^{-\beta}) P_h^\beta W^\beta(t) + L_h^{-\beta} P_h^\beta W^\beta(t)\|_H^p\big] \\ &\leq 3^{p-1}\Big(\|Q_{h,k}^\beta \Pi_h g\|_H^p + \mathbb{E}\big[\|(Q_{h,k}^\beta - L_h^{-\beta}) P_h^\beta W^\beta(t)\|_H^p\big] + \mathbb{E}\big[\|L_h^{-\beta} P_h^\beta W^\beta(t)\|_H^p\big]\Big) \\ &\leq 3^{p-1}\Big(\|Q_{h,k}^\beta\|_{\mathscr{L}(V_h)}^p \|g\|_H^p + C^p e^{-\frac{p\pi^2}{2k}} N_h^{\frac{p}{2}} t^{\frac{p}{2}} \mu_p + \operatorname{tr}(L^{-2\beta})^{\frac{p}{2}} t^{\frac{p}{2}} \mu_p\Big)\end{aligned}$$

holds. Finally, the assertion follows by the boundedness of $Q_{h,k}^\beta$ which is uniform in h and k, the finiteness of $\operatorname{tr}(L^{-2\beta})$, and Assumption 2.1(i). □

Proof (of Theorem 2.1) Owing to the partition (3.17) and the estimates of the error terms (I)–(IV) in Lemmata 3.4 and 3.6–3.8 we can bound the weak error as follows

$$\begin{aligned}\big|\mathbb{E}[\varphi(u)] - \mathbb{E}[\varphi(u_{h,k}^Q)]\big| &\lesssim \Big(h^{\min\{d(2\alpha\beta-1),s\}} + e^{-\frac{\pi^2}{2k}}\Big)\|g\|_\theta \Big(1 + \|g\|_H^{p+1}\Big) \\ &\quad + \sup_{t\in[0,1]} \mathbb{E}\big[\|w_{xx}(t,\widetilde{Y}(t))\|_{\mathscr{L}(H)}\big]\Big(e^{-\frac{\pi^2}{k}} h^{-d} + e^{-\frac{\pi^2}{2k}} + e^{-\frac{\pi^2}{2k}} f_{\alpha,\beta}(h)\Big) \\ &\quad + \sup_{t\in[0,1]} \mathbb{E}\big[\|w_{xx}(t,\widetilde{Y}(t))\|_{\mathscr{L}(H)}\big] h^{\min\{d(2\alpha\beta-1),r,s\}},\end{aligned}$$

since $w_{xx}(t,x) \in \mathscr{L}(H)$ is self-adjoint for every $t \in [0,1]$ and $x \in H$. The application of Lemma 3.2 and of the tower property for conditional expectations yield

$$\begin{aligned}\mathbb{E}[\|w_{xx}(t,\widetilde{Y}(t))\|_{\mathscr{L}(H)}] &= \mathbb{E}[\|\mathbb{E}[D^2\varphi(\widetilde{Y}(t) + Y(1) - Y(t))|\mathscr{F}_t]\|_{\mathscr{L}(H)}] \\ &\leq \mathbb{E}[\|D^2\varphi(\widetilde{Y}(t) + Y(1) - Y(t))\|_{\mathscr{L}(H)}].\end{aligned}$$

By the polynomial growth (2.6) of $D^2\varphi$ and the boundedness of the p-th moments of $Y(t)$ and $\widetilde{Y}(t)$ in Lemmata 3.1 and 3.9, respectively, we obtain that

$$\begin{aligned}\mathbb{E}[\|w_{xx}(t,\widetilde{Y}(t))\|_{\mathscr{L}(H)}] &\lesssim \big(1 + \mathbb{E}[\|\widetilde{Y}(t)\|_H^p] + \mathbb{E}[\|Y(1)\|_H^p] + \mathbb{E}[\|Y(t)\|_H^p]\big) \\ &\lesssim \Big(1 + e^{-\frac{p\pi^2}{2k}} h^{-\frac{pd}{2}} + \|g\|_H^p\Big),\end{aligned}$$

since $\operatorname{tr}(L^{-2\beta}) < \infty$. This completes the proof of the weak error estimate in (2.8). □

Remark 3.2 Note that, if the first and second Fréchet derivatives of φ are bounded, the estimates of the Lemmata 3.1 and 3.9 are not needed and the weak error estimate in (2.8) simplifies to

$$\begin{aligned}&\left|\mathbb{E}[\varphi(u)] - \mathbb{E}[\varphi(u_{h,k}^{Q})]\right| \\ &\qquad \leq C\left(h^{\min\{d(2\alpha\beta-1),r,s\}} + e^{-\frac{\pi^2}{k}}h^{-d} + e^{-\frac{\pi^2}{2k}} + e^{-\frac{\pi^2}{2k}} f_{\alpha,\beta}(h)\right)(1+\|g\|_\theta).\end{aligned}$$

The calibration of the discretization parameters k and h remains as described in Remark 2.2.

4 An application and numerical experiments

In this section we validate the theoretical results of the previous sections within the scope of a simulation study based on the model for Matérn approximations in (1.1) on the domain $\mathscr{D} = (0,1)^d$ for $d = 1, 2$, $\kappa = 0.5$, and $u = 0$ on $\partial\mathscr{D}$, i.e., $L = \kappa^2 - \Delta$ with homogeneous Dirichlet boundary conditions. In this case, the operator L has the following eigenvalue-eigenvector pairs [8, Ch. VI.4]:

$$\lambda_{\mathbf{j}} = \kappa^2 + \pi^2|\mathbf{j}|^2 = \kappa^2 + \pi^2\sum_{i=1}^{d} j_i^2, \qquad e_{\mathbf{j}}(\mathbf{x}) = \prod_{i=1}^{d}\left(\sqrt{2}\sin(\pi\, j_i\, x_i)\right), \tag{4.1}$$

where $\mathbf{j} = (j_1, \ldots, j_d) \in \mathbb{N}^d$ is a d-dimensional multi-index. As already mentioned in Example 2.1, these eigenvalues satisfy (2.2) for $\alpha = 2/d$.

Note that, for every $\mathbf{x} \in \mathscr{D}$, the solution u satisfies $u(\mathbf{x}) \sim \mathscr{N}(0, \sigma(\mathbf{x})^2)$. Following a Karhunen–Loève expansion of u with respect to the eigenfunctions $\{e_{\mathbf{j}}\}_{\mathbf{j}\in\mathbb{N}^d}$ in (4.1), the variance $\sigma(\mathbf{x})^2$ can be expressed explicitly in terms of the eigenvalues and eigenfunctions in (4.1) by

$$\sigma(\mathbf{x})^2 = \mathbb{E}\left|\sum_{\mathbf{j}\in\mathbb{N}^d}\lambda_{\mathbf{j}}^{-\beta}\widetilde{\xi}_{\mathbf{j}}\, e_{\mathbf{j}}(\mathbf{x})\right|^2 = \sum_{\mathbf{j}\in\mathbb{N}^d}\lambda_{\mathbf{j}}^{-2\beta} e_{\mathbf{j}}(\mathbf{x})^2, \tag{4.2}$$

where $\left\{\widetilde{\xi}_{\mathbf{j}}\right\}_{\mathbf{j}\in\mathbb{N}^d}$ are independent $\mathscr{N}(0,1)$-distributed random variables.

Considering continuous evaluation functions $\varphi\colon L_2(\mathscr{D}) \to \mathbb{R}$ of the form

$$\varphi(u) = \int_{\mathscr{D}} f(u(\mathbf{x}))\,\mathrm{d}\mathbf{x}$$

allows us to perform the simulation study without Monte Carlo sampling, since

$$\mathbb{E}[\varphi(u)] = \int_{\mathscr{D}} \mathbb{E}[f(u(\mathbf{x}))]\,\mathrm{d}\mathbf{x},$$

Table 1 Numbers of finite element basis functions and the corresponding numbers of quadrature nodes as a function of β

		β			
	N_h	0.6	0.7	0.8	0.9
$d = 1$	511	146	226	386	866
	1023	180	278	476	1069
	2047	218	337	576	1293
	4095	258	400	685	1538
$d = 2$	225	24	36	60	133
	961	38	58	98	218
	3969	56	86	145	325
	16,129	78	119	203	453

and the value of $\mathbb{E}[f(u(\mathbf{x}))]$ can be derived analytically from $u(\mathbf{x}) \sim \mathcal{N}(0, \sigma(\mathbf{x})^2)$. More precisely, we choose $f(u) = |u|^p$, $p = 2, 3, 4$, and $f(u) = \Phi(20(u - 0.5))$, where $\Phi(\cdot)$ denotes the cumulative distribution function for the standard normal distribution. The motivation of the latter function is given by its correspondence to a probit transform which is often used to approximate step functions (see, e.g., [1]), in this case $\mathbb{1}(u > 0.5)$. These four functions satisfy Assumption 2.1(v) and we obtain for the quantity of interest,

$$\mathbb{E}[\varphi(u)] = \frac{2^{p/2}\Gamma((p+1)/2)}{\sqrt{\pi}} \int_{\mathscr{D}} \sigma(\mathbf{x})^p \, \mathrm{d}\mathbf{x}, \tag{4.3}$$

if $f(u) = |u|^p$, and

$$\mathbb{E}[\varphi(u)] = \int_{\mathscr{D}} \Phi\left(-\frac{a}{\sqrt{c^{-2} + \sigma(\mathbf{x})^2}}\right) \mathrm{d}\mathbf{x}, \tag{4.4}$$

if $f(u) = \Phi(c(u - a))$ for $a \in \mathbb{R}$ and $c > 0$.

We truncate the series in (4.2) in order to approximate the variance $\sigma(\mathbf{x})^2$,

$$\sigma(\mathbf{x})^2 \approx \sum_{j_1=1}^{N_{\text{ok}}} \cdots \sum_{j_d=1}^{N_{\text{ok}}} \lambda_{(j_1,\ldots,j_d)}^{-2\beta} e_{(j_1,\ldots,j_d)}(\mathbf{x})^2.$$

Here, we choose $N_{\text{ok}} = 1 + 2^{18}$ for $d = 1$ and $N_{\text{ok}} = 1 + 2^{11}$ for $d = 2$ so that, in both cases, $N_{\text{ok}}^d \gg N_h$ for all considered finite element spaces with N_h basis functions. This estimate of $\sigma(\mathbf{x})$ is used at N_{ok}^d equally spaced locations $\mathbf{x} \in \mathscr{D}$, and the reference solution $\mathbb{E}[\varphi(u)]$ is then approximated by applying the trapezoidal rule in order to evaluate the integrals in (4.3) and (4.4) numerically.

We consider (1.1) for $\beta = 0.6, 0.7, 0.8, 0.9$ and use a finite element discretization based on continuous piecewise linear basis functions with respect to uniform meshes on $\bar{\mathscr{D}} = [0, 1]^d$. We use four different mesh sizes h in each dimension $d = 1, 2$, and calibrate the quadrature step size k with h for each value of β by $k = -1/(\beta \ln h)$.

This results in the numbers of basis functions and quadrature nodes shown in Table 1. As already pointed out in Example 2.1, the growth exponent of the eigenvalues is in this case $\alpha = 2/d$, and Assumption 2.1 is satisfied for $r = s = q = 2$. This gives the theoretical value $\min\{4\beta - d, 2\}$ for the weak convergence rate.

For the computation of $\mathbb{E}[\varphi(u_{h,k}^Q)]$ we can use the same procedure as for the reference solution in order to avoid Monte Carlo simulations. For this purpose, we have to replace $\sigma(\mathbf{x})^2$ in (4.3) and (4.4) by the variance of the finite element solution, $\sigma_h(\mathbf{x})^2 = \mathrm{Var}(u_{h,k}^Q(\mathbf{x}))$. To this end, we first assemble the matrix

$$\mathbf{Q}_{h,k}^{\beta} = \frac{2k\sin(\pi\beta)}{\pi} \sum_{\ell=-K^-}^{K^+} e^{2\beta y_\ell} (\mathbf{M} + e^{2y_\ell}(\kappa^2\mathbf{M} + \mathbf{S}))^{-1},$$

where $y_\ell := \ell k$ and $\mathbf{M}, \mathbf{S} \in \mathbb{R}^{N_h \times N_h}$ are the mass matrix and the stiffness matrix with respect to the finite element basis $\{\phi_{j,h}\}_{j=1}^{N_h}$ with entries

$$M_{ij} := (\phi_{i,h}, \phi_{j,h})_{L_2(\mathscr{D})}, \quad S_{ij} := (\nabla\phi_{i,h}, \nabla\phi_{j,h})_{L_2(\mathscr{D})}, \quad 1 \le i, j, \le N_h.$$

If we let $\boldsymbol{\phi}_h(\mathbf{x}) := (\phi_{1,h}(\mathbf{x}), \ldots, \phi_{N_h,h}(\mathbf{x}))^T$ denote the vector of the finite element basis functions evaluated at $\mathbf{x} \in \mathscr{D}$ and $\mathbf{b} := ((\mathscr{W}_h^\Phi, \phi_{j,h})_{L_2(\mathscr{D})})_{j=1}^{N_h} \sim \mathscr{N}(\mathbf{0}, \mathbf{M})$, the variance $\sigma_h(\mathbf{x})^2$ is given by

$$\sigma_h(\mathbf{x})^2 = \mathrm{Var}(u_{h,k}^Q(\mathbf{x})) = \mathrm{Var}\Big(\boldsymbol{\phi}_h(\mathbf{x})^T \mathbf{Q}_{h,k}^{\beta}\mathbf{b}\Big) = \boldsymbol{\phi}_h(\mathbf{x})^T \mathbf{Q}_{h,k}^{\beta}\mathbf{M}(\mathbf{Q}_{h,k}^{\beta})^T\boldsymbol{\phi}_h(\mathbf{x}).$$

The computation of $\sigma_h(\mathbf{x})^2$ at the same N_{ok}^d locations as for the reference solution again enables a numerical evaluation of the integrals in (4.3) and (4.4) via the trapezoidal rule for approximating $\mathbb{E}[\varphi(u_{h,k}^Q)]$.

The resulting observed weak errors $\mathrm{err}_\ell := |\mathbb{E}[\varphi(u)] - \mathbb{E}[\varphi(u_{h_\ell,k}^Q)]|$ are shown in Fig. 1. For each function φ and for each value of β, we compute the empirical convergence rate r by a least-squares fit of a line $\mathrm{c} + \mathrm{r}\ln h$ to the data set $\{h_\ell, \mathrm{err}_\ell\}$. The results are shown in Table 2 and can be seen to validate the theoretical rates given in Theorem 2.1 for $d = 1$. For $d = 2$, the observed rates deviate slightly from the theoretical rates for $\beta = 0.9$, which is caused by the fact that we had to use coarser finite element meshes for $d = 2$ than for $d = 1$ in order to be able to assemble the dense matrices $\mathbf{Q}_{h,k}^{\beta} \in \mathbb{R}^{N_h \times N_h}$ for performing the simulation study without Monte Carlo simulations.

5 Conclusion

Gaussian random fields are of great importance as models in spatial statistics. A popular method for reducing the computational cost for operations, which are needed during statistical inference, is to represent the Gaussian field as a solution to an SPDE. In this work, we have investigated a recent extension of this approach to Gaussian random

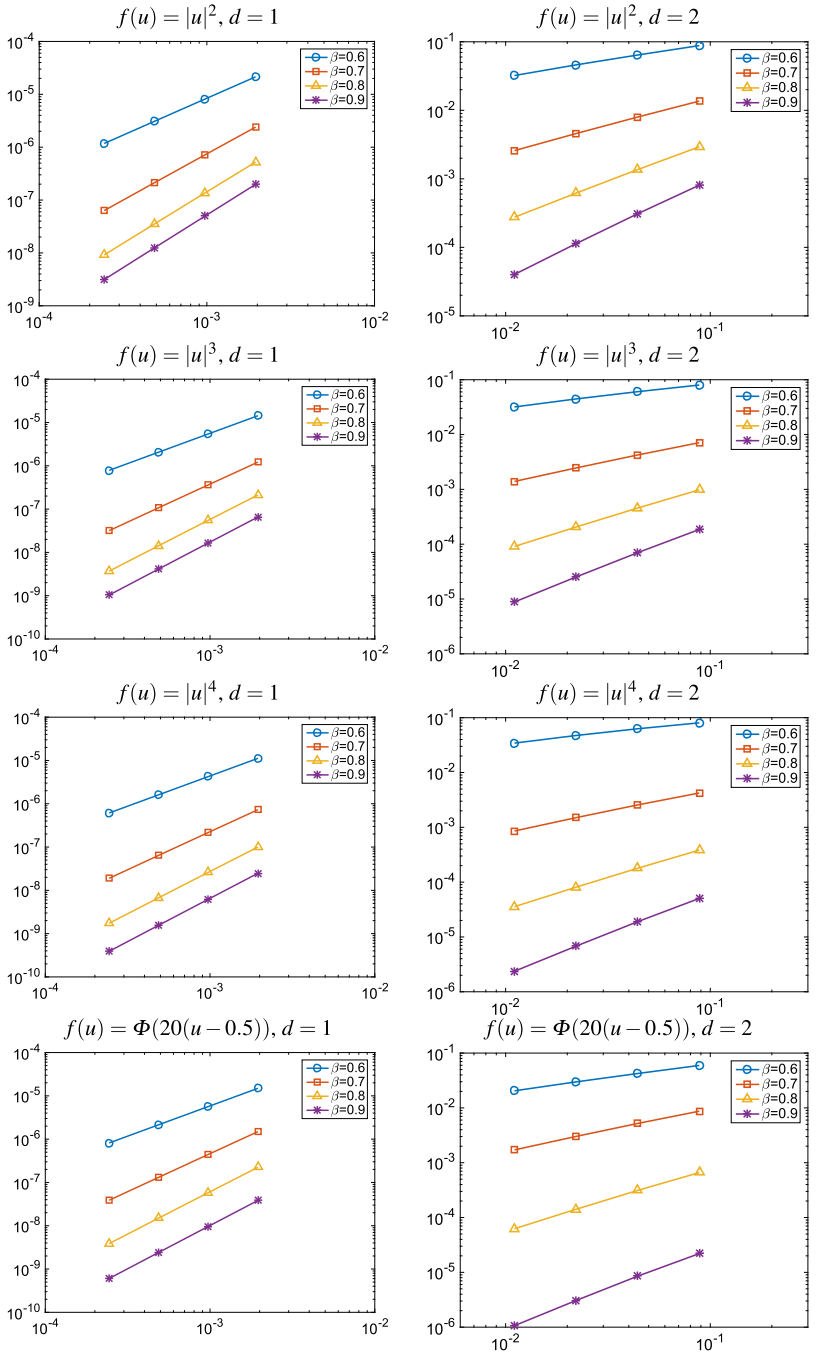

Fig. 1 Observed weak errors for $d = 1, 2$ and different values of β. The errors for the four choices of $\varphi(u) = \int_{\mathscr{D}} f(u(\mathbf{x}))\,\mathrm{d}\mathbf{x}$ are shown as functions of the mesh size h in a log-log scale. The corresponding observed convergence rates are shown in Table 2

Table 2 Observed (resp. theoretical) rates of convergence for the weak errors shown in Fig. 1

		β			
	$f(u)$	0.6	0.7	0.8	0.9
$d = 1$	$\|u\|^2$	1.396 (1.4)	1.748 (1.8)	1.945 (2)	1.994 (2)
	$\|u\|^3$	1.397 (1.4)	1.753 (1.8)	1.949 (2)	1.995 (2)
	$\|u\|^4$	1.398 (1.4)	1.754 (1.8)	1.951 (2)	1.996 (2)
	$\Phi(20(u - 0.5))$	1.398 (1.4)	1.755 (1.8)	1.952 (2)	1.996 (2)
$d = 2$	$\|u\|^2$	0.483 (0.4)	0.800 (0.8)	1.139 (1.2)	1.442 (1.6)
	$\|u\|^3$	0.442 (0.4)	0.783 (0.8)	1.145 (1.2)	1.465 (1.6)
	$\|u\|^4$	0.409 (0.4)	0.768 (0.8)	1.143 (1.2)	1.472 (1.6)
	$\Phi(20(u - 0.5))$	0.512 (0.4)	0.782 (0.8)	1.135 (1.2)	1.458 (1.6)

fields with general smoothness proposed in [4]. The method considers the fractional order equation (2.1) and is based on combining a finite element discretization in space with the quadrature approximation (2.4) of the inverse fractional power operator. This yields an approximate solution $u_{h,k}^Q$ of the SPDE, which in [4] was shown to converge to the solution u of (2.1) in the strong mean-square sense with rate (2.7).

In many applications one is mostly interested in a certain quantity of the random field u which can be expressed by $\varphi(u)$ for some real-valued function φ. For this reason, the focus of the present work has been the weak error $|\mathbb{E}[\varphi(u)] - \mathbb{E}[\varphi(u_{h,k}^Q)]|$. The main outcome of this article, Theorem 2.1, shows convergence of this type of error to zero at an explicit rate for twice continuously Fréchet differentiable functions φ, which have a second derivative of polynomial growth. Notably, the component of the convergence rate stemming from the stochasticity of the problem is doubled compared to the strong convergence rate (2.7) derived in [4]. For proving this result, we have performed a rigorous error analysis in Sect. 3, which is based on an extension of the Eq. (2.1) to a time-dependent problem as well as an associated Kolmogorov backward equation and Itô calculus.

In order to validate the theoretical findings, we have performed a simulation study for the stochastic model problem (1.1) on the domain $\mathscr{D} = (0, 1)^d$ for $d = 1, 2$ in Sect. 4. This model is highly relevant for applications in spatial statistics, since it is often used to approximate Gaussian Matérn fields. We have considered four different functions φ and the fractional orders $\beta = 0.6, 0.7, 0.8, 0.9$. The observed empirical weak convergence rates can be seen to verify the theoretical results. One of the considered functions φ is based on a transformation of the random field by a Gaussian cumulative distribution function. Quantities of this form are particularly important for applications to porous materials, as they are used to model the pore volume fraction of the material, see, e.g., [1]. Thus, we see ample possibilities for applying the outcomes of this work to problems in spatial statistics and related disciplines.

Acknowledgements The authors thank Stig Larsson for valuable comments on the manuscript and an anonymous referee who helped to improve the presentation of the results.

References

1. Barman, S., Bolin, D.: A three-dimensional statistical model for imaged microstructures of porous polymer films. J. Microsc. **269**, 247–258 (2018)
2. Bolin, D.: Spatial Matérn fields driven by non-Gaussian noise. Scand. J. Stat. **41**, 557–579 (2014)
3. Bolin, D., Kirchner, K.: The rational SPDE approach for Gaussian random fields with general smoothness. arXiv preprint, arXiv:1711.04333v2 (2018)
4. Bolin, D., Kirchner, K., Kovács, M.: Numerical solution of fractional elliptic stochastic PDEs with spatial white noise. arXiv preprint, arXiv:1705.06565v2 (2018)
5. Bolin, D., Lindgren, F.: Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. Ann. Appl. Stat. **5**, 523–550 (2011)
6. Bonito, A., Pasciak, J.E.: Numerical approximation of fractional powers of elliptic operators. Math. Comput. **84**, 2083–2110 (2015)
7. Brzeźniak, Z.: Some Remarks on Itô and Stratonovich Integration in 2-Smooth Banach Spaces, in Probabilistic Methods in Fluids, pp. 48–69. World Science Publisher, River Edge (2003)
8. Courant, R., Hilbert, D.: Methods of Mathematical Physics: Vol. I. Interscience Publishers, Inc, New York (1953)
9. Da Prato, G., Zabczyk, J.: Second Order Partial Differential Equations in Hilbert Spaces London Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge (2002)
10. Fuglstad, G.-A., Lindgren, F., Simpson, D., Rue, H.: Exploring a new class of non-stationary spatial Gaussian random fields with varying local anisotropy. Stat. Sin. **25**, 115–133 (2015)
11. Kovács, M., Printems, J.: Weak convergence of a fully discrete approximation of a linear stochastic evolution equation with a positive-type memory term. J. Math. Anal. Appl. **413**, 939–952 (2014)
12. Lindgren, F., Rue, H., Lindström, J.: An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion). J. R. Stat. Soc. Ser. B Stat. Methodol. **73**, 423–498 (2011)
13. Matérn, B.: Spatial variation, Meddelanden från statens skogsforskningsinstitut, vol. 49 (1960)
14. Pazy, A.: Semigroups of Linear Operators and Applications to Partial Differential Equations. Applied Mathematical Sciences. Springer, Berlin (1983)
15. Peszat, S., Zabczyk, J.: Stochastic Partial Differential Equations with Lévy Noise. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge (2007)
16. Rue, H., Held, L.: Gaussian Markov Random Fields: Theory and Applications. Chapman & Hall/CRC Monographs on Statistics and Applied Probability. CRC Press, Boca Raton (2005)
17. Stein, M.L.: Interpolation of Spatial Data: Some Theory for Kriging. Springer Series in Statistics. Springer, New York (1999)
18. Strang, G., Fix, G.: An Analysis of the Finite Element Method. Wellesley-Cambridge Press, Wellesley (2008)
19. Thomée, V.: Galerkin Finite Element Methods for Parabolic Problems. Springer Series in Computational Mathematics. Springer, Berlin (2006)
20. Wallin, J., Bolin, D.: Geostatistical modelling using non-Gaussian Matérn fields. Scand. J. Stat. **42**, 872–890 (2015)
21. Whittle, P.: On stationary processes in the plane. Biometrika **41**, 434–449 (1954)
22. Whittle, P.: Stochastic processes in several dimensions. Bull. Int. Stat. Inst. **40**, 974–994 (1963)