ESAIM: M2AN 42 (2008) 777–819
DOI: 10.1051/m2an:2008027

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

SPARSE FINITE ELEMENT APPROXIMATION OF HIGH-DIMENSIONAL TRANSPORT-DOMINATED DIFFUSION PROBLEMS

Christoph Schwab[1], Endre Süli[2] and Radu Alexandru Todor[1]

Abstract. We develop the analysis of stabilized sparse tensor-product finite element methods for high-dimensional, non-self-adjoint and possibly degenerate second-order partial differential equations of the form $-a : \nabla\nabla u + b \cdot \nabla u + cu = f(x)$, $x \in \Omega = (0,1)^d \subset \mathbb{R}^d$, where $a \in \mathbb{R}^{d\times d}$ is a symmetric positive semidefinite matrix, using piecewise polynomials of degree $p \geq 1$. Our convergence analysis is based on new high-dimensional approximation results in sparse tensor-product spaces. We show that the error between the analytical solution u and its stabilized sparse finite element approximation u_h on a partition of Ω of mesh size $h = h_L = 2^{-L}$ satisfies the following bound in the streamline-diffusion norm $|||\cdot|||_{\mathrm{SD}}$, provided u belongs to the space $\mathcal{H}^{k+1}(\Omega)$ of functions with square-integrable mixed $(k+1)$st derivatives:

$$|||u - u_h|||_{\mathrm{SD}} \leq C_{p,t} d^2 \max\{(2-p)_+, \kappa_0^{d-1}, \kappa_1^d\}(|\sqrt{a}|h_L^t + |b|^{\frac{1}{2}} h_L^{t+\frac{1}{2}} + c^{\frac{1}{2}} h_L^{t+1})|u|_{\mathcal{H}^{t+1}(\Omega)},$$

where $\kappa_i = \kappa_i(p,t,L)$, $i = 0,1$, and $1 \leq t \leq \min(k,p)$. We show, under various mild conditions relating L to p, L to d, or p to d, that in the case of elliptic transport-dominated diffusion problems $\kappa_0, \kappa_1 \in (0,1)$, and hence for $p \geq 2$ the 'error constant' $C_{p,t} d^2 \max\{(2-p)_+, \kappa_0^{d-1}, \kappa_1^d\}$ exhibits exponential decay as $d \to \infty$; in the case of a general symmetric positive semidefinite matrix a, the error constant is shown to grow no faster than $\mathcal{O}(d^2)$. In any case, in the absence of assumptions that relate L, p and d, the error $|||u - u_h|||_{\mathrm{SD}}$ is still bounded by $\kappa_*^{d-1} |\log_2 h_L|^{d-1} \mathcal{O}(|\sqrt{a}|h_L^t + |b|^{\frac{1}{2}} h_L^{t+\frac{1}{2}} + c^{\frac{1}{2}} h_L^{t+1})$, where $\kappa_* \in (0,1)$ for all $L, p, d \geq 2$.

Mathematics Subject Classification. 65N30.

Received March 7, 2007. Received February 11, 2008.
Published online July 30, 2008.

Dedicated to Henryk Woźniakowski, on the occasion of his 60th birthday.

Keywords and phrases. High-dimensional Fokker-Planck equations, partial differential equations with nonnegative characteristic form, sparse finite element method.

[1] Seminar für Angewandte Mathematik, Eidgenössische Technische Hochschule, 8092 Zürich, Switzerland.
schwab@sam.math.ethz.ch; todor@math.ethz.ch

[2] University of Oxford, Computing Laboratory, Wolfson Building, Parks Road, Oxford OX1 3QD, UK.
endre.süli@comlab.ox.ac.uk

Article published by EDP Sciences

© EDP Sciences, SMAI 2008

1. Introduction

Suppose that $\Omega := (0,1)^d$, $d \geq 2$, and that $a = (a_{ij})_{i,j=1}^d$ is a symmetric positive semidefinite matrix with entries $a_{ij} \in \mathbb{R}$, $i,j = 1,\ldots,d$. In other words,

$$a^\top = a \qquad \text{and} \qquad \xi^\top a\,\xi \geq 0 \quad \forall \xi \in \mathbb{R}^d.$$

Suppose further that $b \in \mathbb{R}^d$ and $c \in \mathbb{R}_{\geq 0}$, and let $f \in \mathrm{L}^2(\Omega)$. We shall consider the partial differential equation

$$-a : \nabla\nabla u + b \cdot \nabla u + cu = f(x), \quad x \in \Omega, \tag{1.1}$$

subject to suitable boundary conditions on $\partial\Omega$ that will be stated below. Here $\nabla\nabla u$ is the $d \times d$ Hessian matrix of u whose (i,j) entry is $\partial^2 u/\partial x_i\,\partial x_j$, $i,j = 1,\ldots,d$. For two $d \times d$ matrices A and B, we define their scalar product $A : B := \sum_{i,j=1}^d A_{ij}B_{ij}$. The induced norm, called the Frobenius norm, is defined by $|A| = (A : A)^{\frac{1}{2}}$.

The real-valued polynomial $\alpha \in \mathcal{P}^2(\mathbb{R}^d;\mathbb{R})$ of degree ≤ 2 defined by

$$\xi \in \mathbb{R}^d \;\mapsto\; \alpha(\xi) := \xi^\top a\,\xi \in \mathbb{R}$$

is called the *characteristic polynomial* or *characteristic form* of the differential operator

$$u \;\mapsto\; \mathcal{L}u := -a : \nabla\nabla u + b \cdot \nabla u + cu$$

featuring in (1.1) and, under our hypotheses on the matrix a, the equation (1.1) is referred to as a *partial differential equation with nonnegative characteristic form* (*cf.* Oleĭnik and Radkevič [20]). In order to avoid trivialities, we shall assume throughout that $|a| + |b| > 0$.

For the sake of simplicity of presentation we shall confine ourselves to differential operators $\mathcal{L}$ with constant coefficients. In this case,

$$a : \nabla\nabla u = \nabla \cdot (a\nabla u) = \nabla\nabla : (au) \qquad \text{and} \qquad b \cdot \nabla u = \nabla \cdot (bu).$$

Partial differential equations with nonnegative characteristic form frequently arise as mathematical models in physics and chemistry [28] (*e.g.* in the kinetic theory of polymers [21]; see also [2,3,18]; and coagulation-fragmentation problems [17]), molecular biology [10], and mathematical finance. Important special cases of these equations include the following:

(a) when the diffusion matrix $a = a^\top$ is positive definite, (1.1) is an elliptic partial differential equation;
(b) when $a \equiv 0$ and the transport direction $b \neq 0$, the partial differential equation (1.1) is a first-order hyperbolic equation;
(c) when

$$a = \begin{pmatrix} \alpha & 0 \\ 0 & 0 \end{pmatrix},$$

where α is a $(d-1) \times (d-1)$ symmetric positive definite matrix and $b = (0,\ldots,0,1)^\top \in \mathbb{R}^d$, (1.1) is a parabolic partial differential equation, with time-like direction b.

In addition to these classical types, the family of partial differential equations with nonnegative characteristic form encompasses a range of other linear second-order partial differential equations, such as degenerate elliptic equations and ultra-parabolic equations. According to a result of Hörmander [13] (*cf.* Thm. 11.1.10 on p. 67), second-order hypoelliptic operators have non-negative characteristic form, after possible multiplication by -1, so they too fall into this category.

For classical types of partial differential equations, such as those listed under (a), (b) and (c) above, rich families of reliable, stable and highly accurate numerical techniques have been developed. Yet, only isolated attempts have been made to explore computational aspects of the class of partial differential equations with

nonnegative characteristic form as a whole (*cf.* [14,15]). In particular, there has been only a limited amount of research to date on the numerical analysis of high-dimensional partial differential equations with nonnegative characteristic form (*cf.* Süli [25,26]).

The field of stochastic analysis is a particularly fertile source of equations of this kind (see, for example, [4]): the progressive Kolmogorov equation satisfied by the probability density function $\psi(x_1,\ldots,x_d,t)$ of a d-component vectorial stochastic process $X(t) = (X_1(t),\ldots,X_d(t))^\top$, which is the solution of a system of stochastic differential equations including Brownian noise is a partial differential equation with nonnegative characteristic form in the $d+1$ variables $(x,t) = (x_1,\ldots,x_d,t)$. To be more precise, consider the stochastic differential equation:

$$\mathrm{d}X(t) = b(X(t))\,\mathrm{d}t + \sigma(X(t))\,\mathrm{d}W(t), \qquad X(0) = X,$$

where $W = (W_1,\ldots,W_p)^\top$ is a p-dimensional Wiener process adapted to a filtration $\{\mathcal{F}_t\,,\ t \geq 0\}$, $b \in \mathrm{C}^1_b(\mathbb{R}^d;\mathbb{R}^d)$ is the drift vector, and $\sigma \in \mathrm{C}^2_b(\mathbb{R}^d,\mathbb{R}^{d\times p})$ is the diffusion matrix. Here $\mathrm{C}^k_b(\mathbb{R}^n,\mathbb{R}^m)$ denotes the space of bounded and continuous mappings from $\mathbb{R}^n$ into $\mathbb{R}^m$, $m,n \geq 1$, all of whose partial derivatives of order k or less are bounded and continuous on $\mathbb{R}^n$. When the subscript $_b$ is absent, boundedness is not enforced.

Assuming that the random variable $X(t) = (X_1(t),\ldots,X_d(t))^\top$ has a probability density function $\psi \in \mathrm{C}^{2,1}(\mathbb{R}^d \times [0,\infty),\mathbb{R})$, then ψ is the solution of the initial-value problem

$$\begin{aligned} \frac{\partial\psi}{\partial t}(x,t) &= (A\psi)(x,t), && x \in \mathbb{R}^d,\ t > 0,\\ \psi(x,0) &= \psi_0(x), && x \in \mathbb{R}^d, \end{aligned}$$

where the differential operator $A:\ \mathrm{C}^2(\mathbb{R}^d;\mathbb{R}) \to \mathrm{C}^0(\mathbb{R}^d;\mathbb{R})$ is defined by

$$A\psi := -\sum_{j=1}^{d} \frac{\partial}{\partial x_j}\left(b_j(x)\psi\right) + \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i\,\partial x_j}\left(a_{ij}(x)\psi\right),$$

with $a(x) = \sigma(x)\,\sigma^\top(x) \geq 0$ (see Cor. 5.2.10 on p. 135 in [16]). Thus, ψ is the solution of the initial-value problem

$$\begin{aligned} \frac{\partial\psi}{\partial t} + \sum_{j=1}^{d} \frac{\partial}{\partial x_j}(b_j(x)\psi) &= \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i\,\partial x_j}\left(a_{ij}(x)\psi\right), && x \in \mathbb{R}^d,\ t \geq 0,\\ \psi(x,0) &= \psi_0(x), && x \in \mathbb{R}^d, \end{aligned}$$

where, for each $x \in \mathbb{R}^d$, $a(x)$ is a $d \times d$ symmetric positive semidefinite matrix. The progressive Kolmogorov equation $\frac{\partial\psi}{\partial t} = A\psi$ is a partial differential equation with nonnegative characteristic form, called a Fokker-Planck equation.

The operator A is generally nonsymmetric (since, typically, $b \neq 0$) and degenerate (since, in general, $a(x) = \sigma(x)\sigma^\top(x)$ has nontrivial kernel). In addition, since the (possibly large) number d of equations in the system of stochastic differential equations is equal to the number of components of the independent variable x of the probability density function ψ, the Fokker-Planck equation may be high-dimensional.

The focus of the present paper is the construction and analysis of finite element approximations to *high-dimensional* partial differential equations with non-negative characteristic form. Specifically, our aim is to extend the results from [25,26], developed for the case of sparse tensor-product finite element spaces consisting of piecewise multilinear functions, to polynomials of degree $p \geq 1$. The paper is structured as follows. We shall state in Section 2 the appropriate boundary conditions for the model equation (1.1), derive the weak formulation of the resulting boundary value problem, and show the existence of a unique weak solution. Section 3 is devoted to the construction of a hierarchical finite element space for univariate functions. The tensorization of this space and the subsequent sparsification of the resulting tensor-product space are described in Section 4;

our chief objective is to reduce the computational complexity of the discretization without adversely effecting the approximation properties of the finite element space. In Sections 5 and 6 we build a stabilized finite element method over the sparse tensor-product space, and we explore its stability and convergence.

The convergence analysis relies on new high-dimensional approximation results in sparsified tensor-product spaces, based on continuous piecewise polynomials of degree $p \geq 1$ on a mesh of granularity $h_L = 2^{-L}$, in the L^2 and H^1 norms. We highlight several scenarios in which the error-constants in these approximation results exhibit exponential decay as functions of the dimension d, for all $p \geq 1$. These bounds are related to those in the recent work of Griebel [11], where similar decay of the error constant as a function of d was proved in the H^1 seminorm in the special case of $p = 1$ for "energy-norm-based" sparse-grid-spaces. Using our approximation results, under various mild assumptions that relate L to p, L to d, or p to d, we then show that in the case of elliptic transport-diffusion problems the error constant of the stabilized sparse finite element method exhibits exponential decay as $d \to \infty$ for all $p \geq 2$; more generally, when the characteristic form of the partial differential equation is non-negative, the error constant is shown to grow no faster than $\mathcal{O}(d^2)$. In any case, the stabilized sparse finite element method exhibits an optimal rate of convergence with respect to the mesh size h_L, up to the factor $\kappa_*^{d-1}|\log_2 h_L|^{d-1}$, where $\kappa_* \in (0,1)$ for all $L, p, d \geq 2$; thus, even when the polylogarithmic factor $|\log_2 h_L|^{d-1}$ is genuinely present in the error bound, it is modulated by an exponentially small term κ_*^{d-1}.

Our analysis is fairly general, in the sense that only two basic structural properties of the univariate finite element space are used in the subsequent analysis: namely, (1) that the univariate finite element space is hierarchically stratified, *viz.*, it can be written as a direct sum of so-called increment spaces, and (2) that there exists a projector onto the univariate finite element space that exhibits optimal approximation properties in the L^2 and H^1 norms. The specific choice of basis in the finite element space does not explicitly enter into our error analysis, as it does not affect the asymptotic rate of convergence. Of course, the implementation of the method will necessitate that a choice of basis is made; indeed, the specific choice of basis will strongly influence the sparsity structure and conditioning of the matrix in the resulting linear system. These (and other) questions (such as numerical integration in high-dimensional finite element algorithms) are important and we shall briefly comment on them in the concluding section, although, strictly speaking, they are beyond the scope of the present paper and will be therefore considered in detail elsewhere.

The origins of sparse tensor-product constructions and hyperbolic cross spaces can be traced back to the works of Babenko [1] and Smolyak [24]; we refer to the papers of Temlyakov [27], DeVore *et al.* [8] for the study of high-dimensional approximation problems, to the works of Wasilkowski and Woźniakowski [30] and Novak and Ritter [19] for high-dimensional integration problems and associated complexity questions, to the paper of Zenger [31] for an early contribution to sparse tensor-product finite element methods, to the articles by von Petersdorff and Schwab [29] and Hoang and Schwab [12] for the analysis of sparse-grid methods for high-dimensional parabolic and elliptic multiscale problems, respectively, and to the recent *Acta Numerica* article of Bungartz and Griebel [7] for a detailed survey of the field of sparse-grid methods.

2. Boundary conditions and weak formulation

Before embarking on the construction of the numerical algorithm, we shall introduce the necessary boundary conditions and the weak formulation of the model boundary-value problem on $\Omega = (0,1)^d$ for the equation (1.1).

Let Γ denote the union of all $(d-1)$-dimensional open faces of the domain $\Omega = (0,1)^d$. On recalling that, by hypothesis, $a = a^\top$ and $\alpha(\xi) = \xi^\top a\, \xi \geq 0$ for all $\xi \in \mathbb{R}^d$, we define the subset Γ_0 of Γ by

$$\Gamma_0 := \{x \in \Gamma : \alpha(\nu(x)) > 0\};$$

here $\nu(x)$ denotes the unit outward normal vector to Ω at $x \in \Gamma$. The set Γ_0 can be thought of as the *elliptic part* of Γ. The complement $\Gamma \backslash \Gamma_0$ of Γ_0 relative to Γ is referred to as the *hyperbolic part* of Γ. We note that, by definition, $\alpha = 0$ on $\Gamma \setminus \Gamma_0$.

On introducing the *Fichera function*

$$x \in \Gamma \mapsto \beta(x) := b \cdot \nu(x) \in \mathbb{R}$$

defined on Γ, we partition $\Gamma \backslash \Gamma_0$ as follows:

$$\Gamma_- := \{x \in \Gamma \backslash \Gamma_0 : \beta < 0\}, \qquad \Gamma_+ := \{x \in \Gamma \backslash \Gamma_0 : \beta \geq 0\};$$

the sets Γ_- and Γ_+ are referred to as the (hyperbolic) *inflow* and *outflow* boundary, respectively. Thereby, we obtain the following decomposition of Γ:

$$\Gamma = \Gamma_0 \cup \Gamma_- \cup \Gamma_+.$$

Lemma 2.1. *Each of the sets Γ_0, Γ_-, Γ_+ is a (possibly empty) union of $(d-1)$-dimensional open faces of Ω. Moreover, each pair of opposite $(d-1)$-dimensional faces of Ω is contained either in the elliptic part Γ_0 of Γ or in its complement $\Gamma \setminus \Gamma_0 = \Gamma_- \cup \Gamma_+$, the hyperbolic part of Γ.*

Proof. Since, by hypothesis, $|a| + |b| > 0$, a and b are not simultaneously zero. If $a = 0$, then trivially $\Gamma_0 = \emptyset$. Let us therefore suppose that $a \neq 0$. Since a is a constant matrix and ν is a face-wise constant vector, Γ_0 is a union of (disjoint) $(d-1)$-dimensional open faces of Γ. Indeed, if $x \in \Gamma_0$ and y is any point that lies on the same $(d-1)$-dimensional open face of Ω as x, then $\nu(y) = \nu(x)$ and therefore $\alpha(\nu(y)) = \alpha(\nu(x)) > 0$; hence $y \in \Gamma_0$ also.

A certain $(d-1)$-dimensional open face φ of Ω is contained in Γ_0 if, and only if, the opposite face $\hat{\varphi}$ is also contained in Γ_0. To prove this, let $\varphi \subset \Gamma_0$ and let $x = (x_1, \ldots, x_i, \ldots, x_d) \in \varphi$, with Ox_i signifying the (unique) co-ordinate direction such that $\nu(x) \| Ox_i$; here $O = (0, \ldots, 0)$. In other words, $x_i \in \{0, 1\}$, and the $(d-1)$-dimensional face φ to which x belongs is orthogonal to the co-ordinate direction Ox_i. Hence, the point $\hat{x} = (x_1, \ldots, |x_i - 1|, \ldots, x_d)$ lies on the $(d-1)$-dimensional open face $\hat{\varphi}$ of Ω that is opposite the face φ (*i.e.*, $\hat{\varphi} \| \varphi$), and $\nu(\hat{x}) = -\nu(x)$. As α is a homogeneous function of degree 2 on Γ_0, it follows that

$$\alpha(\nu(\hat{x})) = \alpha(-\nu(x)) = (-1)^2 \, \alpha(\nu(x)) = \alpha(\nu(x)) > 0,$$

which implies that $\hat{x} \in \Gamma_0$. By what we have shown before, we deduce that the entire face $\hat{\varphi}$ is contained in Γ_0.

Similarly, if $b = 0$, then $\Gamma_- = \emptyset$ and $\Gamma_+ = \Gamma \setminus \Gamma_0$. Let us therefore suppose that $b \neq 0$. Since b is a constant vector, each of Γ_- and Γ_+ is a union of $(d-1)$-dimensional open faces of Γ. If a certain $(d-1)$-dimensional open face φ is contained in Γ_-, then the opposite face $\hat{\varphi}$ is contained in the set Γ_+.

We note in passing, however, that if $\varphi \subset \Gamma_+$ then the opposite face $\hat{\varphi}$ need not be contained in Γ_-; indeed, if $\varphi \subset \Gamma_+$ and $\beta = 0$ on φ then $\beta = 0$ on $\hat{\varphi}$ also, so then both φ and the opposite face $\hat{\varphi}$ are contained in Γ_+. Of course, if $\beta > 0$ on $\varphi \subset \Gamma_+$, then $\beta < 0$ on the opposite face $\hat{\varphi}$, and then $\hat{\varphi} \subset \Gamma_-$. □

Lemma 2.1 motivates the following definition.

Definition 2.1. For $i \in \{0, \ldots, d\}$, a co-ordinate direction Ox_i that is orthogonal to a pair of faces of $\Omega = (0,1)^d$ which belong to Γ_0 will be called an elliptic co-ordinate direction. Otherwise, Ox_i will be called a hyperbolic co-ordinate direction.

We consider the following boundary-value problem: Find u such that

$$\mathcal{L}u := -a : \nabla\nabla u + b \cdot \nabla u + cu = f \qquad \text{in } \Omega, \tag{2.1}$$

$$u = 0 \qquad \text{on } \Gamma_0 \cup \Gamma_-. \tag{2.2}$$

Before stating the variational formulation of (2.1), (2.2), we recall the following result from [14].

Lemma 2.2. *Suppose that $M \in \mathbb{R}^{d \times d}$ is a $d \times d$ symmetric positive semidefinite matrix. If $\xi \in \mathbb{R}^d$ satisfies $\xi^\top M \xi = 0$, then $M\xi = 0$.*

Since $\nu^\top a\nu = 0$ on $\Gamma \setminus \Gamma_0$ and $a \in \mathbb{R}^{d\times d}$ is a symmetric positive semidefinite matrix, we deduce from Lemma 2.2 with $M = a$ and $\xi = \nu$ that

$$a\nu = 0 \qquad \text{on } \Gamma \setminus \Gamma_0. \tag{2.3}$$

Let us suppose for a moment that (2.1), (2.2) has a solution u in $\mathrm{H}^2(\Omega)$. Thanks to our assumption that a is a constant matrix, we have that

$$a : \nabla\nabla u = \nabla \cdot (a\nabla u).$$

Furthermore, $a\nabla u \in [\mathrm{H}^1(\Omega)]^d$, which implies that the normal trace $\gamma_{\nu,\partial\Omega}(a\nabla u)$ of $a\nabla u$ on $\partial\Omega$ belongs to $\mathrm{H}^{\frac{1}{2}}(\partial\Omega)$. By virtue of (2.3),

$$\gamma_{\nu,\partial\Omega}(a\nabla u)|_{\Gamma\setminus\Gamma_0} = 0.$$

Note also that $\mathrm{meas}_{d-1}(\partial\Omega \setminus \Gamma) = 0$. Hence

$$\int_{\partial\Omega} \gamma_{\nu,\partial\Omega}(a\nabla u) \cdot \gamma_{0,\partial\Omega}(v)\,\mathrm{d}s = \int_{\Gamma} \gamma_{\nu,\partial\Omega}(a\nabla u)|_\Gamma \cdot \gamma_{0,\partial\Omega}(v)|_\Gamma\,\mathrm{d}s = 0 \tag{2.4}$$

for all $v \in \mathcal{V}$, where

$$\mathcal{V} := \left\{v \in \mathrm{H}^1(\Omega) : \gamma_{0,\partial\Omega}(v)|_{\Gamma_0} = 0\right\}.$$

This observation will be of key importance. On multiplying the partial differential equation (2.1) by $v \in \mathcal{V}$ and integrating by parts, we find that

$$(a\nabla u, \nabla v) - (u, \nabla \cdot (bv)) + (cu, v) + \langle u, v\rangle_{\Gamma_+} = (f, v) \qquad \forall v \in \mathcal{V}, \tag{2.5}$$

where $(\cdot,\cdot)$ denotes the L^2 inner-product over Ω and

$$\langle w, v\rangle_{\Gamma_\pm} := \int_{\Gamma_\pm} |\beta| wv\ \mathrm{d}s,$$

with β signifying the Fichera function $b \cdot \nu$, as before. We note that in the transition to (2.5) the boundary integral term on Γ, which arises in the course of partial integration from the $-\nabla \cdot (a\nabla u)$ term, vanishes by virtue of (2.4), while the boundary integral term on $\Gamma \setminus \Gamma_+ = \Gamma_0 \cup \Gamma_-$ resulting from the $b \cdot \nabla u$ term on partial integration disappears since $u = 0$ on this set by (2.2).

The form of (2.5) serves as motivation for the statement of the weak formulation of (2.1), (2.2) presented below. We consider the inner product $(\cdot,\cdot)_{\mathcal{H}}$ defined by

$$(w, v)_{\mathcal{H}} := (a\nabla w, \nabla v) + (w, v) + \langle w, v\rangle_{\Gamma_-\cup\Gamma_+},$$

and denote by $\mathcal{H}$ the closure of the space $\mathcal{V}$ in the norm $\|\cdot\|_{\mathcal{H}}$ defined by $\|w\|_{\mathcal{H}} := (w, w)_{\mathcal{H}}^{\frac{1}{2}}$. Clearly, $\mathcal{H}$ is a Hilbert space. For $w \in \mathcal{H}$ and $v \in \mathcal{V}$, we now consider the bilinear form $B(\cdot,\cdot) : \mathcal{H} \times \mathcal{V} \to \mathbb{R}$ defined by

$$B(w, v) := (a\nabla w, \nabla v) - (w, \nabla \cdot (bv)) + (cw, v) + \langle w, v\rangle_{\Gamma_+},$$

and for $v \in \mathcal{V}$ we introduce the linear functional $L : \mathcal{V} \to \mathbb{R}$ by

$$L(v) := (f, v).$$

We shall say that $u \in \mathcal{H}$ is a *weak solution* to the boundary-value problem (2.1), (2.2) if

$$B(u, v) = L(v) \qquad \forall v \in \mathcal{V}. \tag{2.6}$$

The existence of a unique weak solution is guaranteed by the following theorem (*cf.* also Thm. 1.4.1 on p. 29 of [20]).

Theorem 2.1. *Suppose that $c \in \mathbb{R}_{>0}$. For each $f \in \mathrm{L}^2(\Omega)$, there exists a unique u in a Hilbert subspace $\hat{\mathcal{H}}$ of $\mathcal{H}$ such that (2.6) holds.*

Proof. For $v \in \mathcal{V}$ fixed, we deduce by means of the Cauchy-Schwarz inequality that

$$B(w,v) \le K_1 \|w\|_{\mathcal{H}} \|v\|_{\mathrm{H}^1(\Omega)} \qquad \forall w \in \mathcal{H},$$

where we have used the trace theorem for $\mathrm{H}^1(\Omega)$. Thus $B(\cdot, v)$ is a bounded linear functional on the Hilbert space $\mathcal{H}$. By the Riesz representation theorem, there exists a unique element $T(v)$ in $\mathcal{H}$ such that

$$B(w,v) = (w, T(v))_{\mathcal{H}} \qquad \forall w \in \mathcal{H}.$$

Since B is bilinear, it follows that $T : v \to T(v)$ is a linear operator from $\mathcal{V}$ into $\mathcal{H}$. Next we show that T is injective. Note that

$$B(v,v) = (a\nabla v, \nabla v) - (v, \nabla \cdot (bv)) + (cv, v) + \langle v, v\rangle_{\Gamma_+} \qquad \forall v \in \mathcal{V}.$$

On integrating by parts in the second term on the right-hand side we deduce that

$$\begin{aligned} B(v,v) &= (a\nabla v, \nabla v) + c\|v\|^2_{\mathrm{L}^2(\Omega)} + \tfrac{1}{2}\langle v, v\rangle_{\Gamma_- \cup \Gamma_+} \\ &\ge K_0 \|v\|^2_{\mathcal{H}} \qquad \forall v \in \mathcal{V}, \end{aligned}$$

where $K_0 = \min\{c, \frac{1}{2}\} > 0$. Hence

$$(v, T(v))_{\mathcal{H}} \ge K_0 \|v\|^2_{\mathcal{H}} \qquad \forall v \in \mathcal{V}. \tag{2.7}$$

Consequently, $T : v \to T(v)$ is an injection from $\mathcal{V}$ onto the range $\mathcal{R}(T)$ of T contained in $\mathcal{H}$. Thus, $T : \mathcal{V} \to \mathcal{R}(T)$ is a bijection. Let $S = T^{-1} : \mathcal{R}(T) \to \mathcal{V}$, and let $\hat{\mathcal{H}}$ denote the closure of $\mathcal{R}(T)$ in $\mathcal{H}$. Since, by (2.7), $\|S(w)\|_{\mathcal{H}} \le (1/K_0)\|w\|_{\mathcal{H}}$ for all $w \in \mathcal{R}(T)$, it follows that $S : \mathcal{R}(T) \to \mathcal{V}$ is a continuous linear operator; therefore, it can be extended, from the dense subspace $\mathcal{R}(T)$ of $\hat{\mathcal{H}}$ to the whole of $\hat{\mathcal{H}}$, as a continuous linear operator $\hat{S} : \hat{\mathcal{H}} \to \mathcal{H}$. Furthermore, since

$$|L(v)| \le \|f\|_{\mathrm{L}^2(\Omega)} \|v\|_{\mathcal{H}} \qquad \forall v \in \mathcal{H},$$

it follows that $L \circ \hat{S} : v \in \hat{\mathcal{H}} \mapsto L(\hat{S}(v)) \in \mathbb{R}$ is a continuous linear functional on $\hat{\mathcal{H}}$. Since $\hat{\mathcal{H}}$ is closed (by definition) in the norm of $\mathcal{H}$, it is a Hilbert subspace of $\mathcal{H}$. Hence, by the Riesz representation theorem, there exists a unique $u \in \hat{\mathcal{H}}$ such that

$$L(\hat{S}(w)) = (u, w)_{\mathcal{H}} \qquad \forall w \in \hat{\mathcal{H}}.$$

Thus, by the definition of $\hat{S}$, $\hat{S}(w) = S(w)$ for all w in $\mathcal{R}(T)$; hence,

$$L(S(w)) = (u, w)_{\mathcal{H}} \qquad \forall w \in \mathcal{R}(T).$$

Equivalently, on writing $v = S(w)$,

$$(u, T(v))_{\mathcal{H}} = L(v) \qquad \forall v \in \mathcal{V}.$$

Thus we have shown the existence of a unique $u \in \hat{\mathcal{H}} (\subset \mathcal{H})$ such that

$$B(u,v) = (u, Tv)_{\mathcal{H}} = L(v) \qquad \forall v \in \mathcal{V}.$$

We note in passing that in the presence of additional structural assumptions on a (*e.g.* when $a \in \mathbb{R}^{d\times d}$ is positive definite) the assumption $c > 0$ can be relaxed to $c \ge 0$. □

The boundary condition $u|_{\Gamma_-} = 0$ on the inflow part Γ_- of the hyperbolic boundary $\Gamma \setminus \Gamma_0 = \Gamma_- \cup \Gamma_+$ is imposed weakly, through the definition of the bilinear form $B(\cdot,\cdot)$, while the boundary condition $u|_{\Gamma_0} = 0$ on the elliptic part Γ_0 of Γ is imposed strongly, through the choice of the function space $\mathcal{H}$. Hence, we deduce from Lemma 2.1 that

$$\bigotimes_{i=1}^{d} \mathrm{H}^1_{(0)}(0,1) := \mathrm{H}^1_{(0)}(0,1) \otimes \cdots \otimes \mathrm{H}^1_{(0)}(0,1) \subset \mathcal{H}, \tag{2.8}$$

where the ith component $\mathrm{H}^1_{(0)}(0,1)$ in the d-fold tensor-product on the left-hand side of the inclusion is taken to be equal to $\mathrm{H}^1_0(0,1)$ if Ox_i is an elliptic co-ordinate direction; otherwise (*i.e.*, when Ox_i is a hyperbolic co-ordinate direction), it is chosen to be equal to $\mathrm{H}^1(0,1)$.

Next, we shall consider the discretization of the weak formulation (2.6). Motivated by the tensor-product structure of the space on the left-hand side of the inclusion (2.8), we shall base our Galerkin discretization on a finite-dimensional subspace of $\mathcal{H}$ that is the tensor-product of finite-dimensional subspaces of $\mathrm{H}^1_{(0)}(0,1)$. Thus, we begin by setting up the necessary notation in the case of the univariate space $\mathrm{H}^1_{(0)}(0,1)$.

3. Univariate approximation results

Let $I = (0,1)$ and consider the sequence of partitions $\{\mathcal{T}^\ell\}_{\ell\geq 0}$, where $\mathcal{T}^0 = \{I\}$ and where the partition $\mathcal{T}^{\ell+1}$ is obtained from the previous partition

$$\mathcal{T}^\ell := \{I^\ell_j \,:\, j = 0, \ldots, 2^\ell - 1\}$$

by halving each of the intervals I^ℓ_j; $I^0_0 := I$. The mesh-size in the partition $\mathcal{T}^\ell$ is $h_\ell := 2^{-\ell}$.

We consider the finite-dimensional linear subspace $\mathcal{V}^{\ell,p}$ of $\mathrm{H}^1(0,1)$ consisting of all continuous piecewise polynomials of degree $p \geq 1$ on the partition $\mathcal{T}^\ell$, $\ell \geq 0$. For $\ell \geq 0$ we also consider the subspace $\mathcal{V}^{\ell,p}_0$ of $\mathcal{V}^{\ell,p}$ defined by $\mathcal{V}^{\ell,p}_0 := \mathcal{V}^{\ell,p} \cap \mathrm{C}_0[0,1] \subset \mathrm{H}^1_0(0,1)$ consisting of all continuous piecewise polynomial functions on $\mathcal{T}^\ell$ of degree p that vanish at both endpoints of the interval $[0,1]$.

Remark 3.1. When $p = 1$ the linear space $\mathcal{V}^{0,p}_0$ is trivial, that is $\mathcal{V}^{0,1}_0 = \{0\}$.

Let us note that the families of spaces $\{\mathcal{V}^{\ell,p}_0\}_{\ell\geq 0}$ and $\{\mathcal{V}^{\ell,p}\}_{\ell\geq 0}$ are nested, *i.e.*,

$$\mathcal{V}^{0,p}_0 \subsetneq \mathcal{V}^{1,p}_0 \subsetneq \mathcal{V}^{2,p}_0 \subsetneq \cdots \subsetneq \mathcal{V}^{\ell,p}_0 \subsetneq \cdots \subsetneq \mathrm{H}^1_0(0,1),$$

and

$$\mathcal{V}^{0,p} \subsetneq \mathcal{V}^{1,p} \subsetneq \mathcal{V}^{2,p} \subsetneq \cdots \subsetneq \mathcal{V}^{\ell,p} \subsetneq \cdots \subsetneq \mathrm{H}^1(0,1),$$

each space in each of the two chains being a proper subspace of the next space in the same chain. As in the previous section, we shall use $\mathrm{H}^1_{(0)}(0,1)$ to denote $\mathrm{H}^1_0(0,1)$ or $\mathrm{H}^1(0,1)$, as the case may be; analogously, we shall use $\mathcal{V}^{\ell,p}_{(0)}$ to denote $\mathcal{V}^{\ell,p}_0$ or $\mathcal{V}^{\ell,p}$. We shall adopt the following hypothesis.

Hypothesis 1$_{(0)}$. *Suppose that $p \geq 1$. For each integer $\ell \geq 0$ there exists a projector (*i.e.*, a linear, idempotent, surjective mapping) $P^{\ell,p}_{(0)} : \mathrm{H}^1_{(0)}(0,1) \to \mathcal{V}^{\ell,p}_{(0)}$.*

Under this hypothesis,

$$\mathcal{V}^{\ell,p}_{(0)} = P^{\ell,p}_{(0)}\, \mathrm{H}^1_{(0)}(0,1), \qquad \ell \geq 0, \quad p \geq 1.$$

Now, let

$$Q^{\ell,p}_{(0)} := \begin{cases} P^{\ell,p}_{(0)} - P^{\ell-1,p}_{(0)}, & \ell \geq 1, \\ P^{0,p}_{(0)}, & \ell = 0. \end{cases}$$

Thus, for any integer $L \geq 0$,

$$P_{(0)}^{L,p} = \sum_{\ell=0}^{L} Q_{(0)}^{\ell,p}.$$

We define the *increment spaces* $\mathcal{W}_{(0)}^{\ell,p}$, $\ell = 0, 1, \ldots$, as follows:

$$\mathcal{W}_{(0)}^{\ell,p} := Q_{(0)}^{\ell,p} \, \mathrm{H}_{(0)}^{1}(0,1).$$

Hence, for any pair of integers $L \geq 0$ and $p \geq 1$,

$$\mathcal{V}_{(0)}^{L,p} = P_{(0)}^{L,p} \, \mathrm{H}_{(0)}^{1}(0,1) = \left(\sum_{\ell=0}^{L} Q_{(0)}^{\ell,p} \right) \mathrm{H}_{(0)}^{1}(0,1) = \sum_{\ell=0}^{L} \left(Q_{(0)}^{\ell,p} \, \mathrm{H}_{(0)}^{1}(0,1) \right) = \sum_{\ell=0}^{L} \mathcal{W}_{(0)}^{\ell,p}. \tag{3.1}$$

In fact, if $Q_{(0)}^{\ell_1,p} \circ Q_{(0)}^{\ell_2,p} = 0$ for all $\ell_1, \ell_2 \in \{1, \ldots, d\}$, $\ell_1 \neq \ell_2$, one can make a stronger statement: for any pair of integers $L \geq 0$ and $p \geq 1$, the vector space $\mathcal{V}_{(0)}^{L,p}$ is the *direct sum* of the increment spaces $\mathcal{W}_{(0)}^{\ell,p}$, $\ell = 0, \ldots, L$:

$$\mathcal{V}_{(0)}^{L,p} = \bigoplus_{\ell=0}^{L} \mathcal{W}_{(0)}^{\ell,p}. \tag{3.2}$$

This is a direct consequence of the following result (see, for example, [5], Thm. 2.5), which we shall return to at the end of Section 3.3.

Proposition 3.1. *Let X be a vector space; then, there exist nontrivial subspaces X_ℓ, $\ell = 0, \ldots, L$, of X such that $X = \bigoplus_{\ell=0}^{L} X_\ell$ if, and only if, there are nonzero linear mappings $q_0, \ldots, q_L : X \to X$ such that*

(1) $\sum_{\ell=0}^{L} q_\ell = \mathrm{Id}_X$;
(2) $q_{\ell_1} \circ q_{\ell_2} = 0_X$ *for all* $\ell_1, \ell_2 \in \{0, \ldots, L\}$, $\ell_1 \neq \ell_2$.

Moreover, each q_ℓ is necessarily a projector and X_ℓ can be chosen to be $\mathrm{Im}(q_\ell)$, $\ell = 0, \ldots, L$.

So far, the choice of the projectors $P_{(0)}^{\ell,p}$ has been fairly arbitrary: the discussion above only appealed to their algebraic properties stated in Hypothesis $1_{(0)}$. Below, we shall be interested in the convergence of tensor-products of univariate projectors. Specifically, we shall investigate the dependence of convergence rates on the level index ℓ, the polynomial degree p, and the dimension d of the domain of definition $\Omega = (0,1)^d$ of the function u to be approximated. To this end, we shall make a second assumption on the projectors.

Hypothesis $2_{(0)}$. *Let $k \geq 1$ and $p \geq 1$ be two integers, $s \in \{0, 1\}$ and $h_\ell = 2^{-\ell}$, where $\ell \geq 0$ is an integer, and suppose that $v \in \mathrm{H}^{k+1}(0,1) \cap \mathrm{H}_{(0)}^{1}(0,1)$. For any integer t such that $1 \leq t \leq \min\{p, k\}$, there exists a positive constant $c_p(s,t)$, independent of v, such that*

$$|v - P_{(0)}^{\ell,p} v|_{\mathrm{H}^s(0,1)} \leq c_p(s,t) h_\ell^{t+1-s} \, |v|_{\mathrm{H}^{t+1}(0,1)}. \tag{3.3}$$

In particular, Hypothesis $2_{(0)}$ implies that $v = \lim_{\ell \to \infty} P_{(0)}^{\ell,p} v$ for all $v \in \mathrm{H}^2(0,1) \cap \mathrm{H}_{(0)}^{1}(0,1)$ and all $p \geq 1$, where the limit is considered in the $\mathrm{H}^s(0,1)$-seminorm for $s \in \{0, 1\}$, with the convention that, for $s = 0$, $|\cdot|_{\mathrm{H}^0(0,1)} = \|\cdot\|_{\mathrm{L}^2(0,1)}$.

3.1. **Examples of univariate projectors**

We present projectors $P_0^{\ell,p}$ and $P^{\ell,p}$ that satisfy our two hypotheses. Consider the projector $P^{\ell,p} : \mathrm{H}^1(0,1) \to \mathcal{V}^{\ell,p}$ defined, for $x \in [0,1]$, by

$$(P^{\ell,p}u)(x) := u(0) + \int_0^x (\Pi^{\ell,p-1}u')(\xi)\,\mathrm{d}\xi, \qquad \ell \geq 0, \quad p \geq 1,$$

where $\Pi^{\ell,p-1} : \mathrm{L}^2(0,1) \to \mathcal{V}^{\ell,p-1}$ is the $\mathrm{L}^2(0,1)$-orthogonal projector onto $\mathcal{V}^{\ell,p-1}$, with the convention that $\mathcal{V}^{\ell,0} \subset \mathrm{L}^2(0,1)$ is the set of piecewise constant functions on $\mathcal{T}^\ell$.

Since $u \in \mathrm{H}^1(0,1) \subset \mathrm{C}[0,1]$, the projector is well-defined. Trivially, $(P^{\ell,p}u)(0) = 0$. If $p = 1$, the projection $\Pi^{\ell,p-1}u'$ is a piecewise constant function on the mesh $\{x_j^\ell := 2^{-\ell}j : j = 0,\ldots,2^\ell\}$, and $u(x_j^\ell) - (P^{\ell,1}u)(x_j^\ell) = (u' - \Pi^{\ell,0}u', \mathbf{1}_{[0,x_j^\ell]}) = 0$, $j = 1,\ldots,2^\ell$. Consequently, for $p = 1$ we have that $P^{\ell,p}u$ is equal to u at all nodes of $\mathcal{T}^\ell$. More generally, $(P^{\ell,p}u)(1) = u(1)$ for all $\ell \geq 0$ and all $p \geq 1$; furthermore,

$$P^{\ell,p}|_{\mathrm{H}_0^1(0,1)} = P_0^{\ell,p}, \qquad \text{where} \quad (P_0^{\ell,p}u)(x) := \int_0^x (\Pi^{\ell,p-1}u')(\xi)\,\mathrm{d}\xi \qquad \text{for all } \ell \geq 0$$

(*cf.* Thm. 3.14 on p. 73 in Schwab [23]). In particular, $P_0^{0,1} = 0$.

In addition, the projector $P^{\ell,p}$ has the following approximation property (*cf.* inequalities (3.3.29) and (3.3.30) in Schwab [23]): For any v in $\mathrm{H}^{k+1}(0,1) \cap \mathrm{H}^1_{(0)}(0,1)$, $k \geq 1$, we have that

$$|v - P^{\ell,p}_{(0)}v|_{\mathrm{H}^s(0,1)} \leq \left(\frac{h_\ell}{2}\right)^{t+1-s} \frac{1}{p^{1-s}} \sqrt{\frac{(p-t)!}{(p+t)!}}\, |v|_{\mathrm{H}^{t+1}(0,1)}, \ 1 \leq t \leq \min\{p,k\}, \tag{3.4}$$

where $h_\ell = 2^{-\ell}$, $\ell \geq 0$, $p \geq 1$, $s \in \{0,1\}$, $t \in \mathbb{N}$, and $\mathbb{N}$ denotes the set of nonnegative integers.

Thus we have shown that the family of finite element spaces $\{\mathcal{V}^{\ell,p}_{(0)}\}_{\ell\geq 0} \subseteq \mathrm{H}^1_{(0)}(0,1)$ and the associated projectors $P^{\ell,p}_{(0)}$, $\ell \geq 0$, satisfy the approximation property

$$|v - P^{\ell,p}_{(0)}v|_{\mathrm{H}^s(0,1)} \leq c_{p,s,t} 2^{-(t+1-s)(\ell+1)} |v|_{\mathrm{H}^{t+1}(0,1)}, \tag{3.5}$$

for all $v \in \mathrm{H}^{k+1}(0,1) \cap \mathrm{H}^1_{(0)}(0,1)$, $k \geq 1$, $\ell \geq 0$, $p \geq 1$, $t \in \mathbb{N}$, $1 \leq t \leq \min\{p,k\}$ and $s \in \{0,1\}$, where

$$c_{p,s,t} := \frac{1}{p^{1-s}} \sqrt{\frac{(p-t)!}{(p+t)!}}. \tag{3.6}$$

Consequently, Hypotheses $1_{(0)}$ and $2_{(0)}$ hold, inequality (3.3) being satisfied with

$$c_p(s,t) := 2^{-(t+1-s)} c_{p,s,t} = \frac{1}{2^{t+1-s}p^{1-s}} \sqrt{\frac{(p-t)!}{(p+t)!}}. \tag{3.7}$$

In fact, when $p = t = 1$ and $\ell = 0$, using Poincaré's inequality, (3.5) can be shown to hold with $c_{1,s,1} = (2/\pi)^{2-s}$, yielding (3.3) with $c_1(s,1) = 1/\pi^{2-s}$, with $s \in \{0,1\}$.

3.2. **Bounds on the incremental projectors for $p \geq 1$**

Let us define, as above, the projection $Q_{(0)}^{\ell,p}$ onto the increment of the hierarchical family $\{\mathcal{V}_{(0)}^{\ell,p}\}_{\ell\geq 0}$ by

$$Q_{(0)}^{\ell,p} := \begin{cases} P_{(0)}^{\ell,p} - P_{(0)}^{\ell-1,p}, & \ell \geq 1, \\ P_{(0)}^{0,p}, & \ell = 0, \end{cases} \tag{3.8}$$

where now $P_{(0)}^{\ell,p}$ signifies the projector introduced in Section 3.1.

Suppose that $v \in \mathrm{H}^{k+1}(0,1) \cap \mathrm{H}^1_{(0)}(0,1)$, $k \geq 1$, $\ell \geq 0$, $p \geq 1$, $t \in \mathbb{N}$, $1 \leq t \leq \min\{p,k\}$ and $s \in \{0,1\}$.

(a) For $\ell \geq 1$, the triangle inequality and the approximation property (3.5) ensure that

$$|Q_{(0)}^{\ell,p} v|_{\mathrm{H}^s(0,1)} \leq \tilde{c}_{p,s,t} 2^{-(t+1-s)\ell} |v|_{\mathrm{H}^{t+1}(0,1)}, \qquad \ell \geq 1, \tag{3.9}$$

with

$$\tilde{c}_{p,s,t} = \left(1 + \frac{1}{2^{t+1-s}}\right) c_{p,s,t}. \tag{3.10}$$

(b) For $\ell = 0$ and $s = 0$, by Poincaré's inequality,

$$\begin{aligned} \|Q_{(0)}^{0,p} u\|_{\mathrm{L}^2(0,1)} &\leq \|u\|_{\mathrm{L}^2(0,1)} + \|u - P_{(0)}^{0,p} u\|_{\mathrm{L}^2(0,1)} \\ &\leq \|u\|_{\mathrm{L}^2(0,1)} + \frac{1}{\pi} |u - P_{(0)}^{0,p} u|_{\mathrm{H}^1(0,1)} \\ &= \|u\|_{\mathrm{L}^2(0,1)} + \frac{1}{\pi} \sqrt{|u|^2_{\mathrm{H}^1(0,1)} - |Q_{(0)}^{0,p} u|^2_{\mathrm{H}^1(0,1)}}, \end{aligned}$$

and therefore since, for $a \geq b \geq 0$, $\frac{1}{\pi}\sqrt{a^2 - b^2} \leq a - b\sqrt{1 - \frac{1}{\pi^2}}$, we deduce that

$$\|Q_{(0)}^{0,p} u\|_{\mathrm{L}^2(0,1)} + \sqrt{1 - \frac{1}{\pi^2}} |Q_{(0)}^{0,p} u|_{\mathrm{H}^1(0,1)} \leq \|u\|_{\mathrm{L}^2(0,1)} + |u|_{\mathrm{H}^1(0,1)} =: \|u\|_{\mathrm{H}^1_*(0,1)}. \tag{3.11}$$

(c) For $\ell = 0$ and $s = 1$ on the other hand, we have that

$$|Q_{(0)}^{0,p} u|_{\mathrm{H}^1(0,1)} = |P_{(0)}^{0,p} u|_{\mathrm{H}^1(0,1)} \leq |u|_{\mathrm{H}^1(0,1)}. \tag{3.12}$$

Also, (3.11) implies that

$$\|Q_{(0)}^{0,p} u\|_{\mathrm{H}^1_*(0,1)} \leq \frac{\pi}{\sqrt{\pi^2 - 1}} \|u\|_{\mathrm{H}^1_*(0,1)}. \tag{3.13}$$

(d) We note that when $\ell = 0$, $s \in \{0,1\}$, $p = 1$ and $u \in \mathrm{H}^1_0(0,1)$, then $P_{(0)}^{0,1} u = P_0^{0,1} u = 0$ and $Q_{(0)}^{0,1} u = Q_0^{0,1} u = 0$, and inequalities (3.11) and (3.12) are thereby trivially satisfied.

For $\ell = 0$ we set

$$\begin{aligned} \hat{c}_{p,0,(0)} &:= \|Q_{(0)}^{0,p}\|_{\mathcal{B}(\mathrm{H}^1_{(0)}(0,1),\mathrm{L}^2(0,1))}, \\ \hat{c}_{p,1,(0)} &:= \|Q_{(0)}^{0,p}\|_{\mathcal{B}(\mathrm{H}^1_{(0)}(0,1),\mathrm{H}^1_{(0)}(0,1))}, \end{aligned} \tag{3.14}$$

with the convention that the norm in $\mathrm{H}^1_0(0,1)$ is the seminorm $|\cdot|_{\mathrm{H}^1(0,1)}$ while the norm in $\mathrm{H}^1(0,1)$ is $\|\cdot\|_{\mathrm{H}^1_*(0,1)}$. It will be clear from the context whether we use zero or nonzero boundary conditions in the spaces.

Example 3.1. If $p \geq 2$ and $\ell = 0$, then $Q_0^{0,p}$ is, for $u \in \mathrm{H}_0^1(0,1)$, given by

$$(Q_0^{0,p}u)(x) = \int_0^x (\Pi^{0,p-1}u')(\xi)\,\mathrm{d}\xi$$

where $\Pi^{0,p-1}$ denotes the $\mathrm{L}^2(0,1)$-projection onto $\mathcal{V}^{0,p-1}$. Now, $Q_0^{0,p} \in \mathcal{B}(\mathrm{H}_0^1(0,1), \mathrm{H}_0^1(0,1))$ with $\hat{c}_{p,1,0} = 1$ and, as the embedding of $\mathrm{H}_0^1(0,1)$ into $\mathrm{L}^2(0,1)$ has norm $1/\pi$ by the Poincaré inequality, $\hat{c}_{p,0,0} \leq 1/\pi$. On the other hand, (3.11) still implies that $\hat{c}_{p,0,(0)} \leq 1$, with $\mathrm{H}^1(0,1)$ equipped by the norm $\|\cdot\|_{\mathrm{H}^1_*(0,1)}$, while (3.13) implies that $\hat{c}_{p,1,(0)} \leq \pi/\sqrt{\pi^2-1}$.

3.3. Refined bounds on the incremental projectors for $p = 1$

We shall now take a closer look at estimating $|Q_{(0)}^{\ell,p}v|_{\mathrm{H}^s(0,1)}$ in the case of $p = 1$ and $\ell \geq 1$. As in Section 3.2, $Q_{(0)}^{\ell,p}$ is defined by (3.8) where $P_{(0)}^{\ell,p}$ is the projector introduced in Section 3.1. Our objective is to sharpen our earlier expression (3.10) for the constant $\tilde{c}_{p,s,t}$ appearing in the detail-size estimate (3.9) in the special case of $p = 1$ and $s \in \{0,1\}$ (note that we necessarily have $t = 1$).

We use the following two simple auxiliary results, the first of which is a discrete version of the Poincaré inequality.

Lemma 3.1. *Suppose that $v \in \mathrm{H}_0^1(0,1)$ is piecewise linear on $\mathcal{T}^1 := \{[0,\frac{1}{2}],[\frac{1}{2},1]\}$; then*

$$\|v\|_{\mathrm{L}^2(0,1)} \leq \frac{1}{\sqrt{12}}\|v'\|_{\mathrm{L}^2(0,1)}. \tag{3.15}$$

Proof. The result follows from a straightforward calculation with v taken to be the standard hat function $\varphi : x \mapsto \frac{1}{2}(1-2|x-\frac{1}{2}|)_+$ defined on $[0,1]$. □

Lemma 3.2. *Suppose that $v \in \mathrm{H}^1(0,1)$; then*

$$\left|\int_0^{\frac{1}{2}} v(t)\,\mathrm{d}t - \int_{1/2}^1 v(t)\,\mathrm{d}t\right| \leq \frac{1}{\sqrt{12}}\|v'\|_{\mathrm{L}^2(0,1)}. \tag{3.16}$$

Proof. Denoting by φ, as in the proof Lemma 3.1, the hat function on $[0,1]$ with $\varphi(\frac{1}{2}) = \frac{1}{2}$, we note that $\mathbf{1}_{[0,\frac{1}{2}]} - \mathbf{1}_{[\frac{1}{2},1]} = \varphi'$ on $[0,1] \setminus \{\frac{1}{2}\}$. Partial integration and the Cauchy-Schwarz inequality yield

$$\begin{aligned}\left|\int_0^{\frac{1}{2}} v(t)\,\mathrm{d}t - \int_{1/2}^1 v(t)\,\mathrm{d}t\right| &= \left|\int_0^1 \varphi'(t)v(t)\,\mathrm{d}t\right| = \left|\int_0^1 \varphi(t)v'(t)\,\mathrm{d}t\right| \\ &\leq \|\varphi\|_{\mathrm{L}^2(0,1)}\|v'\|_{\mathrm{L}^2(0,1)} = \frac{1}{\sqrt{12}}\|v'\|_{\mathrm{L}^2(0,1)}.\end{aligned} \tag{3.17}$$

That completes the proof. □

Remark 3.2. Rescaling Lemmas 3.1 and 3.2 above from $[0,1]$ to $[0,h]$ with $h > 0$ we obtain the following inequalities:

$$\|v\|_{\mathrm{L}^2(0,h)} \leq \frac{h}{\sqrt{12}}\|v'\|_{\mathrm{L}^2(0,h)} \quad \forall v \in \mathrm{H}_0^1(0,h), \text{ piecewise linear on } [0,h/2] \cup [h/2,h]; \tag{3.18}$$

and

$$\left|\int_0^{h/2} v(t)\,\mathrm{d}t - \int_{h/2}^h v(t)\,\mathrm{d}t\right| \leq \frac{h^{\frac{3}{2}}}{\sqrt{12}}\|v'\|_{\mathrm{L}^2(0,h)} \quad \forall v \in \mathrm{H}^1(0,h). \tag{3.19}$$

Proposition 3.2. *Suppose that the projectors $P^{\ell,1}_{(0)}$, $\ell = 0, 1, \ldots$, are defined by*

$$(P^{\ell,1}_{(0)} v)(x) = v(0) + \int_0^x (\Pi^{\ell,0} v')(\xi)\, \mathrm{d}\xi \quad \forall v \in \mathrm{H}^1_{(0)}(0,1),\ \ell \geq 0; \tag{3.20}$$

then, for any $v \in \mathrm{H}^2(0,1) \cap \mathrm{H}^1_{(0)}(0,1)$, we have

$$\|Q^{\ell,1}_{(0)} v\|_{\mathrm{L}^2(0,1)} \leq \frac{1}{3} 2^{-2\ell} \|v''\|_{\mathrm{L}^2(0,1)} \quad \textit{and} \quad |Q^{\ell,1}_{(0)} v|_{\mathrm{H}^1(0,1)} \leq \frac{1}{\sqrt{3}} 2^{-\ell} \|v''\|_{\mathrm{L}^2(0,1)} \quad \forall \ell \geq 1. \tag{3.21}$$

Hence, for the constants $\tilde{c}_{1,0,1}$ and $\tilde{c}_{1,1,1}$ appearing in (3.9), we obtain the upper bounds

$$\tilde{c}_{1,0,1} \leq \frac{1}{3} \quad \textit{and} \quad \tilde{c}_{1,1,1} \leq \frac{1}{\sqrt{3}}. \tag{3.22}$$

Proof. Denote by I^ℓ_k, for $1 \leq k \leq 2^\ell$, the subintervals of $\mathcal{T}^\ell$, of length $h_\ell = 2^{-\ell}$. The nodal exactness of $P^{\ell,1}_{(0)}$ (*cf.* the beginning of Sect. 3.1) ensures that $Q^{\ell,1}_{(0)} v\,|_{I^{\ell-1}_k} \in \mathrm{H}^1_0(I^{\ell-1}_k)$ is a multiple of the standard hat function, rescaled to $I^{\ell-1}_k$, for all $\ell \geq 1$ and $1 \leq k \leq 2^{\ell-1}$. Applying Lemma 3.1 (after rescaling to $I^{\ell-1}_k$), we obtain

$$\|Q^{\ell,1}_{(0)} v\,|_{I^{\ell-1}_k}\|^2_{\mathrm{L}^2(I^{\ell-1}_k)} \leq \frac{1}{12} (h_{\ell-1})^2 \|(Q^{\ell,1}_{(0)} v)'\,|_{I^{\ell-1}_k}\|^2_{\mathrm{L}^2(I^{\ell-1}_k)} \quad \forall v \in \mathrm{H}^2(0,1),\ \ell \geq 1. \tag{3.23}$$

The definition (3.20) ensures that

$$(Q^{\ell,1}_{(0)} v)' = \Pi^{\ell,0} v' - \Pi^{\ell-1,0} v', \qquad \ell \geq 1,$$

so that, since $I^{\ell-1}_k = I^\ell_{2k-1} \cup I^\ell_{2k}$,

$$(Q^{\ell,1}_{(0)} v)'\,|_{I^{\ell-1}_k} = \left(\frac{1}{h_\ell} \int_{I^\ell_{2k-1}} v'(t)\, \mathrm{d}t \right) \mathbf{1}_{I^\ell_{2k-1}} + \left(\frac{1}{h_\ell} \int_{I^\ell_{2k}} v'(t)\, \mathrm{d}t \right) \mathbf{1}_{I^\ell_{2k}} - \left(\frac{1}{h_{\ell-1}} \int_{I^{\ell-1}_k} v'(t)\, \mathrm{d}t \right) \mathbf{1}_{I^{\ell-1}_k}.$$

Using $h_\ell = 2^{-\ell}$ we obtain

$$\begin{aligned} \|(Q^{\ell,1}_{(0)} v)'\,|_{I^{\ell-1}_k}\|^2_{\mathrm{L}^2(I^{\ell-1}_k)} &= \frac{1}{4} \left(\frac{1}{h_\ell} \int_{I^\ell_{2k-1}} v'(t)\, \mathrm{d}t - \frac{1}{h_\ell} \int_{I^\ell_{2k}} v'(t)\, \mathrm{d}t \right)^2 \|\mathbf{1}_{I^{\ell-1}_k}\|^2_{\mathrm{L}^2(I^{\ell-1}_k)} \\ &= \frac{1}{h_{\ell-1}} \left(\int_{I^\ell_{2k-1}} v'(t)\, \mathrm{d}t - \int_{I^\ell_{2k}} v'(t)\, \mathrm{d}t \right)^2, \end{aligned} \tag{3.24}$$

from which we deduce using Lemma 3.2, rescaled to $I^{\ell-1}_k$, that

$$\|(Q^{\ell,1}_{(0)} v)'\,|_{I^{\ell-1}_k}\|^2_{\mathrm{L}^2(I^{\ell-1}_k)} \leq \frac{1}{12} (h_{\ell-1})^2 \|v''\|^2_{\mathrm{L}^2(I^{\ell-1}_k)}. \tag{3.25}$$

The estimates (3.21) now follow from (3.23) and (3.25), upon summing over k from 1 to $2^{\ell-1}$. Finally, (3.22) follows by comparing (3.9) and (3.21). □

We complete this section by showing that the equality (3.2) holds in the case of $p = 1$. To do so, we apply Proposition 3.1 with $X = \mathcal{V}^{L,1}$, $X_\ell = \mathcal{W}^{\ell,1}$ and $q_\ell = Q^{\ell,1}$, $\ell = 0, \ldots, L$, and note that $P^{L,1} = \sum_{\ell=0}^L Q^{\ell,1}$ is the

identity-map in $\mathcal{V}^{L,1}$ and $Q^{\ell_1,1} \circ Q^{\ell_2,1}$ is the zero-map in $\mathcal{V}^{L,1}$ for all $\ell_1, \ell_2 \in \{0, \ldots, L\}$, $\ell_1 \neq \ell_2$; thus, (3.2) follows from (3.1).

When $p = 1$ and $X = \mathcal{V}_0^{L,1}$ a small modification is required since then $\mathcal{W}_0^{0,1} = \mathcal{V}_0^{0,1} = \{0\}$ and $Q_0^{0,1} = P_0^{0,1} = 0$, so Proposition 3.1 does not directly apply with $\ell = 0, \ldots, L$. Instead, we use Proposition 3.1 with $X = \mathcal{V}_0^{L,1}$, $X_\ell = \mathcal{W}_0^{\ell,1}$ and $q_\ell = Q_0^{\ell,1}$, $\ell = 1, \ldots, L$, and note that $P_0^{L,1} = \sum_{\ell=1}^{L} Q_0^{\ell,1}$ is the identity-map in $\mathcal{V}_0^{L,1}$ and $Q_0^{\ell_1,1} \circ Q_0^{\ell_2,1}$ is the zero-map in $\mathcal{V}_0^{L,1}$ for all $\ell_1, \ell_2 \in \{1, \ldots, L\}$, $\ell_1 \neq \ell_2$, to deduce that $\mathcal{V}_0^{L,1} = \bigoplus_{\ell=1}^{L} \mathcal{W}_0^{\ell,1}$. On taking the direct sum of each side in the last equality with $\mathcal{W}_0^{0,1}$, (3.2) follows since $\mathcal{W}_0^{0,1} \oplus \mathcal{V}_0^{L,1} = \mathcal{V}_0^{L,1}$.

Thus we have shown that, for $p = 1$,

$$\mathcal{V}_{(0)}^{L,p} = \bigoplus_{\ell=0}^{L} \mathcal{W}_{(0)}^{\ell,p} = \mathcal{W}_{(0)}^{0,p} \oplus \cdots \oplus \mathcal{W}_{(0)}^{L,p}, \qquad L \geq 0; \tag{3.26}$$

in other words,

$$\mathcal{V}_{(0)}^{\ell,p} = \mathcal{V}_{(0)}^{\ell-1,p} \oplus \mathcal{W}_{(0)}^{\ell,p}, \qquad \ell \geq 1.$$

For $p \geq 2$, we still have that $Q_{(0)}^{\ell_1,p} \circ Q_{(0)}^{\ell_2,p}$ is the zero-map in $\mathcal{V}_{(0)}^{L,p}$ for all $\ell_1, \ell_2 \in \{1, \ldots, L\}$ such that $\ell_1 > \ell_2$, but not for $\ell_1 < \ell_2$, and therefore (3.26) does not follow. This, however, will not affect our subsequent arguments.

4. Sparse finite element discretization

We shall now use the finite-dimensional spaces $\mathcal{V}^{L,p}$ and $\mathcal{V}_0^{L,p}$ of univariate functions to construct a tensor-product space of multivariate functions. We shall then sparsify the resulting tensor-product space with the aim to reduce its dimension without significantly compromising the approximation properties of the original tensor-product space.

4.1. Sparse tensor-product space

Let L and p be positive integers, and consider on $\Omega = (0,1)^d$ the finite-dimensional subspace $V_{(0)}^{L,p}$ of $\bigotimes_{i=1}^{d} \mathrm{H}_{(0)}^1(0,1)$ defined by

$$V_{(0)}^{L,p} := \bigotimes_{i=1}^{d} \mathcal{V}_{(0)}^{L,p} = \mathcal{V}_{(0)}^{L,p} \otimes \cdots \otimes \mathcal{V}_{(0)}^{L,p}, \tag{4.1}$$

where the ith component $\mathcal{V}_{(0)}^{L,p}$ in this tensor-product is chosen to be $\mathcal{V}_0^{L,p}$ if Ox_i is an elliptic co-ordinate direction, and $\mathcal{V}_{(0)}^{L,p}$ is chosen as $\mathcal{V}^{L,p}$ otherwise.

In particular, if $a \equiv 0$ and therefore $\Gamma_0 = \emptyset$, then $\mathcal{V}_{(0)}^{L,p} = \mathcal{V}^{L,p}$ for each component in the tensor-product. Conversely, if a is positive definite, then $\Gamma_0 = \Gamma$ and therefore $\mathcal{V}_{(0)}^{L,p} = \mathcal{V}_0^{L,p}$ for each component of the tensor-product. In general, for $a \geq 0$ that is neither identically zero nor positive definite, $\mathcal{V}_{(0)}^{L,p} = \mathcal{V}_0^{L,p}$ for a certain number i of components in the tensor-product, where $0 < i < d$, and $\mathcal{V}_{(0)}^{L,p} = \mathcal{V}^{L,p}$ for the rest.

For a multi-index $\ell = (\ell_1, \ldots, \ell_d) \in \mathbb{N}_0^d$ we denote its ℓ_∞ and ℓ_1 norms, respectively, by

$$|\ell|_\infty := \max\{\ell_i : 1 \leq i \leq d\} \quad \text{and} \quad |\ell|_1 := \ell_1 + \ldots + \ell_d.$$

Using the hierarchical decomposition (3.1) we have that

$$V_{(0)}^{L,p} = \sum_{|\ell|_\infty \leq L} \mathcal{W}_{(0)}^{\ell_1,p} \otimes \cdots \otimes \mathcal{W}_{(0)}^{\ell_d,p}, \qquad \ell = (\ell_1, \ldots, \ell_d), \tag{4.2}$$

with the convention that $\mathcal{W}_{(0)}^{\ell,p} = \mathcal{W}_0^{\ell,p}$ whenever Ox_i is an elliptic co-ordinate direction, and $\mathcal{W}_{(0)}^{\ell,p} = \mathcal{W}^{\ell,p}$ otherwise.

The space $V_{(0)}^{L,p}$ has $\mathcal{O}(2^{Ld}p^d)$ degrees of freedom, a number that grows exponentially as a function of d. In order to reduce the number of degrees of freedom, we shall replace $V_{(0)}^{L,p}$ with a lower-dimensional subspace $\hat{V}_{(0)}^{L,p}$ defined as follows:

$$\hat{V}_{(0)}^{L,p} := \sum_{|\ell|_1 \le L} \mathcal{W}_{(0)}^{\ell_1,p} \otimes \cdots \otimes \mathcal{W}_{(0)}^{\ell_d,p}, \qquad \ell = (\ell_1, \ldots, \ell_d). \tag{4.3}$$

The space $\hat{V}_{(0)}^{L,p}$ is called a *sparse tensor-product space*.

The space $\hat{V}_{(0)}^{L,p}$ has $\mathcal{O}(2^L L^{d-1} p^d)$ degrees of freedom, which is a considerable reduction compared to the $\mathcal{O}(2^{Ld}p^d)$ degrees of freedom for the space $V_{(0)}^{L,p}$. Replacement of $V^{L,p}$ by $\hat{V}^{L,p}$ is an h-version sparsification (*i.e.*, sparsification with respect to the mesh parameter $L = |\log_2 h_L|$). One could further reduce the size of the space $\hat{V}^{L,p}$ by performing p-version sparsification (*i.e.*, sparsification with respect to the parameter p); here we shall refrain from doing so as we are concerned with h-version sparse finite element methods, with $p \ge 1$ fixed.

Consider the d-dimensional projector

$$P_{(0)}^{L,p} \otimes \cdots \otimes P_{(0)}^{L,p} : \bigotimes_{i=1}^{d} \mathrm{H}_{(0)}^1(0,1) \to \bigotimes_{i=1}^{d} \mathcal{V}_{(0)}^{L,p} = V_{(0)}^{L,p},$$

where the ith component $P_{(0)}^{L,p}$ is equal to $P_0^{L,p}$ if Ox_i is an elliptic co-ordinate direction, and equal to $P^{L,p}$ otherwise. Let us now recall that

$$Q_{(0)}^{\ell,p} = \begin{cases} P_{(0)}^{\ell,p} - P_{(0)}^{\ell-1,p}, & \ell \ge 1, \\ P_{(0)}^{0,p}, & \ell = 0. \end{cases}$$

Thus,

$$P_{(0)}^{L,p} = \sum_{\ell=0}^{L} Q_{(0)}^{\ell,p},$$

where $Q_{(0)}^{\ell,p} = Q_0^{\ell,p}$ when $P_{(0)}^{\ell,p} = P_0^{\ell,p}$ and $Q_{(0)}^{\ell,p} = Q^{\ell,p}$ when $P_{(0)}^{\ell,p} = P^{\ell,p}$. Hence,

$$P_{(0)}^{L,p} \otimes \cdots \otimes P_{(0)}^{L,p} = \sum_{|\ell|_\infty \le L} Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}, \qquad \ell = (\ell_1, \ldots, \ell_d),$$

where $Q_{(0)}^{\ell_i,p}$ is equal to $Q_0^{\ell_i,p}$ when Ox_i is an elliptic co-ordinate direction, and equal to $Q^{\ell_i,p}$ otherwise.

The sparse counterpart $\hat{P}_{(0)}^{L,p}$ of the tensor-product projector $P_{(0)}^{L,p} \otimes \cdots \otimes P_{(0)}^{L,p}$ is defined by truncating the index set $\{\ell : |\ell|_\infty \le L\}$ of the sum to $\{\ell : |\ell|_1 \le L\}$:

$$\hat{P}_{(0)}^{L,p} := \sum_{|\ell|_1 \le L} Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p} : \bigotimes_{i=1}^{d} \mathrm{H}_{(0)}^1(0,1) \to \hat{V}_{(0)}^{L,p}, \qquad \ell = (\ell_1, \ldots, \ell_d),$$

where $Q_{(0)}^{\ell_i,p}$ is equal to $Q_0^{\ell_i,p}$ when Ox_i is an elliptic co-ordinate direction Ox_i, and equal to $Q^{\ell_i,p}$ otherwise.

4.2. **Sparse stabilized finite element method**

Having defined, for $L, p \geq 1$, the finite-dimensional space $\hat{V}_{(0)}^{L,p}$ in which the approximate solution will be sought, we now introduce a stabilized Galerkin finite element method over this finite-dimensional space. The main ingredients of the method are a bilinear form $b_\delta(\cdot,\cdot)$ that approximates the bilinear form $B(\cdot,\cdot)$ from the weak formulation of the boundary value problem and a linear functional $l_\delta(\cdot)$ that approximates the linear functional $L(\cdot)$.

Let us consider the bilinear form

$$b_\delta(w,v) := B(w,v) + \delta_L \sum_{\kappa\in\mathcal{T}^L} (\mathcal{L}w, b\cdot\nabla v)_\kappa.$$

Here, in the light of the fact that in the transport-dominated case $|a| \ll |b|$, the second term in the bilinear form $b_\delta(\cdot,\cdot)$ can be thought of as least-square stabilization in the direction of subcharacteristics ('streamlines'). We shall suppose for ease of presentation that $c \geq 0$.

We also define the linear functional

$$l_\delta(v) := L(v) + \delta_L \sum_{\kappa\in\mathcal{T}^L} (f, b\cdot\nabla v)_\kappa \quad (= L(v) + \delta_L(f, b\cdot\nabla v))\,.$$

Here $\delta_L \in \mathbb{R}_{\geq 0}$ is a ('streamline-diffusion') parameter to be chosen below, and $\kappa \in \mathcal{T}^L$ are d-dimensional axiparallel cubic elements of edge-length h_L in the partition of the computational domain $\Omega = (0,1)^d$. As there are 2^{Ld} such elements κ in $\mathcal{T}^L$, the computation of the *stabilization term* $\delta_L \sum_{\kappa\in\mathcal{T}^L}(\mathcal{L}w, b\cdot\nabla v)_\kappa$ in the definition of $b_\delta(w,v)$ may seem intractable for $d \gg 1$; however, it turns out that this is not so: the sum over the 2^{Ld} terms in the stabilization term can be rewritten as a sum over $2^L d + \frac{1}{2}d(d-1) + 1$ terms only; see Remark 6.2(c).

We consider the finite-dimensional problem: Find $u_h \in \hat{V}_{(0)}^{L,p}$ such that

$$b_\delta(u_h, v_h) = l_\delta(v_h) \qquad \forall v_h \in \hat{V}_{(0)}^{L,p}. \tag{4.4}$$

The idea behind the method (4.4) is to introduce mesh-dependent numerical diffusion into the standard Galerkin finite element method along subcharacteristic directions, with the aim to suppress maximum-principle-violating oscillations on the scale of the mesh, and let $\delta_L \to 0$ with $h_L \to 0$. For an analysis of the method in the case of standard finite element spaces and (low-dimensional) elliptic transport-dominated diffusion equations we refer to the monograph [22].

It would have been more accurate to write u_{h_L} and v_{h_L} instead of u_h and v_h in (4.4). However, to avoid notational clutter, we shall refrain from doing so. Instead, we adopt the convention that the dependence of $h = h_L$ on the index L will be implied, even when not explicitly indicated.

We begin with the stability-analysis of the method. First, we shall show that, with an appropriate choice of the streamline-diffusion parameter δ_L, the bilinear form $b_\delta(\cdot,\cdot)$ is coercive on $V_{(0)}^{L,p} \times V_{(0)}^{L,p}$. To this end,

we begin by noting that

$$\begin{aligned} b_\delta(v_h, v_h) &= (a\nabla v_h, \nabla v_h) - (v_h, \nabla\cdot(bv_h)) + (cv_h, v_h) + \langle v_h, v_h\rangle_{\Gamma_+} + \delta_L \sum_{\kappa\in\mathcal{T}^L} (\mathcal{L}v_h, b\cdot\nabla v_h)_\kappa \\ &= (a\nabla v_h, \nabla v_h) + c\|v_h\|^2_{\mathrm{L}^2(\Omega)} + \delta_L\|b\cdot\nabla v_h\|^2_{\mathrm{L}^2(\Omega)} \\ &\quad + \frac{1}{2}\int_{\Gamma_-\cup\Gamma_+} |\beta|\,|v_h|^2\,\mathrm{d}s + \frac{1}{2}c\delta_L\int_{\Gamma_-\cup\Gamma_+}\beta|v_h|^2\,\mathrm{d}s \\ &\quad + \delta_L\sum_\kappa(-a:\nabla\nabla v_h, b\cdot\nabla v_h)_\kappa \\ &\geq \|\sqrt{a}\,\nabla v_h\|^2_{\mathrm{L}^2(\Omega)} + c\|v_h\|^2_{\mathrm{L}^2(\Omega)} + \frac{1}{2}\delta_L\|b\cdot\nabla v_h\|^2_{\mathrm{L}^2(\Omega)} \\ &\quad + \frac{1}{2}(1+c\delta_L)\int_{\Gamma_+}|\beta||v_h|^2\,\mathrm{d}s + \frac{1}{2}(1-c\delta_L)\int_{\Gamma_-}|\beta||v_h|^2\,\mathrm{d}s \\ &\quad - \frac{1}{2}\delta_L\sum_\kappa\|a:\nabla\nabla v_h\|^2_{\mathrm{L}^2(\kappa)} \qquad \forall v_h\in V^{L,p}_{(0)}, \end{aligned} \tag{4.5}$$

where we have made use of the facts that $\beta = -|\beta|$ on Γ_- and $v_h|_{\Gamma_0} = 0$.

When $p = 1$ and $a \geq 0$ is diagonal, the last term in (4.5) is equal to zero, and the coercivity of $b_\delta(\cdot,\cdot)$ in the streamline-diffusion norm $|||\cdot|||_{\mathrm{SD}}$ (*cf.* [22]), defined by

$$|||v|||^2_{\mathrm{SD}} := \|\sqrt{a}\,\nabla v\|^2_{\mathrm{L}^2(\Omega)} + c\|v\|^2_{\mathrm{L}^2(\Omega)} + \delta_L\|b\cdot\nabla v\|^2_{\mathrm{L}^2(\Omega)} + \frac{1}{2}(1+c\delta_L)\int_{\Gamma_+}|\beta||v|^2\,\mathrm{d}s,$$

then follows immediately, provided that δ_L is chosen so that $0 \leq c\delta_L \leq 1$. If, however, $p > 1$ or if $a \in \mathbb{R}^{d\times d}$ is a general symmetric semidefinite matrix, the final term in (4.5) is generally nonzero. Still, we shall show that, with a somewhat more restrictive choice of δ_L, the final term in (4.5) can be absorbed into the first term on the right-hand side of (4.5), yielding coercivity in the norm $|||\cdot|||_{\mathrm{SD}}$. We shall therefore assume that $a \in \mathbb{R}^{d\times d}$ is a general positive semidefinite matrix and $p \geq 1$, and will only distinguish between the cases $p = 1$ and $p \geq 2$ when the choice of $p = 1$ necessitates special treatment.

We shall require the following inverse inequality for a univariate polynomial (*cf.* Schwab [23], p. 148, Thm. 3.91).

Lemma 4.1. *Let $J \subset \mathbb{R}$ be a bounded open interval with $h := \mathrm{meas}_1(J)$, and let $p \geq 1$; then,*

$$\|v'\|_{\mathrm{L}^2(J)} \leq \sqrt{12}\,\frac{p^2}{h}\|v\|_{\mathrm{L}^2(J)} \qquad \forall v \in \mathcal{P}^p(J),$$

where $\mathcal{P}^p(J)$ denotes the set of all polynomials of degree p or less defined on $\bar{J}$.

Now, letting $w_i := (a\nabla v_h)_i$ we have, for any $\kappa \in \mathcal{T}^L$, that

$$\begin{aligned}
\|a : \nabla\nabla v_h\|^2_{\mathrm{L}^2(\kappa)} &= \|\nabla\cdot(a\nabla v_h)\|^2_{\mathrm{L}^2(\kappa)} = \left\|\sum_{i=1}^d \frac{\partial}{\partial x_i} w_i\right\|^2_{\mathrm{L}^2(\kappa)} \\
&\le \left(\sum_{i=1}^d \left\|\frac{\partial w_i}{\partial x_i}\right\|_{\mathrm{L}^2(\kappa)}\right)^2 \le d\sum_{i=1}^d \left\|\frac{\partial w_i}{\partial x_i}\right\|^2_{\mathrm{L}^2(\kappa)} \le \frac{12dp^4}{h_L^2}\sum_{i=1}^d \|w_i\|^2_{\mathrm{L}^2(\kappa)} \\
&= \frac{12dp^4}{h_L^2}\int_\kappa \sum_{i=1}^d |w_i|^2\,\mathrm{d}x = \frac{12dp^4}{h_L^2}\int_\kappa |w|^2\,\mathrm{d}x = \frac{12dp^4}{h_L^2}\int_\kappa |a\nabla v_h|^2\,\mathrm{d}x \\
&= \frac{12dp^4}{h_L^2}\int_\kappa |\sqrt{a}\,\sqrt{a}\,\nabla v_h|^2\,\mathrm{d}x \;\le \frac{12dp^4}{h_L^2}\int_\kappa |\sqrt{a}\,|^2\,|\sqrt{a}\,\nabla v_h|^2\,\mathrm{d}x \\
&= |\sqrt{a}\,|^2\frac{12dp^4}{h_L^2}\,\|\sqrt{a}\,\nabla v_h\|^2_{\mathrm{L}^2(\kappa)},
\end{aligned}$$

where $|w|$ denotes the ℓ^2 norm of $w \in \mathbb{R}^d$ and $|\sqrt{a}\,|$ again denotes the Frobenius norm of the symmetric positive semidefinite matrix $\sqrt{a} \in \mathbb{R}^{d\times d}$. Hence, after summation over all $\kappa \in \mathcal{T}^L$,

$$\sum_{\kappa\in\mathcal{T}^L} \|a : \nabla\nabla v_h\|^2_{\mathrm{L}^2(\kappa)} \le |\sqrt{a}\,|^2\frac{12dp^4}{h_L^2}\,\|\sqrt{a}\,\nabla v_h\|^2_{\mathrm{L}^2(\Omega)}.$$

Using this bound in (4.5) we deduce that

$$\begin{aligned}
b_\delta(v_h, v_h) \;\ge\; & \left(1 - \delta_L|\sqrt{a}\,|^2\frac{6dp^4}{h_L^2}\right)\|\sqrt{a}\,\nabla v_h\|^2_{\mathrm{L}^2(\Omega)} + c\|v_h\|^2_{\mathrm{L}^2(\Omega)} + \frac{1}{2}\delta_L\|b\cdot\nabla v_h\|^2_{\mathrm{L}^2(\Omega)} \\
& + \frac{1}{2}(1+c\delta_L)\int_{\Gamma_+} |\beta|\,|v_h|^2\,\mathrm{d}s + \frac{1}{2}(1-c\delta_L)\int_{\Gamma_-} |\beta|\,|v_h|^2\,\mathrm{d}s.
\end{aligned}$$

Let us suppose, with the convention $1/0 = \infty$, that $\delta_L \in \mathbb{R}_{\ge 0}$ satisfies

$$0 \le \delta_L \le \min\left(\frac{h_L^2}{12dp^4|\sqrt{a}\,|^2}\,,\ \frac{1}{c}\right).$$

Then,

$$\begin{aligned}
b_\delta(v_h, v_h) \;\ge\; & \frac{1}{2}\|\sqrt{a}\,\nabla v_h\|^2_{\mathrm{L}^2(\Omega)} + c\|v_h\|^2_{\mathrm{L}^2(\Omega)} + \frac{1}{2}\delta_L\|b\cdot\nabla v_h\|^2_{\mathrm{L}^2(\Omega)} \\
& + \frac{1}{2}(1+c\delta_L)\int_{\Gamma_+} |\beta|\,|v_h|^2\,\mathrm{d}s + \frac{1}{2}(1-c\delta_L)\int_{\Gamma_-} |\beta|\,|v_h|^2\,\mathrm{d}s \\
\ge\; & \frac{1}{2}|||v_h|||^2_{\mathrm{SD}} \qquad \forall v_h \in \hat{V}^{L,p}_{(0)}.
\end{aligned} \tag{4.6}$$

Since (4.4) is a linear problem in a finite-dimensional linear space, the coercivity (4.6) of the bilinear form $b_\delta(\cdot,\cdot)$ implies the existence and uniqueness of a solution u_h to (4.4) in $\hat{V}^{L,p}_{(0)}$. Furthermore, if $c > 0$ then

$$\frac{1}{2}|||u_h|||^2_{\mathrm{SD}} \le \left(\frac{1}{c} + \delta_L\right)^{\frac{1}{2}}\|f\|_{\mathrm{L}^2(\Omega)}\,|||u_h|||_{\mathrm{SD}},$$

which, in turn, implies that

$$|||u_h|||_{\mathrm{SD}} \leq (8/c)^{\frac{1}{2}} \|f\|_{\mathrm{L}^2(\Omega)}, \tag{4.7}$$

and hence the stability of the method for all $\delta_L \in \mathbb{R}_{\geq 0}$ such that

$$0 \leq \delta_L \leq \min\left(\frac{h_L^2}{12dp^4|\sqrt{a}\,|^2}, \frac{1}{c}\right).$$

We note here that in the case of $p = 1$ and $a \geq 0$ diagonal, the constant $\frac{1}{2}$ in the coercivity result $b_\delta(v_h, v_h) \geq \frac{1}{2}|||v_h|||^2_{\mathrm{SD}}$ stated in (4.6) can be replaced by 1, under the simpler condition $0 \leq c\delta_L \leq 1$, which does not involve the matrix norm $|\sqrt{a}\,|$ or the dimension d. Consequently, when $c > 0$ the constant $(8/c)^{\frac{1}{2}}$ in the stability inequality (4.7) can then be improved to $(2/c)^{\frac{1}{2}}$, under this same condition on δ_L.

In Section 6 we shall consider the convergence analysis of the method (4.4); we shall require the following multiplicative trace inequality, with explicit dependence on the dimension d.

Lemma 4.2 (multiplicative trace inequality). *Let $\Omega = (0,1)^d$ where $d \geq 2$ and suppose that Γ_+ is the hyperbolic outflow part of Γ. Then,*

$$\int_{\Gamma_+} |v|^2 \, \mathrm{d}s \leq 4d\|v\|_{\mathrm{L}^2(\Omega)} \, \|v\|_{\mathrm{H}^1(\Omega)} \qquad \forall v \in \mathrm{H}^1(\Omega).$$

Proof. We shall prove the inequality for $v \in \mathrm{C}^1(\bar{\Omega})$. For $v \in \mathrm{H}^1(\Omega)$ the result follows from the density of $\mathrm{C}^1(\bar{\Omega})$ in $\mathrm{H}^1(\Omega)$. As we have noted before, Γ_+ is a union of $(d-1)$-dimensional open faces of Ω. Let us suppose without loss of generality that the face $x_1 = 0$ of Ω belongs to Γ_+. Then,

$$v^2(0, x') = v^2(x_1, x') + \int_{x_1}^0 \frac{\partial}{\partial x_1} v^2(\xi, x') \, \mathrm{d}\xi, \qquad x' = (x_2, \ldots, x_n).$$

Hence, on integrating this over $x = (x_1, x') \in (0,1) \times (0,1)^{d-1} = \Omega$,

$$\begin{aligned}\int_{x' \in (0,1)^{d-1}} v^2(0, x') \, \mathrm{d}x' &= \int_0^1 \int_{x' \in (0,1)^{d-1}} v^2(x_1, x') \, \mathrm{d}x' \, \mathrm{d}x_1 \\ &\quad + 2\int_0^1 \int_{x' \in (0,1)^{d-1}} \int_{x_1}^0 v(\xi, x') \frac{\partial}{\partial x_1} v(\xi, x') \, \mathrm{d}\xi \, \mathrm{d}x' \, \mathrm{d}x_1 \\ &\leq \|v\|^2_{\mathrm{L}^2(\Omega)} + 2\|v\|_{\mathrm{L}^2(\Omega)} \, \|v_{x_1}\|_{\mathrm{L}^2(\Omega)}.\end{aligned}$$

In the generic case when $\beta > 0$ on the whole of Γ_+, the set Γ_+ will contain at most d of the $2d$ faces of Ω – at most one complete face of Ω orthogonal to the ith co-ordinate direction, $i = 1, \ldots, d$. Otherwise, if $\beta = 0$ on certain faces that belong to Γ_+, the set Γ_+ may contain as many as $2d - 1$ of the $2d$ faces of Ω. Thus, in the worst case,

$$\int_{\Gamma_+} |v|^2 \, \mathrm{d}s \leq (2d-1)\|v\|^2_{\mathrm{L}^2(\Omega)} + 4\|v\|_{\mathrm{L}^2(\Omega)} \sum_{i=1}^d \|v_{x_i}\|_{\mathrm{L}^2(\Omega)}. \tag{4.8}$$

Therefore,

$$\int_{\Gamma_+} |v|^2 \, \mathrm{d}s \leq 2d\sqrt{2} \max\left(1, \frac{2}{d^{\frac{1}{2}}}\right) \|v\|_{\mathrm{L}^2(\Omega)} \, \|v\|_{\mathrm{H}^1(\Omega)} \leq 4d\|v\|_{\mathrm{L}^2(\Omega)} \, \|v\|_{\mathrm{H}^1(\Omega)}.$$

Hence the required result. □

Remark 4.1. It follows from (4.8) that by altering the definition of the $\mathrm{H}^1(\Omega)$ norm in a similar manner as in (3.11), the constant in Lemma 4.2 can be slightly improved:

$$\int_{\Gamma_+} |v|^2 \, \mathrm{d}s \le 2d\|v\|_{\mathrm{L}^2(\Omega)} \, \|v\|_{\mathrm{H}^1_*(\Omega)}, \qquad \text{where} \qquad \|v\|_{\mathrm{H}^1_*(\Omega)} := \|v\|_{\mathrm{L}^2(\Omega)} + \sum_{i=1}^{d} \|v_{x_i}\|_{\mathrm{L}^2(\Omega)}.$$

5. Approximation rates of sparse finite element spaces

In Section 6 we develop the convergence analysis of the stabilized sparse finite element method. Our error bounds will include constants whose precise dependence on the dimension d will be explicitly tracked. For this purpose, we require optimal approximation results from the sparse finite element space, with explicit dependence of the error constants on d. To this end we first prove, in Section 5.1, some combinatorial bounds. A second key ingredient in our argument is a result, established in Section 5.2, concerning linear operators, which are bounded in semi-norms, on tensor-products of separable Hilbert spaces. As before, $\mathbb{N}$ will denote the set of all non-negative integers and $\mathbb{N}_{>0}$ will signify the set of all positive integers.

5.1. Some combinatorial bounds

Lemma 5.1. *For $d \in \mathbb{N}_{>0}$ and $x > 1$ we have that*

$$\sup_{m\in\mathbb{N}} \sum_{\substack{\ell\in\mathbb{N}^d \\ |\ell|_1=m}} x^{|\ell|_\infty - m} = d\left(1 + \frac{1}{x-1}\right)^{d-1}. \tag{5.1}$$

Proof. The case $d = 1$ being trivial, we assume without loss of generality that $d \ge 2$. Let us denote by $\Sigma(m, x, d)$ the sum in (5.1) and rewrite it as

$$\Sigma(m,x,d) := \sum_{k=0}^{\infty} \sum_{\substack{\ell\in\mathbb{N}^d \\ |\ell|_1=m,\, |\ell|_\infty=k}} x^{k-m} = \sum_{k=0}^{\infty} |\mathcal{S}(m,k,d)| x^{k-m}, \tag{5.2}$$

where, for $m, k \in \mathbb{N}$, the set $\mathcal{S}(m, k, d)$ is defined by

$$\mathcal{S}(m,k,d) := \{\ell \in \mathbb{N}^d \,:\, |\ell|_1 = m,\, |\ell|_\infty = k\}. \tag{5.3}$$

We deduce from (5.9) in Lemma 5.2 below that

$$d \sum_{m/2<k\le m} \binom{m-k+d-2}{d-2} x^{k-m} \le \Sigma(m,x,d) \le d \sum_{k=0}^{m} \binom{m-k+d-2}{d-2} x^{k-m}. \tag{5.4}$$

The statement of the theorem will follow once we have shown that the suprema over $m \in \mathbb{N}$ of both the lower and the upper bound in (5.4) are equal to the right-hand side of (5.1).

We begin by considering the upper bound in (5.4), which can be written, after substituting $m - k$ by k, as

$$d \sum_{k=0}^{m} \binom{k+d-2}{d-2} \left(\frac{1}{x}\right)^k.$$

The supremum over $m \in \mathbb{N}$ is thus attained for $m \to \infty$ and equals

$$d\left(\frac{1}{1-1/x}\right)^{d-1}. \tag{5.5}$$

Note that here we have used the identity

$$\frac{1}{(1-q)^{n+1}} = \sum_{k=0}^{\infty} \binom{k+n}{n} q^k \qquad \forall n \in \mathbb{N},\ \forall q \in (-1,1),$$

which follows by differentiating n times with respect to q the identity $(1-q)^{-1} = 1 + q + q^2 + \dots$.

Now we use a similar argument to compute the supremum over $m \in \mathbb{N}$ of the lower bound in (5.4), which can be written, again after substituting $m-k$ by k, as

$$d \sum_{0 \le k < m/2} \binom{k+d-2}{d-2} \left(\frac{1}{x}\right)^k, \qquad x > 1.$$

The supremum over $m \in \mathbb{N}$ is attained again for $m \to \infty$ and equals (5.5). □

In particular, it is a simple consequence of this theorem that, for any $d, m \in \mathbb{N}_{>0}$ and $x > 1$, we have that

$$d \cdot x^m \le \sum_{\substack{\ell \in \mathbb{N}^d \\ |\ell|_1 = m}} x^{|\ell|_\infty} \le d \left(1 + \frac{1}{x-1}\right)^{d-1} \cdot x^m, \tag{5.6}$$

the lower bound being trivial. It remains to prove the following lemma that was used in (5.4).

Lemma 5.2. *Consider the sets $\mathcal{S}(m,k,d)$ defined, for $d \in \mathbb{N}_{>0}$ and $m, k \in \mathbb{N}$ in (5.3). Then,*

$$\mathcal{S}(m,k,d) = \emptyset \qquad \forall k > m, \tag{5.7}$$

$$\sum_{k=0}^{\infty} |\mathcal{S}(m,k,d)| = \binom{m+d-1}{d-1}, \tag{5.8}$$

$$|\mathcal{S}(m,k,d)| \le d \binom{m-k+d-2}{d-2} \qquad \forall d \ge 2, \tag{5.9}$$

with equality for $k > m/2$.

Proof. We note that (5.7) is obvious, whereas (5.8) follows from the fact that for fixed m, d, the sets $(\mathcal{S}(m,k,d))_{0 \le k \le m}$ are disjoint and

$$\bigcup_{k=0}^{m} \mathcal{S}(m,k,d) = \{\ell \in \mathbb{N}^d : |\ell|_1 = m\}.$$

To prove (5.9) we consider, for fixed k, m with $0 \le k \le m$, the mapping

$$\{1, 2, \dots, d\} \times \bigcup_{j=0}^{k} \mathcal{S}(m-k, j, d-1) \xrightarrow{\varphi} \mathcal{S}(m,k,d)$$

given by

$$\varphi(q, (l_1, l_2, \dots, l_{d-1})) = (l_1, l_2, \dots, l_{q-1}, k, l_q, \dots, l_{d-1}).$$

Obviously, φ is surjective, so that we obtain, using (5.8),

$$|\mathcal{S}(m,k,d)| \quad \leq \quad |\{1,2,\ldots,d\}| \cdot \sum_{j=0}^{k} |\mathcal{S}(m-k,j,d-1)| \tag{5.10}$$

$$\leq \quad d\binom{m-k+d-2}{d-2}. \tag{5.11}$$

For $k > m/2$ the mapping φ is also injective, which ensures equality in (5.10). Also (5.11) holds with equality for $k > m/2$ due to (5.7) and (5.8). □

Remark 5.1. Of particular interest in the H^1-seminorm bounds below is the case $x = 2$, for which (5.1) becomes

$$\sum_{\substack{\ell\in\mathbb{N}^d \\ |\ell|_1=m}} 2^{|\ell|_\infty} \leq d \cdot 2^{d-1+m} \qquad \forall m \in \mathbb{N},\ \forall d \in \mathbb{N}_{>0}. \tag{5.12}$$

Before stating our next set of combinatorial bounds on lattice sums, we define, for $L \in \mathbb{N}$, $d \in \mathbb{N}_{>0}$, $x > 1$ and $\alpha, \beta > 0$, the following expressions:

$$A(L,d,x) \quad := \quad \sum_{\substack{\ell\in\mathbb{N}^d \\ |\ell|_1>L}} x^{-|\ell|_1}, \tag{5.13}$$

$$B(L,d,x,\alpha,\beta) \quad := \quad \sum_{k=1}^{d} \binom{d}{k} \alpha^k \beta^{d-k} A(L,k,x). \tag{5.14}$$

In our analysis of these quantities, we will often use that

$$A(L,k,x) = \sum_{m=L+1}^{\infty} \binom{m+k-1}{k-1} x^{-m}, \tag{5.15}$$

which follows on noting that, by (5.3) and (5.8), we have

$$\begin{aligned} A(L,d,x) \quad &= \quad \sum_{m=L+1}^{\infty} \sum_{\substack{\ell\in\mathbb{N}^d \\ |\ell|_1=m}} x^{-|\ell|_1} = \sum_{m=L+1}^{\infty} \sum_{k=0}^{\infty} \sum_{\substack{\ell\in\mathbb{N}^d \\ |\ell|_1=m,\ |\ell|_\infty=k}} x^{-|\ell|_1} \\ &= \quad \sum_{m=L+1}^{\infty} \sum_{k=0}^{\infty} |S(m,k,d)|\, x^{-m} = \sum_{m=L+1}^{\infty} \binom{m+d-1}{d-1} x^{-m}. \end{aligned}$$

The roles of α and β in (5.14) above will be played by the estimated error constants in the approximation properties of the univariate FE projectors appearing in (3.9), (3.10) and (3.14). The following lemma will be used in order to establish a sharp upper bound on

$$A(L,d,2^t) = \sum_{\substack{\ell\in\mathbb{N}^d \\ |\ell|_1>L}} 2^{-t|\ell|_1}. \tag{5.16}$$

Lemma 5.3. *For $L, n \in \mathbb{N}$ and $x > 1$ we have*

$$A(L,n+1,x) \leq \frac{1+\theta^{-1}}{1-x^{-1}} \cdot \left(1+\frac{x-1}{1+\theta}\right)^{L+1} \left(1+\frac{1+\theta}{x-1}\right)^{n} \cdot x^{-(L+1)}, \tag{5.17}$$

where

$$\theta := (x-1)\frac{L+1}{n}, \tag{5.18}$$

with the convention that if $n = 0$, *then* $\theta = \infty$ *in* (5.18); *in this case the right-hand side of* (5.17) *is to be understood as the limiting value* $x^{-(L+1)}/(1-x^{-1})$ *when* $n \to 0$.

Proof. Using (5.15) and the summation formula

$$\sum_{m=L+1}^{\infty} \binom{m+n}{n} x^{-m} = \frac{1}{n!} \cdot \frac{\mathrm{d}^n}{\mathrm{d}z^n} \left(\frac{z^{L+n+1}}{1-z} \right) \Big|_{z=x^{-1}},$$

and the Cauchy inequality for holomorphic functions

$$|f^{(n)}(z_0)| \le n! \cdot \frac{\max_{\{z:\,|z-z_0|=r\}} |f(z)|}{r^n}$$

with $z_0 = x^{-1}$, $f(z) = z^{L+n+1}/(1-z)$ and $r \in (0, 1-x^{-1})$ arbitrary, we deduce that

$$\begin{aligned} \sum_{m=L+1}^{\infty} \binom{m+n}{n} x^{-m} &\le \frac{(r+x^{-1})^{L+n+1}}{(1-r-x^{-1})r^n} = \frac{(1+rx)^{L+n+1}}{(rx)^n(1-x^{-1}(1+rx))} \cdot x^{-(L+1)} \\ &= \frac{1}{1-x^{-1}(1+rx)} \cdot (1+rx)^{L+1} \left(1+\frac{1}{rx}\right)^n \cdot x^{-(L+1)}. \end{aligned} \tag{5.19}$$

The optimal value of r to be used in the right-hand side of (5.19) can be obtained by minimizing ($y := rx$)

$$y \in (0, x-1) \mapsto \frac{(1+y)^{L+n+1}}{y^n(1-x^{-1}(1+y))} \in (0,\infty).$$

Elementary arguments show that the minimum of (5.19) is attained for y equal to the smallest solution of

$$Ly^2 - (n(1+\theta)+1)y + n^2\theta/(L+1) = 0,$$

with θ as in (5.18). More precisely, we have

$$\begin{aligned} y &= \frac{2n^2\theta/(L+1)}{n(1+\theta)+1+\sqrt{(n(1+\theta)+1)^2 - 4Ln^2\theta/(L+1)}} \\ &\in \left[\frac{n^2\theta/(L+1)}{n(1+\theta)+1}, 2\frac{n^2\theta/(L+1)}{n(1+\theta)+1}\right]. \end{aligned} \tag{5.20}$$

From (5.20) we deduce that $y \sim n\theta/(L+1)(1+\theta)$, so for simplicity we choose in (5.19)

$$r = x^{-1}\frac{n\theta}{(L+1)(1+\theta)} = \frac{1-x^{-1}}{1+\theta} \in (0, 1-x^{-1}).$$

Simple algebraic manipulations of (5.19) then lead to the desired estimate (5.17). □

To derive bounds for $B(L,d,x,\alpha,\beta)$ in (5.14), we will use Lemma 5.3.

Lemma 5.4. *For* $L \in \mathbb{N}$, $d \in \mathbb{N}_{>0}$, $\alpha, \beta > 0$, *and* $x \ge 2$,

$$B(L,d,x,\alpha,\beta) \le \frac{de\alpha x}{x-1} \cdot (\alpha(L+1)\mathrm{e}^{1/(L+1)} + \beta)^{d-1} \cdot x^{-(L+1)}. \tag{5.21}$$

Proof. Noting (5.15) and using Lemma 5.3, with $n := k-1$ and $\theta := (x-1)(L+1)/(k-1)$, we obtain

$$B(L,d,x,\alpha,\beta) \;\le\; \sum_{k=1}^{d}\binom{d}{k}\alpha^k\beta^{d-k}\cdot\frac{1+\theta^{-1}}{1-x^{-1}}\left(1+\frac{x-1}{1+\theta}\right)^{L+1}\left(1+\frac{1+\theta}{x-1}\right)^{k-1}\cdot x^{-(L+1)}.$$

In (5.18), for $n = k-1$ with $k=1$, we formally take $\theta=\infty$, so the entry in the sum on the right-hand side corresponding to $k=1$ should be understood as the limiting value $\binom{d}{1}\alpha\beta^{d-1}\cdot\frac{x}{x-1}\cdot x^{-(L+1)}$, as $k\to 1$.

Introducing the notations $y := 1/(x-1)$ and $m := (L+1)/(k-1)$ we write

$$\begin{aligned}
\left(1+\frac{x-1}{1+\theta}\right)^{L+1}\left(1+\frac{1+\theta}{x-1}\right)^{k-1} &= \left(1+\frac{k-1}{L+1}\cdot\frac{\theta}{1+\theta}\right)^{L+1}\left(1+\frac{1}{x-1}+\frac{L+1}{k-1}\right)^{k-1}\\
&\le \mathrm{e}^{\theta(k-1)/(1+\theta)}(1+y+m)^{k-1}\\
&\le \frac{\mathrm{e}^{k-1}}{(1+1/(1+\theta))^{k-1}}(1+y+m)^{k-1}\\
&= \mathrm{e}^{k-1}\left(\frac{1+y+m}{1+y/(y+m)}\right)^{k-1}\\
&\overset{x\ge 2}{\le} \mathrm{e}^{k-1}(1+m)^{k-1},
\end{aligned}\tag{5.22}$$

with equalities in the special case of $k=1$. Additionally,

$$\frac{1+\theta^{-1}}{1-x^{-1}} \le (k-1)x/(x-1)^2 + x/(x-1) \overset{x\ge 2}{\le} kx/(x-1),$$

again with equalities in the special case of $k=1$, so that

$$\begin{aligned}
B(L,d,x,\alpha,\beta) &\le \frac{x}{x-1}\sum_{k=1}^{d}k\binom{d}{k}\alpha^k\beta^{d-k}\mathrm{e}^{k-1}\left(\frac{L+k}{k-1}\right)^{k-1}\cdot x^{-(L+1)}\\
&= \frac{d\alpha x}{x-1}\sum_{k=1}^{d}\binom{d-1}{k-1}(\alpha\mathrm{e})^{k-1}\beta^{d-1-(k-1)}\left(\frac{L+k}{k-1}\right)^{k-1}\cdot x^{-(L+1)}\\
&= \frac{d\alpha x}{x-1}\sum_{k=1}^{d}\binom{d-1}{k-1}(\alpha\gamma\mathrm{e})^{k-1}\beta^{d-1-(k-1)}\left(\frac{L+k}{\gamma(k-1)}\right)^{k-1}\cdot x^{-(L+1)}
\end{aligned}$$

for $\gamma>0$ to be chosen later. Using the elementary inequality $(a/x)^x \le \mathrm{e}^{a/\mathrm{e}}$, $a,x>0$, we obtain (with the same convention as above in the case $k=1$) that

$$\begin{aligned}
B(L,d,x,\alpha,\beta) &\le \frac{d\alpha x}{x-1}\sum_{k=1}^{d}\binom{d-1}{k-1}(\alpha\gamma\mathrm{e})^{k-1}\beta^{d-1-(k-1)}\mathrm{e}^{(L+k)/\gamma\mathrm{e}}\cdot x^{-(L+1)}\\
&= \frac{d\alpha x}{x-1}\mathrm{e}^{(L+1)/\gamma\mathrm{e}}(\alpha\gamma\mathrm{e}^{1+1/\gamma\mathrm{e}}+\beta)^{d-1}\cdot x^{-(L+1)}.
\end{aligned}$$

The proof is now concluded by choosing $\gamma = (L+1)/\mathrm{e}$. □

The next result shows the existence of a *preasymptotic domain* $[0, L_0(\alpha,\beta,d)]$ of the mesh refinement level L where $B := B(L,d,x,\alpha,\beta)$ is free of powers of L, *i.e.*,

$$B \le C_0\, x^{-(L+1)}$$

with a constant $C_0 = C_0(d)$ decaying exponentially in d, uniformly with respect to $L \in [0, L_0(\alpha, \beta, d)]$ where L_0 *increases linearly with* d.

Lemma 5.5. *Let* $L \in \mathbb{N}$, $d \in \mathbb{N}_{>0}$, $\alpha, \beta > 0$, *and* $x \geq 2$. *If*

$$\gamma := \alpha x/(x-1) + \beta < 1, \tag{5.23}$$

then there exist constants $c_{1,x,\gamma} > 0, c_{2,x,\gamma} \in (0,1)$ *such that for all*

$$d \geq 2 \text{ and } L + 1 \leq c_{1,x,\gamma}(d-1) \tag{5.24}$$

we have

$$B(L,d,x,\alpha,\beta) \leq \frac{d\alpha x}{x-1} \cdot c_{2,x,\gamma}^{d-1} \cdot x^{-(L+1)}. \tag{5.25}$$

Proof. As in the proof of Lemma 5.4 and with the same convention as above in the case of $k = 1$, we write

$$B(L,d,x,\alpha,\beta) \leq \sum_{k=1}^{d} \binom{d}{k} \alpha^k \beta^{d-k} \cdot \frac{1+\theta^{-1}}{1-x^{-1}} \left(1 + \frac{x-1}{1+\theta}\right)^{L+1} \left(1 + \frac{1+\theta}{x-1}\right)^{k-1} \cdot x^{-(L+1)}, \tag{5.26}$$

where

$$\frac{1+\theta^{-1}}{1-x^{-1}} \leq (k-1)x/(x-1)^2 + x/(x-1) \overset{x\geq 2}{\leq} kx/(x-1).$$

With the notations $B := B(L,d,x,\alpha,\beta)$, $y := 1/(x-1)$, $m := (L+1)/(k-1)$, we rewrite the right-hand side of (5.26) as

$$\begin{aligned} B &\leq \frac{x}{x-1} \sum_{k=1}^{d} k \binom{d}{k} \alpha^k \beta^{d-k} \left(1 + \frac{1}{m+y}\right)^{L+1} (1+m+y)^{k-1} \cdot x^{-(L+1)} \\ &= \frac{d\alpha x}{x-1} \sum_{k=1}^{d} \binom{d-1}{k-1} \alpha^{k-1} \beta^{d-1-(k-1)} \left(1 + \frac{1}{m+y}\right)^{L+1} (1+m+y)^{k-1} \cdot x^{-(L+1)}. \end{aligned} \tag{5.27}$$

For $k = 1$ we have $m = \infty$ and the corresponding entries in the sums above are again to be understood as the limiting values for $k \to 1$. Now, fixing $m_0 > 0$ to be chosen later, we split the summation over k as follows

$$B \leq \frac{d\alpha x}{x-1} \left(\sum_{\substack{k=1 \\ m<m_0}}^{d} \ldots + \sum_{\substack{k=1 \\ m\geq m_0}}^{d} \ldots \right) =: \frac{d\alpha x}{x-1} \left(\mathrm{S}_1 + \mathrm{S}_2\right).$$

For the first sum, S_1, the inequality $m < m_0$ ensures that $L + 1 < m_0(k-1)$, so that

$$\begin{aligned} \mathrm{S}_1 &\leq \sum_{k=1}^{d} \binom{d-1}{k-1} \alpha^{k-1} \beta^{d-1-(k-1)} \left(1 + \frac{1}{m_0+y}\right)^{m_0(k-1)} (1+m_0+y)^{k-1} \cdot x^{-(L+1)} \\ &\leq \left[\alpha(1+m_0+y)\left(1 + \frac{1}{m_0+y}\right)^{m_0} + \beta\right]^{d-1} \cdot x^{-(L+1)}. \end{aligned} \tag{5.28}$$

Due to (5.23) we can now choose $m_0 > 0$ small enough depending on x, α and β to ensure that the expression in square brackets on the right-hand side of (5.28) does not exceed 1.

To estimate the second sum, S_2, note that $m \geq m_0$ implies, with the same convention as above for the case of $k = 1$, that

$$\begin{aligned}
\mathrm{S}_2 &\leq \sum_{\substack{k=1 \\ m \geq m_0}}^{d} \binom{d-1}{k-1} \alpha^{k-1} \beta^{d-1-(k-1)} \left(1 + \frac{1}{m_0 + y}\right)^{L+1} (1+y)^{k-1} \left(1 + m/(1+y)\right)^{k-1} \cdot x^{-(L+1)} \\
&\leq (\alpha(1+y) + \beta)^{d-1} \left(1 + \frac{1}{1+y} \frac{L+1}{d-1}\right)^{d-1} \left(1 + \frac{1}{m_0 + y}\right)^{L+1} \cdot x^{-(L+1)} \\
&\leq \left[(\alpha(1+y) + \beta) \left(1 + \frac{c_{1,x,\gamma}}{1+y}\right) \left(1 + \frac{1}{m_0 + y}\right)^{c_{1,x,\gamma}}\right]^{d-1} \cdot x^{-(L+1)}. \qquad (5.29)
\end{aligned}$$

To obtain the second inequality above we have also used the monotonicity of

$$t \in (0, \infty) \mapsto (1 + a/t)^t \in (1, \infty), \qquad a > 0.$$

The proof is then concluded by noting that if $c_{1,x,\gamma}$ is small enough depending on m_0 and y (that is, on x, α and β), then, under our assumption that $\gamma < 1$, the expression in square brackets on the right-hand side of (5.29) is also less than 1. □

Remark 5.2. *In the case of $d = 1$ the conclusion (5.25) of Lemma 5.5 holds with equality for any $L \in \mathbb{N}$, and without the additional assumptions (5.23) and (5.24), since*

$$A(L, 1, x) = \frac{x}{x-1} \cdot x^{-(L+1)} \quad \textit{and} \quad B(L, 1, x, \alpha, \beta) = \frac{\alpha x}{x-1} \cdot x^{-(L+1)}, \quad x > 1.$$

5.2. **Tensorization of seminorms**

Next we develop some auxiliary results concerning tensorization of seminorms. These will then be used in the derivation of the approximation properties of the d-dimensional sparse tensor-product space, built from the univariate finite element space scale $\{\mathcal{V}_{(0)}^{\ell,p}\}_{\ell \geq 0}$.

Let $(\mathrm{H}, \langle\cdot,\cdot\rangle_{\mathrm{H}})$ and $(\mathrm{K}, \langle\cdot,\cdot\rangle_{\mathrm{K}})$ be two Hilbert spaces and $T \in \mathcal{B}(\mathrm{H}, \mathrm{K})$ a bounded linear operator. Clearly,

$$|u|_T := \|Tu\|_{\mathrm{K}}, \quad u \in \mathrm{H}, \qquad (5.30)$$

defines a seminorm on H.

Now, considering four Hilbert spaces $(\mathrm{H}_i, \langle\cdot,\cdot\rangle_{\mathrm{H}_i})$, $(\mathrm{K}_i, \langle\cdot,\cdot\rangle_{\mathrm{K}_i})$, $i = 1, 2$, as well as two bounded linear operators $T_i \in \mathcal{B}(\mathrm{H}_i, \mathrm{K}_i)$, $i = 1, 2$, it is natural to define *via*

$$|u|_{T_1 \otimes T_2} := \|(T_1 \otimes T_2)u\|_{\mathrm{K}_1 \otimes \mathrm{K}_2}, \quad u \in \mathrm{H}_1 \otimes \mathrm{H}_2,$$

a seminorm on the tensor-product $\mathrm{H}_1 \otimes \mathrm{H}_2$ of the spaces H_1 and H_2.

Next we define bounded linear operators with respect to seminorms of the type (5.30) and investigate their tensor-products.

Definition 5.1. Let $(\mathrm{H}, \langle\cdot,\cdot\rangle_{\mathrm{H}})$, $(\mathrm{K}, \langle\cdot,\cdot\rangle_{\mathrm{K}})$, $(\tilde{\mathrm{H}}, \langle\cdot,\cdot\rangle_{\tilde{\mathrm{H}}})$, $(\tilde{\mathrm{K}}, \langle\cdot,\cdot\rangle_{\tilde{\mathrm{K}}})$ be four Hilbert spaces and consider the bounded linear operators $T \in \mathcal{B}(\mathrm{H}, \mathrm{K})$, $\tilde{T} \in \mathcal{B}(\tilde{\mathrm{H}}, \tilde{\mathrm{K}})$ and $Q \in \mathcal{B}(\mathrm{H}, \tilde{\mathrm{H}})$. We say that Q is $(T, \tilde{T})$-bounded if there exists $c \geq 0$ such that

$$|Qu|_{\tilde{T}} \leq c|u|_T \quad \forall u \in \mathrm{H}. \qquad (5.31)$$

We further denote by $|Q|_{T,\tilde{T}}$ the infimum over all constants $c \geq 0$ satisfying (5.31).

Example 5.1. We give some examples based on the bounds in Section 3.2.

(a) We use the terminology from Definition 5.1, with $\mathrm{H} := \mathrm{H}^{t+1}(0,1) \cap \mathrm{H}^1_{(0)}(0,1)$, and let $\tilde{\mathrm{H}} := \mathrm{H}^s(0,1)$, $\mathrm{K} = \tilde{\mathrm{K}} := \mathrm{L}^2(0,1)$, $T := \partial^{t+1}$, $\tilde{T} := \partial^s$, with $t \geq 1$ and $s \in \{0,1\}$. The approximation property (3.5) shows that the linear operator $\mathrm{Id}_{\mathrm{H}} - P^{\ell,p}_{(0)}$ is $(\partial^{t+1}, \partial^s)$-bounded. Thus, by (3.9), the projector $Q^{\ell,p}_{(0)}$ is also $(\partial^{t+1}, \partial^s)$-bounded for all $\ell \geq 1$ and $p \geq 1$ (with $\partial^0 := \mathrm{Id}_{\mathrm{L}^2(0,1)}$).

(b) Trivially, on taking $\mathrm{H} = \tilde{\mathrm{H}} := \mathrm{H}^1_{(0)}(0,1)$ and $\mathrm{K} = \tilde{\mathrm{K}} := \mathrm{L}^2(0,1)$, the projector $Q^{\ell,p}_{(0)}$ is (∂^1, ∂^1)-bounded for all $\ell \geq 0$ and $p \geq 1$.

(c) Finally, on taking $\mathrm{H} = \mathrm{K} := \mathrm{H}^1_{(0)}(0,1)$ (equipped with the norm $\|\cdot\|_{\mathrm{H}^1_*(0,1)}$) and $\tilde{\mathrm{H}} = \tilde{\mathrm{K}} := \mathrm{L}^2(0,1)$, we see that $Q^{0,p}_{(0)}$ is $(\mathrm{Id}_{\mathrm{H}^1_{(0)}(0,1)}, \mathrm{Id}_{\mathrm{L}^2(0,1)})$-bounded for all $p \geq 1$. In particular, on taking $\mathrm{H} := \mathrm{H}^1_0(0,1)$, $\tilde{\mathrm{H}} := \mathrm{L}^2(0,1)$ and $\mathrm{K} = \tilde{\mathrm{K}} := \mathrm{L}^2(0,1)$ we see that $Q^{0,p}_0$ is $(\partial^1, \mathrm{Id}_{\mathrm{L}^2(0,1)})$-bounded for all $p \geq 1$.

Proposition 5.1. *Let $(\mathrm{H}_i, \langle\cdot,\cdot\rangle_{\mathrm{H}_i})$, $(\mathrm{K}_i, \langle\cdot,\cdot\rangle_{\mathrm{K}_i})$, $(\tilde{\mathrm{H}}_i, \langle\cdot,\cdot\rangle_{\tilde{\mathrm{H}}_i})$, $(\tilde{\mathrm{K}}_i, \langle\cdot,\cdot\rangle_{\tilde{\mathrm{K}}_i})$ for $i = 1,2$ be separable Hilbert spaces. Let $T_i \in \mathcal{B}(\mathrm{H}_i, \mathrm{K}_i), \tilde{T}_i \in \mathcal{B}(\tilde{\mathrm{H}}_i, \tilde{\mathrm{K}}_i)$ and $Q_i \in \mathcal{B}(\mathrm{H}_i, \tilde{\mathrm{H}}_i)$ be bounded linear operators, and assume that Q_i is $(T_i, \tilde{T}_i)$-bounded for $i = 1,2$. Then $Q_1 \otimes Q_2$ is $(T_1 \otimes T_2, \tilde{T}_1 \otimes \tilde{T}_2)$-bounded, and*

$$|Q_1 \otimes Q_2|_{T_1 \otimes T_2, \tilde{T}_1 \otimes \tilde{T}_2} \leq |Q_1|_{T_1, \tilde{T}_1} |Q_2|_{T_2, \tilde{T}_2}.$$

In other words, if $\|\tilde{T}_i Q_i v_i\|_{\tilde{\mathrm{K}}_i} \leq c_i \|T_i v_i\|_{\mathrm{K}_i}$ for all $v_i \in \mathrm{H}_i$, $i = 1,2$, then

$$\|(\tilde{T}_1 \otimes \tilde{T}_2)(Q_1 \otimes Q_2)u\|_{\tilde{\mathrm{K}}_1 \otimes \tilde{\mathrm{K}}_2} \leq c_1 c_2 \|(T_1 \otimes T_2)u\|_{\mathrm{K}_1 \otimes \mathrm{K}_2} \qquad \forall u \in \mathrm{H}_1 \otimes \mathrm{H}_2.$$

Proof. For any $u \in \mathrm{H}_1 \otimes \mathrm{H}_2$ we have

$$\begin{aligned} |(Q_1 \otimes Q_2)u|_{\tilde{T}_1 \otimes \tilde{T}_2} &= \|(\tilde{T}_1 \otimes \tilde{T}_2)(Q_1 \otimes Q_2)u\|_{\tilde{\mathrm{K}}_1 \otimes \tilde{\mathrm{K}}_2} \\ &= \|(\tilde{T}_1 Q_1 \otimes \mathrm{Id}_{\tilde{\mathrm{K}}_2})(\mathrm{Id}_{\mathrm{H}_1} \otimes \tilde{T}_2 Q_2)u\|_{\tilde{\mathrm{K}}_1 \otimes \tilde{\mathrm{K}}_2}. \end{aligned} \tag{5.32}$$

Denoting $v := (\mathrm{Id}_{\mathrm{H}_1} \otimes \tilde{T}_2 Q_2)u \in \mathrm{H}_1 \otimes \tilde{\mathrm{K}}_2$ and considering an orthonormal basis $(e_i)_{i \in I}$ in $\tilde{\mathrm{K}}_2$, where $I \subset \mathbb{N}$ is a countable index set, we expand $v = \sum_{i \in I} v_i \otimes e_i$, so that

$$\begin{aligned} \|(\tilde{T}_1 Q_1 \otimes \mathrm{Id}_{\tilde{\mathrm{K}}_2})v\|^2_{\tilde{\mathrm{K}}_1 \otimes \tilde{\mathrm{K}}_2} &= \sum_{i \in I} \|\tilde{T}_1 Q_1 v_i\|^2_{\tilde{\mathrm{K}}_1} \\ &\overset{(5.31)}{\leq} c_1^2 \sum_{i \in I} \|T_1 v_i\|^2_{\mathrm{K}_1} \\ &= c_1^2 \|(T_1 \otimes \mathrm{Id}_{\tilde{\mathrm{K}}_2})v\|^2_{\mathrm{K}_1 \otimes \tilde{\mathrm{K}}_2}, \end{aligned} \tag{5.33}$$

where $c_1 = |Q_1|_{T_1, \tilde{T}_1}$. We now note that

$$(T_1 \otimes \mathrm{Id}_{\tilde{\mathrm{K}}_2})v = (T_1 \otimes \mathrm{Id}_{\tilde{\mathrm{K}}_2})(\mathrm{Id}_{\mathrm{H}_1} \otimes \tilde{T}_2 Q_2)u = (\mathrm{Id}_{\mathrm{K}_1} \otimes \tilde{T}_2 Q_2)(T_1 \otimes \mathrm{Id}_{\mathrm{H}_2})u,$$

so that defining $w := (T_1 \otimes \mathrm{Id}_{\mathrm{H}_2})u \in \mathrm{K}_1 \otimes \mathrm{H}_2$ and arguing as in (5.33) to estimate the norm of $(\mathrm{Id}_{\mathrm{K}_1} \otimes \tilde{T}_2 Q_2)w$, we obtain

$$\|(\mathrm{Id}_{\mathrm{K}_1} \otimes \tilde{T}_2 Q_2)w\|_{\mathrm{K}_1 \otimes \tilde{\mathrm{K}}_2} \leq c_2 \|(\mathrm{Id}_{\mathrm{K}_1} \otimes T_2)w\|_{\mathrm{K}_1 \otimes \mathrm{K}_2} = c_2 \|(T_1 \otimes T_2)u\|_{\mathrm{K}_1 \otimes \mathrm{K}_2}, \tag{5.34}$$

where $c_2 = |Q_2|_{T_2, \tilde{T}_2}$. From (5.32), (5.33), (5.34) we obtain

$$|(Q_1 \otimes Q_2)u|_{\tilde{T}_1 \otimes \tilde{T}_2} \leq c_1 c_2 \|(T_1 \otimes T_2)u\|_{\mathrm{K}_1 \otimes \mathrm{K}_2} = c_1 c_2 |u|_{T_1 \otimes T_2},$$

and the desired result follows by recalling the definitions of the constants c_1, c_2. □

5.3. **Approximation from sparse tensor-product spaces**

We are now ready to embark on the study of the approximation properties of the sparse tensor-product spaces. In order to track the dependence of the constants in the error bounds on the polynomial degree p, the Sobolev regularity t and the dimension d, we consider

$$\Omega := (0,1)^d.$$

This domain has, for any d, d-dimensional Lebesgue measure 1.

To characterize the regularity of the function u to be approximated, we introduce, for $I \subset \{1,2,\dots,d\}$ with $|I| = k \geq 1$, $I = \{i_1, i_2, \dots, i_k\}$, the notation $\mathrm{H}^{\alpha,\beta,I}(\Omega)$ for the tensor-product space consisting of d factors, each of them being either $\mathrm{H}^{\alpha}_{(0)}(0,1)$ (in the jth co-ordinate, if $j \in I$), or $\mathrm{H}^{\beta}_{(0)}(0,1)$ (in the jth co-ordinate, if $j \notin I$).

Given $I = \{i_1, i_2, \dots, i_k\} \subset \{1,2,\dots,d\}$, let $I^c = \{j_1, j_2, \dots, j_{d-k}\}$ denote the (possibly empty) complement of I with respect to $\{1,2,\dots,d\}$; for non-negative integers α and β we then denote by $|u|_{\mathrm{H}^{\alpha,\beta,I}(\Omega)}$ the seminorm

$$\sum_{(\alpha)_1 \leq \alpha_1 \leq \alpha} \cdots \sum_{(\alpha)_k \leq \alpha_k \leq \alpha} \sum_{(\beta)_1 \leq \beta_1 \leq \beta} \cdots \sum_{(\beta)_{d-k} \leq \beta_{d-k} \leq \beta} \left\| \left(\frac{\partial^{\alpha_1}}{\partial x_{i_1}^{\alpha_1}} \cdots \frac{\partial^{\alpha_k}}{\partial x_{i_k}^{\alpha_k}} \right) \left(\frac{\partial^{\beta_1}}{\partial x_{j_1}^{\beta_1}} \cdots \frac{\partial^{\beta_{d-k}}}{\partial x_{j_{d-k}}^{\beta_{d-k}}} \right) u \right\|_{\mathrm{L}^2(\Omega)},$$

where, for $i = 1, \dots, k$,

$$(\alpha)_i := \begin{cases} \alpha & \text{if } Ox_i \text{ is an elliptic co-ordinate direction,} \\ 0 & \text{if } Ox_i \text{ is a hyperbolic co-ordinate direction,} \end{cases}$$

with analogous definition of $(\beta)_j$, $j = 1, \dots, d-k$.

Let $\mathrm{C}^{\infty}_{(0)}(\bar{\Omega})$ denote the set of all functions in $\mathrm{C}^{\infty}(\bar{\Omega})$ that vanish on $\bar{\Gamma}_0$, and let $\mathcal{H}^{t+1}(\Omega)$ denote the closure of $\mathrm{C}^{\infty}_{(0)}(\bar{\Omega})$ in the seminorm $|\cdot|_{\mathcal{H}^{t+1}(\Omega)}$ defined by

$$|u|_{\mathcal{H}^{t+1}(\Omega)} := \max_{s \in \{0,1\}} \max_{1 \leq k \leq d} \left(\max_{\substack{I \subseteq \{1,2,\dots,d\} \\ |I| = k}} |u|_{\mathrm{H}^{t+1,s,I}(\Omega)} \right).$$

Theorem 5.1. *Let $\Omega = (0,1)^d$, $s \in \{0,1\}$, $k \in \mathbb{N}_{>0}$, and let a polynomial degree $p \geq 1$ be given. Then, for $1 \leq t \leq \min\{p,k\}$, there exist constants $\underline{c}_{p,t}$, $\kappa_{(0)}(p,t,s,L) > 0$, independent of d, such that, for any $u \in \mathcal{H}^{k+1}(\Omega)$ and for any $L \geq 1$ and any $d \geq 2$, we have*

$$|u - \hat{P}^{L,p}_{(0)} u|_{\mathrm{H}^s(\Omega)} \leq d^{1+\frac{s}{2}} \underline{c}_{p,t} (\kappa_{(0)}(p,t,s,L))^{d-1+s}\, 2^{-(t+1-s)L}\, |u|_{\mathcal{H}^{t+1}(\Omega)}, \tag{5.35}$$

where, for $s = 0$, the seminorm $|\cdot|_{\mathrm{H}^s(\Omega)}$ is understood to coincide with the $\mathrm{L}^2(\Omega)$-norm while for $s = 1$ the seminorm $|\cdot|_{\mathrm{H}^s(\Omega)}$ is the $\mathrm{H}^1(\Omega)$-seminorm, and

$$\kappa_{(0)}(p,t,s,L) := \begin{cases} \tilde{c}_{p,0,t}(L+1)\mathrm{e}^{1/(L+1)} + \hat{c}_{p,0,(0)}, & s = 0, \\ 2\tilde{c}_{p,0,t} + \hat{c}_{p,0,(0)}, & s = 1. \end{cases} \tag{5.36}$$

Moreover, in the case of $s = 0$, if

$$\gamma_{(0)}(t,p) := \tilde{c}_{p,0,t} 2^{t+1}/(2^{t+1} - 1) + \hat{c}_{p,0,(0)} < 1,$$

then there exists a positive constant $c_{t,p}$, independent of L and d, such that $\kappa_{(0)}(p,t,0,L) < 1$ in (5.35) for all $L \geq 1$ and $d \geq 2$ satisfying $L + 1 \leq c_{t,p}(d-1)$.

Proof. Let us define

$$\mathrm{C}^\infty_{(0)}(\bar\Omega) := \bigotimes_{i=1}^{d} \mathrm{C}^\infty_{(0)}(0,1) = \mathrm{C}^\infty_{(0)}(0,1) \otimes \cdots \otimes \mathrm{C}^\infty_{(0)}(0,1),$$

where the ith component $\mathrm{C}^\infty_{(0)}(0,1)$ in the d-fold tensor-product is taken to be equal to $\mathrm{C}^\infty_0(0,1)$ if Ox_i is an elliptic co-ordinate direction; otherwise (*i.e.*, when Ox_i is a hyperbolic co-ordinate direction), it is chosen to be equal to $\mathrm{C}^\infty[0,1]$.

For $u \in \mathrm{C}^\infty_{(0)}(\bar\Omega) \subset \mathrm{H}^1_{(0)}(\Omega)$, the following identity holds in $\mathrm{H}^1(\Omega)$:

$$u = \sum_{\ell\in\mathbb{N}^d} \left(Q^{\ell_1,p}_{(0)} \otimes \cdots \otimes Q^{\ell_d,p}_{(0)}\right) u.$$

We estimate, for $s \in \{0,1\}$, the approximation error as a sum of details, *i.e.*,

$$\left|u - \hat P^{L,p}_{(0)} u\right|_{\mathrm{H}^s(\Omega)} \le \sum_{\ell\in\mathbb{N}^d,\ |\ell|_1>L} \left|\left(Q^{\ell_1,p}_{(0)} \otimes \cdots \otimes Q^{\ell_d,p}_{(0)}\right) u\right|_{\mathrm{H}^s(\Omega)} \tag{5.37}$$

provided that the right-hand side is finite. We discuss the two cases, $s = 0$ and $s = 1$, separately.

For $s = 1$ and any $\ell = (\ell_1, \ell_2, \ldots, \ell_d) \in \mathbb{N}^d$ with $\mathrm{supp}(\ell) = I$ (that is, $\ell_j \ne 0$ iff $j \in I$) and $|I| = k$, $I \subseteq \{1, \ldots, d\}$, we have to estimate the solution details

$$\left|\left(Q^{\ell_1,p}_{(0)} \otimes \cdots \otimes Q^{\ell_d,p}_{(0)}\right) u\right|^2_{\mathrm{H}^1(\Omega)} = \sum_{j=1}^{d} \left|\left(Q^{\ell_1,p}_{(0)} \otimes \cdots \otimes Q^{\ell_d,p}_{(0)}\right) u\right|^2_{\mathrm{H}^{1,0,\{j\}}(\Omega)} =: (\star)$$

for $\ell \in \mathbb{N}^d$.

Using Proposition 5.1 and the notation ∂ for the differentiation operator in dimension 1, we obtain the following chain of inequalities:

$$\begin{aligned}
(\star) \quad &\le \quad \sum_{j\in I} \prod_{\substack{j'\in I\\ j'\ne j}} |Q^{\ell_{j'},p}_{(0)}|^2_{(\partial^{t+1},\mathrm{Id}_{\mathrm{L}^2(0,1)})} \cdot |Q^{\ell_j,p}_{(0)}|^2_{(\partial^{t+1},\partial^1)} |Q^{0,p}_{(0)}|^{2(d-k)}_{(\mathrm{Id}_{\mathrm{H}^1_{(0)}(0,1)},\mathrm{Id}_{\mathrm{L}^2(0,1)})} |u|^2_{\mathrm{H}^{t+1,1,I}(\Omega)} \\
&\quad + \sum_{j\notin I} \prod_{j'\in I} |Q^{\ell_{j'},p}_{(0)}|^2_{(\partial^{t+1},\mathrm{Id}_{\mathrm{L}^2(0,1)})} \cdot |Q^{0,p}_{(0)}|^2_{(\partial^1,\partial^1)} |Q^{0,p}_{(0)}|^{2(d-k-1)}_{(\mathrm{Id}_{\mathrm{H}^1_{(0)}(0,1)},\mathrm{Id}_{\mathrm{L}^2(0,1)})} |u|^2_{\mathrm{H}^{t+1,1,I}(\Omega)} \\
&\le \quad \sum_{j\in I} \tilde c^{2(k-1)}_{p,0,t} \tilde c^2_{p,1,t} 4^{-(t+1)|\ell|_1+\ell_j} \hat c^{2(d-k)}_{p,0,(0)} |u|^2_{\mathrm{H}^{t+1,1,I}(\Omega)} \\
&\quad + \sum_{j\notin I} \tilde c^{2k}_{p,0,t} 4^{-(t+1)|\ell|_1} \hat c^2_{p,1,(0)} \hat c^{2(d-k-1)}_{p,0,(0)} |u|^2_{\mathrm{H}^{t+1,1,I}(\Omega)}. \\
&\le \quad \tilde c^{2(k-1)}_{p,0,t} 4^{-(t+1)|\ell|_1} \hat c^{2(d-k-1)}_{p,0,(0)} |u|^2_{\mathrm{H}^{t+1,1,I}(\Omega)} \left(\tilde c^2_{p,1,t} \hat c^2_{p,0,(0)} \sum_{j\in I} 4^{\ell_j} + (d-k) \tilde c^2_{p,0,t} \hat c^2_{p,1,(0)}\right) \\
&\le \quad d \bar c_{p,t} \tilde c^{2(k-1)}_{p,0,t} 4^{|\ell|_\infty-(t+1)|\ell|_1} \hat c^{2(d-k-1)}_{p,0,(0)} |u|^2_{\mathrm{H}^{t+1,1,I}(\Omega)},
\end{aligned} \tag{5.38}$$

where

$$\bar c_{p,t} := \max\left(\tilde c^2_{p,1,t} \hat c^2_{p,0,(0)},\ \tilde c^2_{p,0,t} \hat c^2_{p,1,(0)}\right), \tag{5.39}$$

with $\tilde c_{p,s,t}$ defined in (3.6), (3.10) and $\hat c_{p,s,(0)}$ defined in (3.14).

We note in passing that in the (important) special case when $\Gamma = \Gamma_0$, and thereby $\mathrm{H}^1_{(0)}(0,1) = \mathrm{H}^1_0(0,1)$ in each of the d co-ordinate directions, the factor $|Q^{0,p}_{(0)}|_{(\mathrm{Id}_{\mathrm{H}^1_{(0)}(0,1)},\mathrm{Id}_{\mathrm{L}^2(0,1)})}$ in the first two lines of (5.38) above can be replaced by $|Q^{0,p}_{(0)}|_{(\partial^1,\mathrm{Id}_{\mathrm{L}^2(0,1)})}$.

We thus have,

$$\sum_{\substack{\ell\in\mathbb{N}^d,\,|\ell|_1>L\\ \mathrm{supp}(\ell)=I}} \left|\left(Q^{\ell_1,p}_{(0)}\otimes\cdots\otimes Q^{\ell_d,p}_{(0)}\right)u\right|_{\mathrm{H}^1(\Omega)} \le \sqrt{d\bar{c}_{p,t}}\tilde{c}^{k-1}_{p,0,t}\hat{c}^{d-k-1}_{p,0,(0)} \sum_{\substack{\ell\in\mathbb{N}^d,\,|\ell|_1>L\\ \mathrm{supp}(\ell)=I}} 2^{|\ell|_\infty-(t+1)|\ell|_1}|u|_{\mathrm{H}^{t+1,1,I}(\Omega)}$$
$$\le \sqrt{d\bar{c}_{p,t}}\tilde{c}^{k-1}_{p,0,t}\hat{c}^{d-k-1}_{p,0,(0)} \sum_{\ell\in\mathbb{N}^k,\,|\ell|_1>L} 2^{|\ell|_\infty-(t+1)|\ell|_1}|u|_{\mathrm{H}^{t+1,1,I}(\Omega)}\,.$$

In passing from the second to the third line in the estimate above we have dropped all $d-k$ trivial entries from the indexing of ℓ.

We now use, with arbitrary $l > L$, the estimate (5.12), *i.e.*, $\sum_{\ell\in\mathbb{N}^k,\,|\ell|_1=l} 2^{|\ell|_\infty} \le k\cdot 2^{k-1+l}$, and obtain

$$\begin{aligned}\sum_{\substack{\ell\in\mathbb{N}^d,|\ell|_1>L\\ \mathrm{supp}(\ell)=I}} \left|\left(Q^{\ell_1,p}_{(0)}\otimes\cdots\otimes Q^{\ell_d,p}_{(0)}\right)u\right|_{\mathrm{H}^1(\Omega)} &\le k\sqrt{d\bar{c}_{p,t}}\tilde{c}^{k-1}_{p,0,t}\hat{c}^{d-k-1}_{p,0,(0)}2^{k-1}\left(\sum_{l>L}2^{-tl}\right)|u|_{\mathrm{H}^{t+1,1,I}(\Omega)}\\ &= k\sqrt{d\bar{c}_{p,t}}(1-2^{-t})^{-1}\tilde{c}^{k-1}_{p,0,t}\hat{c}^{d-k-1}_{p,0,(0)}2^{k-1}2^{-t(L+1)}|u|_{\mathrm{H}^{t+1,1,I}(\Omega)}\\ &= d^{\frac12}\underline{c}_{p,t}k(2\tilde{c}_{p,0,t})^k\hat{c}^{d-k}_{p,0,(0)}2^{-tL}|u|_{\mathrm{H}^{t+1,1,I}(\Omega)},\end{aligned} \tag{5.40}$$

where

$$\underline{c}_{p,t} := \frac12\sqrt{\bar{c}_{p,t}}((2^t-1)\tilde{c}_{p,0,t}\hat{c}_{p,0,(0)})^{-1}. \tag{5.41}$$

Now, summing (5.40) over $I\subseteq\{1,2,\dots,d\}$ we deduce that

$$\begin{aligned}&\sum_{k=1}^d\sum_{\substack{I\subseteq\{1,2,\dots,d\}\\|I|=k}}\sum_{\substack{\ell\in\mathbb{N}^d,|\ell|_1>L\\ \mathrm{supp}(\ell)=I}}\left|\left(Q^{\ell_1,p}_{(0)}\otimes\cdots\otimes Q^{\ell_d,p}_{(0)}\right)u\right|_{\mathrm{H}^1(\Omega)}\\ &\qquad\le d^{\frac12}\underline{c}_{p,t}2^{-tL}\sum_{k=1}^d\binom{d}{k}(2\tilde{c}_{p,0,t})^k\hat{c}^{d-k}_{p,0,(0)}\cdot k\max_{\substack{I\subseteq\{1,2,\dots,d\}\\|I|=k}}|u|_{\mathrm{H}^{t+1,1,I}(\Omega)}\\ &\qquad\le d^{\frac32}\underline{c}_{p,t}(\kappa_{(0)}(p,t,1,L))^d2^{-tL}\cdot\max_{1\le k\le d}\left(\max_{\substack{I\subseteq\{1,2,\dots,d\}\\|I|=k}}|u|_{\mathrm{H}^{t+1,1,I}(\Omega)}\right),\end{aligned} \tag{5.42}$$

where

$$\kappa_{(0)}(p,t,1,L) := 2\tilde{c}_{p,0,t}+\hat{c}_{p,0,(0)},\qquad p\ge1,\quad 1\le t\le p,\quad L\ge1.$$

This completes the proof in the case of $s=1$.

For $s=0$, we bound (5.37) as a sum of details as follows:

$$\begin{aligned}\|u-\hat{P}^{L,p}_{(0)}u\|_{\mathrm{L}^2(\Omega)} &\le \sum_{\ell\in\mathbb{N}^d,\ |\ell|_1>L}\left\|\left(Q^{\ell_1,p}_{(0)}\otimes\cdots\otimes Q^{\ell_d,p}_{(0)}\right)u\right\|_{\mathrm{L}^2(\Omega)}\\ &= \sum_{k=1}^d\sum_{\substack{I\subset\{1,2,\dots,d\}\\|I|=k}}\sum_{\substack{\ell\in\mathbb{N}^d,\ |\ell|_1>L\\ \mathrm{supp}(\ell)=I}}\left\|\left(Q^{\ell_1,p}_{(0)}\otimes\cdots\otimes Q^{\ell_d,p}_{(0)}\right)u\right\|_{\mathrm{L}^2(\Omega)}.\end{aligned}$$

Next, we estimate the size of the entry with multi-index $\ell \in \mathbb{N}^d$ in the above sum. To this end, we define

$$(*) := \left\| \left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p} \right) u \right\|_{\mathrm{L}^2(\Omega)}^2 .$$

Using $I = \operatorname{supp}(\ell)$ and that $|I| = k$, we get

$$\begin{aligned} (*) &\leq \left\{ \prod_{j \in I} |Q_{(0)}^{\ell_j,p}|^2_{(\partial^{t+1}, \mathrm{Id}_{\mathrm{L}^2(0,1)})} \right\} |Q_{(0)}^{0,p}|^{2(d-k)}_{(\mathrm{Id}_{\mathrm{H}^1_{(0)}(0,1)}, \mathrm{Id}_{\mathrm{L}^2(0,1)})} |u|^2_{\mathrm{H}^{t+1,0,I}(\Omega)} \\ &= \tilde{c}_{p,0,t}^{2k} \hat{c}_{p,0,(0)}^{2(d-k)} 2^{-2(t+1)|\ell|_1} |u|^2_{\mathrm{H}^{t+1,0,I}(\Omega)} . \end{aligned}$$

Summing this bound over all $I \subseteq \{1, 2, \ldots, d\}$ with $|I| = k$ implies

$$\|u - \hat{P}_0^{L,p} u\|_{\mathrm{L}^2(\Omega)} \leq \sum_{k=1}^{d} \binom{d}{k} \tilde{c}_{p,0,t}^{k} \hat{c}_{p,0,(0)}^{d-k} \left\{ \sum_{\substack{\ell \in \mathbb{N}^k \\ |\ell|_1 > L}} 2^{-(t+1)|\ell|_1} \right\} \max_{1 \leq k \leq d} \left(\max_{\substack{I \subset \{1,2,\ldots,d\} \\ |I| = k}} |u|_{\mathrm{H}^{t+1,0,I}(\Omega)} \right) .$$

From Lemmas 5.4 and 5.5 with $x := 2^{t+1} \geq 2$ for $t \geq 0$, $\alpha := \tilde{c}_{p,0,t}$, and $\beta := \hat{c}_{p,0,(0)}$ we obtain

$$\|u - \hat{P}_{(0)}^{L,p} u\|_{\mathrm{L}^2(\Omega)} \leq 2d\,\mathrm{e}\,\tilde{c}_{p,0,t} \cdot \kappa_{(0)}(p,t,0,L)^{d-1} \cdot 2^{-(t+1)(L+1)} |u|_{\mathcal{H}^{t+1}(\Omega)} \tag{5.43}$$

where

$$\kappa_{(0)}(p,t,0,L) := \tilde{c}_{p,0,t}(L+1)\mathrm{e}^{1/(L+1)} + \hat{c}_{p,0,(0)}, \qquad p \geq 1, \quad 1 \leq t \leq p, \quad L \geq 1.$$

Hence the required bound for $s = 0$, with $\underline{c}_{p,t} = 2^{-t}\mathrm{e}\,\tilde{c}_{p,0,t}$.

Moreover, by Lemma 5.5, (5.23)–(5.25), if

$$\gamma_{(0)}(t,p) := \tilde{c}_{p,0,t} 2^{t+1}/(2^{t+1} - 1) + \hat{c}_{p,0,(0)} < 1, \tag{5.44}$$

then there exists a constant $c_{t,p} > 0$, independent of L and d, such that $\kappa_{(0)}(p,t,0,L) < 1$ for all $L \geq 1$ and $d \geq 2$ satisfying $L + 1 \leq c_{t,p}(d-1)$. □

Setting $s = 0$ and $t = p$ in (5.35), corresponding to the error bound in the $\mathrm{L}^2(\Omega)$ norm, we obtain the optimal convergence rate $\mathcal{O}(h_L^{p+1})$ up to the polylogarithmic term $L^{d-1} = |\log_2 h_L|^{d-1}$, in the asymptotic limit of $h_L \to 0$. It is by now well accepted by sparse-grid practitioners that in low dimensions such polylogarithmic terms are indeed present and they dominate the convergence behaviour. However, as we shall now show, in high dimensions, where necessarily $L < d$, the situation is much more favourable in this respect. This somewhat surprising phenomenon is discussed in Remarks 5.3–5.5 below. We shall develop conditions under which, in the practically relevant *preasymptotic regime*, the positive constants $\kappa_{(0)}(p,t,s,L)$, $s \in \{0,1\}$, on the right-hand side of (5.35) are strictly less than 1. In such instances, the error constant in (5.35) will exhibit exponential decay as a function of the dimension d, and the case of $p \geq 2$ will be shown to be more favourable in terms of the rate of decay than $p = 1$.

Remark 5.3. Note that the factor $\kappa_{(0)}(p,t,s,L)^{d-1+s}$ appearing in the bound (5.35), with $\kappa_{(0)}(p,t,s,L)$ defined in (5.36) for $s = 0$ and $s = 1$, decreases exponentially as $d \to \infty$, if

$$\hat{c}_{p,0,(0)} < 1$$

and

$$\tilde{c}_{p,0,t} < \begin{cases} (1 - \hat{c}_{p,0,(0)})/((L+1)\exp(1/(L+1))) & \text{when } s = 0, \\ (1 - \hat{c}_{p,0,(0)})/2 & \text{when } s = 1. \end{cases} \tag{5.45}$$

For the projector considered in Example 3.1, we showed in Section 3.1 that

$$\tilde{c}_{p,0,t} = \left(1 + \frac{1}{2^{t+1-s}}\right) \frac{1}{p} \sqrt{\frac{(p-t)!}{(p+t)!}}$$

for all $p \geq 1$, $t \in \mathbb{N}$, $1 \leq t \leq p$, $s \in \{0,1\}$, and we also showed in Section 3.2 that in the case of $p = 1$ the following refined bound holds:

$$\tilde{c}_{1,0,1} \leq \frac{1}{3}.$$

In the case of homogeneous Dirichlet boundary condition on the whole of Γ (*viz.* $\Gamma = \Gamma_0$ by virtue of $a = (a_{ij})_{i,j=1}^d$ being positive definite), when $\mathrm{H}^1_{(0)}(\Omega) = \mathrm{H}^1_0(\Omega)$ and writing κ_0 instead of $\kappa_{(0)}$, by Example 3.1 we have that $\hat{c}_{p,0,(0)} = \hat{c}_{p,0,0} \leq 1/\pi$ (< 1) for $p \geq 2$, while for $p = 1$, $\hat{c}_{p,0,(0)} = \hat{c}_{1,0,0} = 0$ (since $\mathcal{V}^{0,1}_{(0)} = \mathcal{V}^{0,1}_0 = \{0\}$). By scanning the range of validity of (5.45), we then deduce that, for $s = 1$,

$$\kappa_0(p,p,1,L) < 1 \qquad \forall p \geq 1, \quad L \geq 1,$$

thus ensuring exponential decay of the term $\kappa_0(p,p,1,L)^d$ in (5.35) as $d \to \infty$, for all $p \geq 1$, $s = 1$ and $L \geq 1$.

When $s = 0$, we still have $\kappa_0(p,p,0,L) < 1$, provided that $p = 2$ and $L \leq 3$ (corresponding to $h_L \geq 2^{-3}$), or $p = 3$ and $L \leq 49$ (corresponding to $h_L \geq 2^{-49}$), or $p = 4$ and $L \leq 528$ (corresponding to $h_L \geq 2^{-528}$), or $p = 5$ and $L \leq 6390$ (corresponding to $h_L \geq 2^{-6390}$), and so on. For the sake of comparison, recall that machine epsilon in IEEE double precision is 2^{-52}. Thus, once $p \geq 3$, the potentially harmful polylogarithmic factor $L^{d-1} = |\log_2 h_L|^{d-1}$, hidden in $(\kappa_0(p,p,0,L))^{d-1}$, has no detrimental effect on the approximation error from the sparse tensor product space on any mesh that might conceivably arise in practice.

The final part of the proof of the above theorem also indicates that $\kappa_0(p,p,0,L) < 1$ provided $\gamma_0(p,p) < 1$ and there exists a positive constant $c_{p,p}$ such that $L + 1 \leq c_{p,p}(d-1)$; if so, then, again, the polylogarithmic factor $L^{d-1} = |\log_2 h_L|^{d-1}$, hidden in $(\kappa_0(p,p,0,L))^{d-1}$, has no effect on the approximation error. As it happens, the first of these inequalities holds for all $p \geq 1$. Concerning the second inequality, simple computations reveal the existence of $c_{p,p}$ for all $p \geq 1$; for example, $c_{1,1} = 0.6$ corresponding to $p = 1$ (and $m_0 = 0.6$ in the proof of Lem. 5.5), $c_{2,2} = 0.71$ corresponding to $p = 2$ (and $m_0 = 1.69$), $c_{3,3} = 1.846$ corresponding to $p = 3$ (and $m_0 = 18.5$), $c_{4,4} = 2.161$ corresponding to $p = 4$ (and $m_0 = 194.8$), $c_{5,5} = 2.169$ corresponding to $p = 5$ (and $m_0 = 2351$), and so on.

Remark 5.4. A result analogous to that contained in (5.35), in the special case of $s = 1$ and $p = 1$, and with $\kappa_0(1,1,1,L) < 1$ (in our notation) was stated in Theorem 2 in [11]. There, however, an "energy-norm-based" sparse-grid-space was used that is strictly included in $\hat{V}^L_0$. The result contained in [11] is restricted to the case of $s = 1$ and $p = 1$ and does not cover either $s = 0$ or $p \geq 2$.

Remark 5.5. If $\Gamma_0 \subsetneq \Gamma$ (*i.e.*, the hyperbolic part $\Gamma_- \cup \Gamma_+$ of the boundary Γ is nonempty), and therefore $\mathrm{H}^1_0(\Omega) \subsetneq \mathrm{H}^1_{(0)}(\Omega)$, then we still have $\hat{c}_{p,0,(0)} \leq 1$ by (3.11).

Concerning the case of $s = 1$, if

$$\left(1 + \frac{1}{2^p}\right) \frac{1}{p\sqrt{(2p)!}} \leq \frac{c_*}{d}, \tag{5.46}$$

which is a very mild condition on the minimum size of p in terms of d, then we have that

$$(\kappa_{(0)}(p,p,1,L))^d \leq \left(1 + \frac{2c_*}{d}\right)^d \leq \mathrm{e}^{2c_*},$$

which, in turn, ensures that $(\kappa_{(0)}(p,p,1,L))^d$ remains uniformly bounded for $d \gg 1$. For example, taking $c_* = 1$, for $d \leq 7$ the condition (5.46) requires $p = 2$, for $8 \leq d \leq 71$ taking $p = 3$ will suffice, while for for $71 \leq d \leq 755$ taking $p = 4$ will be sufficient.

The discussion in the previous paragraph presupposed that the number of hyperbolic co-ordinate directions is equal to, or is very close to, d. If, however, the number of such directions is small relative to d, and can be regarded as being bounded as $d \to \infty$, then we expect that the factor $(\kappa_{(0)}(p,p,1,L))^d$ will exhibit exponential decay as $d \to \infty$ without the extra hypothesis (5.46), just as in the case when $\Gamma = \Gamma_0$. The proof of this would require a selective treatment of the constant $\hat{c}_{p,0,(0)}$ in the proof of Theorem 5.1 when $s = 1$, to monitor whether a particular factor of $\hat{c}_{p,0,(0)}$ in lines 3 and 4 of (5.38) arises from a univariate bound on $Q^{0,p}_{(0)}$ in an elliptic or in a hyperbolic co-ordinate direction. An altogether different approach to removing the condition (5.46) in the case of $\Gamma_0 \subsetneq \Gamma$ and $s = 1$ would be to show that $\hat{c}_{p,0,(0)} < 1$, uniformly in p. These lines of investigation are, however, beyond the scope of the present paper.

Similar comments apply in the case of $s = 0$, assuming, instead, the existence of a constant c_* such that

$$(L+1)\left(1+\frac{1}{2^{p+1}}\right)\frac{1}{p\sqrt{(2p)!}} \leq \frac{c_*}{d}.$$

If no assumption relating L, p and d is made, then there still exists $\kappa_* \in (0,1)$ such that

$$\kappa_{(0)}(p,p,0,L)^{d-1} < (L+1)^{d-1}\left(\frac{1}{L+1}+\frac{2}{p\sqrt{(2p)!}}\right)^{d-1} \leq (L+1)^{d-1}\kappa_*^{d-1}$$

for all $L \geq 1$, $p \geq 2$ and $d \geq 2$; as a matter of fact, the larger L and p are the smaller the value of κ_*. Thus, the growth of $(L+1)^d$ is compensated by the exponential decay of κ_*^{d-1}.

Remark 5.6. When $\Gamma_0 = \Gamma$, the function u to be approximated enters into the right-hand side of the estimate (5.35) in a nonstandard, yet favourable manner: through the L^2 norm of exactly one mixed derivative – rather than through a *sum* of L^2 norms of mixed derivatives as would have been the case had we used a more conventional seminorm on the space of functions with square-integrable mixed highest derivatives.

6. Convergence of the sparse stabilized method

Our goal in this section is to estimate the size of the error between the analytical solution $u \in \mathcal{H}$ and its approximation $u_h \in \hat{V}^{L,p}_{(0)}$. We shall assume throughout that $f \in \mathrm{L}^2(\Omega)$ and the corresponding solution $u \in \mathcal{H}^{k+1}(\Omega) \cap \mathrm{H}^2(\Omega) \cap \bigotimes_{i=1}^d \mathrm{H}^1_{(0)}(0,1) \subset \mathcal{H}$, $k \geq 1$ and $1 \leq t \leq \min\{p,k\}$. Clearly,

$$b_\delta(u-u_h,v_h) = B(u,v_h) - L(v_h) + \delta_L \sum_{\kappa \in \mathcal{T}^L} (\mathcal{L}u - f, b\cdot\nabla v_h)_\kappa$$

for all $v_h \in \hat{V}^{L,p}_{(0)} \subset \mathcal{V}$. Hence we deduce from (2.6) the following *Galerkin orthogonality* property:

$$b_\delta(u-u_h,v_h) = 0 \qquad \forall v_h \in \hat{V}^{L,p}_{(0)}. \tag{6.1}$$

Let us decompose the error $u - u_h$ as follows:

$$u - u_h = (u - \hat{P}^{L,p}_{(0)}u) + (\hat{P}^{L,p}_{(0)}u - u_h) = \eta + \xi,$$

where $\eta := u - \hat{P}^{L,p}_{(0)}u$ and $\xi := \hat{P}^{L,p}_{(0)}u - u_h$. By the triangle inequality,

$$|||u-u_h|||_{\mathrm{SD}} \leq |||\eta|||_{\mathrm{SD}} + |||\xi|||_{\mathrm{SD}}. \tag{6.2}$$

We begin by bounding $|||\xi|||_{\mathrm{SD}}$. By (4.6) and (6.1), we have that

$$|||\xi|||_{\mathrm{SD}}^2 \le b_\delta(\xi,\xi) = b_\delta(u-u_h,\xi) - b_\delta(\eta,\xi) = -b_\delta(\eta,\xi).$$

Therefore,

$$|||\xi|||_{\mathrm{SD}}^2 \le |b_\delta(\eta,\xi)|. \tag{6.3}$$

Now,

$$\begin{aligned} b_\delta(\eta,\xi) &= (a\nabla\eta,\nabla\xi) - (\eta, b\cdot\nabla\xi) + (c\eta,\xi) + \int_{\Gamma_+} |\beta|\eta\,\xi\,\mathrm{d}s \\ &\quad + \delta_L \sum_{\kappa\in\mathcal{T}^L} (-a:\nabla\nabla\eta + b\cdot\nabla\eta + c\eta, b\cdot\nabla\xi)_\kappa \\ &= \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV} + (\mathrm{V} + \mathrm{VI} + \mathrm{VII})\,. \end{aligned}$$

For the terms I to III and V to VII we have:

$$\begin{aligned} \mathrm{I} &\le \left(|\sqrt{a}\,|\,\|\nabla\eta\|_{\mathrm{L}^2(\Omega)}\right)|||\xi|||_{\mathrm{SD}}, \\ \mathrm{II} &\le \left(\delta_L^{-\frac{1}{2}}\|\eta\|_{\mathrm{L}^2(\Omega)}\right)|||\xi|||_{\mathrm{SD}}, \\ \mathrm{III} &\le \left(c^{\frac{1}{2}}\|\eta\|_{\mathrm{L}^2(\Omega)}\right)|||\xi|||_{\mathrm{SD}}, \\ \mathrm{V} &\le \left(\delta_L^{\frac{1}{2}}|a|\left(\sum_{\kappa\in\mathcal{T}^L}|\eta|_{\mathrm{H}^2(\kappa)}^2\right)^{\frac{1}{2}}\right)|||\xi|||_{\mathrm{SD}}, \\ \mathrm{VI} &\le \left(\delta_L^{\frac{1}{2}}|b|\,\|\nabla\eta\|_{\mathrm{L}^2(\Omega)}\right)|||\xi|||_{\mathrm{SD}}, \\ \mathrm{VII} &\le \left(c\delta_L^{\frac{1}{2}}\|\eta\|_{\mathrm{L}^2(\Omega)}\right)|||\xi|||_{\mathrm{SD}}. \end{aligned}$$

Here $|a|$ is the Frobenius norm of the matrix a and $|b|$ is the Euclidean norm of the vector b. It remains to estimate IV:

$$\begin{aligned} \mathrm{IV} &\le \left(\frac{2|b|}{1+c\delta_L}\right)^{\frac{1}{2}}\left(\int_{\Gamma_+}|\eta|^2\,\mathrm{d}s\right)^{\frac{1}{2}}|||\xi|||_{\mathrm{SD}} \\ &\le (2|b|)^{\frac{1}{2}}(4d)^{\frac{1}{2}}\|\eta\|_{\mathrm{L}^2(\Omega)}^{\frac{1}{2}}\|\eta\|_{\mathrm{H}^1(\Omega)}^{\frac{1}{2}}|||\xi|||_{\mathrm{SD}}, \end{aligned}$$

where in the transition to the last line we used the multiplicative trace inequality from Lemma 4.2. Hence, by (6.3),

$$\begin{aligned} |||\xi|||_{\mathrm{SD}} &\le |\sqrt{a}\,|\,\|\nabla\eta\|_{\mathrm{L}^2(\Omega)} + \delta_L^{-\frac{1}{2}}\|\eta\|_{\mathrm{L}^2(\Omega)} + c^{\frac{1}{2}}\|\eta\|_{\mathrm{L}^2(\Omega)} + (8d|b|)^{\frac{1}{2}}\|\eta\|_{\mathrm{L}^2(\Omega)}^{\frac{1}{2}}\|\eta\|_{\mathrm{H}^1(\Omega)}^{\frac{1}{2}} \\ &\quad + \delta_L^{\frac{1}{2}}|\sqrt{a}\,|^2\left(\sum_{\kappa\in\mathcal{T}^L}|\eta|_{\mathrm{H}^2(\kappa)}^2\right)^{\frac{1}{2}} + \delta_L^{\frac{1}{2}}|b|\,\|\nabla\eta\|_{\mathrm{L}^2(\Omega)} + c\delta_L^{\frac{1}{2}}\|\eta\|_{\mathrm{L}^2(\Omega)}. \end{aligned} \tag{6.4}$$

The bounds on $\|\eta\|_{\mathrm{L}^2(\Omega)}$ and $\|\nabla\eta\|_{\mathrm{L}^2(\Omega)}$ will follow from Theorem 5.1. However, the fifth term in the sum on the right-hand side of (6.4) is nonstandard and needs to be dealt with separately (except when $p=1$ and a is diagonal; then term V is bounded by $\delta_L^{\frac{1}{2}}|a||u|_{\mathrm{H}^2(\Omega)}$ and requires no further estimation). Let us note that

$$\begin{aligned}
\sum_{\kappa\in\mathcal{T}^L}|\eta|^2_{\mathrm{H}^2(\kappa)} &= \sum_{i,j=1}^{d}\sum_{\kappa\in\mathcal{T}^L}\int_\kappa\left|\frac{\partial^2\eta}{\partial x_i\partial x_j}\right|^2\mathrm{d}x\\
&= \sum_{i=1}^{d}\sum_{\kappa\in\mathcal{T}^L}\int_\kappa\left|\frac{\partial^2\eta}{\partial x_i^2}\right|^2\mathrm{d}x+\sum_{\substack{i,j=1\\ i\neq j}}^{d}\sum_{\kappa\in\mathcal{T}^L}\int_\kappa\left|\frac{\partial^2\eta}{\partial x_i\partial x_j}\right|^2\mathrm{d}x\\
&= \sum_{i=1}^{d}\sum_{\kappa\in\mathcal{T}^L}|\eta|^2_{\mathrm{H}^{2,0,\{i\}}(\kappa)}+\sum_{\substack{i,j=1\\ i\neq j}}^{d}|\eta|^2_{\mathrm{H}^{1,0,\{i,j\}}(\Omega)}\\
&=: \mathrm{A}^2+\mathrm{B}^2.
\end{aligned}$$

Here, we made use of the fact that

$$\frac{\partial^2\eta}{\partial x_i\partial x_j}\in\mathrm{L}^2(\Omega)\qquad\forall i,j\in\{1,2,\dots,d\},\qquad i\neq j.$$

Let us first estimate

$$\mathrm{A}^2=\sum_{i=1}^{d}\sum_{\kappa\in\mathcal{T}^L}|\eta|^2_{\mathrm{H}^{2,0,\{i\}}(\kappa)}=\sum_{i=1}^{d}\sum_{\kappa\in\mathcal{T}^L}\int_\kappa\left|\frac{\partial^2\eta}{\partial x_i^2}\right|^2\mathrm{d}x=\sum_{i=1}^{d}\sum_{j=1}^{2^L}|\eta|^2_{\mathrm{H}^{2,0,\{i\}}(K_j^i)},$$

where K_j^i denotes the d-dimensional slab

$$K_j^i=(0,1)\times\cdots\times(0,1)\times(\xi_{j-1},\xi_j)\times(0,1)\times\cdots\times(0,1)\tag{6.5}$$

with the interval $(\xi_{j-1},\xi_j)=(x_{j-1}^L,x_j^L)$ entering at position i. The reason for agglomerating the elements $\kappa\in\mathcal{T}^L$ into the slabs K_j^i, $j=1,\dots,2^L$, in this way is that the function $\partial^2\eta/\partial x_i^2$ involves no derivatives in the co-ordinate directions Ox_k for $k\neq i$. In other words, it only needs to be considered piecewise in the ith co-ordinate direction; in the other $d-1$ co-ordinate directions it is defined on the whole of $(0,1)^{d-1}$ as an H^1 function.

Let us define the seminorms $|||\cdot|||_{2,i}$, $i=1,\dots,d$, and $|||\cdot|||_{2,*}$, by

$$|||v|||^2_{2,i}:=\sum_{j=1}^{2^L}|v|^2_{\mathrm{H}^{2,0,\{i\}}(K_j^i)}\qquad\text{and}\qquad|||v|||^2_{2,*}:=\sum_{i=1}^{d}|||v|||^2_{2,i}.$$

With this notation, we have that

$$\mathrm{A}^2=|||\eta|||^2_{2,*}=\sum_{i=1}^{d}|||\eta|||^2_{2,i}.$$

When $p=1$, term A is dealt with easily on recalling that $\eta=u-\hat{P}^{L,1}_{(0)}u$:

$$\mathrm{A}=\left(\sum_{i=1}^{d}\left\|\frac{\partial^2u}{\partial x_i^2}\right\|^2_{\mathrm{L}^2(\Omega)}\right)^{\frac{1}{2}}\le d^{\frac{1}{2}}\max_{1\le k\le d}\left\|\frac{\partial^2u}{\partial x_k^2}\right\|_{\mathrm{L}^2(\Omega)}\le d^{\frac{1}{2}}|u|_{\mathcal{H}^2(\Omega)}.$$

In order to bound A in the case of $p \geq 2$, we first observe that, as a consequence of Lemma 4.1,

$$|v|_{\mathrm{H}^2(J)} \leq \sqrt{12}\,(p^2/h_L)|v|_{\mathrm{H}^1(J)} \qquad \forall v \in \mathcal{P}^p(J), \tag{6.6}$$

where $J \in \{(x_{j-1}^{\ell_i}, x_j^{\ell_i}) : i = 1, \ldots, d, \quad j = 1, \ldots, 2^L\}$. Hence, on recalling that

$$\eta = u - \hat{P}_{(0)}^{L,p} u = \sum_{\ell \in \mathbb{N}^d\,:\,|\ell|_1 > L} \left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u, \tag{6.7}$$

for a fixed $i \in \{1, 2, \ldots, d\}$ we deduce from (6.6) with $h_L = 2^{-L}$ that

$$\left|\left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u\right|_{\mathrm{H}^{2,0,\{i\}}(\kappa)} \leq \sqrt{12}\,p^2\,2^L \left|\left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u\right|_{\mathrm{H}^{1,0,\{i\}}(\kappa)}.$$

Now, we square the last bound, and sum over all elements $\kappa \in \mathcal{T}^L$ that are contained in the d-dimensional slab K_j^i defined in (6.5) to deduce that

$$\left|\left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u\right|^2_{\mathrm{H}^{2,0,\{i\}}(K_j^i)} \leq 12\,p^4\,2^{2L} \left|\left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u\right|^2_{\mathrm{H}^{1,0,\{i\}}(K_j^i)}.$$

Hence,

$$\begin{aligned}\sum_{j=1}^{2^L} \left|\left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u\right|^2_{\mathrm{H}^{2,0,\{i\}}(K_j^i)} &\leq 12\,p^4\,2^{2L} \sum_{j=1}^{2^L} \left|\left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u\right|^2_{\mathrm{H}^{1,0,\{i\}}(K_j^i)} \\ &= 12\,p^4\,2^{2L} \left|\left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u\right|^2_{\mathrm{H}^{1,0,\{i\}}(\Omega)}.\end{aligned}$$

This implies that

$$\left|\!\left|\!\left|\left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u\right|\!\right|\!\right|^2_{2,i} \leq 12\,p^4\,2^{2L} \left|\left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u\right|^2_{\mathrm{H}^{1,0,\{i\}}(\Omega)},$$

and, on summing over $i = 1, \ldots, d$, and taking square-roots of both sides, we get

$$\begin{aligned}\left|\!\left|\!\left|\left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u\right|\!\right|\!\right|_{2,*} &\leq \sqrt{12}\,p^2\,2^L \left(\sum_{i=1}^{d} \left|\left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u\right|^2_{\mathrm{H}^{1,0,\{i\}}(\Omega)}\right)^{\frac{1}{2}} \\ &= \sqrt{12}\,p^2\,2^L \left|\left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u\right|_{\mathrm{H}^1(\Omega)}.\end{aligned}$$

Hence, by (6.7) and the proof of (5.35) in the case of $s = 1$ (*cf.* (5.37)–(5.42)),

$$\begin{aligned}\mathrm{A} = |||\eta|||_{2,*} &\leq \sum_{\ell \in \mathbb{N}^d,\ |\ell|_1 > L} \left|\!\left|\!\left|\left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u\right|\!\right|\!\right|_{2,*} \\ &\leq \sqrt{12}\,p^2\,2^L \sum_{\ell \in \mathbb{N}^d,\ |\ell|_1 > L} \left|\left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u\right|_{\mathrm{H}^1(\Omega)} \\ &\leq \sqrt{12}\,p^2\,2^L d^{\frac{3}{2}} \underline{c}_{p,t}\,(\kappa_{(0)}(p,t,1,L))^d\,2^{-tL} \cdot |u|_{\mathcal{H}^{t+1}(\Omega)}.\end{aligned}$$

Thus we have shown that

$$\mathrm{A} \leq \sqrt{12}\,p^2\,d^{\frac{3}{2}} \underline{c}_{p,t}\,(\kappa_{(0)}(p,t,1,L))^d\,2^{-(t-1)L} \cdot |u|_{\mathcal{H}^{t+1}(\Omega)} \tag{6.8}$$

for $1 \le t \le \min\{p,k\}$, $p \ge 2$; and

$$A \le d^{\frac{1}{2}} \cdot |u|_{\mathcal{H}^2(\Omega)}$$

when $p = t = 1$.

Now, let us bound term $\mathrm{B}^2 = \sum_{\substack{i,j=1\\ i\neq j}}^{d} |\eta|^2_{\mathrm{H}^{1,0,\{i,j\}}(\Omega)}$. We define the seminorm $|||\cdot|||_{2,**}$ by

$$|||v|||^2_{2,**} := \sum_{\substack{i,j=1\\ i\neq j}}^{d} |v|^2_{\mathrm{H}^{1,0,\{i,j\}}(\Omega)};$$

then, $\mathrm{B}^2 = |||\eta|||^2_{2,**}$. Now, since

$$\eta = u - \hat{P}_0^{L,p} u = \sum_{\ell\in\mathbb{N}^d,\ |\ell|_1 > L} \left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u,$$

it follows that

$$|||\eta|||_{2,**} \le \sum_{\ell\in\mathbb{N}^d,\ |\ell|_1 > L} \left\| \left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u \right\|_{2,**}.$$

Given $\ell = (\ell_1, \ell_2, \ldots, \ell_d) \in \mathbb{N}^d$ with $\operatorname{supp}(\ell) = I$ (that is, $\ell_j \neq 0$ iff $j \in I$) and $|I| = k$, we have to estimate

$$\left\| \left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u \right\|^2_{2,**} = \sum_{\substack{i,j=1\\ i\neq j}}^{d} \left| \left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p}\right) u \right|^2_{\mathrm{H}^{1,0,\{i,j\}}(\Omega)} =: (\star\star).$$

Using Proposition 5.1 and the notation ∂ for the univariate differentiation operator, we obtain the following inequality:

$$\begin{aligned}
(\star\star) \quad \le \quad & \sum_{\substack{i,j\in I\\ i\neq j}} \prod_{\substack{j'\in I\\ j'\notin\{i,j\}}} |Q_{(0)}^{\ell_{j'},p}|^2_{(\partial^{t+1},\mathrm{Id}_{\mathrm{L}^2(0,1)})} \\
& \cdot |Q_{(0)}^{\ell_i,p}|^2_{(\partial^{t+1},\partial^1)} |Q_{(0)}^{\ell_j,p}|^2_{(\partial^{t+1},\partial^1)} |Q_{(0)}^{0,p}|^{2(d-k)}_{(\mathrm{Id}_{\mathrm{H}^1_{(0)}(0,1)},\mathrm{Id}_{\mathrm{L}^2(0,1)})} |u|^2_{\mathrm{H}^{t+1,1,I}(\Omega)} \\
& + \sum_{i\in I} \sum_{j\notin I} \prod_{\substack{j'\in I\\ j'\neq i}} |Q_{(0)}^{\ell_{j'},p}|^2_{(\partial^{t+1},\mathrm{Id}_{\mathrm{L}^2(0,1)})} \\
& \cdot |Q_{(0)}^{\ell_i,p}|^2_{(\partial^{t+1},\partial^1)} |Q_{(0)}^{0,p}|^2_{(\partial^1,\partial^1)} |Q_{(0)}^{0,p}|^{2(d-k-1)}_{(\mathrm{Id}_{\mathrm{H}^1_{(0)}(0,1)},\mathrm{Id}_{\mathrm{L}^2(0,1)})} |u|^2_{\mathrm{H}^{t+1,1,I}(\Omega)} \\
& + \sum_{i\notin I} \sum_{j\in I} \prod_{\substack{j'\in I\\ j'\neq j}} |Q_{(0)}^{\ell_{j'},p}|^2_{(\partial^{t+1},\mathrm{Id}_{\mathrm{L}^2(0,1)})} \\
& \cdot |Q_{(0)}^{0,p}|^2_{(\partial^1,\partial^1)} |Q_{(0)}^{\ell_j,p}|^2_{(\partial^{t+1},\partial^1)} |Q_{(0)}^{0,p}|^{2(d-k-1)}_{(\mathrm{Id}_{\mathrm{H}^1_{(0)}(0,1)},\mathrm{Id}_{\mathrm{L}^2(0,1)})} |u|^2_{\mathrm{H}^{t+1,1,I}(\Omega)} \\
& + \sum_{\substack{i,j\notin I\\ i\neq j}} \prod_{j'\in I} |Q_{(0)}^{\ell_{j'},p}|^2_{(\partial^{t+1},\mathrm{Id}_{\mathrm{L}^2(0,1)})} \\
& \cdot |Q_{(0)}^{0,p}|^2_{(\partial^1,\partial^1)} |Q_{(0)}^{0,p}|^2_{(\partial^1,\partial^1)} |Q_{(0)}^{0,p}|^{2(d-k-2)}_{(\mathrm{Id}_{\mathrm{H}^1_{(0)}(0,1)},\mathrm{Id}_{\mathrm{L}^2(0,1)})} |u|^2_{\mathrm{H}^{t+1,1,I}(\Omega)}. \qquad (6.9)
\end{aligned}$$

Hence,

$$
\begin{aligned}
(\star\star) \quad \leq \quad & \sum_{\substack{i,j\in I\\ i\neq j}} \tilde{c}_{p,0,t}^{2(k-2)} 4^{\ell_i+\ell_j-(t+1)|\ell|_1} \tilde{c}_{p,1,t}^2 \tilde{c}_{p,1,t}^2 \hat{c}_{p,0,(0)}^{2(d-k)} |u|^2_{\mathrm{H}^{t+1,1,I}(\Omega)} \\
& + \sum_{i\in I} \sum_{j\notin I} \tilde{c}_{p,0,t}^{2(k-1)} 4^{\ell_i-(t+1)|\ell|_1} \tilde{c}_{p,1,t}^2 \hat{c}_{p,1,(0)}^2 \hat{c}_{p,0,(0)}^{2(d-k-1)} |u|^2_{\mathrm{H}^{t+1,1,I}(\Omega)} \\
& + \sum_{i\notin I} \sum_{j\in I} \tilde{c}_{p,0,t}^{2(k-1)} 4^{\ell_j-(t+1)|\ell|_1} \hat{c}_{p,1,(0)}^2 \tilde{c}_{p,1,t}^2 \hat{c}_{p,0,(0)}^{2(d-k-1)} |u|^2_{\mathrm{H}^{t+1,1,I}(\Omega)} \\
& + \sum_{\substack{i,j\notin I\\ i\neq j}} \tilde{c}_{p,0,t}^{2k} 4^{-(t+1)|\ell|_1} \hat{c}_{p,1,(0)}^2 \hat{c}_{p,1,(0)}^2 \hat{c}_{p,0,(0)}^{2(d-k-2)} |u|^2_{\mathrm{H}^{t+1,1,I}(\Omega)}. \qquad (6.10)
\end{aligned}
$$

Thus we deduce that

$$
\begin{aligned}
(\star\star) \quad \leq \quad & \tilde{c}_{p,0,t}^{2(k-2)} \hat{c}_{p,0,(0)}^{2(d-k-2)} \cdot 4^{-(t+1)|\ell|_1} |u|^2_{\mathrm{H}^{t+1,1,I}(\Omega)} \\
& \times \left(\tilde{c}_{p,1,t}^4 \hat{c}_{p,0,(0)}^4 \sum_{\substack{i,j\in I\\ i\neq j}} 4^{\ell_i+\ell_j} + 2\tilde{c}_{p,1,t}^2 \hat{c}_{p,0,(0)}^2 \tilde{c}_{p,0,t}^2 \hat{c}_{p,1,(0)}^2 \sum_{i\in I, j\notin I} 4^{\ell_i} + \tilde{c}_{p,0,t}^4 \hat{c}_{p,1,(0)}^4 \sum_{\substack{i,j\notin I\\ i\neq j}} 1 \right) \\
\leq \quad & \tilde{c}_{p,0,t}^{2(k-2)} \hat{c}_{p,0,(0)}^{2(d-k-2)} \cdot 4^{-(t+1)|\ell|_1} |u|^2_{\mathrm{H}^{t+1,1,I}(\Omega)} \\
& \quad \times \Big(\tilde{c}_{p,1,t}^4 \hat{c}_{p,0,(0)}^4 k^2 4^{|\ell|_1} + 2\tilde{c}_{p,1,t}^2 \hat{c}_{p,0,(0)}^2 \tilde{c}_{p,0,t}^2 \hat{c}_{p,1,(0)}^2 k(d-k) 4^{|\ell|_1} \\
& \qquad + \tilde{c}_{p,0,t}^4 \hat{c}_{p,1,(0)}^4 [(d-k)^2 - (d-k)] \Big) \\
\leq \quad & d^2 \bar{\bar{c}}_{p,t} \tilde{c}_{p,0,t}^{2(k-2)} \hat{c}_{p,0,(0)}^{2(d-k-2)} \cdot 4^{-t|\ell|_1} |u|^2_{\mathrm{H}^{t+1,1,I}(\Omega)},
\end{aligned}
$$

where

$$
\bar{\bar{c}}_{p,t} := \max \left(\tilde{c}_{p,1,t}^4 \hat{c}_{p,0,(0)}^4, \tilde{c}_{p,1,t}^2 \hat{c}_{p,0,(0)}^2 \tilde{c}_{p,0,t}^2 \hat{c}_{p,1,(0)}^2, \tilde{c}_{p,0,t}^4 \hat{c}_{p,1,(0)}^4 \right).
$$

Therefore, we have that

$$
\begin{aligned}
& \sum_{\substack{\ell\in\mathbb{N}^d\,:\,|\ell|_1>L\\ \mathrm{supp}(\ell)=I}} \left\| \left(Q_{(0)}^{\ell_1,p} \otimes \cdots \otimes Q_{(0)}^{\ell_d,p} \right) u \right\|_{2,**} \\
& \quad \leq d\sqrt{\bar{\bar{c}}_{p,t}}\, \tilde{c}_{p,0,t}^{k-2}\, \hat{c}_{p,0,(0)}^{d-k-2} \cdot \sum_{\substack{\ell\in\mathbb{N}^d\,:\,|\ell|_1>L\\ \mathrm{supp}(\ell)=I}} 2^{-t|\ell|_1} |u|_{\mathrm{H}^{t+1,1,I}(\Omega)} \\
& \quad \leq d\sqrt{\bar{\bar{c}}_{p,t}}\, \tilde{c}_{p,0,t}^{k-2}\, \hat{c}_{p,0,(0)}^{d-k-2} \cdot \left(\sum_{\ell\in\mathbb{N}^k\,:\,|\ell|_1>L} 2^{-t|\ell|_1} \right) |u|_{\mathrm{H}^{t+1,1,I}(\Omega)}.
\end{aligned}
$$

Once again, we note in passing that in the (important) special case when $\Gamma = \Gamma_0$, and thereby $\mathrm{H}^1_{(0)}(0,1) = \mathrm{H}^1_0(0,1)$ in each of the d co-ordinate directions, the factor $|Q_{(0)}^{0,p}|_{(\mathrm{Id}_{\mathrm{H}^1_{(0)}(0,1)}, \mathrm{Id}_{\mathrm{L}^2(0,1)})}$ in the lines above can be replaced by $|Q_{(0)}^{0,p}|_{(\partial^1, \mathrm{Id}_{\mathrm{L}^2(0,1)})}$.

Hence, upon summation and using Lemma 5.4,

$$\begin{aligned}
&\sum_{k=1}^{d} \sum_{\substack{I\subset\{1,2,\dots,d\}\\ |I|=k}} \sum_{\substack{\ell\in\mathbb{N}^d\,:\,|\ell|_1>L\\ \operatorname{supp}(\ell)=I}} \left\| \left(Q_{(0)}^{\ell_1,p}\otimes\cdots\otimes Q_{(0)}^{\ell_d,p}\right) u \right\|_{2,**} \\
&\quad \le d\sqrt{\bar{\bar{c}}_{p,t}} \sum_{k=1}^{d} \binom{d}{k} \tilde{c}_{p,0,t}^{k-2}\, \hat{c}_{p,0,(0)}^{d-k-2} \left(\sum_{\ell\in\mathbb{N}^k\,:\,|\ell|_1>L} 2^{-t|\ell|_1} \right) \max_{1\le k\le d} \left(\max_{\substack{I\subset\{1,2,\dots,d\}\\ |I|=k}} |u|_{\mathrm{H}^{t+1,1,I}(\Omega)} \right) \\
&\quad \le \frac{d^2 \mathrm{e}\sqrt{\bar{\bar{c}}_{p,t}}}{\tilde{c}_{p,0,t}\,\hat{c}_{p,0,(0)}^2} \frac{2^t}{2^t-1} (\kappa_{(0)}(p,t,0,L))^{d-1} 2^{-t(L+1)} |u|_{\mathcal{H}^{t+1}(\Omega)}.
\end{aligned}$$

We deduce that

$$\mathrm{B} \le d^2 \underline{\underline{c}}_{p,t} (\kappa_{(0)}(p,t,0,L))^{d-1} 2^{-tL} |u|_{\mathcal{H}^{t+1}(\Omega)}, \tag{6.11}$$

where

$$\underline{\underline{c}}_{p,t} := \frac{\mathrm{e}\,\sqrt{\bar{\bar{c}}_{p,t}}}{\tilde{c}_{p,0,t}\,\hat{c}_{p,0,(0)}^2 (2^t-1)}.$$

Combining the bound (6.8) on A with the bound (6.11) on B yields

$$\begin{aligned}
\left(\sum_{\kappa\in\mathcal{T}^L} |\eta|_{\mathrm{H}^2(\kappa)}^2 \right)^{\frac12} \quad \le \quad & \left(\sqrt{12}\, p^2\, d^{\frac32} \underline{c}_{p,t} (\kappa_{(0)}(p,t,1,L))^d + d^2 \underline{\underline{c}}_{p,t} (\kappa_{(0)}(p,t,0,L))^{d-1} h_L \right) \\
& \times h_L^{t-1} |u|_{\mathcal{H}^{t+1}(\Omega)},
\end{aligned} \tag{6.12}$$

for $1 \le t \le \min\{p,k\}$, $p \ge 2$, $k \ge 2$. For $p = 1$, we have

$$\left(\sum_{\kappa\in\mathcal{T}^L} |\eta|_{\mathrm{H}^2(\kappa)}^2 \right)^{\frac12} \quad \le \quad \left(d^{\frac12} + d^2 \underline{\underline{c}}_{p,t} (\kappa_{(0)}(1,1,0,L))^{d-1} h_L \right) |u|_{\mathcal{H}^2(\Omega)}. \tag{6.13}$$

We also know from Theorem 5.1 that, for $1 \le t \le \min\{p,k\}$, $p \ge 1$, $k \ge 1$,

$$\|\eta\|_{\mathrm{L}^2(\Omega)} \le d \underline{c}_{p,t} (\kappa_{(0)}(p,t,0,L))^{d-1}\, h_L^{t+1}\, |u|_{\mathcal{H}^{t+1}(\Omega)}, \tag{6.14}$$

$$|\eta|_{\mathrm{H}^1(\Omega)} \le d^{\frac32} \underline{c}_{p,t} (\kappa_{(0)}(p,t,1,L))^{d}\, h_L^{t}\, |u|_{\mathcal{H}^{t+1}(\Omega)}. \tag{6.15}$$

Let us introduce, for ease of writing, the notation

$$\kappa_0 := \kappa_{(0)}(p,t,0,L) \qquad \text{and} \qquad \kappa_1 := \kappa_{(0)}(p,t,1,L), \tag{6.16}$$

and absorb all constants that depend on p and t only into a generic constant $C_{p,t}$. In particular, $C_{p,t}$ is independent of d and L and the coefficients a, b, c and the right-hand side f of the partial differential equation.

Remark 6.1. Since (6.12), (6.14), (6.15) and all of our earlier bounds are completely explicit in p and t (as well as in d and L), one could track the actual value of $C_{p,t}$ in our argument below. For clarity of presentation we shall however refrain from doing so, particularly since the emphasis here is on h-version rather than p- or hp-version finite element methods.

With these notational conventions (6.12)–(6.15) become, for $p \geq 1$ and $d \geq 2$:

$$\|\eta\|_{\mathrm{L}^2(\Omega)} \leq C_{p,t}\, d\, \kappa_0^{d-1} h_L^{t+1} |u|_{\mathcal{H}^{t+1}(\Omega)}, \tag{6.17}$$

$$|\eta|_{\mathrm{H}^1(\Omega)} \leq C_{p,t}\, d^{\frac{3}{2}} \kappa_1^d\, h_L^t |u|_{\mathcal{H}^{t+1}(\Omega)}, \tag{6.18}$$

$$\left(\sum_{\kappa\in\mathcal{T}^L} |\eta|^2_{\mathrm{H}^2(\kappa)}\right)^{\frac{1}{2}} \leq \begin{cases} C_{p,t}(d^{\frac{3}{2}}\kappa_1^d + d^2\kappa_0^{d-1}h_L) h_L^{t-1} |u|_{\mathcal{H}^{t+1}(\Omega)}, & p \geq 2, \\ C_{p,t}(d^{\frac{1}{2}} + d^2\kappa_0^{d-1}h_L)|u|_{\mathcal{H}^2(\Omega)}, & p = 1. \end{cases} \tag{6.19}$$

Using (6.17), (6.18) and (6.19) in the definition of $|||\cdot|||_{\mathrm{SD}}$ and selecting

$$\delta_L := K_\delta \min\left(\frac{h_L^2}{12dp^4|\sqrt{a}|^2}, \frac{h_L}{|b|}, \frac{1}{c}\right), \tag{6.20}$$

with $K_\delta \in \mathbb{R}_{>0}$ a constant, independent of h_L and d, we then deduce that

$$|||\eta|||_{\mathrm{SD}} \leq C_{p,t} d^2 \max\{\kappa_0^{d-1}, \kappa_1^d\} \left(|\sqrt{a}| h_L^t + |b|^{\frac{1}{2}} h_L^{t+\frac{1}{2}} + c^{\frac{1}{2}} h_L^{t+1}\right) |u|_{\mathcal{H}^{t+1}(\Omega)},$$

with $1 \leq t \leq \min\{p, k\}$, $p \geq 2$, $k \geq 2$. An identical bound holds for $p = 1$, $k \geq 1$, with $\max\{\kappa_0^{d-1}, \kappa_1^d\}$ replaced by $\max\{1, \kappa_0^{d-1}, \kappa_1^d\}$.

Similarly, using (6.17), (6.18) and (6.19) in (6.4) with δ_L as above,

$$|||\xi|||_{\mathrm{SD}} \leq C_{p,t} d^2 \max\{\kappa_0^{d-1}, \kappa_1^d\} \left(|\sqrt{a}| h_L^t + |b|^{\frac{1}{2}} h_L^{t+\frac{1}{2}} + c^{\frac{1}{2}} h_L^{t+1}\right) |u|_{\mathcal{H}^{t+1}(\Omega)},$$

with $1 \leq t \leq \min\{p, k\}$, $p \geq 2$, $k \geq 2$. An identical bound holds for $p = 1$, $k \geq 1$, with $\max\{\kappa_0^{d-1}, \kappa_1^d\}$ replaced by $\max\{1, \kappa_0^{d-1}, \kappa_1^d\}$.

Inserting the bounds on $|||\xi|||_{\mathrm{SD}}$ and $|||\eta|||_{\mathrm{SD}}$ in the right-hand side of the triangle inequality (6.2), we deduce the following theorem.

Theorem 6.1. *Suppose that* $f \in \mathrm{L}^2(\Omega)$ *in* $\Omega = (0,1)^d$, *that* $c > 0$ *and assume the regularity* $u \in \mathcal{H}^{k+1}(\Omega) \cap \mathrm{H}^2(\Omega) \cap \bigotimes_{i=1}^d \mathrm{H}^1_{(0)}(0,1)$, $k \in \mathbb{N}_{>0}$.

Then, for $p \geq 1$ *and* $1 \leq t \leq \min\{p, k\}$, *the following bound holds for the error* $u - u_h$ *between the analytical solution* u *of* (2.6) *and its stabilized sparse finite element approximation* $u_h \in \hat{V}^{L,p}_{(0)}$ *defined by* (4.4), *with* $L \geq 1$ *and* $h = h_L = 2^{-L}$:

$$|||u - u_h|||_{\mathrm{SD}} \leq C_{p,t} d^2 \max\{(2-p)_+, \kappa_0^{d-1}, \kappa_1^d\} \left(|\sqrt{a}| h_L^t + |b|^{\frac{1}{2}} h_L^{t+\frac{1}{2}} + c^{\frac{1}{2}} h_L^{t+1}\right) |u|_{\mathcal{H}^{t+1}(\Omega)} \tag{6.21}$$

where κ_0 *and* κ_1 *are defined in* (6.16) *and the stabilization parameter* δ_L *is given by* (6.20).

Remark 6.2. We close with some remarks on Theorem 6.1 and on possible extensions of the results presented here. We begin by noting that, save for the potential presence of a polylogarithmic factor on the right-hand side of (6.21), the definition of δ_L and the structure of the error bound in the $|||\cdot|||_{\mathrm{SD}}$ norm are exactly the same as if we used the full tensor-product finite element space $V^{L,p}_{(0)}$ instead of the sparse tensor-product space $\hat{V}^{L,p}_{(0)}$ (*cf.* Houston and Süli [14]). On the other hand, as we have commented earlier, through the use of the sparse space $\hat{V}^{L,p}_{(0)}$ (discounting the effect of $p \geq 1$ on the computational cost, since we are interested in h-version methods here with p fixed at a relatively low value), computational complexity has been reduced from $\mathcal{O}(2^{Ld})$ to $\mathcal{O}(2^L(\log_2 2^L)^{d-1})$. Hence, in comparison with a streamline-diffusion method based on the full tensor-product space, a substantial computational saving can been achieved at the cost of only marginal loss in accuracy.

(a) In the transport-dominated case, *i.e.*, when $|a| \approx 0$ and $|b| \approx 1$, we take $\delta_L = K_\delta h_L/|b|$, so the error in the streamline-diffusion norm is $\max\{(2-p)_+, \kappa_0^{d-1}, \kappa_1^d\}\mathcal{O}(h_L^{p+\frac{1}{2}})$.

In the diffusion-dominated case, that is when $\xi^\top a\xi \geq c_a|\xi|^2$, with $c_a \approx 1$ and $|b| \approx 0$, we see from Theorem 6.1 that the error in the streamline-diffusion norm $|||\cdot|||_{\mathrm{SD}}$ is $\max\{(2-p)_+, \kappa_0^{d-1}, \kappa_1^d\}\mathcal{O}(h_L^p)$ as h_L tends to zero, provided that the streamline-diffusion parameter is chosen as $\delta_L = K_\delta h_L^2/(12dp^4|\sqrt{a}|^2)$. When the matrix $a = (a_{ij})_{i,j=1}^d$ is positive definite, we have that $\Gamma_0 = \Gamma$ and therefore $u \in \mathrm{H}^1_{(0)}(\Omega) = \mathrm{H}^1_0(\Omega)$. Thus, under the conditions stated in Remark 5.3, the constant $\max\{(2-p)_+, \kappa_0^{d-1}, \kappa_1^d\}$ appearing in (6.21) for $t = p$ decays to zero exponentially as $d \to \infty$, for all $p \geq 2$.

When $a \geq 0$ and $\Gamma_0 \subsetneq \Gamma$, the constant $\max\{(2-p)_+, \kappa_0^{d-1}, \kappa_1^d\}$ remains uniformly bounded under the conditions of Remark 5.5.

In the absence of assumptions relating L, p and d, the error $|||u - u_h|||_{\mathrm{SD}}$ is still bounded by $\kappa_*^{d-1}|\log_2 h_L|^{d-1}\mathcal{O}(|\sqrt{a}|h_L^t + |b|^{\frac{1}{2}}h_L^{t+\frac{1}{2}} + c^{\frac{1}{2}}h_L^{t+1})$, where $\kappa_* \in (0,1)$ for all $L, p, d \geq 2$.

As long as the basis of the univariate space from which the sparse finite element space is constructed is a hierarchical basis on a uniform mesh, its specific choice (*viz.* whether it is a wavelet basis as in [29], or a standard hierarchical finite element basis) does not affect our final result. Thus we believe that the presence of the exponentially decreasing error constant is generic, and will be observed for error bounds in various norms. Note that the smallness of $\kappa_{(0)}(p,p,0,L)$ and $\kappa_{(0)}(p,p,1,L)$ does *not* require particularly high regularity of u as expressed by the parameter $t = p$.

If $b = 0$, streamline-diffusion stabilization is absent from the method. If, in addition, a is symmetric positive definite and $c = 0$ it follows from Céa's lemma and (5.35) with $s = 1$ that

$$|u - u_h|_{\mathrm{H}^1(\Omega)} \leq d^{\frac{3}{2}}\underline{c}_{p,t}(\kappa_0(p,t,1,L))^d\, h_L^t\, |u|_{\mathcal{H}^{t+1}(\Omega)}, \qquad 1 \leq t \leq \min\{p,k\}. \tag{6.22}$$

As has been noted in Remark 5.3, $\kappa_0(p,p,1,L) < 1$ for all $p \geq 1$ and all $L > 1$ (unconditionally). Therefore, the error constant in (6.22) exhibits exponential decay for all $p \geq 1$ and all $L \geq 1$ as $d \to \infty$, as long as $u \in \mathcal{H}^{p+1}(\Omega) \cap \mathrm{H}^1_0(\Omega)$. Under the mild conditions from Remark 5.3, relating L to p or L to d, a similar statement can be made when $c > 0$.

(b) For the sake of simplicity, we have restricted ourselves to *uniform* tensor-product partitions of $[0,1]^d$. Numerical experiments indicate that, in the presence of boundary-layers, the accuracy of the proposed sparse streamline-diffusion method can be improved by using high-dimensional versions of Shishkin-type boundary-layer-fitted tensor-product nonuniform partitions.

(c) It is important to note that the stabilization term $\delta_L \sum_{\kappa\in\mathcal{T}^L}(\mathcal{L}w, b\cdot\nabla v)_\kappa$ in the definition of the bilinear form $b_\delta(w,v)$ can be rewritten as

$$\delta_L \sum_{i=1}^d \sum_{j=1}^{2^L} \left(a_{ii}\frac{\partial^2 w}{\partial x_i^2}, b\cdot\nabla v\right)_{K_j^i} + \delta_L \sum_{\substack{i=1,j=1\\ i\neq j}}^d \left(a_{ij}\frac{\partial^2 w}{\partial x_i \partial x_j}, b\cdot\nabla v\right) + \delta_L(b\cdot\nabla w + cw, b\cdot\nabla v).$$

Here K_j^i, $i = 1,\ldots,d$, $j = 1,\ldots,2^L$, are the d-dimensional slabs defined in (6.5). Thus, instead of summing over $|\mathcal{T}^L| = 2^{Ld}$ entries we can realize the computation of the stabilization term by summing over $2^L d + \frac{1}{2}d(d-1) + 1$ terms only. The evaluation of the inner products will involve high-dimensional numerical quadrature (*cf.* [9,19,30], and the survey paper [7] for pointers to the relevant literature).

(d) For technical details concerning the efficient implementation of sparse-grid finite element methods, we refer to Zumbusch [32] and Bungartz and Griebel [7]. The work of Bungartz [6] is specifically devoted to the implementation and computational assessment of high-order sparse grid methods.

Acknowledgements. We wish to express our sincere gratitude to Adri Olde Daalhuis (University of Edinburgh), Christoph Ortner (University of Oxford) and the anonymous referees for numerous helpful suggestions.

References

[1] K. Babenko, Approximation by trigonometric polynomials is a certain class of periodic functions of several variables. *Soviet Math. Dokl.* **1** (1960) 672–675. Russian original in *Dokl. Akad. Nauk SSSR* **132** (1960) 982–985.

[2] J.W. Barrett and E. Süli, Existence of global weak solutions to kinetic models of dilute polymers. *Multiscale Model. Simul.* **6** (2007) 506–546.

[3] J.W. Barrett, C. Schwab and E. Süli, Existence of global weak solutions for some polymeric flow models. *Math. Models Methods Appl. Sci.* **15** (2005) 939–983.

[4] R.F. Bass, *Diffusion and Elliptic Operators*. Springer-Verlag, New York (1997).

[5] T.S. Blyth and E.F. Robertson, *Further Linear Algebra*. Springer-Verlag, London (2002).

[6] H.-J. Bungartz, *Finite elements of higher order on sparse grids*. Habilitation thesis, Informatik, TU München, Aachen: Shaker Verlag (1998).

[7] H.-J. Bungartz and M. Griebel, Sparse grids. *Acta Numer.* **13** (2004) 1–123.

[8] R. DeVore, S. Konyagin and V. Temlyakov, Hyperbolic wavelet approximation. *Constr. Approx.* **14** (1998) 1–26.

[9] J. Dick, I.H. Sloan, X. Wang and H. Woźniakowski, Good lattice rules in weighted Korobov spaces with general weights. *Numer. Math.* **103** (2006) 63–97.

[10] J. Elf, P. Lötstedt and P. Sjöberg, Problems of high dimension in molecular biology, in *Proceedings of the 19th GAMM-Seminar Leipzig*, W. Hackbusch Ed. (2003) 21–30.

[11] M. Griebel, Sparse grids and related approximation schemes for higher dimensional problems, in *Foundations of Computational Mathematics 2005*, L.-M. Pardo, A. Pinkus, E. Süli, M. Todd Eds., Cambridge University Press (2006) 106–161.

[12] V.H. Hoang and C. Schwab, High dimensional finite elements for elliptic problems with multiple scales. *Multiscale Model. Simul.* **3** (2005) 168–194.

[13] L. Hörmander, *The Analysis of Linear Partial Differential Operators II: Differential Operators with Constant Coefficients*. Springer-Verlag, Berlin, Reprint of the 1983 edition (2005).

[14] P. Houston and E. Süli, Stabilized *hp*-finite element approximation of partial differential equations with non-negative characteristic form. *Computing* **66** (2001) 99–119.

[15] P. Houston, C. Schwab and E. Süli, Discontinuous *hp*-finite element methods for advection-diffusion-reaction problems. *SIAM J. Numer. Anal.* **39** (2002) 2133–2163.

[16] B. Lapeyre, É. Pardoux and R. Sentis, *Introduction to Monte-Carlo Methods for Transport and Diffusion Equations*, *Oxford Texts in Applied and Engineering Mathematics*. Oxford University Press, Oxford (2003).

[17] P. Laurençot and S. Mischler, The continuous coagulation fragmentation equations with diffusion. *Arch. Rational Mech. Anal.* **162** (2002) 45–99.

[18] C. Le Bris and P.-L. Lions, Renormalized solutions of some transport equations with $W^{1,1}$ velocities and applications. *Annali di Matematica* **183** (2004) 97–130.

[19] E. Novak and K. Ritter, The curse of dimension and a universal method for numerical integration, in *Multivariate Approximation and Splines*, G. Nürnberger, J. Schmidt and G. Walz Eds., *International Series in Numerical Mathematics*, Birkhäuser, Basel (1998) 177–188.

[20] O.A. Oleĭnik and E.V. Radkevič, *Second Order Equations with Nonnegative Characteristic Form*. American Mathematical Society, Providence, RI (1973).

[21] H.-C. Öttinger, *Stochastic Processes in Polymeric Fluids*. Springer-Verlag, New York (1996).

[22] H.-G. Roos, M. Stynes and L. Tobiska, *Numerical Methods for Singularly Perturbed Differential Equations. Convection-Diffusion and Flow Problems*, *Springer Series in Computational Mathematics* **24**. Springer-Verlag, New York (1996).

[23] C. Schwab, *p- and hp-Finite Element Methods: Theory and Applications in Solid and Fluid Mechanics, Numerical Methods and Scientific Computation.* Clarendon Press, Oxford (1998).

[24] S. Smolyak, Quadrature and interpolation formulas for products of certain classes of functions. *Soviet Math. Dokl.* **4** (1963) 240–243. Russian original in *Dokl. Akad. Nauk SSSR* **148** (1963) 1042–1045.

[25] E. Süli, Finite element approximation of high-dimensional transport-dominated diffusion problems, in *Foundations of Computational Mathematics 2005*, L.-M. Pardo, A. Pinkus, E. Süli, M. Todd Eds., Cambridge University Press (2006) 343–370. Available at: `http://web.comlab.ox.ac.uk/oucl/publications/natr/index.html`

[26] E. Süli, Finite element algorithms for transport-diffusion problems: stability, adaptivity, tractability, in *Invited Lecture at the International Congress of Mathematicians*, Madrid, 22–30 August 2006. Available at: `http://web.comlab.ox.ac.uk/work/endre.suli/Suli-ICM2006.pdf`

[27] V. Temlyakov, Approximation of functions with bounded mixed derivative, in *Proc. Steklov Inst. of Math.* **178**, American Mathematical Society, Providence, RI (1989).

[28] N.G. van Kampen, *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam (1992).

[29] T. von Petersdorff and C. Schwab, Numerical solution of parabolic equations in high dimensions. *ESAIM: M2AN* **38** (2004) 93–128.

[30] G. Wasilkowski and H. Woźniakowski, Explicit cost bounds of algorithms for multivariate tensor product problems. *J. Complexity* **11** (1995) 1–56.

[31] C. Zenger, Sparse grids, in *Parallel Algorithms for Partial Differential Equations*, W. Hackbusch Ed., *Notes on Numerical Fluid Mechanics* **31**, Vieweg, Braunschweig/Wiesbaden (1991).
[32] G.W. Zumbusch, A sparse grid PDE solver, in *Advances in Software Tools for Scientific Computing*, H.P. Langtangen, A.M. Bruaset and E. Quak Eds., *Lecture Notes in Computational Science and Engineering* **10**, Springer, Berlin (Proceedings SciTools '98) (2000) 133–177.