Working Paper

Sparse finite elements for elliptic problems with stochastic data

Author(s):
Schwab, Christoph; Todor, R.-A.

Publication Date:
2002

Permanent Link:

https://doi.org/10.3929/ethz-a-004339381 →

Sparse Finite Elements for Elliptic Problems with Stochastic Data

C. Schwab and R.-A. Todor

Research Report No. 2002-05
April 2002

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Sparse Finite Elements for Elliptic Problems with Stochastic Data

C. Schwab and R.-A. Todor

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Research Report No. 2002-05 April 2002

Abstract

We formulate elliptic boundary value problems with stochastic input data in a domain D. We show well-posedness of the problem in stochastic Sobolev spaces. We derive a deterministic elliptic PDE in $D \times D$ for the spatial correlations, respectively the covariance, of the solution. We show well-posedness and regularity results for this PDE in a scale of weighted Sobolev spaces with mixed highest order derivatives. We prove that discretization with sparse tensor products of hierarchic FE-spaces in D yields optimal asymptotic rates of convergence for the second moments even in the presence of singularities or for spatially completely uncorrelated data. Multilevel preconditioning in $D \times D$ by means of sparse wavelet basis and iterative solution of the finite element equations for the correlation functions is shown to lead to an algorithm of logarithmic-linear complexity. Numerical experiments confirm the theory.

1 Problem formulation

Due to the rapid development of scientific computing in recent years, accurate numerical solution of boundary value problems for partial differential equations is now possible in many applications. For given problem data, such as domains, coefficients and boundary data, the solution can be computed to high accuracy. Often, however, the problem data is either incompletely known or uncertain which implies that highly accurate numerical solutions are of limited use. One way to deal with such uncertainty is to describe the problem data as random fields which turns the problem into a stochastic differential equation. The formulation and numerical solution of stochastic differential equations has received increasing interest in recent years. We mention here only [16], [17], [14] and the references there on stochastic ordinary differential equations and [11], [12], [7] on stochastic partial differential equations. In engineering simulations, uncertainty in coefficients and loadings has been dealt with by means of the stochastic finite element method in structural mechanics (see [13] and the references there) and by the related first order, second moment perturbation technique in subsurface flow models.

The solution of a stochastic differential equation is, in general, a random field which takes values in a suitable function space. Complete description of this random field requires knowledge of its joint probability densities. In applications, however, one is often only interested in the first moments of the random solution. These moments can be computed e.g. by the Monte-Carlo (MC) Method, where numerous 'samples' of the random input data are generated according to prescribed, often empirical, distributions and each MC sample entails the solution of a deterministic boundary value problem. From the computed solutions, the mean and covariance then give estimates for the first moments of the random solution. This approach is costly – due to the generally slow convergence of MC methods, numerous samples must be taken until a satisfactory accuracy of the computed solution has been reached. Nevertheless, in the context of stochastic ordinary differential equations, this technique is frequently employed (e.g. [14]) with good success. For partial differential equations, one could discretize in the spatial variables first, e.g. by the Finite Element Method (FEM). This will then lead to large linear systems with random stiffness and mass matrices, the so-called stochastic FEM [13]. The cost of this approach is often prohibitive, particularly in 3-d.

Alternatively one can directly compute the moments of interest for the random solution and this is the approach which we follow here. This approach consists in deriving deterministic partial differential equations for the moments of the random solution, thereby eliminating the need for MC simulations. This advantage is bought, however, at the expense of a high dimensionality in the deterministic problem for the moments: if the differential equation is posed in the physical domain $D \subset \mathbb{R}^d$, the k-th moment of the solution is a function of k variables in $D^k \subset \mathbb{R}^{kd}$. We show in the present paper for elliptic partial differential equations with stochastic input data that the deterministic equations for the moments have a very special structure. We exploit this structure for anisotropic regularity estimates which in turn show that finite element approximations of the moments of the solution in D^k can be computed in essentially the same complexity as FE-solutions of the deterministic problem in D.

We now specify the problems to be considered. Let (Ω, Σ, P) be a $\sigma -$ finite probability space and $D \subset \mathbb{R}^d$ a bounded open set with Lipschitz boundary ∂D. Let $A \in L^\infty(D, \mathbb{R}^{d\times d}_{sym})$ satisfy

$$\exists\, \alpha, \beta > 0 \text{ s.t. } \alpha\|\xi\|^2 \leq \xi^\top A(x)\xi \leq \beta\|\xi\|^2 \quad \forall \xi \in \mathbb{R}^d \text{ and } \lambda - \text{a.e. } x \in D. \tag{1.1}$$

We define a random field on a submanifold M of $\mathbb{R}^d$ (it will be always D or some part of its boundary) as a jointly measurable function from $M \times \Omega$ to $\mathbb{C}$. Suppose $\partial D = \Gamma_0 \cup \Gamma_1$ is a disjoint union of subsets, where Γ_0 has positive surface measure and let f, g and h be random

fields on D, Γ_0 and Γ_1 respectively. We consider the following model problem:

$$Pu(x,\omega) := \left\{ \begin{array}{c} L(\partial_x)u \\ \gamma_0(u) \\ \gamma_n(u) \end{array} \right\} = \left\{ \begin{array}{r} -\mathrm{div}(A(x)\nabla u(x,\omega)) \\ u(x,\omega)\mid_{\Gamma_0} \\ n^\top A(x)\nabla u(x,\omega)\mid_{\Gamma_1} \end{array} \right\} = \left\{ \begin{array}{ll} f(x,\omega) & \text{in } D \\ g(x,\omega) & \text{on } \Gamma_0 \\ h(x,\omega) & \text{on } \Gamma_1 \end{array} \right\}, \qquad (1.2)$$

where the operators involved in the boundary conditions should be thought of as stochastic counterparts of the classical trace on Γ_0 or Γ_1 and distributional conormal derivative operators, γ_0, γ_1 and γ_n respectively. In Section 2, we first introduce appropriate function spaces of data (f,g,h) and solutions u in such a way that (1.2) becomes a well-posed problem. In Section 3 we define the 'statistics of u', and we derive deterministic partial differential equations which describe them. The 'statistics of u' that we are interested in, are the moments of first and second order of the random solution $u(x,\omega)$ to (1.2), sometimes referred to as the expectation and the covariance of $u(x,\omega)$, and defined by

$$E_u(x) := \int_\Omega u(x,\omega)dP(\omega) \quad \text{and} \quad C_u(x,y) := \int_\Omega u(x,\omega)\overline{u(y,\omega)}dP(\omega),$$

respectively, whenever these quantities exist. Section 4 addresses the regularity of the covariance equation, in particular also in polygonal domains. Section 5 discusses the finite element approximation of the covariance equation. We describe a particular FE-space which allows to achieve, in terms of the number of degrees of freedom, essentially the same convergence rates as for the problem in domain D. Section 6 addresses the preconditioning and the iterative solution of the linear system of equations for the covariance problem. Also, a detailed complexity analysis is given and logarithmic linear complexity of the algorithm is established. Finally, Section 7 presents numerical experiments which confirm the theoretical estimates.

2 Preliminaries

2.1 Stochastic Sobolev spaces

The most appropriate tools for the study of (1.2) are the stochastic Sobolev spaces, which we shall define as tensor products of usual function spaces. Within the setting of the previous section, we consider $L^2(\Omega, dP)$, the Hilbert space of all random variables on Ω with finite second-order moments, equipped with the usual inner product

$$\langle u, v\rangle := \int_\Omega u(\omega)\overline{v(\omega)}dP(\omega).$$

Our convention will be that whenever H is a Sobolev space, $\mathcal{H}$ will denote its stochastic counterpart, that is, the tensor product of H and $L^2(\Omega, dP)$. For instance, we shall employ the following notations

$$\mathcal{H}^k(D) := H^k(D) \otimes L^2(\Omega, dP), \;\; \mathcal{L}^2(D)^d := L^2(D)^d \otimes L^2(\Omega, dP)$$
$$\mathcal{H}^{1/2}(\Gamma_0) := H^{1/2}(\Gamma_0) \otimes L^2(\Omega, dP), \;\; \mathcal{H}^{-1/2}(\Gamma_1) := H^{-1/2}(\Gamma_1) \otimes L^2(\Omega, dP) \text{ etc.}$$

The stochastic Sobolev spaces introduced above are then equipped with natural Hilbert structures induced from the tensor product factors. Embedding and trace theorems similar to the usual ones hold also true on account of the fact that bounded linear operators between Hilbert spaces can be tensorised. We shall use the following notations for deterministic operators involved and their stochastic counterparts. Note that $\mathcal{B}$ stands for the space of bounded linear operators between two Hilbert spaces.

$$\nabla \in \mathcal{B}(H^1(D), L^2(D)^d),\ \nabla \otimes \mathrm{Id} \in \mathcal{B}(\mathcal{H}^1(D), \mathcal{L}^2(D)^d)$$
$$-\mathrm{div}\ \in \mathcal{B}(L^2(D)^d, H^{-1}(D))\,,\ -\mathrm{div} \otimes \mathrm{Id} \in \mathcal{B}(\mathcal{L}^2(D)^d, \mathcal{H}^{-1}(D))$$
$$\gamma_j\ \in \mathcal{B}(H^1(D), H^{1/2}(\Gamma_j))\,,\ \gamma_j \otimes \mathrm{Id}\ \in \mathcal{B}(\mathcal{H}^1(D), \mathcal{H}^{1/2}(\Gamma_j)),\ j = 0, 1$$
$$H^1_{(0)}(D) := \mathrm{Ker}\,\gamma_0 = \{u \in H^1(D) \,|\, \gamma_0\, u = 0\}\,,\ \mathcal{H}^1_{(0)}(D) := \mathrm{Ker}\,\gamma_0 \otimes \mathrm{Id}$$
$$\mathcal{H}^{-1}(D) := (H^1_{(0)}(D) \otimes L^2(\Omega, dP))^{\,*} \simeq H^{-1}(D) \otimes L^2(\Omega, dP).$$

Here, as usual, we have employed the classical notation $\mathcal{K}^\star$ for the dual space of the Hilbert space $\mathcal{K}$ and we have identified $L^2(D)$ with its dual, via the Riesz isomorphism.
As for the coefficient $A(x)$, the condition (1.1) ensures that $A \in \mathcal{B}(L^2(D)^d)$, with lower bound α and this in turn implies $A \otimes \mathrm{Id} \in \mathcal{B}(\mathcal{L}^2(D)^d)$, as well as

$$\alpha \|u\|^2_{\mathcal{L}^2(D)^d} \leq \langle (A \otimes \mathrm{Id})u, u\rangle_{\mathcal{L}^2(D)^d} \quad \forall u \in \mathcal{L}^2(D)^d. \tag{2.1}$$

Regarding the stochastic Sobolev spaces, we shall also need the following result and its straightforward consequence :

Lemma 2.1 *If H and K are two Hilbert spaces, then the topology of $H \otimes K$ depends only on the topology, and not on the choice of the inner products, of H and K.*

Proof. Simple argument based upon the Open Mapping Theorem. ◇

Corollary 2.2 *$\|\cdot\|_1 := \|(\nabla \otimes \mathrm{Id})\cdot\|_{\mathcal{L}^2(D)^d}$ defines a norm on $\mathcal{H}^1_{(0)}(D)$, equivalent to the usual one.*

Proof. Take $H = H^1_{(0)}(D)$ and $K = L^2(\Omega, dP)$ in the previous Lemma and note that the inner product associated to $\|\cdot\|_1$ is exactly the tensor product of the one corresponding to $\|\nabla \cdot\|_{L^2(D)^d}$ (and which gives the usual topology on $H^1_{(0)}(D)$) and $\langle\cdot,\cdot\rangle_{L^2(\Omega,dP)}$. ◇

The following result justifies the terminology 'random fields' for the elements of the tensor-product spaces introduced above. We shall avoid Bochner Integrals (see [20]) whenever this is possible and work instead with tensor products.

Proposition 2.3 *We have the canonical isomorphisms*

$$H^k(D; L^2(\Omega, dP)) \simeq \mathcal{H}^k(D) := H^k(D) \otimes L^2(\Omega, dP) \simeq L^2(\Omega, dP; H^k(D)). \tag{2.2}$$

Since this result is standard, we simply recall here the definitions of the spaces and isomorphisms in (2.2) (see also [20]). If H is, say, a separable Hilbert space and (S, Υ, m) a measure space, then

$$L^2(S, dm; H) := \{f : S \to H \,|\, f \text{ is strongly measurable and } \int_S \|f(x)\|^2_H < \infty\}.$$

If $S = D$, Υ is the family of Borel sets in D and m is the Lebesgue measure, then

$$H^k(D; H) := \{f \in L^2(S, dm; H) \mid \forall\, |\alpha| \leq k\ \exists\, f_\alpha \in L^2(S, dm; H) \text{ s.t.}$$
$$\forall \phi \in C^\infty_0(D; H)\,, \int_D \langle f(x), \partial^\alpha \phi(x)\rangle_H\, dx = (-1)^{|\alpha|} \int_D \langle f_\alpha(x), \phi(x)\rangle_H\, dx\}. \tag{2.3}$$

The functions f_α are called the generalized derivatives of f, they are uniquely defined by (2.3) and $H^k(D; H)$ has a natural Hilbert structure:

$$\langle f, g\rangle_{H^k(D;H)} := \sum_{|\alpha| \leq k} \int_D \langle f_\alpha(x), g_\alpha(x)\rangle_{L^2(\Omega,dP)}\, dx.$$

Spaces on the left and right of (2.2) are then obtained by choosing $H = L^2(\Omega, dP)$ and $H = H^k(D)$ in (2.3), respectively.

2.2 Random solutions of the boundary value problem

Now we can give a mathematical formulation of our problem (1.2).

Proposition 2.4 *Assume $f \in \mathcal{H}^{-1}(D)$, $g \in \mathcal{H}^{1/2}(\Gamma_0)$ and $h \in \mathcal{H}^{-1/2}(\Gamma_1)$. Then there exists a unique random solution $u \in \mathcal{H}^1(D)$ such that $(\gamma_0 \otimes \mathrm{Id})u = g$ and:*

$$\langle (A \otimes \mathrm{Id})(\nabla \otimes \mathrm{Id})u, (\nabla \otimes \mathrm{Id})v\rangle_{\mathcal{L}^2(D)^d} = \langle f, v\rangle_{\mathcal{H}^{-1}(D),\mathcal{H}^1_{(0)}(D)} + \langle h, (\gamma_1 \otimes \mathrm{Id})v\rangle_{\mathcal{H}^{-1/2}(\Gamma_1),\mathcal{H}^{1/2}(\Gamma_1)} \tag{2.4}$$

for all $v \in \mathcal{H}^1_{(0)}(D)$.

Proof. Since $H^1(D)/H^1_{(0)}(D) \simeq H^{1/2}(\Gamma_0)$ as topological spaces, there exists $u_1 \in \mathcal{H}^1(D)$ such that $(\gamma_0 \otimes \mathrm{Id})(u_1) = g$ and our problem reduces to the existence and uniqueness of $u_0 \in \mathcal{H}^1_{(0)}(D)$ satisfying:

$$\mathcal{A}(u_0, v) := \langle (A \otimes \mathrm{Id})(\nabla \otimes \mathrm{Id})u_0, (\nabla \otimes \mathrm{Id})v\rangle_{\mathcal{L}^2(D)^d} = -\langle (A \otimes \mathrm{Id})(\nabla \otimes \mathrm{Id})u_1, (\nabla \otimes \mathrm{Id})v\rangle_{\mathcal{L}^2(D)^d}$$

$$+\langle f, v\rangle_{\mathcal{H}^{-1}(D),\mathcal{H}^1_{(0)}(D)} + \langle h, (\gamma_1 \otimes \mathrm{Id})v\rangle_{\mathcal{H}^{-1/2}(\Gamma_1),\mathcal{H}^{1/2}(\Gamma_1)} \quad \forall v \in \mathcal{H}^1_{(0)}(D). \tag{2.5}$$

And this is a simple consequence of Lax-Milgram Lemma in $\mathcal{H}^1_{(0)}(D)$, as soon as we note that, on account of Corollary 2.2 and (2.1), the sesquilinear form $\mathcal{A}$ defined by the l.h.s. of (2.5) is bounded and coercive on $\mathcal{H}^1_{(0)}(D)$, while the r.h.s. is a continuous antilinear functional on the same space. ◇

Remark 2.5 If we choose $(e)_{i\geq 1}$ to be an ONB in $L^2(\Omega; dP)$ and if we expand $f = \sum_i f_i \otimes e_i$ with $\sum_i \|f_i\|^2_{L^2(D)} \leq \infty$, as well as g and h accordingly, then the solution of (1.2) can be written as a series $u = \sum_i u_i \otimes e_i$ which converges absolutely in $\mathcal{H}^1(D)$ and whose coefficient functions u_i solve the deterministic mixed boundary value problem:

$$Pu_i = \left\{ \begin{array}{c} L(\partial_x)u_i \\ \gamma_0(u_i) \\ \gamma_n(u_i) \end{array} \right\} = \left\{ \begin{array}{l} f_i \text{ in } D, \\ g_i \text{ on } \Gamma_0, \\ h_i \text{ on } \Gamma_1. \end{array} \right\}. \tag{2.6}$$

This can be seen by choosing the test function in (2.4) of the form $v = w \otimes e_i$, with $w \in H^1_{(0)}(D)$. Note that the deterministic character of A is essential in this decomposition.

3 Statistics of u

In this section our interest will be focused on finding deterministic equations for the expectation and the covariance of the random solution u respectively. While the expectation $E_u(x)$ of the random solution $u(x, \omega)$ at $x \in D$ is obviously of interest, the covariance (or spatial correlation) $C_u(x, x')$ allows to obtain the variance of the random solution $u(x, \omega)$ at $x \in D$ via

$$\mathrm{Var}^2(u(x, \cdot)) = (E_u(x))^2 - (C_u(x, x))^2.$$

3.1 Second order moments

We shall first give the definition of the covariance of a pair (u, v) when $u, v \in \mathcal{H}^1(D)$ and we shall introduce the expectation of u as the covariance of the pair $(u, 1)$ where $1 \in \mathcal{H}^1(D)$ is the tensor product of constant functions equal to 1 on D and Ω respectively.

Proposition 3.1 *Let u and v be elements of $\mathcal{H}^1(D)$ and let $(e_i)_{i\geq 1}$ be an ONB in $L^2(\Omega; dP)$, so that $u = \sum_i u_i \otimes e_i$, where $u_i \in H^1(D)\ \forall i \geq 1$ and $\sum_i \|u_i\|^2_{H^1(D)} < \infty$. We define v_i similarly. Then $\sum_i u_i \otimes \bar{v}_i$ converges in $H^1(D) \otimes H^1(D)$ and the limit does not depend on the choice of the basis $(e_i)_{i\geq 1}$.*

Proof. The convergence of the series $C_{u,v} := \sum_i u_i \otimes \bar{v_i}$ follows from the obvious inequality:

$$\|\sum_{i=k}^{n} u_i \otimes \bar{v_i}\|_{H^1(D)\otimes H^1(D)} \leq \sum_{i=k}^{n} \|u_i\|_{H^1(D)} \|\bar{v_i}\|_{H^1(D)} \leq (\sum_{i=k}^{n} \|u_i\|^2_{H^1(D)})^{1/2} (\sum_{i=k}^{n} \|v_i\|^2_{H^1(D)})^{1/2}. \tag{3.1}$$

To prove the second claim, we identify the tensor-product spaces above with some concrete function spaces, taking into account the first isomorphism in (2.2) and that

$$L^2(D) \otimes L^2(D) \overset{can}{\simeq} L^2(D \times D). \tag{3.2}$$

Recall that the latter isometric isomorphism is the unique linear and continuous extension of the function which maps $f \otimes g$ onto the function defined a.e. on $D \times D$ by $f(x)g(y)$, where $x, y \in D$. As from every L^2-convergent sequence we can extract an a.e. convergent subsequence, there exists a strictly increasing sequence of positive integers $(i_k)_{k\geq 1}$ such that

$$\sum_{i=1}^{i_k} u_i(x)\overline{v_i(y)} \to C_{u,v}(x,y) \quad \text{a.e.}\,(x,y) \in D \times D, \tag{3.3}$$

as $k \to \infty$. Because $\sum_{i=1}^{i_k} u_i \otimes e_i$ converges to u also in $\mathcal{L}^2(D)$ when $k \to \infty$, it follows, passing again to some subsequence if necessary, that

$$(\sum_{i=1}^{i_k} u_i \otimes e_i)(x) \to u(x) \text{ in } L^2(\Omega; dP) \quad \text{a.e. } x \in D.$$

(the result invoked above holds for Hilbert-space-valued functions, too, see [20]). We can also assume, extracting a further subsequence of $(i_k)_{k\geq 1}$, that $(\sum_{i=1}^{i_k} v_i \otimes e_i)(x)$ converges in $L^2(\Omega; dP)$ to $v(x)$ a.e. $x \in D$ too, and this implies:

$$\sum_{i=1}^{i_k} u_i(x)\overline{v_i(y)} = \langle \sum_{i=1}^{i_k} u_i(x) \otimes e_i\,, \sum_{i=1}^{i_k} v_i(y) \otimes e_i \rangle_{L^2(\Omega;dP)} \to \langle u(x), v(y) \rangle_{L^2(\Omega;dP)}, \tag{3.4}$$

as $k \to \infty$, a.e. $(x,y) \in D \times D$. From (3.3) and (3.4) we conclude that:

$$C_{u,v}(x,y) = \langle u(x), v(y) \rangle_{L^2(\Omega;dP)} \quad \text{a.e.}\,(x,y) \in D \times D. \tag{3.5}$$

Note that in the previous argument we identified $\sum_{i=1}^{i_k} u_i \otimes e_i\,, \sum_{i=1}^{i_k} v_i \otimes e_i\,, u$, v and $C_{u,v}$ with their images via (2.2) and (3.2), respectively. ⋄

The previous result justifies the following

Definition 3.2 *If u and v are elements of $\mathcal{H}^1(D)$, then the series $C_{u,v}$ defined in Proposition* 3.1 *is called the* covariance of the pair (u,v). *If $u = v$ we write C_u instead of $C_{u,u}$ and speak about the* covariance of u.

Later we shall also need an extension of this definition:

Remark 3.3 *From the proof it follows also that if H, H_1, H_2 are separable Hilbert spaces, and $u \in H_1 \otimes H, v \in H_2 \otimes H$, then we can define in a similar way the covariance $C_{u,v}$ as an element of $H_1 \otimes H_2$. The examples we have in mind are, first, $H_1 = H^{-1}(D)$ and $H_2 = L^2(S; dm)$ where (S, Υ, m) is a $\sigma-$ finite measure space and second, $H_1 = H_2 = H^{-1}(D)$. These particular choices enable us to construct, for instance, the covariances of the pairs (f,h) and (f,f) with f and h as in Proposition 2.4. Note also that we shall assume in the following that h is more regular, namely $h \in \mathcal{L}^2(\Gamma_1)$, which will allow us to use the Hilbert structure of $\mathcal{L}^2(\Gamma_1)$.*

3.2 Equation for E_u

We define the expectation of u by $E_u := C_{u,1} = \int_\Omega u(x,\omega)dP(\omega)$. The expectation E_u of the random solution $u(x,\omega)$ satisfies a deterministic boundary value problem which is easily derived. We choose $(e_i)_{i\geq 1}$ an ONB in $L^2(\Omega; dP)$ with $e_1 = 1$ (the constant function equal to 1 on Ω), so that $E_u = u_1$ is the unique solution of a mixed boundary value problem with data $f_1 = C_{f,1} =: E_f,\ g_1 = C_{g,1} =: E_g,\ h_1 = C_{h,1} =: E_h$, as follows from Remark 2.5

$$P(E_u) = \left\{ \begin{array}{c} L(\partial_x)E_u \\ \gamma_0(E_u) \\ \gamma_n(E_u) \end{array} \right\} = \left\{ \begin{array}{l} E_f \text{ in } D, \\ E_g \text{ on } \Gamma_0, \\ E_h \text{ on } \Gamma_1. \end{array} \right\}. \tag{3.6}$$

For future reference, we recall here also the variational formulation:
Find $E_u \in \{E_g\} + H^1_{(0)}(D)$ such that

$$q(E_u, v) = l(v) \quad \forall v \in H^1_{(0)}(D), \tag{3.7}$$

where

$$q(u,v) := \langle A\nabla u, \nabla v\rangle_{L^2(D)^d}, \tag{3.8a}$$
$$l(v) := \langle E_f, v\rangle_{H^{-1}(D),H^1_{(0)}(D)} + \langle E_h, v\rangle_{L^2(\Gamma_1)}. \tag{3.8b}$$

3.3 Equation for C_u

To give a weak deterministic equation for the covariance function, let us introduce, following [1], the anisotropic Sobolev spaces on $D \times D$ by

$$H^{k,l}(D\times D) := H^k(D)\otimes H^l(D)\ ,\ H^{k,l}_{(0)}(D\times D) := H^k_{(0)}(D)\otimes H^l_{(0)}(D), \tag{3.9}$$

for all integers $k, l \geq 1$. We define also

$$L^2(D\times D)^{d\times d} := L^2(D)^d \otimes L^2(D)^d.$$

Due to (3.2), this definition is unambiguous. We let also the following operators act on these spaces

$$\begin{gathered} \nabla_{x,y} := \nabla_x \otimes \nabla_y \in \mathcal{B}(H^{1,1}(D\times D), L^2(D\times D)^{d\times d}) \\ \gamma_{j,x,y} := \gamma_{j.x} \otimes \gamma_{j,y} \in \mathcal{B}(H^{1,1}(D\times D), L^2(\Gamma_j \times \Gamma_j)) \text{ for } j = 0, 1. \\ A_{x,y} := A_x \otimes A_y \in \mathcal{B}(L^2(D\times D)^{d\times d}). \end{gathered}$$

Next we prove that the covariance of u given by (2.4) satisfies a fourth-order elliptic equation in $D \times D$ and that the sesquilinear form involved is coercive on the appropriate space (see also [6]). As it can be easily seen, if u solves (2.4), then C_u satifies also the following boundary conditions

$$(\gamma_0 \otimes \mathrm{Id})\, C_u = C_{g,u} \text{ and } (\mathrm{Id} \otimes \gamma_0)\, C_u = C_{u,g}. \tag{3.10}$$

on $\Gamma_0 \times D$ and $D \times \Gamma_0$ respectively. Since we are primarily interested in approximating C_u in terms of the statistics of the data, and not in terms of u, we shall assume $g = 0$. In view of the fact that the trace operator of $D \times D$ on $(\Gamma_0 \times \bar{D}) \cup (\bar{D} \times \Gamma_0)$ is $(\gamma_0 \otimes \mathrm{Id}) \oplus (\mathrm{Id} \otimes \gamma_0)$, (3.10) means, in the case $g = 0$, that $C_u \in H^{1,1}_{(0)}(D \times D)$. We can actually be more specific:

Proposition 3.4 *Assume that u is the solution of* (2.4) *with $h \in \mathcal{L}^2(\Gamma_1)$ and $g = 0$. Then the correlation C_u of the random solution $u(x,\omega)$ is the unique solution in $H^{1,1}_{(0)}(D \times D)$ of*

$$C_u \in H^{1,1}_{(0)}(D \times D): \quad \mathcal{Q}(C_u, C_v) = \mathcal{L}(C_v) \quad \forall C_v \in H^{1,1}_{(0)}(D \times D), \tag{3.11}$$

where

$$\begin{aligned} \mathcal{Q}(C_u, C_v) &:= \langle A_{x,y} \nabla_{x,y} C_u, \nabla_{x,y} C_v \rangle_{L^2(D\times D)^{d\times d}}, && \text{(3.12a)} \\ \mathcal{L}(C_v) &:= \langle C_f, C_v \rangle_{H^{-1}(D)\otimes H^{-1}(D), H^{1,1}_{(0)}(D\times D)} \\ &\quad + \langle C_{h,f}, (\gamma_1 \otimes \mathrm{Id}) C_v \rangle_{L^2(\Gamma_1)\otimes H^{-1}(D), L^2(\Gamma_1)\otimes H^1_{(0)}(D)} && \text{(3.12b)} \\ &\quad + \langle C_{f,h}, (\mathrm{Id} \otimes \gamma_1) C_v \rangle_{H^{-1}(D)\otimes L^2(\Gamma_1), H^1_{(0)}(D)\otimes L^2(\Gamma_1)} \\ &\quad + \langle C_h, (\gamma_1 \otimes \gamma_1) C_v \rangle_{L^2(\Gamma_1\times\Gamma_1)}. \end{aligned}$$

Proof. Expand $C_v = \sum_i w_i \otimes v_i$ where $(v_i)_{i\geq 1}$ is an ONB in $H^1_{(0)}(D)$ and $(w_i)_{i\geq 1} \subset H^1_{(0)}(D)$ with $\sum_i \|w_i\|^2_{H^1_{(0)}(D)} < \infty$. Then $\mathcal{Q}(C_u, C_v)$ equals

$$\begin{aligned} &\langle \sum_{i,j} A_x \nabla_x u_i \otimes A_y \nabla_y \bar{u}_i, \nabla_x w_j \otimes \nabla_y v_j \rangle = \sum_{i,j} \langle A_x \nabla_x u_i, \nabla_x w_j \rangle_{L^2(D)} \langle A_y \nabla_y \bar{u}_i, \nabla_y v_j \rangle_{L^2(D)} \\ &= \langle \sum_i f_i \otimes \bar{f}_i, \sum_j w_j \otimes v_j \rangle_{-1,-1,1,1} + \langle \sum_i h_i \otimes \bar{f}_i, \sum_j \gamma_1 w_j \otimes v_j \rangle_{0,-1,0,1} \\ &+ \langle \sum_i f_i \otimes \bar{h}_i, \sum_j w_j \otimes \gamma_1 v_j \rangle_{-1,0,1,0} + \langle \sum_i h_i \otimes \bar{h}_i, \sum_j \gamma_1 w_j \otimes \gamma_1 v_j \rangle_{0,0,0,0} \\ &= \langle C_f, C_v \rangle_{H^{-1}(D)\otimes H^{-1}(D), H^{1,1}_{(0)}(D\times D)} + \langle C_{h,f}, (\gamma_1 \otimes \mathrm{Id}) C_v \rangle_{L^2(\Gamma_1)\otimes H^{-1}(D), L^2(\Gamma_1)\otimes H^1_{(0)}(D)} \\ &\quad + \langle C_{f,h}, (\mathrm{Id} \otimes \gamma_1) C_v \rangle_{H^{-1}(D)\otimes L^2(\Gamma_1), H^1_{(0)}(D)\otimes L^2(\Gamma_1)} + \langle C_h, (\gamma_1 \otimes \gamma_1) C_v \rangle_{L^2(\Gamma_1\times\Gamma_1)}, \end{aligned}$$

where all the series are absolutely convergent in appropriate spaces and the fourfold indices denote, in short, obvious duality products.
As (3.12b) defines a continuous antilinear functional on $H^{1,1}_{(0)}(D\times D)$, we have to check, in order to ensure the uniqueness of a solution for (3.11), only the boundedness and coercivity in the same space of the sesquilinear form (3.12a). To this end, we note first that $A_{x,y}$ is a bounded and strictly positive operator in $L^2(D\times D)^{d\times d}$, with lower and upper bounds α^2 and β^2. Second, using again Lemma 2.1 with $H = K = H^1_{(0)}(D)$, we see that $\|\cdot\|_1 := \|\nabla_{x,y} \cdot\|_{L^2(D\times D)^{d\times d}}$ is a norm on $H^{1,1}_{(0)}(D \times D)$, equivalent to the usual one, and this proves the claim. ⋄

Remark 3.5 *As it is readily seen, the superposition principle does not hold for* (3.11) *due to the non-linearity of the covariances $C_f, C_{f,h}, C_{h,f}$. If $C_{f,h}$ does not vanish identically, f and h are said to be* correlated.

We conclude this section with a description of the smallest closed linear subspace of $H^{1,1}_{(0)}(D\times D)$ which contains all the covariances C_u as u runs in $\mathcal{H}^1_{(0)}(D)$. We shall see that this subspace consists exactly of the functions which are hermitian in x and y. To this end, define $\Xi \in \mathcal{B}(L^2(D) \otimes L^2(D))$ by $\Xi(f \otimes g) = \bar{g} \otimes \bar{f}$. It is easy to check that Ξ is well-defined and that if we use the isomorphism (3.2), then Ξ sends the class of, say, $(x,y) \to h(x,y)$ to the class of $(x,y) \to \overline{h(y,x)}$. Denote by $\mathcal{F}$ the subspace of the fixed points of Ξ.

Proposition 3.6

$$\overline{Span}\,\{C_u \,|\, u \in \mathcal{H}^1_{(0)}(D)\} = H^{1,1}_{(0)}(D \times D) \cap \mathcal{F}. \tag{3.13}$$

Here the closure is taken in $H^{1,1}_{(0)}(D \times D)$.

Proof. As $\mathcal{F}$ is a closed subspace of $L^2(D) \otimes L^2(D)$, the r.h.s. of (3.13) is also closed in $H^{1,1}_{(0)}(D \times D) = H^1_{(0)}(D) \otimes H^1_{(0)}(D)$. Using the decomposition $C_u = \sum_i u_i \otimes \bar{u}_i$ and the fact that $u_i \otimes \bar{u}_i \in \mathcal{F}$, the inclusion $\subset$ follows at once. Let now $h \in H^1_{(0)}(D) \otimes H^1_{(0)}(D) \cap \mathcal{F}$. Expand as we have already done many times $h = \sum_i a_i \otimes b_i$, where $(b_i)_{i \geq 1}$ is an ONB in $H^1_{(0)}(D)$. As $h \in \mathcal{F}$, we can write equally $h = \sum_i (a_i \otimes b_i + \bar{b}_i \otimes \bar{a}_i)$. It is therefore enough to prove that $a_i \otimes b_i + \bar{b}_i \otimes \bar{a}_i$ is an element of the l.h.s. of (3.13). But this follows at once from the obvious identity $a_i \otimes b_i + \bar{b}_i \otimes \bar{a}_i = (a_i + \bar{b}_i) \otimes (b_i + \bar{a}_i) - a_i \otimes \bar{a}_i - \bar{b}_i \otimes b_i$, if we note that for all $c \in H^1_{(0)}(D)$, $c \otimes \bar{c} = C_{c \otimes e}$, with e an arbitrary unit vector in $L^2(\Omega; dP)$. $\diamond$

Remark 3.7 *As a consequence of the previous result, the closure of the space spanned by all covariances does not depend on the stochastic data space $L^2(\Omega; dP)$.*

4 Regularity

Here, we derive a regularity result for the weak solution of the covariance equation (3.11). As we shall see, this follows from the standard elliptic regularity, applied in a suitable fashion. We therefore recapitulate first this standard result for the elliptic problem (3.7).

4.1 Regularity of E_u

It is well-known that for elliptic boundary problems (3.7) holds a shift-theorem. We say that the following boundary value problem satisfies a shift-theorem at order $s \geq 0$ if

$$P(u) = \left\{ \begin{array}{c} L(\partial_x)u \\ \gamma_0(u) \\ \gamma_n(u) \end{array} \right\} = \left\{ \begin{array}{l} f \text{ in } D, \\ g \text{ on } \Gamma_0, \\ h \text{ on } \Gamma_1. \end{array} \right\}, \tag{4.1}$$

with $f \in H^{s-1}(D), g \in H^{s+1/2}(\Gamma_0), h \in H^{s-1/2}(\Gamma_1)$ implies $u \in H^{s+1}(D)$.

Proposition 4.1 *If problem* (4.1) *admits a shift theorem at order $s \geq 0$ and if $E_f \in H^{s-1}(D)$, $E_g \in H^{s+1/2}(\Gamma_0)$ and $E_h \in H^{s-1/2}(\Gamma_1)$, then $E_u \in H^{s+1}(D)$.*

Sufficient conditions for a shift theorem at order $s \geq 0$ are given e.g. in [8]:

Proposition 4.2 *Assume that $\partial D \in C^\infty$, $\bar{\Gamma}_0 \cap \bar{\Gamma}_1 = \emptyset$ and that the entries of A are of class $C^{s,1}(\bar{D})$ with $s > 0$. Then the problem* (4.1) *above admits a shift theorem at order s.*

If ∂D is not smooth, problem (4.1) admits a shift theorem at order s only for $0 \leq s < s^*$ with a small $s^* > 0$ (depending on the smoothness of ∂D and A). Nevertheless, in such situations we also have a shift theorem at order $s \geq s^*$ in weighted Sobolev spaces. We exemplify this in dimension $d = 2$. Let $D \subset \mathbb{R}^2$ be a bounded polygon with M vertices A_i, $i = 1, \ldots, M$ and straight sides Γ_i, $i = 1, \ldots, M$ connecting A_i and A_{i+1} (we set $A_{M+1} = A_1$). Denote by ω_i the size of the interior angle at vertex A_i. For $x \in D$, $r_i(x)$ is the distance from x to A_i and we associate with each A_i an exponent $\beta_i \in (0, 1)$. We write $\underline{\beta} = (\beta_1, \beta_2, \ldots, \beta_M)$ and, for $k \in \mathbb{Z}$, $\underline{\beta} + k := (\beta_1 + k, \beta_2 + k, \ldots, \beta_M + k)$. We define further the weight functions by

$$\omega_{\underline{\beta}+k}(x) := \prod_{i=1}^{M} (r_i(x))^{\beta_i + k}. \tag{4.2}$$

The weighted Sobolev spaces $H^{k,l}_\beta(D)$ are defined as closures of $C^\infty(\overline{D})$ under the norms

$$\|u\|^2_{H^{k,l}_\beta(D)} := \|u\|^2_{H^{l-1}(D)} + \sum_{|\underline{\alpha}|=l}^{k} \|\omega_{\underline{\beta}+|\alpha|-l} \mathrm{D}^\alpha u\|^2_{L^2(D)}, \tag{4.3}$$

if $k \geq l \geq 0$. Then it holds (see [2])

Theorem 4.3 *Assume that $D \subset \mathbb{R}^2$ is a polygon with M straight sides and that $A_{ij}(x) \in C^\infty(\overline{D})$. Assume further that the boundary data g, h in (4.1) admits liftings $G \in H^{s+1,2}_\beta(D), H \in H^{s,1}_\beta(D)$ for some $s \geq 0$. Then there exist numbers $\beta_i \in [0,1), i = 1, \ldots, M$ such that for any $k \in \mathbb{N}_0$ and $f \in H^{k,0}_\beta(D)$ the solution u of (4.1) belongs to $H^{k+2,2}_\beta(D)$. Moreover, denoting $s := k+1$, there holds a shift theorem at order s in weighted spaces:*

$$\|u\|_{H^{s+1,2}_\beta(D)} \leq C(s)\{\|f\|_{H^{s-1,0}_\beta(D)} + \|G\|_{H^{s+1,2}_\beta(D)} + \|H\|_{H^{s,1}_\beta(D)}\}. \tag{4.4}$$

4.2 Regularity of C_u

4.2.1 Smooth data

Here we assume, for convenience, that $g = 0$.

Proposition 4.4 *Suppose that (4.1) satisfies the shift theorem at order s. Then also for the covariance problem (3.11) holds the shift at order s, in the sense that if $C_f \in H^{s-1,s-1}(D \times D)$, $C_{f,h} \in H^{s-1,s-1/2}(D \times \Gamma_1), C_{h,f} \in H^{s-1/2,s-1}(\Gamma_1 \times D)$ and $C_h \in H^{s-1/2,s-1/2}(\Gamma_1 \times \Gamma_1)$, then $C_u \in H^{s+1,s+1}(D) \cap H^{1,1}_{(0)}(D \times D)$.*

Proof. Note first the following sharper version of the standard shift theorem at order s mentioned above, in the case ($g = 0$): The operator P^{-1} which associates with each element of $H^{s-1}(D) \oplus H^{s-1/2}(\Gamma_1)$ the corresponding solution of the problem (4.1) is a homeomorphism on $H^{s+1}(D) \cap H^1_{(0)}(D)$. We deduce that $P^{-1} \otimes P^{-1}$ is a homeomorphism from $H := (H^{s-1}(D) \otimes H^{s-1}(D)) \oplus (H^{s-1}(D) \otimes H^{s-1/2}(\Gamma_1)) \oplus (H^{s-1/2}(\Gamma_1) \otimes H^{s-1}(D)) \oplus (H^{s-1/2}(\Gamma_1) \otimes H^{s-1/2}(\Gamma_1))$ onto its range in $H^{s+1,s+1}(D) \cap H^{1,1}_{(0)}(D \times D)$. We still have to check that $P^{-1} \otimes P^{-1}$ sends the quadruple $(C_f, C_{f,h}, C_{h,f}, C_h)$ into the solution C_u of the corresponding problem (3.11). It is enough to prove this for $(f_1, h_1) \otimes (f_2, h_2)$, in view of the density of the span of such elements in H. To this end, we note that $(P^{-1} \otimes P^{-1})((f_1, h_1) \otimes (f_2, h_2)) = u_1 \otimes u_2$, where u_1 and u_2 solve the classical boundary value problem (2.6) with data $(f_1, 0, h_1)$ and $(f_2, 0, h_2)$ respectively. Upon multiplying the variational formulations of these two problems we obtain the desired conclusion. ◇

4.2.2 Corner singularities

We assume in what follows that $D \subset \mathbb{R}^2$ is a polygon with straight sides. Since solution singularities can only appear on a measure zero subset of ∂D (i.e. at vertices), a trace operator Tr on Γ_1 will be defined in the following as an $L^0(\Gamma_1)$-valued linear operator on $H^{s,1}_\beta(D)$, where by $L^0(\Gamma_1)$ we denote the space of measurable functions on Γ_1. Set, for all $n \in \mathbb{N}$,

$$D_n := D \setminus \bigcup_{i=1}^{M} B(A_i, \frac{1}{n}).$$

(here $B(a, r)$ stands for the ball of center a and radius r) and note that the restriction to D_n of an arbitrary $u \in H^{s,1}_\beta(D)$ is of class $H^s(D_n)$. Recall that $s = k + 1 \geq 1$, which implies the existence of the trace of $u|_{D_n}$ on ∂D_n. Obviously, the definitions agree almost everywhere on $\partial D_n \cap \partial D_m \cap \Gamma_1$ for all $n, m \in \mathbb{N}$, and it can be seen that the kernel of this trace operator is closed. To this end, it suffices to consider a Cauchy sequence u_k in the kernel of Tr, with limit u in $H^{s,1}_\beta(D)$, and to note that, due to the continuous imbedding $H^{s,1}_\beta(D) \hookrightarrow H^s(D_n)$ and to the boundedness of the usual trace operator in D_n, the trace of u on $\partial D_n \cap \Gamma_1$ vanishes for all n, which in turn implies that the trace of u in the sense explained above vanishes, too. This enables us to define further

$$H^{s-1/2,1/2}_\beta(\Gamma_1) := H^{s,1}_\beta(D) \,/\, \mathrm{Ker}(\mathrm{Tr}), \tag{4.5}$$

as a Banach-space, with the usual inf-norm. Passing in (4.4) to the infimum over all $H \in H^{s,1}_\beta(D)$ with the same trace h, we obtain that the operator which associates to each pair (f,h) the solution u of (4.1) with $g=0$ is a homeomorphism from $H^{s-1,0}_\beta(D) \otimes H^{s-\frac{1}{2},\frac{1}{2}}_\beta(\Gamma_1)$ to $H^{s+1,2}_\beta(D) \cap H^1_{(0)}(D)$. In view of the fact that a tensor product of linear homeomorphisms between Banach spaces is again a homeomorphism, we obtain, using the same argument as in Proposition 4.4, the following regularity result.

Proposition 4.5 *Assume that $D \subset \mathbb{R}^2$ is a polygon with straight sides and that the problem (4.1) admits a shift estimate (4.4) at order $s \geq 0$ in weighted spaces. Assume further that the data are sufficently regular, namely that for some positive $s \in \mathbb{R}$ holds*

$$C_{f,f} \in H^{s-1,0}_\beta(D) \otimes H^{s-1,0}_\beta(D),\ C_{f,h} \in H^{s-1,0}_\beta(D) \otimes H^{s-1/2,1/2}_\beta(\Gamma_1),$$

$$C_{h,f} \in H^{s-1/2,1/2}_\beta(\Gamma_1) \otimes H^{s-1,0}_\beta(D),\ C_{h,h} \in H^{s-1/2,1/2}_\beta(\Gamma_1) \otimes H^{s-1/2,1/2}_\beta(\Gamma_1).$$

Then the correlation function C_u of the random solution $u(x,\omega)$ satisfies

$$C_u \in H^{s+1,2}_\beta(D) \otimes H^{s+1,2}_\beta(D).$$

We conclude this section with two frequently used examples of spatial correlation functions.

4.2.3 Exponential covariance

We consider a second order process f with covariance function

$$C_f(x,y) = e^{-c|x-y|}, \quad (x,y) \in D \times D, \tag{4.6}$$

where $c > 0$ is a parameter and the domain $D \subset \mathbb{R}^d$ is smooth. Note that this covariance kernel can be used to characterize the well-known Markovian processes. (For various examples of processes that can be modelled as Markovian ones, we refer the reader to [19].) To deduce the regularity of C_f given by (4.6), we use the following two auxiliary results

Lemma 4.6 *Let $u : \mathbb{R}^d \to \mathbb{R}$ be defined by $u(x) = \exp(-|x|)$. Then it holds:*

$$u \in H^s(\mathbb{R}^d), \quad \forall s < d/2 + 1. \tag{4.7}$$

Proof. Recalling that $H^s(\mathbb{R}^d) = \{u \in L^2(\mathbb{R}^d) \,|\, (1+|\xi|^2)^{s/2}|\hat{u}(\xi)| \in L^2_\xi(\mathbb{R}^d)\}$, we note also that the Fourier transform of u can be explicitly computed, (see [18], page 50) and takes the form:

$$\hat{u}(\xi) = C(d) \cdot (1+|\xi|^2)^{-(d+1)/2}. \tag{4.8}$$

Hence $u \in H^s(\mathbb{R}^d)$ is equivalent to $(1+|\xi|^2)^{s/2-(d+1)/2} \in L^2_\xi(\mathbb{R}^d)$, that is $-4(s/2-(d+1)/2) > d$, or $s < d/2+1$, concluding the proof. ◇

Lemma 4.7 *If $s = p+q$ with $p,q \geq 0$ and $f \in H^s(\mathbb{R}^d)$, then the function $u : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined by*

$$u(x,y) := f(x-y) \quad \text{a.e. } (x,y) \in \mathbb{R}^d \times \mathbb{R}^d \tag{4.9}$$

belongs to $H^{p,q}_{\mathrm{loc}}(\mathbb{R}^d \times \mathbb{R}^d)$.

Proof. We have to show that if $\phi, \psi \in C^\infty_0(\mathbb{R}^d)$, the function $v : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined a.e. by $v(x,y) := \phi(x)\psi(y)f(x-y)$ belongs to $H^{p,q}(\mathbb{R}^d \times \mathbb{R}^d)$. We remark that it suffices to show that the function $w : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined a.e. by $w(x,y) := \psi(y)f(x-y)$ belongs to $H^{p,q}(\mathbb{R}^d \times \mathbb{R}^d)$, since the multiplication operator by ϕ is bounded in $H^p(\mathbb{R}^d)$ and tensorizing it by the identity of $H^q(\mathbb{R}^d)$ produces again a bounded operator, this time in $H^{p,q}(\mathbb{R}^d \times \mathbb{R}^d)$. In

view of the fact that the Fourier transform and tensor product commute, all we have to check is ($\langle\xi\rangle := (1+|\xi|^2)^{1/2}$):

$$\langle\xi\rangle^p\langle\eta\rangle^q\hat{w}(\xi,\eta) \in L^2_{\xi,\eta}(\mathbb{R}^d\times\mathbb{R}^d). \tag{4.10}$$

Explicit computation of the Fourier transform of w in terms of those of f and ψ shows that

$$\hat{w}(\xi,\eta) = \hat{\psi}(\xi+\eta)\hat{f}(\xi). \tag{4.11}$$

Using (4.11), (4.10) can be then written

$$\frac{\langle\xi\rangle^p\langle\eta\rangle^q}{\langle\xi+\eta\rangle^q\langle\xi\rangle^s}\ \cdot\langle\xi+\eta\rangle^q\hat{\psi}(\xi+\eta)\ \cdot\langle\xi\rangle^s\hat{f}(\xi) \in L^2_{\xi,\eta}(\mathbb{R}^d\times\mathbb{R}^d). \tag{4.12}$$

But this follows if we note that, by assumption on f, the last of the three terms in (4.12) belongs to $L^2_\xi(\mathbb{R}^d)$, the second one belongs to $L^2_\eta(\mathbb{R}^d)$, uniformly in ξ ($\hat{\psi}\in\mathcal{S}(\mathbb{R}^d)$), while the first is bounded uniformly in ξ and η, since $s=p+q$ and the inequality $\sqrt{2}\langle x\rangle\langle y\rangle > \langle x+y\rangle$ holds for all $x,y\in\mathbb{R}^d$. ◇

As a direct consequence of the previous two lemmas and the boundedness of the restriction operator from $H^s_{\rm loc}(\mathbb{R}^d)$ to $H^s(D)$ for all $s\geq 0, d\in\mathbb{N}^*$, the data we have chosen in (4.6) satisfy

$$C_f(x,y) = e^{-c|x-y|} \in H^{(d+2)/4-\epsilon,(d+2)/4-\epsilon}(D\times D) \text{ for arbitrary } \epsilon>0.$$

This regularity of the data enables us to apply Proposition 4.4, with $\Gamma_1=\emptyset$ and to deduce

Proposition 4.8 *If problem* (4.1) *admits a shift theorem at level $s\geq 0$ in $H^{s+1}(D)$ and if the covariance of the data f is $C_f(x,y)=e^{-c|x-y|}$ for some $c>0$, then the solution C_u of (3.11) with $g=0$ and $\Gamma_1=\emptyset$ belongs, for any $\epsilon>0$, to $H^{t,t}(D\times D)$, where*

$$t = \min((d+10)/4-\epsilon, s+1),$$

and

$$\|C_u\|_{H^{t,t}(D\times D)} \leq C(t,d)\|C_f\|_{H^{t-2,t-2}(D\times D)}. \tag{4.13}$$

This result shows that the regularity of C_u in a polygon $D\subset\mathbb{R}^2$, measured in $H^s(D)$, with $C_f=e^{-c|x-y|}$, is determined by corner singularities, since usually $s<2$.

4.2.4 Vanishing spatial correlation

Here we consider that D is a bounded Lipschitz domain in $\mathbb{R}^d$, with $d\leq 3$ and $\Gamma_1=\emptyset$ which ensures $H^{1,1}_{(0)}(D\times D) = H^{1,1}_0(D\times D)$. We denote further by Δ_D the diagonal set of $D\times D$, and we consider also an arbitrary function $k\in L^2(\Delta_D)$. We let then $k\cdot\delta(x-y)$ be the distribution defined by

$$\langle k\cdot\delta(x-y),\phi\rangle = \int_{\Delta_D} k(x)\phi(x,x)dx \quad \forall\ \phi\in C_0^\infty(D\times D). \tag{4.14}$$

Note that we can obtain the correlation kernel $\delta(x-y)$ as a limiting case of exponential-type covariances described in the previous subsection, as soon as we remark that

$$c^d\cdot\Big(\int_{\mathbb{R}^d} e^{-|z|}\,dz\Big)^{-1}\cdot e^{-c|x-y|} \longrightarrow \delta(x-y) \quad \text{as } c\to\infty, \quad \text{in } \mathcal{D}'(\mathbb{R}^d\times\mathbb{R}^d). \tag{4.15}$$

Recalling also the notations of the previous section, vanishing spatial correlations lead formally to the problem

$$\text{Find } C_u\in H^{1,1}_{(0)}(D\times D) \text{ s.t. } \mathcal{Q}(C_u,C_v) = \langle k\cdot\delta(x-y),C_v\rangle \quad \forall\, C_v\in H^{1,1}_{(0)}(D\times D) \tag{4.16}$$

We show now that it can be treated using the formalism introduced above, being of the same type as (3.11). Here the covariance kernel (4.14) means that the data is spatially uncorrelated. The solvability of (4.16) depends on the admissibility of the data (4.14). In view of the Cauchy-Schwarz inequality, and the density of $C_0^\infty(D\times D)$ in each anisotropic Sobolev space, it suffices therefore to prove

Lemma 4.9 *If D is a bounded Lipschitz domain in $\mathbb{R}^d$ with $d\leq 3$, the trace operator*

$$R: C_0^\infty(D\times D)\longrightarrow L^2(\Delta_D),\quad R(\phi)(x)=\phi(x,x)\quad \forall\, x\in D \tag{4.17}$$

has a unique linear continuous extension to $H^{1,1}_{(0)}(D\times D)$.

Proof. We consider a hypercube $\mathcal{C}=(-a,a)^d$ with a large enough, so that $\bar{D}\subset\mathcal{C}$. Supposing that the claim is true for the domain $\mathcal{C}$, the conclusion in the case of an arbitrary domain D is then reached by use of the following sequence of continuous embeddings and restrictions

$$H^{1,1}_{(0)}(D\times D)\subset H^{1,1}_{(0)}(\mathcal{C}\times\mathcal{C})\xrightarrow{R} L^2(\Delta_{\mathcal{C}})\xrightarrow{|_{\Delta_D}} L^2(\Delta_D).$$

It follows that it suffices to prove the assertion if D is a hypercube and, without loss of generality, we assume $D=(0,1)^d$.
We shall actually prove that for each pair of d-tuples

$$\alpha,\beta\in {\mathbb{R}_+^*}^d\quad \text{s.t. } \alpha_i+\beta_i>\frac{1}{2}\quad \forall 1\leq i\leq d, \tag{4.18}$$

R has a continuous extension to the closure of $C_0^\infty(D\times D)$ in

$$H^{\alpha,\beta}(D\times D):=\bigotimes_{i=1}^d H^{\alpha_i}((0,1))\otimes H^{\beta_i}((0,1)).$$

To this end, we remark first that the operator R defined by (4.17) in the case $d=1$ has a unique continuous extension to the closure of $C_0^\infty((0,1)\times(0,1))$ in $H^{1/2+\epsilon}((0,1))\otimes L^2((0,1))$, for all $\epsilon>0$. We expand an arbitrary element $u\in C_0^\infty((0,1)\times(0,1))$ as a convergent series in $L^2((0,1))\otimes L^2((0,1))$

$$u=\sum_{k,m\geq 1} u_{k,m}\cdot\sin(k\pi x)\cdot\sin(m\pi y), \tag{4.19}$$

where $(u_{k,m})_{k,m\geq 1}\in \ell^2_{\mathbb{N}^*\times\mathbb{N}^*}$.
Denoting by s the quantity $1/2+\epsilon$, the norm af u in the anisotropic space $H^s((0,1))\otimes L^2((0,1))$ equals

$$\|u\|_{s,0}:=\sum_{k,m\geq 1} k^{2s}\cdot u_{k,m}^2<\infty. \tag{4.20}$$

We have to check that the norm of the following series in $L^2((0,1))$

$$\sum_{k,m\geq 1} u_{k,m}\cdot\sin(k\pi x)\cdot\sin(m\pi x)$$

can be bounded in terms of the norm $\|u\|_{s,0}$ given by (4.20). To this end, let N be a positive integer. Using elementary computations, we can write the remainder after truncation of the series at N-th term as

$$\sum_{k,m\geq N} u_{k,m}\cdot\sin(k\pi x)\cdot\sin(m\pi x)=\frac{1}{2}\cdot\sum_{p\geq 0} v_p\cdot\ \cos(p\pi x), \tag{4.21}$$

where

$$v_p = \sum_{k\geq N+p} u_{k,k-p} + \sum_{k\geq N+p} u_{k-p,k} - \sum_{N\leq k\leq p-N} u_{k,p-k}. \tag{4.22}$$

The squared L^2-norm of the l.h.s. in (4.21) can be then majorized by

$$\frac{3}{16}\cdot\sum_{p\geq 0}\{(\sum_{k\geq N+p} u_{k,k-p})^2 + (\sum_{k\geq N+p} u_{k-p,k})^2 + (\sum_{N\leq k\leq p-N} u_{k,p-k})^2\}. \tag{4.23}$$

We can estimate further, using Cauchy-Schwarz inequality, the first sum in (4.23) as follows

$$\begin{aligned}\sum_{p\geq 0}(\sum_{k\geq N+p} u_{k,k-p})^2) &\leq \sum_{p\geq 0}(\sum_{k\geq N} k^{-2s})\cdot(\sum_{k\geq N+p} k^{2s}\cdot u_{k,k-p}^2)\\ &\leq (\sum_{k\geq N} k^{-2s})\cdot(\sum_{k,m\geq N} k^{2s}\cdot u_{k,m}^2).\end{aligned} \tag{4.24}$$

Due to (4.20) and to the fact that $s > 1/2$, it follows that the l.h.s. of (4.24) tends to 0 as $N \to \infty$. Similar arguments hold for the other two series in (4.23), showing that (4.23) vanishes too, as $N \to \infty$. This of course implies then the convergence we are interested in, as well as the claim concerning the extension of R in the case $D = (0,1)$.
Symmetrically, there exists an extension of R to the closure of $C_0^\infty((0,1)\times(0,1))$ in $L^2((0,1))\otimes H^{1/2+\epsilon}((0,1))$ and a simple interpolation argument ensures the existence of an extension to the closure of $C_0^\infty((0,1)\times(0,1))$ in $H^{\alpha_i}((0,1))\otimes H^{\beta_i}((0,1))$ for all $1\leq i\leq d$. Tensorizing these operators over the index i and taking into account that the algebraic tensor product of dense subspaces is again dense, we get the desired extension to $D = (0,1)^d$.
To conclude, we note that we can continuously embed $H^{1,1}_{(0)}(D\times D)$ into the closure of $C_0^\infty(D\times D)$ in $H^{\alpha,\alpha}(D\times D)$ with $\alpha = (1/d)\cdot(1,\ldots,1)$. If $d\leq 3$, condition (4.18) is satisfied and the trace can be defined. ◇
The previous lemma leads at once to

Proposition 4.10 *If a shift theorem at level $s\geq 0$ holds for problem (1.2) in $D\subset\mathbb{R}^d$, $d\leq 3$, then there exists a unique weak solution C_u solution of (4.16) and it belongs, for any $\epsilon>0$, to $H^{t,t}(D\times D)$, where*

$$t = \min(2 - d/4 - \epsilon, s+1).$$

Moreover, the following a-priori estimate holds

$$\|C_u\|_{H^{t,t}(D\times D)} \leq C(t,d)\cdot\|k\|_{L^2(\Delta)}\cdot\|R\|_{\mathcal{B}(H^{2-t,2-t}(D\times D),L^2(\Delta))}. \tag{4.25}$$

5 Discretization

5.1 FE-spaces and approximation properties

We shall now investigate the approximation of the statistics of u, using the standard Finite Element Method for the elliptic equations (3.7) and (3.11).
We start by defining general FE spaces. Let $\{\mathbf{S}_L\}_{L=0}^\infty$ be a dense sequence of finite dimensional (therefore closed) subspaces of $H^1_{(0)}(D)$. We assume also that this sequence has hierarchical structure, that is:

$$\mathbf{S}_0\subset\mathbf{S}_1\subset\ldots\subset\mathbf{S}_L\subset H^1_{(0)}(D), \tag{5.1}$$

where $N_L = \dim(\mathbf{S}_L) < \infty$ for all L (here L stands for the level), and that the following *approximation property* holds:

$$\min_{v\in\mathbf{S}_L}\|u-v\|_{H^1_{(0)}(D)} \leq \Phi(N_L,s)\|u\|_{H^{s+1}(D)},\quad \forall u\in H^{s+1}(D)\cap H^1_{(0)}(D), \tag{5.2}$$

where $\Phi(N,s) \to 0$ for $s > 0$ as $N \to \infty$. For regular solutions the usual FE-spaces based on quasiuniform, shape regular meshes are suitable.

Example 5.1 *Let $\{\mathcal{T}^{(L)}\}_{L\in\mathbb{N}}$ be a nested sequence of regular triangulations of the domain D of meshwidth $h_L = h_{L-1}/2$, $\forall L \geq 1$ and let $p \geq 1$ be a polynomial degree. Then*

$$\mathbf{S}_L := S^p(D, \mathcal{T}^{(L)}) := \{u \in C^0(\bar{D}) : u|_K \in \mathcal{P}^p(K) \quad \forall K \in \mathcal{T}^{(L)}\} \tag{5.3}$$

satisfies (5.1), as well as (5.2) in the form:

$$\min_{v\in S^p(D,\mathcal{T}^{(L)})} \|u - v\|_{H^1(D)} \leq C \left(\frac{h_L}{p}\right)^{\min\{p,s\}} \|u\|_{H^{s+1}(D)}. \tag{5.4}$$

where C is independent of p, h_L and depends only on $s \geq 0$.

Note that, expressed in terms of the number of degrees of freedom $N = N_L$, the convergence rate reads:

$$\Phi(N,s) = O(N^{-\frac{\min\{p,s\}}{d}}) = O(N^{-\delta}), \quad \delta := \frac{\min\{p,s\}}{d}. \tag{5.5}$$

for fixed p and $L \to \infty$.
We conclude this introductory part recalling that estimates similar to (5.5) also hold for p-version or spectral element methods, i.e. on fixed $\mathcal{T}$ as $p = p_L \to \infty$.

In view of Proposition 4.5, we briefly discuss next the FE-approximation in the case of a nonsmooth domain.

Remark 5.2 If $D \subset \mathbb{R}^2$ is a polygon with M straight sides, problem (4.1) admits a shift theorem at order s in the spaces $H^{s+1}(D)$ only for small values of s (often $1/2 < s < 1$). In this case, however, for smooth data in (4.1), we still have a shift theorem at order $s \geq 0$ in the weighted spaces $H^{1+s,2}_{\underline{\beta}}(D)$ with some $\underline{\beta} \in (0,1)^M$, i.e. the weight function $\omega_{\underline{\beta}+k}(x)$ introduced in Section 4.1 compensates for the corner singularities of the solution u. To the weighted spaces $H^{k,l}_{\underline{\beta}}(D)$ correspond FE-approximations on sequences of graded meshes $\{\mathcal{T}^n_\gamma\}_n$ with shape-regular elements which satisfy in dimension 2:

$$|\mathcal{T}^n_\gamma| := \#\text{ of triangles in } \mathcal{T}^n_\gamma = O(n^2), \tag{5.6}$$

$$\forall\, T \in \mathcal{T}^n_\gamma : h_T := \text{ diam } (T) \leq C\omega(x)^{1-1/\gamma} n^{-1}, \tag{5.7}$$

for $\gamma \geq 1$. Clearly, $\gamma = 1$ corresponds to quasiuniform triangulations of meshwidth $h = O(n^{-1})$, whereas $\gamma \gg 1$ corresponds to strong refinement near the vertices. Then, for any $u \in H^{1+s,2}_{\underline{\beta}}(D)$ we have, as $n \to \infty$:

$$\inf_{v\in S^p(D,\mathcal{T}^n_\gamma)} \|u - v\|_{H^1(D)} \leq CN^{-\delta}, \tag{5.8}$$

with δ as in (5.5), provided that $\gamma > \frac{\min\{p,s\}}{\mathrm{Re}\lambda}$, where $\mathrm{Re}\lambda > 0$ denotes the real part of the smallest singularity exponent of the solution u in the polygon D.

5.2 Rate of convergence for E_u

From the ellipticity (1.1) of the coefficients and the approximation error estimates (5.4) we deduce:

Proposition 5.3 *Assume that the mixed boundary value problem (3.6) for E_u satisfies the shift theorem at order $s \geq 0$. Then the FE-approximation E_u^L of E_u, the solution of (3.6) with data $E_f \in H^{s-1}(D)$, $E_g \in H^{s+1/2}(\Gamma_0)$, $E_h \in H^{s-1/2}(\Gamma_1)$, reads:*

$$E_u^L \in S^p(D, \mathcal{T}^{(L)}): \quad q(E_u^L, E_v) = l(E_v) \quad \forall E_v \in S^p(D, \mathcal{T}^{(L)}). \tag{5.9}$$

Then, as $N_L \to \infty$ we have the following error estimate

$$\|E_u - E_u^L\|_{H^1(D)} \leq C {N_L}^{-\delta} \, \|E_u\|_{H^{s+1}(D)}. \tag{5.10}$$

The result is a consequence of the ellipticity, the regularity of the solution and the approximation properties of the FE-space. We remark that the constant involved in the estimate above depends only on s, α, p.

5.3 Rate of convergence for C_u

Within the abstract setting described at the beginning of the section 5.1, we follow [21] and introduce further the so called *hierarchic excess* of the scale (5.1), by

$$W_L := \mathbf{S}_L \ominus \mathbf{S}_{L-1} \quad L \geq 0, \tag{5.11}$$

where we set $\mathbf{S}_{-1} := \{0\}$ and the complement is taken with respect to some Hilbert structure of $H^1_{(0)}(D)$ equivalent to the usual one. The corresponding norm will be denoted throughout this section by $|\cdot|$. Further, P_L will be the orthogonal projection w.r.t. $|\cdot|$ on $\mathbf{S}_L$.
We remark that (5.2) still holds for $L = -1$, by choosing $\Phi(N_{-1}, s)$ to be the embedding constant of $H^1(D)$ in $H^{s+1}(D)$.
Noting that with this definition $\mathbf{S}_L$ decomposes as an orthogonal sum

$$\mathbf{S}_L = \bigoplus_{0 \leq i \leq L} W_i, \tag{5.12}$$

we can define the *full tensor product FE-spaces* in $D \times D$ by

$$\mathbf{S}_{L,L} := \mathbf{S}_L \otimes \mathbf{S}_L = \bigoplus_{0 \leq i,j \leq L} (W_i \otimes W_j) \subset H^{1,1}_{(0)}(D \times D), \tag{5.13}$$

for all $L \in \mathbb{N}$. However, due to the fact that the number of degrees of freedom necessary for the FEM based on $\mathbf{S}_{L,L}$ grows very fast when the mesh is refined, we shall employ the *sparse tensor product FE-spaces* given by

$$\hat{\mathbf{S}}_{L,L} := \bigoplus_{0 \leq i+j \leq L} (W_i \otimes W_j). \tag{5.14}$$

For a given $C_u \in H^1_{(0)}(D) \otimes H^1_{(0)}(D)$ we define C_u^L, the *sparse interpolant* of C_u in $\hat{\mathbf{S}}_{L,L}$, as the projection of C_u onto $\hat{\mathbf{S}}_{L,L}$, w.r.t. the usual Hilbert structure of $H^{1,1}_{(0)}(D \times D)$. We shall need next also $\tilde{C}_u^L$, the *modified sparse interpolant* of C_u in $\hat{\mathbf{S}}_{L,L}$, obtained by projecting C_u onto $\hat{\mathbf{S}}_{L,L}$, w.r.t. the Hilbert structure of $H^{1,1}_{(0)}(D \times D)$ obtained by tensorizing $|\cdot|$ with itself:

$$\tilde{C}_u^L := \sum_{0 \leq i+j \leq L} (P_i - P_{i-1}) \otimes (P_j - P_{j-1}) C_u. \tag{5.15}$$

With these notations, the following estimate (see also [10], [15]) will enable us to deduce from (5.2) an approximation property of the sparse scale $(\hat{\mathbf{S}}_{L,L})_{L \in \mathbb{N}}$, too.

Proposition 5.4 *Assume that the sequence (5.1) of FE-spaces $\{\mathbf{S}_L\}_L$ has the approximation property (5.2) and $s,t>0$. Then there exists $C>0$ such that for all $C_u \in H^{1,1}_{(0)}(D\times D)\cap H^{s+1,t+1}(D\times D)$ the sparse interpolant C_u^L approximates C_u with the error:*

$$\begin{aligned}\|C_u - C_u^L\|_{H^{1,1}_{(0)}(D\times D)} &\leq C\left(\left[\sum_{i=0}^{L+1}\Phi^2(N_{i-1},s)\Phi^2(N_{L+1-i},t)\right]^{1/2}\|C_u\|_{H^{s+1,t+1}(D\times D)}\right.\\ &+ \left.\left[\sum_{i=L+1}^{\infty}\Phi^2(N_i,s)\right]^{1/2}\|C_u\|_{H^{s+1,1}(D\times D)}\right). \qquad (5.16)\end{aligned}$$

Here C is a constant depending on the second Hilbert structure of $H^{1,1}_{(0)}(D\times D)$ given by $|\cdot|$.

Proof. Due to the equivalence of the two norms $|\cdot|$ and $\|\cdot\|$, and to the hierarchical structure of the scale $(\hat{\mathbf{S}}_{L,L})_{L\in\mathbf{N}}$, we can write:

$$\begin{aligned}\|C_u - C_u^L\|^2_{H^{1,1}_{(0)}(D\times D)} &= \min_{C_v\in\hat{\mathbf{S}}_{L,L}}\|C_u - C_v\|^2_{H^{1,1}_{(0)}(D\times D)}\\ &\leq C\cdot\min_{C_v\in\hat{\mathbf{S}}_{L,L}}|C_u - C_v|^2_{H^{1,1}_{(0)}(D\times D)}\\ &= C\cdot|C_u - \tilde{C}_u^L|^2_{H^{1,1}_{(0)}(D\times D)}\\ &= C\cdot|\sum_{i+j\geq L+1}(P_i - P_{i-1})\otimes(P_j - P_{j-1})C_u|^2_{H^{1,1}_{(0)}(D\times D)}\\ &= C\cdot|\sum_{i=0}^{\infty}\sum_{j\geq\max\{L+1-i,0\}}(P_i - P_{i-1})\otimes(P_j - P_{j-1})C_u|^2_{H^{1,1}_{(0)}(D\times D)}\\ &= C\cdot\left(|\sum_{i=0}^{L+1}(P_i - P_{i-1})\otimes(\mathrm{Id} - P_{L-i})C_u|^2_{H^{1,1}_{(0)}(D\times D)}\right.\\ &+ \left.|\sum_{i=L+2}^{\infty}(P_i - P_{i-1})\otimes\mathrm{Id}\,C_u|^2_{H^{1,1}_{(0)}(D\times D)}\right). \qquad (5.17)\end{aligned}$$

From (5.2) and the equivalence of the norms again, we deduce

$$|(\mathrm{Id} - P_i)u|_{H^1_{(0)}(D)} \leq C\cdot\Phi(N_i,s)\|u\|_{H^{s+1}(D)} \quad \forall u\in H^1_{(0)}(D)\cap H^{s+1}(D). \qquad (5.18)$$

and similarly with s replaced by t. Taking into account that $(\mathrm{Id} - P_L)_{L\in\mathbb{N}}$ is a decreasing sequence of projections and tensorizing it by a projection produces also a decreasing sequence of projections, the estimate (5.16) follows by using (5.18) in (5.17). ⋄

We use next the sparse tensor product space $\hat{\mathbf{S}}_{L,L}$ built as above from the usual FE-spaces $\mathbf{S}_L$ in example 5.1. We obtain that the FE-approximation $\hat{C}_u^L \in \hat{\mathbf{S}}_{L,L}$ requires, despite the high dimensionality of the problem, essentially not more degrees of freedom than the FE solution of the deterministic problem in D.

Proposition 5.5 *Assume that the mixed boundary value problem (3.6) satisfies the shift theorem at order $s\geq 0$ and that the correlation functions of the data satisfy $C_f \in H^{s-1,s-1}(D\times D)$, $C_{f,h}\in H^{s-1,s-1/2}(D\times\Gamma_1)$, $C_{h,f}\in H^{s-1/2,s-1}(\Gamma_1\times D)$ and $C_h\in H^{s-1/2,s-1/2}(\Gamma_1\times\Gamma_1)$. Then the sparse FE-approximation $\hat{C}_u^L$ of the correlation function C_u which is defined by*

$$\hat{C}_u^L\in\hat{\mathbf{S}}_{L,L}: \quad \mathcal{Q}(\hat{C}_u^L, C_v) = \mathcal{L}(C_v) \quad \forall C_v\in\hat{\mathbf{S}}_{L,L}, \qquad (5.19)$$

converges, as $L \to \infty$, with the rate

$$\|C_u - \hat{C}_u^L\|_{H^{1,1}_{(0)}(D\times D)} \lesssim (\log N_L)^{1/2} \cdot N_L^{-\delta} \ \|C_u\|_{H^{s+1,s+1}(D\times D)}, \tag{5.20}$$

where δ is given by (5.5).

Proof. The coercivity of the sesquilinear form $\mathcal{Q}$ defined in (3.12a) has been proved in Proposition 3.4. A direct consequence of this fact is the quasi-optimality of the FE-approximation $\hat{C}_u^L$ ($\|\cdot\|_{1,1}$ stands here for the norm in $H^{1,1}_{(0)}(D\times D)$)

$$\|C_u - \hat{C}_u^L\|_{1,1} \leq C(\alpha,\beta) \min_{\hat{C}_v \in \hat{\mathbf{S}}_{L,L}} \|C_u - \hat{C}_v\|_{1,1} = C(\alpha,\beta)\|C_u - C_u^L\|_{1,1}. \tag{5.21}$$

The quadruple $(C_f, C_{f,h}, C_{h,f}, C_h)$ satisfies the regularity assumptions that enable us to apply Proposition 4.4 and to deduce that $C_u \in H^{s+1,s+1}(D\times D)$. Taking into account that $N_j = O(2^{dj})$ and using (5.5) in (5.16), with $s = t$, we obtain

$$\|C_u - C_u^L\|_{H^{1,1}_{(0)}(D\times D)} \lesssim (\log N_L)^{1/2} \cdot N_L^{-\delta} \ \|C_u\|_{H^{s+1,s+1}(D\times D)}. \tag{5.22}$$

From (5.21) and (5.22) follows the claimed estimate. Note also, for later use, that in (5.20), $N_L = \dim(\mathbf{S}_L) = O(2^{dL})$. Also, due to (5.14), $N_3 := \dim(\hat{\mathbf{S}}_{L,L}) = O(L\cdot 2^{dL})$, which enables us to rephrase (5.20) in terms of the number of degrees of freedom N_3:

$$\|C_u - \hat{C}_u^L\|_{H^{1,1}_{(0)}(D\times D)} = O((\log N_3)^{1/2+\delta} \cdot N_3^{-\delta}) \tag{5.23}$$

$\diamond$

Remark 5.6 It is easy to see that the FE-approximation $\overline{C}_u^L$ of C_u,

$$\overline{C}_u^L \in \mathbf{S}_{L,L}: \quad \mathcal{Q}(\overline{C}_u^L, C_v) = \mathcal{L}(C_v) \quad \forall C_v \in \mathbf{S}_{L,L}, \tag{5.24}$$

based on the full tensor product space $\mathbf{S}_{L,L}$ in (5.13) satisfies, under the regularity assumptions in Proposition 5.5,

$$\|C_u - \overline{C}_u^L\|_{H^{1,1}_{(0)}(D\times D)} \leq C \cdot N_L^{-\delta/2} \|C_u\|_{H^{s+1,s+1}(D\times D)}. \tag{5.25}$$

We see that for a given regularity of the data, the rate (5.25) in terms of the number of degrees of freedom is essentially half that achieveable with sparse grids (5.20).

Remark 5.7 If D is nonsmooth, it follows from Proposition 4.5 and Remark 5.2 that the influence of corner singularities in $D\times D$ can be compensated by forming sparse tensor-products of FE-spaces in D with judicious mesh refinement towards the vertices of D. Once good meshes for the solution E_u of (3.6) have been determined, the sparse FE-space for C_u based on these meshes will also give optimal rates of convergence for $\hat{C}_u^L$, provided $C_f, C_{f,h}, C_{h,f}, C_h$ in Proposition 4.5 are sufficiently regular.

6 Implementation and complexity

Recall that the discretized problem for the covariance equation consists in solving a linear system

$$\hat{S}^{L,L} \underline{C}_u^L = \underline{C}_f^L, \tag{6.1}$$

where $\hat{S}^{L,L}$ denotes the stiffness matrix of (5.19) with respect to some basis of the sparse tensor product space $\hat{\mathbf{S}}_{L,L} \subset H^{(1,1)}_{(0)}(D\times D)$. Since the solution of (6.1) by Cholesky decomposition

requires a too large computational effort, we solve (6.1) by the standard *conjugate gradient method* (see for example [9]). As it is well-known, the conjugate gradient method generates an approximating sequence $(\underline{C}^L_{u,l})_{l\geq 0}$ satisfying

$$\|\underline{C}^L_u - \underline{C}^L_{u,l}\|_2 \leq 2 \cdot \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^l \|\underline{C}^L_u - \underline{C}^L_{u,0}\|_2, \quad l = 0,1,2\ldots, \tag{6.2}$$

where $\kappa = \text{cond}_2(\hat{S}^{L,L})$ and $\|\cdot\|_2$ denotes the Euclidian vector norm.
The first question to be addressed is therefore the boundedness of $\text{cond}_2(\hat{S}^{L,L})$ with respect to L. This can be achieved by a wavelet preconditioning procedure which we shall discuss next. We then estimate the number of flops required by one cg iteration for (6.1).

6.1 FE space in D and multilevel preconditioning

As we have seen in section 5.3, the sparse FE-space depends on the choice of the complements $(W_i)_{i\in\mathbb{N}}$. To ensure the approximation property of the sparse FE-space, W_i must be the orthogonal complement of S_{i-1} in S_i w.r.t. some suitable Hilbert structure equivalent to the usual topology of $H^1_0(D)$. The following stability condition suffices in this respect. Here and in what follows, $\mathcal{F}$ denotes a family of double indices running in $\mathbb{N}^d \times \mathbb{N}^d$.

Assumption 6.1 *There exist a family $(\psi_{i,Q})_{(i,Q)\in\mathcal{F}} \subset H^1_0(D)$ and constants $C_1, C_2 > 0$ such that each $u \in H^1_0(D)$ can be expanded as a convergent series in $H^1_0(D)$*

$$u = \sum_{(i,Q)\in\mathcal{F}} c_{i,Q}\psi_{i,Q} \tag{6.3}$$

and the following 'stability condition' is fulfilled

$$C_1 \sum_{(i,Q)\in\mathcal{F}} |c_{i,Q}|^2 \leq \| \sum_{(i,Q)\in\mathcal{F}} c_{i,Q}\psi_{i,Q}\|^2_{H^1_0(D)} \leq C_2 \sum_{(i,Q)\in\mathcal{F}} |c_{i,Q}|^2, \tag{6.4}$$

Example 6.2 *If $D = (0,1)$, we choose $\mathcal{F} := \{(i,Q) \mid 0 \leq Q,\ 1 \leq i \leq 2^Q\}$ and we can easily obtain a family with the properties mentioned above by collecting the well-known* 'hat-functions'*, (see, for example, [3]), which are piecewise linear and satisfy*

$$\psi_{(i,Q)}(\frac{j}{2^{Q+1}}) = 2^{-Q/2} \cdot \delta_{2i-1,j} \quad \forall\ 0 \leq j \leq 2^{Q+1}. \tag{6.5}$$

Example 6.3 *We can check Assumption 6.1 with the same choice for $\mathcal{F}$ in the case of the* 'prewavelet basis' *given (see also [3]) by the following family of piecewise linear functions on $D = (0,1)$*

$$\begin{aligned}
\psi_{(i,Q)}(\frac{j}{2^{Q+1}}) &= 2^{-Q/2}(-\frac{1}{2}\delta_{2i-2,j} + \delta_{2i-1,j} - \frac{1}{2}\delta_{2i,j}) \quad \forall\ 2 \leq i \leq 2^Q - 1,\\
\psi_{(1,Q)}(\frac{j}{2^{Q+1}}) &= 2^{-Q/2}(\delta_{1,j} - \frac{1}{2}\delta_{2,j}),\\
\psi_{(2^Q,Q)}(\frac{j}{2^{Q+1}}) &= 2^{-Q/2}(-\frac{1}{2}\delta_{2^Q-2,j} + \delta_{2^Q-1,j}).
\end{aligned} \tag{6.6}$$

Remark 6.4 *For each of the two examples above, Span$\{\psi_{i,Q} \mid 0 \leq Q \leq L\}$ equals the space of continuous functions on $D = (0,1)$ vanishing on the boundary and which are piecewise linear with respect to a regular mesh of width 2^{-L-1}.*

Remark 6.5 *Higher order functions satisfying Assumption 6.1 exist as well (see, e.g. [4]). In the case D is an arbitrary polygon or polyhedron, bases satisfying Assumption 6.1 can also be constructed. See [5] for examples.*

Example 6.6 *By tensorizing a 'prewavelet basis' $(\psi_{i,Q})_{(i,Q)\in\mathcal{F}} \subset H_0^1((0,1))$ we shall build next a family of functions on $D = (0,1)^d$ which still satisfies Assumption 6.1 after choosing the family $\mathcal{F}$ properly. However, we shall not use the coresponding inequalities* (6.4) *explicitly, therefore we omit the proof.*

The stable basis will then be used to construct the FE space as well as the stiffness matrix of the discretized covariance problem (5.19).
From now on i, j, k, m, Q, M, N, P will always denote d-tuples of positive integers, subject to the restrictions

$$1 \le i \le 2^Q,\ \ 1 \le j \le 2^M,\ \ 1 \le k \le 2^N,\ \ 1 \le m \le 2^P,$$

inequalities which have to be understood componentwise. For future use, we set up now also the following notations involving an arbitrary d-tuple $R = (R_q)_{1\le q\le d}$:

$$|R|_\infty := \max_{1\le q\le d} R_q, \quad |R| := \sum_{q=1}^{d} R_q. \tag{6.7}$$

Recalling the piecewise linearity of the functions in the examples above, we deduce that the next definition introduces piecewise multilinear functions on D, with respect to a regular mesh of width 2^{-L}:

$$\psi_{i,Q}(x) = \prod_{q=1}^{d} \psi_{(i_q,Q_q)}(x_q) \quad \forall\, x = (x_q)_{1\le q\le d} \in D.$$

Collecting all these multilinear functions, we obtain a family $(\psi_{i,Q})_{(i,Q)\in\mathcal{F}}$ satisfying Assumption 6.1 for $D = (0,1)^d$, where

$$\mathcal{F} := \{(i,Q) \in \mathbb{N}^d \times \mathbb{N}^d \mid 0 \le Q,\ 1 \le i \le 2^Q\} \tag{6.8}$$

Formally, the FE-space we use in D to calculate E_u is defined as follows:

$$\mathbf{S}_L = \ \mathrm{Span}\{\psi_{i,Q} |\ 0 \le Q \le L\}. \tag{6.9}$$

Then the entries of the stiffness matrix in D at level L are

$$(S^L)_{(i,Q)(j,M)} = \int_D \nabla_x \psi_{j,M} \cdot A(x) \nabla_x \psi_{i,Q}\, dx. \tag{6.10}$$

Recalling also the assumption (1.1) we have made on the coefficient A, we deduce that $S^L \in \mathbb{R}^{N_1\times N_1}_{sym}$ is a positive definite matrix, where $N_1 = (2^{L+1}-1)^d$.

6.2 Sparse FE-space in $D \times D$

The full tensor product space in $D \times D$ is $\mathbf{S}_{L,L} = \mathbf{S}_L \otimes \mathbf{S}_L$. The entries of the corresponding stiffness matrix for the covariance problem in $D \times D$, at level L, have product structure. More precisely, it holds

$$\begin{aligned} (S^{L,L})_{(i,Q)(j,M)(k,N)(m,P)} &:= \int_{D\times D} \nabla_{x,y}(\psi_{k,N} \otimes \psi_{m,P}) \cdot A(x) \otimes A(y) \nabla_{x,y}(\psi_{i,Q} \otimes \psi_{j,M})\, dx\, dy \\ &= S^L_{(i,Q)(k,N)} \cdot S^L_{(j,M)(m,P)}, \end{aligned} \tag{6.11}$$

for all $0 \leq Q, M, N, P \leq L$.
Symbolically, $S^{L,L} = S^L \otimes S^L$. Note again that $S^{L,L} \in \mathbb{R}^{N_2 \times N_2}_{sym}$ is positive definite, where $N_2 = (N_1)^2 = (2^{L+1} - 1)^{2d}$.
As we have seen in Proposition 5.5, however, optimal convergence rates are already achieveable with a sparse FE-space. To estimate the work when using a sparse FE space, we remark that $\mathbf{S}_{L,L}$ can be viewed as a hierarchical space, the corresponding hierarchy being induced by the $|\cdot|_\infty$-norm defined by (6.7). The notations are therefore consistent with those in section 5.3 concerning the hierarchical excess

$$\mathbf{S}_L = \bigoplus_{l=0}^{L} W_l, \text{ where } W_l := \text{ Span}\{\psi_{i,Q} \mid |Q|_\infty = l\}. \tag{6.12}$$

To match the setting of the previous section 5.3, we note that Assumption 6.1 entails the existence of a norm $|\cdot|$ on $H_0^1(D)$, equivalent to the usual one, such that the previous decomposition becomes orthogonal. Indeed, with each $u \in H_0^1(D)$ we associate the unique decomposition (6.3) and we define

$$|u|^2_{H_0^1(D)} := \sum_{(i,Q)\in\mathcal{F}} |c_{i,Q}|^2. \tag{6.13}$$

It is easy to check that (6.13) defines a scalar product equivalent, due to (6.4), to the usual one on $H_0^1(D)$, and w.r.t. which the family $(\psi_{i,Q})_{(i,Q)\in\mathcal{F}}$ is an orthonormal basis.

The sparse FE-space which shall be used to numerically approximate the solution of the covariance problem in $D \times D$ will be, according to the terminology of section 5.3, associated to the hierarchy $(\mathbf{S}_L)_{L\geq 0}$. Formally,

$$\hat{\mathbf{S}}_{L,L} := \bigoplus_{0\leq l_1+l_2\leq L} W_{l_1} \otimes W_{l_2}.$$

Hence the stiffness matrix $\hat{S}^{L,L}$ in (6.1) is simply the restriction of $S^{L,L}$ to $\hat{\mathbf{S}}_{L,L}$. In terms of the entries,

$$\hat{S}^{L,L}_{(i,Q)(j,M)(k,N)(m,P)} = S^{L,L}_{(i,Q)(j,M)(k,N)(m,P)}$$
$$\forall\ 0 \leq Q, M, N, P \leq L \text{ such that } |Q|_\infty + |M|_\infty \leq L,\ |N|_\infty + |P|_\infty \leq L. \tag{6.14}$$

(recall that all inequalities involving d-tuples have to be understood componentwise).
Note also that $\hat{S}^{L,L} \in \mathbb{R}^{N_3 \times N_3}_{sym}$ is still symmetric and positive definite, where $N_3 = (L\cdot 2^{L+1}+1)^d$.

As we aim at comparing the computational efforts needed to obtain E_u and C_u, we remark now and use later that, asymptotically, as $L \to \infty$,

$$N_3 \cong N_1 \cdot L^d \cong N_1 \cdot (\log N_1)^d. \tag{6.15}$$

Finally, we capture in the next result one main feature of the matrices constructed above.

Lemma 6.7 *Under Assumption 6.1, the condition numbers of the matrices $S^L, S^{L,L}, \hat{S}^{L,L}$ remain bounded as $L \to \infty$.*

Proof. The claim follows at once if we remark that in view of the positivity condition (1.1), the 'stability' assumption can be rephrased in terms of the spectrum σ of S^L as

$$\sigma(S^L) \subset [\alpha C_1, \beta C_2] \quad \forall L \geq 0,$$

so that $\sigma(S^{L,L}) \subset [\alpha^2 C_1^2, \beta^2 C_2^2]$. Since $\hat{S}^{L,L}$ is the restriction of the positive matrix $S^{L,L}$ to the sparse tensor space, we conclude

$$\sigma(\hat{S}^{L,L}) \subset [\alpha^2 C_1^2, \beta^2 C_2^2] \quad \forall L \geq 0.$$

$\diamond$

Corollary 6.8 *Under Assumption 6.1, the cg-procedure for solving* (6.1) *yields an approximate solution of the linear system* (5.19) *with accuracy of the order of the discretization error in* $O(L)$ *steps.*

6.3 Complexity

Recalling the framework of the previous paragraph, we investigate next the cost of one iteration of the conjugate gradient method, i.e. the number of flops required by the matrix-vector multiplication

$$x \longrightarrow \hat{S}^{L,L} x. \tag{6.16}$$

In order to be able to estimate the complexity of the algorithm, we make the following additional

Assumption 6.9 *The family of hierarchic basis functions* $(\psi_{i,Q})_{(i,Q)\in\mathcal{F}} \subset H_0^1(D)$ *is indexed over* $\mathcal{F}$ *given in* (6.8)*, and has* 'local support'*, that is, there exists* $p \in \mathbb{N}^*$ *such that for all* $(i,Q) \in \mathcal{F}$ *and* $M \in \mathbb{N}^d$*, the set* $supp\,(\psi_{i,Q}) \cap supp\,(\psi_{j,M})$ *has nonempty interior for at most*

$$p^d \cdot \prod_{i=1}^{d} \max(1, 2^{M_i - Q_i})$$

values of j*.*

Remark 6.10 *One can easily check the local support condition for the families in Examples 6.2 and 6.3 (by choosing* $p = 1$ *and* $p = 4$ *respectively), as well as for the one constructed in Example 6.6.*

As a consequence of Assumption 6.9, we shall prove first a result concerning the sparsity of the matrices involved.

Lemma 6.11 *Denoting by* nnz*(X) the number of nonzero entries of a matrix X, the following upper estimates hold:*

$$\mathrm{nnz}(S^L) \leq C^d \cdot p^d \cdot L^d \cdot 2^{Ld}, \tag{6.17}$$

$$\mathrm{nnz}(S^{L,L}) \leq C^{2d} \cdot p^{2d} \cdot L^{2d} \cdot 2^{2Ld}, \tag{6.18}$$

$$\mathrm{nnz}(\hat{S}^{L,L}) \leq C^d \cdot p^{2d} \cdot 2^{2Ld}. \tag{6.19}$$

Moreover, the estimates are optimal as $L \to \infty$*, while the constant* C *does not depend on* p, d *or* L*.*

Proof. We shall only prove (6.19) as well as its optimality, the other two estimates being obtained by analogous elementary computations based on the local support assumption. To this end, we remark first that for fixed Q, N, the equation

$$S^L_{(i,Q)(k,N)} \neq 0 \tag{6.20}$$

has at most $p^d \cdot 2^{\sum_q \max(Q_q,N_q)}$ solutions (i,k). Taking into account (6.11) and (6.14), it follows that for fixed Q, M, N, P there are at most $p^{2d} \cdot 2^{\sum_q \max(Q_q,N_q)+\sum_q \max(M_q,P_q)}$ quadruples (i,j,k,m) such that

$$\hat{S}^{L,L}_{(i,Q)(j,M)(k,N)(m,P)} \neq 0. \tag{6.21}$$

This in turn implies

$$\begin{aligned} \mathrm{nnz}(\hat{S}^{L,L}) &\leq p^{2d} \cdot \sum_{\substack{|Q|_\infty+|M|_\infty\leq L \\ |N|_\infty+|P|_\infty\leq L}} 2^{\sum_q \max(Q_q,N_q)+\sum_q \max(M_q,P_q)} \\ &\leq p^{2d} \cdot \Big(\sum_{\substack{0\leq a+b\leq L, \\ 0\leq c+e\leq L}} 2^{\max(a,c)+\max(b,e)}\Big)^d. \end{aligned} \tag{6.22}$$

We claim that

$$\sum_{\substack{0\leq a+b\leq L, \\ 0\leq c+e\leq L}} 2^{\max(a,c)+\max(b,e)} \leq C \cdot 2^{2L}. \tag{6.23}$$

To prove this, we write

$$\sum_{\substack{0\leq a+b\leq L, \\ 0\leq c+e\leq L}} 2^{\max(a,c)+\max(b,e)} = \sum_{l=0}^{2L} F(l,L)2^l, \tag{6.24}$$

where $F(l,L)$ denotes the number of quadruples of positive integers (a,b,c,e) satisfying $\max(a,c)+\max(b,e) = l$ with the restrictions $0 \leq a+b, c+e \leq L$.
The rough estimate $F(l,L) \leq (L+1)^4$ which holds for all $0 \leq l \leq L$ ensures that the first half of the sum in the r.h.s. of (6.24) is absorbed by the r.h.s. of (6.23).
As for $L < l \leq 2L$, we note that

$$\begin{aligned} F(l,L) &\leq 2 \sum_{\substack{a+e=l \\ 0\leq a,e\leq L}} (L+1-a)(L+1-e) \\ &\leq \sum_{\substack{x+y=2L-l \\ 0\leq x,y}} (x+1)(y+1) = G(2L-l), \end{aligned} \tag{6.25}$$

where the last equality stands for the definition of the polynomial G. Hence we can estimate also the second half of the sum in the r.h.s. of (6.24) as follows

$$\sum_{l=L}^{2L} F(l,L)2^l \leq 2 \cdot 2^{2L} \cdot \Big(\sum_{l=0}^{2L} G(l) \cdot 2^{-l}\Big) \leq C \cdot 2^{2L}. \tag{6.26}$$

The existence of a constant C in the last inequality is due to the fact that, G being a polynomial, the convergence radius of the series with coefficients $\{G(l)\}_{l\geq 0}$ is 1.
The optimality of (6.19) follows as soon as we note that in the case of an arbitrary coefficient A, the entries $\hat{S}^{L,L}_{(i,Q)(j,M)(k,N)(m,P)}$ are in general (see Examples 6.2, 6.3 above) nontrivial for all admissible i,j,k,m, if $P = Q = (L,L,\ldots,L)$ and $M = N = (0,0,\ldots,0)$. ⋄

Remark 6.12 *Rephrased in terms of the degrees of freedom N_1, N_2 and N_3 respectively, the results of the previous lemma show that only S^L and $S^{L,L}$ are sparse, while $\hat{S}^{L,L}$ is rather densely populated. Namely, the corresponding estimates read as follows:*

$$\mathrm{nnz}(S^L) \leq O((\log N_1)^d \cdot N_1), \tag{6.27}$$

$$\mathrm{nnz}(S^{L,L}) \leq O((\log N_2)^{2d} \cdot N_2), \tag{6.28}$$

$$\mathrm{nnz}(\hat{S}^{L,L}) \leq O((\log N_3)^{-2d} N_3^2). \tag{6.29}$$

Using (6.15) in the r.h.s. of (6.29) we deduce that the number of nonzero entries of $\hat{S}^{L,L}$ is $O(N_1^{2d})$ and this estimate is optimal. At first sight, therefore, it appears that the solution of (6.1) with the sparse FE-space $\hat{\mathbf{S}}_{L,L}$ is more costly than the solution of (5.9), with the usual FE-space $\mathbf{S}_L$. We shall now show that the structure of $\hat{S}^{L,L}$ can be exploited to reduce the complexity of (6.16).

Algorithm 6.13 *To compute the matrix-vector multiplication* (6.16) *at a fixed level L, proceed as follows:*

1. Store S^L as sparse matrix and x; % that is, $O(L^d \cdot 2^{Ld})$ numbers
2. For $|Q|_\infty + |M|_\infty \leq L$
   for $1 \leq i \leq 2^Q$ and $1 \leq j \leq 2^M$
   $(\hat{S}^{L,L}x)_{(i,Q)(j,M)} := 0$;
   for $|N|_\infty + |P|_\infty \leq L$
   if $|Q|_\infty + |P|_\infty \leq |M|_\infty + |N|_\infty$
   compute $y_{(i,Q),N,P,(j,M)} := \sum_m \left(\sum_k S^L_{(i,Q)(k,N)} \cdot x_{(k,N)(m,P)} \right) \cdot S^L_{(j,M)(m,P)}$;
   else
   compute $y_{(i,Q),N,P,(j,M)} := \sum_k S^L_{(i,Q)(k,N)} \cdot \left(\sum_m x_{(k,N)(m,P)} \cdot S^L_{(j,M)(m,P)} \right)$;
   end;
   update $(\hat{S}^{L,L}x)_{(i,Q)(j,M)} := (\hat{S}^{L,L}x)_{(i,Q)(j,M)} + y_{(i,Q),N,P,(j,M)}$;
   end;
   end;
   end;

Lemma 6.14 *The complexity of Algorithm 6.13 is log-linear. More precisely,*

$$\#flops\,(x \longrightarrow \hat{S}^{L,L}x) \;\leq\; C \cdot (\log N_3)^{3d+1} \cdot N_3 = O((\log N_1)^{4d+1} \cdot N_1), \tag{6.30}$$

where the constant C depends on p and d only.

Proof. Note first that the asymptotic equivalence in the r.h.s. of (6.30) follows at once from (6.15). Now, due to (6.11) and (6.14) we can write

$$(\hat{S}^{L,L}x)_{(i,Q)(j,M)} = \sum_{|N|_\infty + |P|_\infty \leq L} \sum_{k,m} S^L_{(i,Q)(k,N)} \cdot x_{(k,N)(m,P)} \cdot S^L_{(j,M)(m,P)}. \tag{6.31}$$

Considering Q, M, N, P fixed at this moment, we can choose to perform the inner sum in (6.31) either first over k and then over m, or reversely, according to the amount of flops required by each of these procedures. Of course, we shall always prefer the cheapest one and this idea has been implemented in our Algorithm 6.13 above. Therefore we shall estimate next the number of flops needed to compute the inner sum first over k and then over m for all i, j.
To this end let us start by remarking that due to the local support assumption, the equation

$$S^L_{(i,Q)(k,N)} \neq 0 \tag{6.32}$$

has at most $p^d \cdot 2^{\sum_q \max(Q_q, N_q)}$ solutions (i, k). It follows that

$$\sum_{i,m} \#flops(x \longrightarrow \sum_k S^L_{(i,Q)(k,N)} \cdot x_{(k,N)(m,P)}) \leq p^d \cdot 2^{\sum_{q=1}^d \max(Q_q, N_q) + |P|}. \tag{6.33}$$

The result of the operations we have counted up to now is the set of numbers given by

$$\{v_{i,m} := \sum_k S^L_{(i,Q)(k,N)} \cdot x_{(k,N)(m,P)} \mid \forall\; i, m\}. \tag{6.34}$$

The inner sum in (6.31) can be then written as

$$\sum_m S^L_{(j,M)(m,P)} v_{i,m}, \tag{6.35}$$

and an argument similar to the one above shows that

$$\sum_i \sum_j \#flops(v \longrightarrow \sum_m S^L_{(j,M)(m,P)} \cdot v_{i,m}) \leq p^d \cdot 2^{\sum_{q=1}^d \max(M_q,P_q)+|Q|}. \tag{6.36}$$

From (6.33) and (6.36) we deduce that

$$\sum_{i,j} \#flops(\sum_m \sum_k) \leq p^d \cdot (2^{\sum_{q=1}^d \max(Q_q,N_q)+|P|} + 2^{\sum_{q=1}^d \max(M_q,P_q)+|Q|}). \tag{6.37}$$

Analogously we estimate the computational effort requested by the reversed inner sum in (6.31) (first over m, then over k).

$$\sum_{i,j} \#flops(\sum_k \sum_m) \leq p^d \cdot (2^{\sum_{q=1}^d \max(M_q,P_q)+|N|} + 2^{\sum_{q=1}^d \max(Q_q,N_q)+|M|}). \tag{6.38}$$

From (6.37) and (6.38) we conclude

$$\sum_{i,j} \#flops(\sum_{k,m}) \leq 2 \cdot p^d \cdot 2^{f(Q,M,N,P)}, \tag{6.39}$$

where the function f is defined in the following manner:

$$\begin{aligned} f(Q,M,N,P) \quad = \quad & \min\{\max(\sum_{q=1}^d \max(Q_q,N_q)+|P|,\ \sum_{q=1}^d \max(M_q,P_q)+|Q|), \\ & \max(\sum_{q=1}^d \max(M_q,P_q)+|N|,\ \sum_{q=1}^d \max(Q_q,N_q)+|M|)\}. \end{aligned} \tag{6.40}$$

Next we claim that the function f satisfies the upper estimate

$$f(Q,M,N,P) \leq d \cdot \max(|Q|_\infty + |M|_\infty, |N|_\infty + |P|_\infty). \tag{6.41}$$

To see this, we first observe that, by definition, f trivially satisfies

$$\begin{aligned} f(Q,M,N,P) \quad \leq \quad & d \cdot \min\{\max(|Q|_\infty + |P|_\infty, |N|_\infty + |P|_\infty, |M|_\infty + |Q|_\infty), \\ & \max(|M|_\infty + |N|_\infty, |P|_\infty + |N|_\infty, |Q|_\infty + |M|_\infty)\}. \end{aligned} \tag{6.42}$$

Now, if (6.41) were not true, it would follow from (6.42)

$$\begin{aligned} |Q|_\infty + |P|_\infty \quad > \quad & \max(|Q|_\infty + |M|_\infty, |N|_\infty + |P|_\infty), \\ |M|_\infty + |N|_\infty \quad > \quad & \max(|Q|_\infty + |M|_\infty, |N|_\infty + |P|_\infty), \end{aligned} \tag{6.43}$$

Since these two inequalities should hold simultaneously, we reach easily the contradiction $|Q|_\infty > |N|_\infty$ and $|N|_\infty > |Q|_\infty$. Therefore (6.41) is proved.
Note that the minimum in (6.42) is attained either by the first term, in case $|Q|_\infty + |P|_\infty \leq |M|_\infty + |N|_\infty$, or by the second one, if the opposite inequality holds: $|Q|_\infty + |P|_\infty \geq |M|_\infty + |N|_\infty$. This justifies the if-else selection on which our Algorithm 6.13 relies.

From (6.31), (6.39) and (6.41) we conclude that

$$\#flops\,(x \longrightarrow \hat{S}^{L,L}x) \;\leq\; 2\cdot p^d \cdot \sum_{\substack{|N|_\infty+|P|_\infty\leq L\\ |M|_\infty+|Q|_\infty\leq L}} 2^{d\cdot\max(|Q|_\infty+|M|_\infty,|N|_\infty+|P|_\infty)}, \tag{6.44}$$

which in turn implies the desired (but not optimal, as far as the exponent of L is concerned) upper bound

$$\#flops\,(x \longrightarrow \hat{S}^{L,L}x) \;\leq\; 2\cdot(\frac{p}{4})^d \cdot L^{4d+1}\cdot 2^{L\cdot d}. \tag{6.45}$$

$\diamond$

Summing up the results contained in (5.23), (6.2) and (6.30), we deduce the efficiency of the algorithm we have presented. Roughly speaking, the FE-approximation of C_u in $D\times D$ can be obtained in the same complexity, ignoring logarithmic terms, as the approximation of E_u, which solves a classical mixed boundary problem in D.

Theorem 6.15 *The computation of a numerical solution $C^L_{u,l}$ of (3.11) in the case $D=(0,1)^d$, $\Gamma_1=\emptyset$, under the regularity assumption $C_u\in H^{s+1,s+1}_{(0)}(D\times D)$, for some $s>0$, and with the relative accuracy*

$$\begin{aligned}\|C_u - C^L_{u,l}\|_{H^{1,1}_{(0)}(D\times D)} &\lesssim (\log N_3)^{1/2+\delta}\cdot N_3^{-\delta}\cdot\|C_u\|_{H^{s+1,s+1}_{(0)}(D\times D)} &(6.46)\\ &\leq O((\log N_1)^{1/2+\delta-\delta d}N_1^{-\delta}) &(6.47)\end{aligned}$$

can be performed, by the cg-method, using at most

$$C_2\cdot(\log N_3)^{3d+2}\cdot N_3 = O((\log N_1)^{4d+2}\cdot N_1) \tag{6.48}$$

floating point operations. Here $N_1=(2^{L+1}-1)^d$ denotes the number of degrees of freedom used to approximate E_u in D, while $N_3=(L\cdot 2^{L+1}+1)^d$ is the number of degrees of freedom used in $D\times D$.

Remark 6.16 *The storage requirements of the algorithm we have presented are of the same order as the required number of operations. More precisely, the amount of information to be saved at each step when using the cg-method can be estimated by $O(N_1\cdot(\log N_1)^d)$. Indeed, we need to store first the matrix S^L, which contains, due to (6.27), at most $O(N_1\cdot(\log N_1)^d)$ nontrivial entries. The positioning of these entries is necessary, but this is easily achieved when using an explicit prewavelet basis as in Examples 6.2 or 6.3. This supplementary amount of information is bounded again by $O(N_1\cdot(\log N_1)^d)$. Next we have to store the first r.h.s. of (6.1), which is a vector with $N_3=N_1\cdot(\log N_1)^d$ components. The recursive procedure of the cg-method preserves then, up to a multiplicative constant, the memory requirements.*

7 Numerical experiments

We present here elementary numerical results that are to be compared with the theoretical ones we have obtained in Sections 5 and 6. We include the two examples introduced in Section 4, involving exponential and Dirac covariance (for simplicity we assume that $D=(-1,1)$ and $A=1$). We investigate then a third situation in which the coefficient A is non-constant. We mention that each figure presents two curves: the one corresponding to the theoretical result (dashed) and the one obtained numerically (solid). We also mention that in all these cases, the hat-function basis from Example 6.2 has been used to perform numerical algorithms.

The first example to be considered is therefore the Dirichlet problem

$$\begin{cases} L(\partial_x)L(\partial_y)C_u &= e^{-|x-y|} & \text{in } L^2((-1,1)^2) \\ \gamma_0(C_u) &= 0 & \text{on } \partial(-1,1)^2, \end{cases} \tag{7.1}$$

with $A(x) = \mathrm{Id}_{\mathbb{R}}, \forall x \in D = (-1,1)$, that is, $L(\partial_x) = -\Delta_x$.
Figure 1 shows the convergence of the FE-solution in this simple case, with non-singular (but also non-smooth) data and constant coefficient A. From theoretical results (Proposition 4.8 and (5.23)) follows that for this particular choice of data the rate of convergence equals $(\log N)^{3/2} \cdot N^{-1}$, where throughout this section N stands for the number of degrees of freedom.

We can also consider the slightly more general situation $C_f(x,y) = e^{-c|x-y|}$, where c is some large real parameter and let c tend to infinity. In this way we obtain, after a proper rescaling (see (4.15)), the convergence of the FE-solution with the singular r.h.s. $C_f = \delta(x-y)$ as a limiting case of the exponential covariance. The problem reads

$$\begin{cases} L(\partial_x)L(\partial_y)C_u &= \delta(x-y) & \text{in the dual space of } H_0^{1,1}((-1,1)^2) \\ \gamma_0(C_u) &= 0 & \text{on } \partial(-1,1)^2, \end{cases} \tag{7.2}$$

where $L(\partial_x) = -\Delta_x$ and Figure 2 shows the convergence of the FE-solution. The theoretical rate of convergence is, in this case, $(\log N)^{5/4} \cdot N^{-3/4}$, as a consequence of Proposition 4.10 and (5.23).
And we conclude this final section with a new example, in which all data are again smooth but the coefficient A is no longer constant. More precisely, we choose the coefficent A and the solution C_u as follows:

$$A(x) = 2 + \sin(\pi x), \quad \forall x \in D = (-1,1) \quad \text{and} \quad C_u(x,y) = (1-x^2)(1-y^2)e^{xy}. \tag{7.3}$$

The numerical results are shown in Figure 3. As in the first example, the error decays as fast as $(\log N)^{3/2} \cdot N^{-1}$ when $N \to \infty$. This curve, as well as those we have plotted before, does not have the appearance of a straight line and this is due to the logarithmic terms arising in the error estimates by sparse grids. Of course, asymptotically as $N \to \infty$, these terms do not play an essential role, but it turns out that their influence is rather strong, within the available computational range.
Finally, Figure 4 shows the performance of our algorithm 6.13 matching the theoretical estimate concerning the computational effort given by (6.48), namely $\#flops = C \cdot (\log N)^5 \cdot N$. We mention that this analysis has been done for the same example (7.3). We finally note that a slight improvement in this respect can follow from the fact that solving the discretized problem at level L already provides us with a good approximation of the solution at level $L+1$. Using the solution at level L as starting value in the conjugate gradient method at level $L+1$ enables one to reduce the computational effort above by a logarithmic factor, to $\#flops = C \cdot (\log N)^4 \cdot N$.

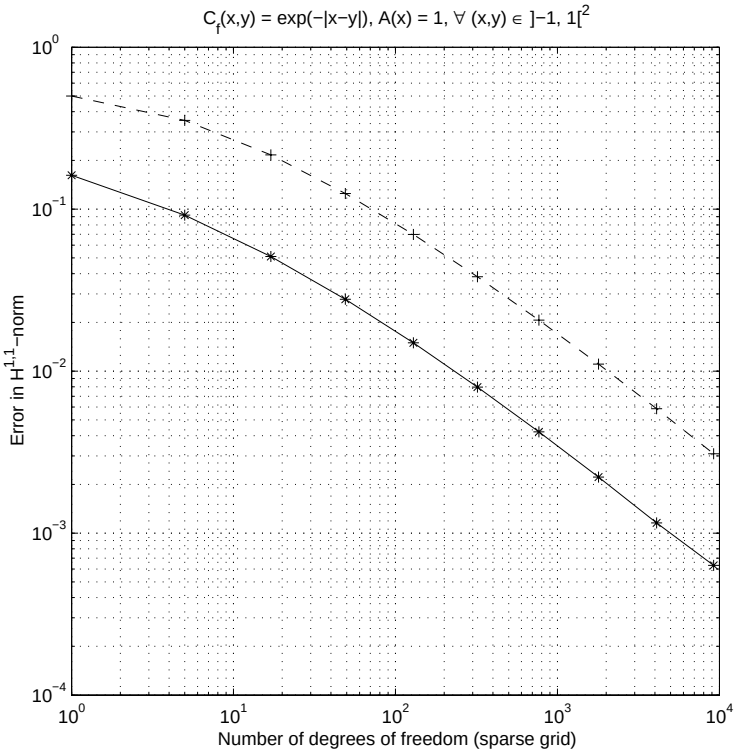

Figure 1: Convergence in the case of exponential r.h.s. and constant coefficient $A = 1$ (solid) and the bound (6.46), with $\delta = 1$ (dashed).

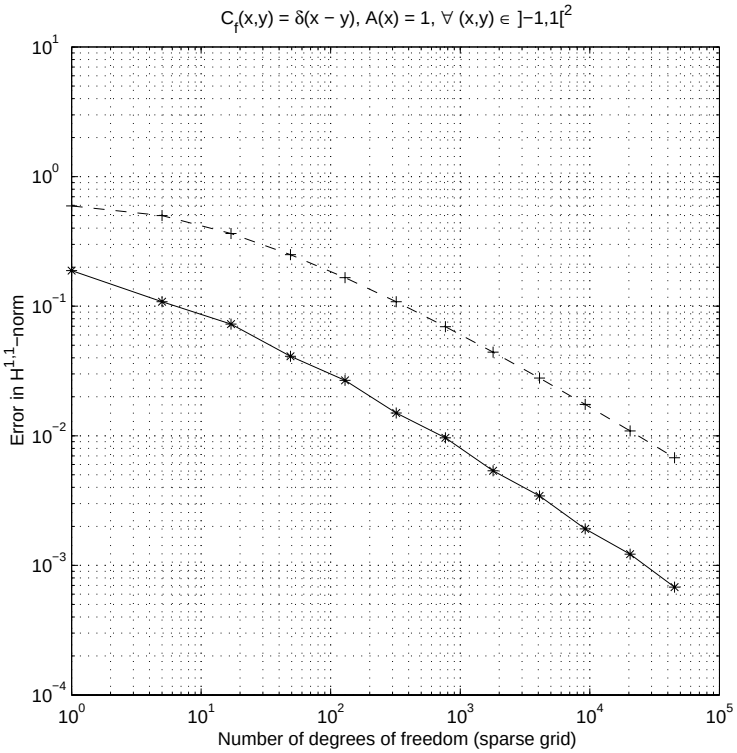

Figure 2: Convergence in the case of singular r.h.s. and constant coefficient $A = 1$ (solid) and the bound (6.46), with $\delta = 3/4$ (dashed).

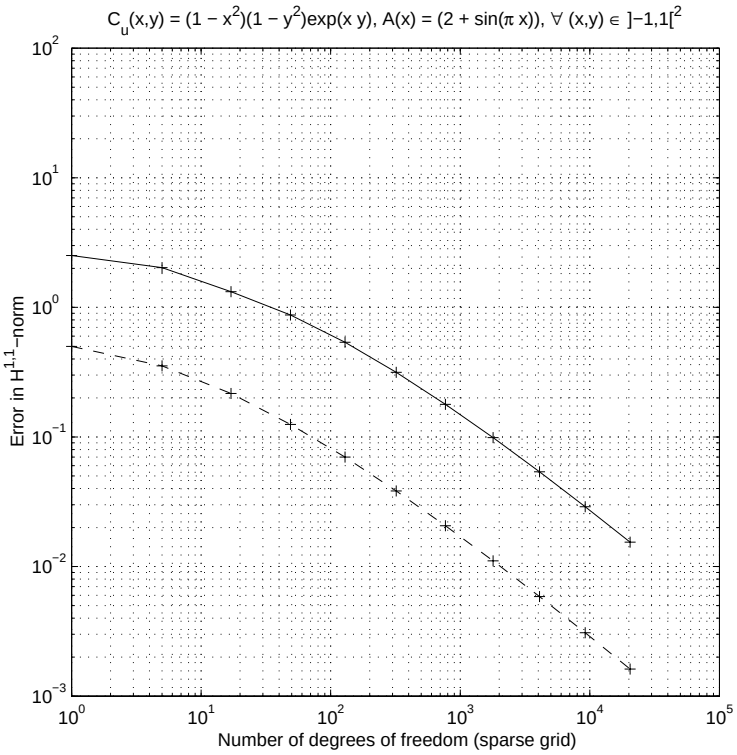

Figure 3: Convergence in the case of non-constant coefficient A (solid) and the bound (6.46), again with $\delta = 1$ (dashed) .

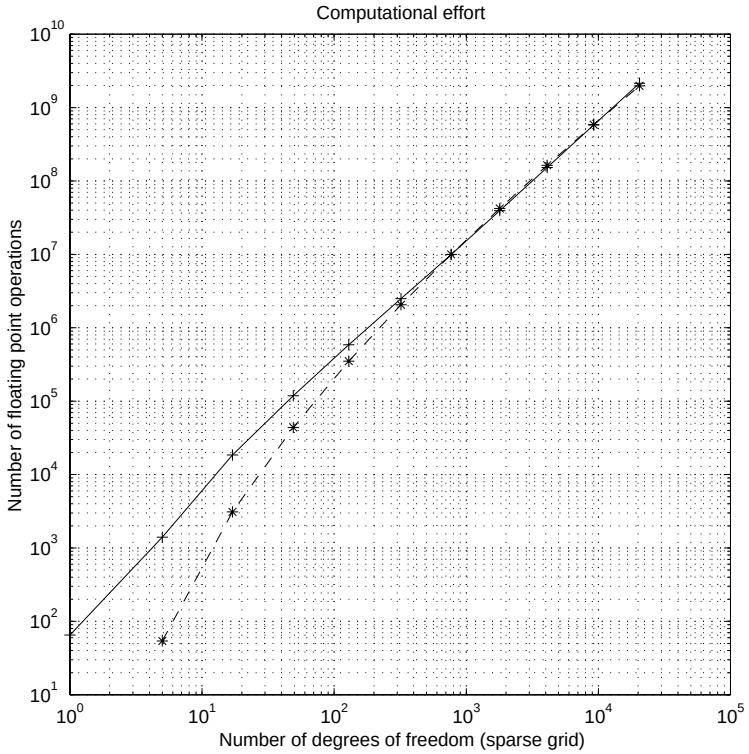

Figure 4: Comparison between the effort required by the standard cg-method based on Algorithm 6.13 (solid) and its theoretical estimate given in Theorem 6.15 by (6.48) (dashed).

References

[1] I. Babuška, *"On Randomised Solutions of Laplace's Equation"*, Časopis pro Pěstováni Matematiky, **86** (1961) pp. 269-275.

[2] I. Babuška and B. Q. Guo, *"Regularity of the solution of elliptic problems with piecewise analytic data. Part I. Boundary value problems for linear elliptic equation of second order"*, SIAM J.Math.Anal., **19** (1988) pp.172-203.

[3] W. Dahmen, *"Wavelet and multiscale methods for operator equations"*, Acta Numerica (1997), pp. 55-228.

[4] W. Dahmen, A. Kunoth, K. Urban, *"Biorthogonal Spline-Wavelets on the Interval - Stability and Moment Conditions"*, Appl. Comp. Harm. Anal. 6 (1999), pp. 132-196.

[5] W. Dahmen, R. Stevenson, *"Element-by-Element Construction of Wavelets Satisfying Stability and Moment Conditions"*, SIAM J. Numer. Anal. 37 (1999), pp. 319-352.

[6] M.K. Deb, I. Babuška, J.T. Oden, *"Solution of Stochastic Partial Differential Equations Using Galerkin Finite Element Techniques"*, Preprint, University of Texas at Austin, 2001.

[7] R. G. Ghanem, P. D. Spanos, *"Stochastic finite elements: a spectral approach"* , Springer-Verlag, 1991.

[8] D. Gilbarg, N.S. Trudinger, *"Elliptic Partial Differential Equations of Second Order"* , Grundlehren, **224**, Springer-Verlag, 1977.

[9] G. Golub, C.F. Van Loan, *"Matrix Computations"*, 4th edition, Johns Hopkins University Press, 1996.

[10] M. Griebel, P. Oswald, and T. Schiekofer, *"Sparse grids for boundary integral equations"*, Numer. Mathematik, 83(2), 1999, pp. 279-312.

[11] H. Holden, B. Oksendal, J. Uboe, T. Zhang, *"Stochastic Partial Differential Equations: A Modeling, White Noise Functional Approach"*, Birkhäuser, 1996.

[12] J. Kampé de Fériet, *"Random Solutions of Partial Differential Equations"* , Proc. of the Third Berkeley Symp. on Math. Statistics and Probability, III pp. 199-208.

[13] M. Kleiber, T.D. Hien, *"The Stochastic Finite Element Method"*, John Wiley & Sons, 1992.

[14] P.E. Kloeden, E. Platen, *"Numerical solution of stochastic differential equations"*, 3rd edition, Springer-Verlag, 1999.

[15] A.M. Matache, Ch. Schwab, *"Sparse Two Scale FEM for Homogenization Problems"*, to appear in Journal of Scientific Computing.

[16] B. Oksendal, *"Stochastic differential equations: an introduction with applications"*, 3rd edition, Springer-Verlag, 1992.

[17] P. Protter, *"Stochastic integration and differential equations: a new approach"*, 3rd edition, Springer-Verlag, 1995.

[18] R. S. Strichartz, *"A guide to distribution theory and Fourier transforms"*, CRC Press, Boca Raton, 1994.

[19] A.M. Yaglom, *"An Introduction to the Theory of Stationary Random Functions"*, Prentice-Hall, Englewood Cliffs, New Jersey, 1962.

[20] K. Yosida, *"Functional Analysis"* , Grundlehren, **123**, Springer-Verlag, 1964.

[21] Ch. Zenger, *"Sparse Grids"*, in "Parallel Algorithms for PDE's ", Proceedings of the 6th GAMM-Seminar, Kiel, 1990.