Ioannis Toulopoulos*

SUPG space-time scheme on anisotropic meshes for general parabolic equations

https://doi.org/10.1515/jnma-2024-0053
Received April 10, 2024; accepted December 6, 2024; published online June 4, 2025

Abstract: In this paper, a stabilized space-time finite element scheme on anisotropic quadrilateral meshes for general linear parabolic problems is considered. The scheme is devised on the basis of a unified space-time variational formulation, and uses continuous piece-wise polynomial spaces. The stabilization is achieved by incorporating Streamline-Upwind Petrov–Galerkin (SUPG) techniques. Defining appropriately the stability coefficients, we initially show anisotropic interpolation estimates and then *a priori* error estimates by following a classical finite element methodology. A series of numerical examples illustrates the theoretical findings.

Keywords: general linear parabolic equations; continuous space-time finite element discretizations; quadrilateral anisotropic meshes; a priori error estimates; asymptotic convergence rates

MSC 2010: 65N30; 65J15; 65N15; 65N12

1 Introduction

Many transport physical problems are often described by using general second order parabolic equations of the form $u_t + Lu = f$, where $Lu := -\operatorname{div}(\varepsilon \nabla_x u) + \boldsymbol{\beta} \cdot \nabla_x u + ru$ is the second order differential operator, and $\nabla_x u$ is the spatial gradient of u, $\boldsymbol{\beta}$ is a constant vector representing the convection velocity and the parameters $\varepsilon > 0$, $r \geqslant 0$ represent the diffusion and reaction coefficients, respectively, [1]. The numerical solution of these problems has been a subject of investigation by many authors in the past decades (cf. [2]). Standard Galerkin Finite Element Methods (FEMs) with continuous spaces may appear numerical instabilities, which are produced due to the advection character of the problem, (advection dominant case). Very often Streamline-Upwind (SU) stabilization terms are added to treat the problem numerically and to ensure the stability of the FE discretization, see, e.g., [3], [4], and in refs. [2], [5], [6] for an overview and computational results for SU methods. The full discretization of the problem is completed by applying a time-stepping scheme, e.g., Runge–Kutta, which results to sequential approximations of the solution in time (see, e.g., [7]–[9]). These approaches typically impose a restriction on the time step relative to the spatial mesh size, which can lead to additional difficulties when highly refined meshes are required.

In contrast to these methods, the last proposed space-time finite element methods (STFEMs) discretize time evolution problems by applying a unified and simultaneous finite element discretization in space and in time directions, [10]. The main idea is to see the time variable t as another spatial variable, let's say, x_{d_x+1}, if $x_1, \ldots, x_{d_x}$, are the spatial variables, and the time derivative u_t as a convection in the direction x_{d_x+1}. In view of this, an associated global space-time variational formulation is derived and the time-dependent problem is considered as a stationary problem into a domain (i.e., the space-time cylinder) with one higher dimension. This idea is not

***Corresponding author: Ioannis Toulopoulos**, Department of Informatics, University of Western Macedonia, Fourka Area, 52100 Kastoria, Greece, E-mail: itoulopoulos@uowm.gr

new and has been used by many authors for developing different STFE discretizations for time evolution problems. Specifically, in the past few years the proposed STFEMs are mainly based on a corresponding space time variational formulation, see, e.g., [11]–[14], and [15], [16] for alternative stabilization approaches, also [17] for nonlinear problems. In refs. [18], [19], space-time methods the Isogeometric Analysis framework are discussed and see the survey paper [20] for applications to engineering problems. The proposed STFEMs offer some further advantages compared to the traditional discretization methods, as for example, the direct use of existing FEM solvers developed for stationary problems, the generalization of adaptivity techniques which have been studied for stationary problems, the use of coarse/fine unstructured space-time meshes in order to compute highly accurate and efficient solutions without having time step limitations, and lastly are more flexible to be implemented in parallel environments and to perform computations in such architectures. We refer to refs. [21], [22] for a discussion on different parallelization approaches for STFEMs. However, many of these techniques do not directly carry over when solving problems in four-dimensional (4D) space-time domains. The construction of unstructured meshes in 4D space-time domains seems to be a challenging task, [23].

In this paper, inspired by the ideas presented in ref. [24], where SU stabilized finite element methods for steady advection–diffusion problems are analysed, we devise a stable STFEM for solving the general parabolic problem mentioned above. The method is considered on anisotropic quadrilateral meshes, which are aligned to the coordinate axes. The aim is to derive error estimates in the related energy norm uniformly with respect to the diffusion parameter ε, by taking into account the anisotropic mesh sizes of the triangulation of the space-time cylinder. Animated by the interpolation results given in ref. [25], anisotropic interpolation error estimates are derived here, where the associated interpolation constants depend on the directional stretching properties of the mesh. The additional SU stability terms appearing in the final space-time scheme are weighted by a numerical parameter, see (3.18), which is accordingly formed by the anisotropic character of the mesh. This parameter is determined through the stability of the resulting bilinear form and includes the local, i.e., per element, spatial mesh sizes. The numerical results confirm the theoretical findings. It is known that the solution of the aforementioned general parabolic problem can exhibit interior boundary layers. In practical applications, the resolution of this layer is of main interest and can be typically treated by applying an anisotropic meshing technique, [26], [27]. This work is the first step to devise STFEMs in this direction. Extensions of the proposed work to more general fluid flow problems with boundary layers are under preparation.

The outline of the paper can be stated as follows. In Section 2, some preliminaries together with the notation of the related Sobolev spaces are given. In Section 3, the general parabolic problem is given, and the weak space-time formulation is described. In the last part of Section 3 the ST-FE discretization in presented and the discretization error analysis is developed. Finally, in Section 4 we show a series of numerical examples for verifying the theoretical results. The paper closes with the conclusions.

2 Preliminaries

2.1 Notations

Let Q be a bounded Lipschitz domain in $\mathbb{R}^d$, $d = 2, \dots, 4$, with boundary $\Gamma = \partial Q$. For any multi-index $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)$ of non-negative integers $\alpha_1, \dots, \alpha_d$, we use the following notations, (i) $|\mathbf{a}| = \sum_{i=1}^{d} a_i$, (ii) set $\mathbf{x}^{\mathbf{a}} = x_1^{a_1} x_2^{a_2} \dots x_d^{a_d}$ for $\mathbf{x} \in \mathbb{R}^d$, (iii) moreover introduce the differential operator $D^{\boldsymbol{\alpha}} := \partial_x^{\mathbf{a}} = \partial_{x_1}^{\alpha_1} \dots \partial_{x_d}^{\alpha_d}$, with $\partial_{x_j}(\cdot) = \partial(\cdot)/\partial x_j$, $j = 1, \dots, d$. Let $1 \leqslant p \leqslant \infty$ be fixed and ℓ be a non-negative integer. As usual, $L^p(Q)$ denotes the Lebesgue spaces for which $\int_Q |\varphi(x)|^p \, \mathrm{d}x < \infty$, endowed with the norm $\|\varphi\|_{L^p(Q)} = \left(\int_Q |\varphi(x)|^p \, \mathrm{d}x \right)^{1/p}$, and $W^{\ell,p}(Q)$ is the Sobolev space, which consists of the functions $\varphi : Q \to \mathbb{R}$ such that their weak derivatives $D^{\boldsymbol{\alpha}}\varphi$ with $|\boldsymbol{\alpha}| \leqslant \ell$ exist and belong to $L^p(Q)$. If $\varphi \in W^{\ell,p}(Q)$, then its norm is defined by

$$\|\varphi\|_{W^{\ell,p}(Q)} = \left(\sum_{0 \leqslant |\boldsymbol{\alpha}| \leqslant \ell} \|D^{\mathbf{a}}\varphi\|_{L^p(Q)}^p \right)^{1/p}, \qquad \|\varphi\|_{W^{\ell,\infty}(Q)} = \max_{0 \leqslant |\boldsymbol{\alpha}| \leqslant \ell} \|D^{\mathbf{a}}\varphi\|_{\infty}$$

for $1 \leqslant p < \infty$ and $p = \infty$, respectively. We further define the spaces

$$W_0^{\ell,p}(Q) := \left\{\varphi \in W^{\ell,p}(Q) \text{ such that } \varphi|_{\partial Q} = 0\right\}, \tag{2.1a}$$

$$W_{\Sigma^*}^{\ell,p}(Q) := \left\{\varphi \in W^{\ell,p}(Q) \text{ such that } \varphi|_{\Sigma^* \subset \partial Q} = 0\right\}. \tag{2.1b}$$

Remark 2.1. If $p = 2$ we usually use the notation $H^\ell(Q) := W^{\ell,2}(Q)$, $\ell = 0, 1, 2, \ldots$

Let the fixed integer $\ell \geqslant 1$ and let the set A of all multi-indices having the form $A := \{\mathbf{a}: \mathbf{a} = (\ldots, m_i e_i, \ldots)\}$ where $0 \leqslant m_i \leqslant \ell$, $i = 1, \ldots, d$, and e_i is the unit vector in the i-th direction. For $\varphi \in W^{\ell,p}(Q)$ denote

$$\sum_{\mathbf{a}\in\mathbf{A}} \| D^{\mathbf{a}}\varphi\|_{L^p(Q)} = \sum_{i=1}^{d} \|\partial_{x_i}^{m_i}\varphi\|_{L_p(Q)}. \tag{2.2}$$

We refer the reader to ref. [28] for more details about Sobolev spaces.

2.1.1 Differential operators on the space-time domain

Next, we define certain differential operators which are related to the time and the spatial variables. Let $J = (0, T]$ be the time interval with some final time $T > 0$ and let Ω be a bounded domain in $\mathbb{R}^{d_x}$, $d_x = 1, 2$ or 3. For later use, we consider the space-time cylinder $Q \subset \mathbb{R}^d$ with $d = d_x + 1$, defined by $Q = J \times \Omega$, and its boundary parts $\Sigma = \partial\Omega \times J$, $\Sigma_T = \Omega \times \{T\}$, and $\Sigma_0 = \Omega \times \{0\}$ such that $\partial Q = \Sigma \cup \overline{\Sigma}_0 \cup \overline{\Sigma}_T$. Accordingly to the definition of $\partial_x^{\boldsymbol{\alpha}}$, we now define the operator $\partial_x^{\boldsymbol{\alpha}_{d_x}}$ and also define the spatial gradient $\nabla_x\varphi = (\partial_{x_1}\varphi, \ldots, \partial_{x_{d_x}}\varphi)$, and the whole gradient $\nabla\varphi := (\partial_t\varphi, \nabla_x\varphi)$.

2.2 Known inequalities and identities

The following inequalities are going to be used in several places in the text. Hölder's and Young's inequalities read: For any δ, $0 < \delta < \infty$, and $1 \leqslant p, q \leqslant \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$, for $f \in L^p(Q)$ and $g \in L^q(Q)$, there holds

$$\left| \int_Q fg \, \mathrm{d}x \right| \leqslant \|f\|_{L^p(Q)}\|g\|_{L^q(Q)} \leqslant \frac{\delta}{p}\|f\|_{L^p(Q)}^p + \frac{\delta^{-q/p}}{q}\|g\|_{L^q(Q)}^q. \tag{2.3a}$$

Poincaré–Friedrichs inequality, see [28], [29]: Let $Q \subset \mathbb{R}^d$ be a parallelepiped (cuboid) and let the face $\Sigma^* \subset \partial Q$ vertical to the $x_j, 1 \leqslant j \leqslant d$, coordinate plane. Then for any $f \in W_{\Sigma^*}^{1,p}(Q)$, it holds

$$\int_Q |f|^p \, \mathrm{d}x \leqslant C(Q) \int_Q |\partial_{x_i} f|^p \, \mathrm{d}x, \quad 1 \leqslant i \leqslant d. \tag{2.3b}$$

Let the vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)$, the function $f \in W^{1,p}(Q)$ and the outward normal vector $\mathbf{n}$ to ∂Q. In several places we will use the identities:

$$\nabla \cdot (\boldsymbol{\beta} f) = \boldsymbol{\beta} \cdot \nabla f + (\nabla \cdot \boldsymbol{\beta}) f, \tag{2.4a}$$

$$2\int_Q \boldsymbol{\beta} \cdot \nabla f \; f \, \mathrm{d}x = -\int_Q \nabla \cdot \boldsymbol{\beta} \; f^2 \, \mathrm{d}x + 2\int_Q \boldsymbol{\beta} \cdot \mathbf{n} \; f^2 \, \mathrm{d}s. \tag{2.4b}$$

3 The continuous problem

Let Ω be a bounded cuboid domain in $\mathbb{R}^{d_x}$, with $d_x = 1, 2, 3$, with smooth boundary $\Gamma = \partial\Omega$. We define the space-time cylinder $\bar{Q}_T := \bar{\Omega} \times [0, T]$, where T is the final time, and boundary $\partial Q_T = \Sigma \cup \bar{\Sigma}_0 \cup \bar{\Sigma}_T$, where $\Sigma := \Gamma \times (0, T)$ is the lateral boundary, $\Sigma_0 := \Omega \times \{0\}$ and $\Sigma_T := \Omega \times \{T\}$. Consider the differential operator L in the form

$$Lu := -\operatorname{div}(\varepsilon \nabla_x u) + \boldsymbol{\beta} \cdot \nabla_x u + ru, \tag{3.1}$$

where $\nabla_x u$ is the spatial gradient of u, and $\varepsilon > 0$ is the constant diffusion coefficient. The constant vector $\boldsymbol{\beta} := (\beta_x, \beta_y, \beta_z)$ takes values in $\mathbb{R}^{d_x}$ and the reaction parameter r takes values in $\mathbb{R}_+$. We consider the following initial-boundary value problem: Find $u(x, t)$: $\bar{Q}_T \to \mathbb{R}$ such that

$$u_t + Lu = f \qquad \text{in } Q_T, \tag{3.2a}$$

$$u = u_\Sigma = 0 \qquad \text{on } \Sigma, \tag{3.2b}$$

$$u(x, 0) = u_0(x) \qquad \text{on } \Omega, \tag{3.2c}$$

where f, u_0 are given functions. For simplicity, we only consider homogeneous Dirichlet boundary conditions on Σ. However, the analysis presented in this work can easily be generalized to other constellations of boundary conditions, cf., [10].

In literature, usually the main point of the concept for defining a weak formulation of (3.2) is to consider that u_t lives in the dual space (or in some sub-space) of the space where u lives, [1]. Anyway, as we mentioned before, in recent years appropriate space-time weak formulations for parabolic problems similar to (3.2) have been presented, where the regularity of the solution is considered uniformly in all the space time cylinder, i.e., $u \in W^{2,p}(Q_T)$, see, e.g., [10] and the reference therein.

3.1 Weak space-time form

Assume that $u_0 \in H_0^1(\Omega)$ and $f \in L^2(Q_T)$. Multiplying (3.2a) by a smooth function $v(x, t)$ vanishing on $\Sigma \cup \Sigma_T$, and after applying Green's theorem and an integration by parts with respect to both t and x, and the usage of the boundary conditions, we can obtain

$$\int_{Q_T} \{-uv_t + \varepsilon \nabla_x u \cdot \nabla_x v + \boldsymbol{\beta} \cdot \nabla_x u\, v + ru\, v\} \,\mathrm{d}x\,\mathrm{d}t = \int_{Q_T} fv \,\mathrm{d}x\,\mathrm{d}t + \int_\Omega u_0(x) v(x, 0) \,\mathrm{d}x. \tag{3.3}$$

Introducing the appropriate regularity properties for the data, global regularity properties can be shown for the generalized solution u of (3.3) in Q_T, i.e., $u_t \in L^2(0, T;\ L^2(\Omega))$, [1], and furthermore, it can be inferred by using embeddings that $u \in W^{1,p}(Q_T)$, cf., [30], [31]. Applying a formal integration by parts with respect to time variable we can arrive at the space-time weak formulation: Find $u \in H_\Sigma^1(Q_T)$ with $u(0, x) = u_0(x)$, such that

$$\int_{Q_T} \{\, u_t v + \varepsilon \nabla_x u \cdot \nabla_x v + \boldsymbol{\beta} \cdot \nabla_x u\, v + ru\, v\} \,\mathrm{d}x\,\mathrm{d}t = \int_{Q_T} fv \,\mathrm{d}x\,\mathrm{d}t, \quad v \in H_\Sigma^1(Q_T). \tag{3.4}$$

Assumption 3.1. For the solution u of (3.4), assume that $u \in V$, with $V = H_\Sigma^1(Q_T) \cap H^\ell(Q_T),\ \ell \geqslant 2$.

Remark 3.1. The space-time variational formulation (3.4) has a unique solution, see, e.g., analysis in refs. [30], [31], and also [12] for considerations in Gelfand triple spaces. In these works, beside existence and uniqueness results, one can also find useful a priori estimates and regularity results.

In view of (3.4), we define

$$B(u,v) := \int_{Q_T} \{\, u_t v + \varepsilon \nabla_x u \cdot \nabla_x v + \boldsymbol{\beta} \cdot \nabla_x u\, v + r u\, v \} \,\mathrm{d}x\,\mathrm{d}t, \tag{3.5a}$$

$$\ell_f(v) := \int_{Q_T} f v \,\mathrm{d}x\,\mathrm{d}t \tag{3.5b}$$

for $v \in H^1_{\Sigma}(Q_T)$. Working in a formalistic way, and setting $u = v$ in (3.4), and then using (2.4b) and the fact that $\nabla_x \cdot \boldsymbol{\beta} = 0$, we can deduce that

$$\begin{aligned} B(u,u) &= \int_{\Sigma_T} \frac{1}{2} u^2(s) \,\mathrm{d}s + 2 \int_{\Sigma} \boldsymbol{\beta} \cdot \mathbf{n}\, u^2 \,\mathrm{d}s + \int_{Q_T} \varepsilon |\nabla_x u|^2 + r u^2 \,\mathrm{d}x\,\mathrm{d}t \\ &\lesssim \int_{Q_T} |fu| \,\mathrm{d}x\,\mathrm{d}t + \int_{\Omega} u_0^2(x) \,\mathrm{d}x \overset{(2.31)(2.3b)}{\lesssim} c(\delta) \int_{Q_T} |f|^2 \,\mathrm{d}x\,\mathrm{d}t + \delta \int_{Q_T} |\nabla_x u|^p \,\mathrm{d}x\,\mathrm{d}t. \end{aligned} \tag{3.6}$$

Recalling (3.2b), and choosing sufficiently small, i.e., $\delta = \varepsilon/4$ in (3.6), we can have the a priory bound

$$\int_{\Sigma_T} u^2(s) \,\mathrm{d}s + \varkappa(\varepsilon) \int_{Q_T} |\nabla_x u|^2 \,\mathrm{d}x\,\mathrm{d}t \lesssim c(\delta) \|f\|^2_{L^2(Q_T)} + \|u_0\|^2_{L^2(\Omega)}. \tag{3.7}$$

Working in the same spirit for the case $\beta = 0$, make test with $v := u_t(x,t)$ in (3.4), and note that $\frac{\mathrm{d}}{\mathrm{d}t} \int_{Q_T} |\nabla_x u|^2 = \int_{Q_T} \nabla_x u \cdot \nabla_x u_t$, then

$$\begin{aligned} &\int_{Q_T} |u_t|^2 \,\mathrm{d}x\,\mathrm{d}t + \varepsilon \int_{\Sigma_T} |\nabla_x u|^2 \,\mathrm{d}x + r \int_{Q_T} u u_t \,\mathrm{d}x\,\mathrm{d}t \\ &\qquad \lesssim c_2(\delta) \int_{Q_T} |f|^2 \,\mathrm{d}x\,\mathrm{d}t + \delta \int_{Q_T} |u_t|^2 \,\mathrm{d}x\,\mathrm{d}t + \int_{\Omega} \varepsilon |\nabla_x u_0|^2 \,\mathrm{d}x \end{aligned} \tag{3.8}$$

provided that $\nabla_x u \in C([0,T], L^2(\Omega))$. Choosing δ appropriate small, we obtain

$$c_3(\delta) \int_{Q_T} |u_t|^2 \,\mathrm{d}x\,\mathrm{d}t + \varepsilon \int_{\Sigma_T} \frac{r}{2} u^2 + |\nabla_x u|^2 \,\mathrm{d}x \lesssim c_2(\delta) \int_{Q_T} |f|^2 \,\mathrm{d}x\,\mathrm{d}t + \int_{\Omega} \frac{r}{2} u_0^2 + \varepsilon |\nabla_x u_0|^2 \,\mathrm{d}x. \tag{3.9}$$

Remark 3.2. The estimate in (3.8) does not provide any bound for the u_t term. On the other hand the estimate in (3.9) gives a bound for u_t in $L^2(Q_T)$. This aligns with the idea of employing streamline stabilization techniques, as described below, to develop a stable space-time discretization scheme for (3.4). Other ideas for producing stable space-time discretizations have been developed by means of Petrov–Galerkin techniques after an appropriate selection of trial and test space, see, e.g., [12], [13], [32].

3.2 The space-time finite element approximation

3.2.1 Basic concepts

We start by introducing the discrete setting. Let $\mathcal{T}_h := \{E^l\}_{l=1,\dots,N}$ be a conforming mesh partition, of the space-time cylinder Q_T into closed rectangular mesh elements, such that

$$\bar{Q}_T = \bigcup_l E^l, \quad E^{\mathrm{o},l_1} \cap E^{\mathrm{o},l_2} = \varnothing, \quad 1 \leqslant l_1 \neq l_2 \leqslant N, \tag{3.10}$$

where $E^{\mathrm{o},l}$ is the interior of the mesh element. Denote $\mathbf{h}_E := (h_{1,E}, \dots, h_{i,E}, \dots, h_{d,E})$ to be the vector with the i-th directional mesh widths of the element $E \in \mathcal{T}_h$. For the analysis below some further mesh width quantities must be introduced. Define $\mathbf{h} = (h_1, \dots, h_i, \dots, h_d) := (\max_E\{h_{1,E}\}, \dots, \max_E\{h_{i,E}\}, \dots, \max_E\{h_{d,E}\})$ and

$\mathbf{h}_{\min} = (h_{\min,1}, \dots, h_{\min,i}, \dots, h_{\min,d}) := (\min_E\{h_{1,E}\}, \dots, \min_E\{h_{i,E}\}, \dots, \min_E\{h_{d,E}\})$. The diameter of every $E^l \in \mathcal{T}_h$ is denoted by h_{E^l} and we set $h := \max_{E^l} h_{E^l}$, $h_t := h_1$, $h_{\boldsymbol{x}} := \max_{i\geqslant 2}\{h_i\}$, $h_{E,\ \min} := \min_{i\geqslant 2}\{h_{i,E}\}$. In the sequel we write $E \in \mathcal{T}_h$ instead of $E^l \in \mathcal{T}_h$. Note that the mesh $\mathcal{T}_h$ forms a Cartesian grid in Q_T and every mesh element $E \in \mathcal{T}_h$ is affine equivalent to reference element $\hat{E} := [0,1]^d$ though the transformation

$$T_E\colon \hat{E} \to E, \quad T_E(\hat{\mathbf{x}}) := \mathbf{x}(\hat{\mathbf{x}}) = \mathcal{B}_E\hat{\mathbf{x}} + \mathbf{b}_E, \tag{3.11}$$

with $\mathbf{b}_E \in \mathbb{R}^d$ and $\mathcal{B}_E \in \mathbb{R}^{d\times d}$ to be a diagonal matrix of the form $\mathcal{B}_{E,ii} = h_{i,E},\ i = 1, \dots, d$.

Assumption 3.2. The partition $\mathcal{T}_h$ is quasi-uniform, in the sense that there are positive constants σ_M, σ_m independent of h such that $\sigma_M \geqslant h_i/h_{\min,i} \geqslant \sigma_m$ for all $i = 1, \dots, d$.

Remark 3.3. The quasi-uniformity properties Assumption 3.2 of $\mathcal{T}_h$ allows the use of anisotropic mesh widths between the different axial directions, and furthermore to replace $\mathbf{h}_E$ by $\mathbf{h}$ in the analysis below.

Remark 3.4. Consider a function $f \in W^{1,2}(E)$ and the function $\hat{f} = f(T_E(\hat{\mathbf{x}}))$. It can be concluded by (3.11) that $\nabla_{\hat{\mathbf{x}}}\hat{f} = \mathcal{B}_E\nabla_{\mathbf{x}} f$. By applying the change of variables (3.11), it can be shown that $\int_{\hat{E}}|\partial_{\hat{x}_i}\hat{f}|^p\,\mathrm{d}\hat{x} = |\mathcal{B}_E|^{-1}|h_{i,E}^p\int_E|\partial_{x_i} f|^p\,\mathrm{d}x$, $i = 1, \dots, d$, where $p > 1$, and $|\mathcal{B}_E|$ is the determinant of $\mathcal{B}_E$.

On $\mathcal{T}_h$ we define the finite dimensional space

$$V_h^k = \left\{\varphi_h \in C^0(\bar{Q})\colon \varphi_h|_E \in \mathbb{Q}^k(E) \quad \forall\, E \in \mathcal{T}_h, \quad \varphi_h = 0 \quad \text{on} \ \ \Sigma \cup \Sigma_0\right\}. \tag{3.12}$$

Here $\mathbb{Q}^k(E)$ is the space of polynomials on E composed by tensor products of uni-variate Lagrange polynomials of degree at most k with respect to each variable, i.e., $\mathbb{Q}^k(E) = \bigotimes_{i=1}^d \mathbb{P}^k(x_i)$, where $\mathbb{P}^k(x_i),\ i = 1, \dots, d$ is the univariate Lagrange polynomials. In the analysis below, we consider the case of $k = 1$.

Proposition 3.1. *Consider a polynomial function $\varphi_h \in \mathbb{Q}^k(E),\ E \in \mathcal{T}_h$ such that $\hat{\varphi}_h = \varphi_h(T_E(\hat{\mathbf{x}})) \in \mathbb{Q}^k(\hat{E})$. There is a constant c_{inv} independent of $\boldsymbol{h}_E$ such that*

$$\sum_{i=1}^d \|\partial_{x_i}\varphi_h\|_{L^2(E)} \leqslant c_{inv}\sum_{i=1}^d h_{i,E}^{-1}\|\varphi_h\|_{L^2(E)}. \tag{3.13}$$

Proof. Since all the norms of $\hat{\varphi}_h \in \mathbb{Q}^k(\hat{E})$ are equivalent, there is a $c > 0$ such that

$$\|\partial_{x_i}\hat{\varphi}_h\|_{L^2(\hat{E})} \leqslant c\|\hat{\varphi}_h\|_{L^2(\hat{E})}. \tag{3.14}$$

Utilizing the form of the transformation (3.11) and applying the chain rule $\partial_{x_i}\varphi_h = \partial_{\hat{x}_i}\hat{\varphi}_h\frac{\partial\hat{x}_i}{\partial x_i}$, we can obtain

$$\begin{aligned}\int_E (\partial_{x_i}\varphi_h)^2\,\mathrm{d}\mathbf{x} &= h_{i,E}^{-2}|\mathcal{B}_E|^{-1}\int_{\hat{E}}(\partial_{\hat{x}_i}\hat{\varphi}_h)^2\,\mathrm{d}\hat{\mathbf{x}}\\ &\overset{(3.14)}{\leqslant} ch_{iE}^{-2}|\mathcal{B}_E|^{-1}\int_{\hat{E}}(\hat{\varphi}_h)^2\,\mathrm{d}\hat{\mathbf{x}} \leqslant ch_{i,E}^{-2}\int_E(\varphi_h)^2\,\mathrm{d}\mathbf{x},\end{aligned} \tag{3.15}$$

from where we can deduce (3.13) by summing over $i = 1, \dots, d$. □

Corollary 3.1. *By inequality* (3.13) *we can infer*

$$\sum_{i=1}^d \|D_{x_i}\varphi_h\|_{L^2(E)}^2 \leqslant c_{inv}\sum_{i=1}^d h_{E,\min}^{-2}\|\varphi_h\|_{L^2(E)}^2. \tag{3.16}$$

3.2.2 The unified space-time FE scheme

Assumption 3.3. For simplicity in the discretization error analysis we suppose that $u_{0,h} = u_0 := 0$. For problems with $u_0 \neq 0$ we refer to refs. [10], [19].

Based on (3.4) and using (3.5) we consider the following discretization of problem (3.4): Find $u_h \in V_h^k$ such that $u_h = u_{0,h}$ on Σ_0 and

$$B(u_h, v_h) = \ell_f(v_h) \quad \forall\, v_h \in V_h^k. \tag{3.17}$$

In order to obtain stable solutions the scheme (3.17) is modified by adding an upwind stabilization term, and the final discrete problem is written: Find $u_h \in V_h^k$ such that $u_h = u_{0,h}$ on Σ_0 and

$$\begin{aligned} B_s(u_h, v_h) :=& B(u_h, v_h) + S(u_h, v_h) \\ &+ R_s(u_h, v_h) = \ell_f\big(v_h + \tau^\lambda \boldsymbol{\beta}_{Q_T} \cdot \nabla v_h\big) \quad \forall\, v_h \in V_h^k, \end{aligned} \tag{3.18a}$$

where the vector $\boldsymbol{\beta}_{Q_T} := (1, \boldsymbol{\beta}) \in \mathbb{R}^d$, and the streamline-upwind (SU) term S is

$$S(u_h, v_h) := \sum_{E \in \mathcal{T}_h} \int_E \tau^\lambda \big(\boldsymbol{\beta}_{Q_T} \cdot \nabla u_h\ \boldsymbol{\beta}_{Q_T} \cdot \nabla v_h\big)\, \mathrm{d}x\, \mathrm{d}t. \tag{3.18b}$$

Here $\tau^\lambda := \vartheta h_{E,\min}^\lambda$, with $\vartheta > 0$, $\lambda > 1$ fixed parameters, which will be specified below. The residual term has the form

$$R_s(u_h, v_h) := \sum_{E \in \mathcal{T}_h} \int_E -\tau^\lambda \big(\varepsilon \Delta u_h - r u_h\big) \boldsymbol{\beta}_{Q_T} \cdot \nabla v_h\, \mathrm{d}x\, \mathrm{d}t \tag{3.18c}$$

and the linear form

$$\ell_f\big(v_h + \tau^\lambda \boldsymbol{\beta}_{Q_T} \cdot \nabla v_h\big) := \sum_{E \in \mathcal{T}_h} \int_E f(v_h + \tau^\lambda \boldsymbol{\beta}_{Q_T} \cdot \nabla v_h)\, \mathrm{d}x\, \mathrm{d}t. \tag{3.19}$$

Remark 3.5. Note that, in case of working with linear spaces, i.e., $V_h^{k=1}$, the residual terms take the form $R_s(u_h, v_h) = \sum_{E \in \mathcal{T}_h} \int_E \tau^\lambda r\ u_h\ \boldsymbol{\beta}_{Q_T} \cdot \nabla v_h\, \mathrm{d}x\, \mathrm{d}t$.

Remark 3.6 (consistency). Under the Assumption 3.1, the following localized variational form

$$\begin{aligned} B_s(u, v_h) :=& B(u, v_h) + S(u, v_h) \\ &+ R_s(u, v_h) = \ell_f\big(v_h + \tau^\lambda \boldsymbol{\beta}_{Q_T} \cdot \nabla v_h\big) \quad \forall\, v_h \in V_h^k \end{aligned} \tag{3.20}$$

holds for the weak solution u.

In view of (3.20) and (3.18a), we have the following equation.

Corollary 3.2. *Let u be the solution of problem* (3.4) *and u_h the solution of problem* (3.18a)*. Then the following error equation holds for $v_h \in V_h^k$,*

$$B_s(u - u_h, v_h) = 0. \tag{3.21}$$

Below the coercivity and boundedness properties of $B_s(\cdot, \cdot)$ are discussed.

Lemma 3.1. *Let $v_h \in V_h^k$ and the Assumption* 3.3*. Then*

$$\begin{aligned} B_s(v_h, v_h) \geqslant& \frac{1}{2} \|v_h\|^2_{L^2(\Sigma_T)} \\ &+ \sum_{E \in \mathcal{T}_h} \left\{ \frac{\varepsilon}{2} \|\nabla_x v_h\|^2_{L^2(E)} + \frac{r}{2} \|v_h\|^2_{L^2(E)} + \frac{\tau^\lambda}{2} \|\boldsymbol{\beta}_{Q_T} \cdot \nabla v_h\|^2_{L^2(E)} \right\}. \end{aligned} \tag{3.22}$$

Proof. For $E_i, \in \mathcal{T}_h$, let $\mathbf{n}_{E_i} = (n_{t,E_i}, \mathbf{n}_{x,E_i})$ be the unit normal vector on ∂E_i. Let $\mathbf{n}_{e,ij} = (n_{t,ij}, \mathbf{n}_{x,ij})$ be the unit normal vector on the common faces $e_{ij} = \partial E_i \cap \partial E_j$ for $E_i,\ E_j \in \mathcal{T}_h$. Using (2.4b) and the identity $2v(\boldsymbol{\beta}_{Q_T} \cdot \nabla v) = \boldsymbol{\beta}_{Q_T} \cdot \nabla v^2 = \nabla \cdot (\boldsymbol{\beta}_{Q_T}\, v^2)$, it follows immediately that

$$\begin{aligned}\sum_{E_i\in\mathcal{T}_h}\int_{E_i} v_h(\boldsymbol{\beta}_{Q_T}\cdot\nabla v_h)\,\mathrm{d}x\,\mathrm{d}t &= \frac{1}{2}\sum_{E_i\in\mathcal{T}_h}\int_{E_i}\nabla\cdot(\boldsymbol{\beta}_{Q_T}\, v_h)\,\mathrm{d}x\,\mathrm{d}t = \frac{1}{2}\sum_{E_i\in\mathcal{T}_h}\int_{\partial E_i}\mathbf{n}_{E_i}\cdot\boldsymbol{\beta}_{Q_T} v_h^2\,\mathrm{d}S\\ &= \frac{1}{2}\sum_{e_{ij}}\int_{e_{ij}}\boldsymbol{\beta}_{Q_T}\cdot\mathbf{n}_{e,ij}\Big(v_h^2|_{E_i} - v_h^2|_{E_j}\Big)\,\mathrm{d}S = \frac{1}{2}\int_{\Sigma_T} v_h^2\,\mathrm{d}S. \end{aligned} \tag{3.23}$$

Recalling the terms of $B_s(\cdot,\cdot)$ in (3.18) and following the same steps as in (3.6) and using (3.23), we can derive the bound

$$\begin{aligned} B_s(v_h, v_h) &= \sum_{E\in\mathcal{T}_h}\int_E \big[(\boldsymbol{\beta}_{Q_T}\cdot\nabla v_h)v_h + \varepsilon\nabla_x v_h\cdot\nabla_x v_h + r v_h^2\big]\,\mathrm{d}x\,\mathrm{d}t\\ &\quad + \sum_{E\in\mathcal{T}_h}\int_E \tau^\lambda\big[-\varepsilon\Delta v_h + \boldsymbol{\beta}_{Q_T}\cdot\nabla v_h + r v_h\big]\boldsymbol{\beta}_{Q_T}\cdot\nabla v_h\,\mathrm{d}x\,\mathrm{d}t\\ &\geqslant \frac{1}{2}\|v_h\|^2_{L^2(\Sigma_T)} + \sum_{E\in\mathcal{T}_h}\Big\{ r\|v_h\|^2_{L^2(E)} + \varepsilon\|\nabla_x v_h\|^2_{L^2(E)} + \tau^\lambda\|\boldsymbol{\beta}_{Q_T}\cdot\nabla v_h\|^2_{L^2(E)}\Big\}\\ &\quad + \sum_{E\in\mathcal{T}_h}\int_E \tau^\lambda\big[-\varepsilon\Delta v_h + r v_h\big]\boldsymbol{\beta}_{Q_T}\cdot\nabla v_h\,\mathrm{d}x\,\mathrm{d}t. \end{aligned} \tag{3.24}$$

Now, for the last sum in (3.24), we apply (2.3a) and obtain

$$\begin{aligned} &\left|\sum_{E\in\mathcal{T}_h}\int_E \tau^\lambda\big[-\varepsilon\Delta v_h + r v_h\big]\boldsymbol{\beta}_{Q_T}\cdot\nabla v_h\,\mathrm{d}x\,\mathrm{d}t\right|\\ &\quad\leqslant \sum_{E\in\mathcal{T}_h}\left\{\left|\int_E -\tau^{\lambda/2}\varepsilon\Delta v_h\ \tau^{\lambda/2}\boldsymbol{\beta}_{Q_T}\cdot\nabla v_h\,\mathrm{d}x\,\mathrm{d}t\right| + \left|\int_E \tau^{\lambda/2} r v_h\ \tau^{\lambda/2}\boldsymbol{\beta}_{Q_T}\cdot\nabla v_h\,\mathrm{d}x\,\mathrm{d}t\right|\right\}\\ &\quad\leqslant \sum_{E\in\mathcal{T}_h}\left\{\tau^\lambda\varepsilon^2\|\Delta v_h\|^2_{L^2(E)} + \tau^\lambda\, r^2\|v_h\|^2_{L^2(E)} + \frac{\tau^\lambda}{2}\|\boldsymbol{\beta}_{Q_T}\cdot\nabla v_h\|^2_{L^2(E)}\right\}. \end{aligned} \tag{3.25}$$

Here, we make use of the discrete-inverse inequality, $\|\Delta v_h\|^2_{L^2(E)} \leqslant \frac{c^2_{\mathrm{inv}}}{h^2_{E,\min}}\|\nabla_x v_h\|^2_{L^2(E)}$, see (3.13) and (3.16), and the identity $ab \leqslant a^2 + \frac{1}{4}b^2$ to find

$$\begin{aligned} &\left|\sum_{E\in\mathcal{T}_h}\int_E \tau^\lambda\big[-\varepsilon\Delta v_h + r v_h\big]\boldsymbol{\beta}_{Q_T}\cdot\nabla v_h\,\mathrm{d}x\,\mathrm{d}t\right|\\ &\quad\leqslant \sum_{E\in\mathcal{T}_h}\left\{\varepsilon^2\,\tau^\lambda\left(\frac{c_{\mathrm{inv}}}{h_{E,\min}}\right)^2\|\nabla_x v_h\|^2_{L^2(E)} + \tau^\lambda\, r^2\|v_h\|^2_{L^2(E)} + \frac{\tau^\lambda}{2}\|\boldsymbol{\beta}_{Q_T}\cdot\nabla v_h\|^2_{L^2(E)}\right\}. \end{aligned} \tag{3.26}$$

Now, introducing the result (3.26) into (3.24), and using that $0 < \varepsilon \leqslant 1$, we find

$$\begin{aligned} B_s(v_h, v_h) \geqslant & \frac{1}{2}\|v_h\|^2_{L^2(\Sigma_T)} + \sum_{E\in\mathcal{T}_h}\Big\{ r\|v_h\|^2_{L^2(E)} + \varepsilon\|\nabla_x v_h\|^2_{L^2(E)} + \tau^\lambda\|(\boldsymbol{\beta}_{Q_T}\cdot\nabla v_h)\|^2_{L^2(E)}\Big\} \\ & - \sum_{E\in\mathcal{T}_h}\left\{ \varepsilon^2\,\tau^\lambda\left(\frac{c_{\text{inv}}}{h_{E,\min}}\right)^2\|\nabla v_h\|^2_{L^2(E)} + \tau^\lambda\, r^2\|v_h\|^2_{L^2(E)} + \frac{\tau^\lambda}{2}\|\boldsymbol{\beta}_{Q_T}\cdot\nabla v_h\|^2_{L^2(E)}\right\} \\ = & \frac{1}{2}\|v_h\|^2_{L^2(\Sigma_T)} + \sum_{E\in\mathcal{T}_h}\left\{\left(\varepsilon - \varepsilon^2\,\tau^\lambda\left(\frac{c_{\text{inv}}}{h_{E,\min}}\right)^2\right)\|\nabla_x v_h\|^2_{L^2(E)} + (r - \tau^\lambda\, r^2)\|v_h\|^2_{L^2(E)}\right. \\ & \left. + \left(\tau^\lambda - \frac{\tau^\lambda}{2}\right)\|\boldsymbol{\beta}_{Q_T}\cdot\nabla v_h\|^2_{L^2(E)}\right\}. \end{aligned} \tag{3.27}$$

Making the appropriate choice for the parameter $\vartheta = \frac{1}{2}\min\left(\frac{1}{2c^2_{\text{inv}}}, \frac{1}{2r}\right)$, and setting $\lambda = 2$, we can infer the bound (3.22). □

In view of (3.22), we introduce the mesh-dependent norms

$$\|v\|^2_s = \frac{1}{2}\|v\|^2_{L^2(\Sigma_T)} + \sum_{E\in\mathcal{T}_h}\left\{\frac{\varepsilon}{2}\|\nabla_x v\|^2_{L^2(E)} + \frac{r}{2}\|v\|^2_{L^2(E)} + \frac{\tau^\lambda}{2}\|\boldsymbol{\beta}_{Q_T}\cdot\nabla v\|^2_{L^2(E)}\right\}, \tag{3.28a}$$

$$\|v\|^2_{h,s} := \|v\|^2_s + \sum_{E\in\mathcal{T}_h}\tau^{-\lambda}\|v\|^2_{L^2(E)} + \sum_{E\in\mathcal{T}_h}\left\{\tau^\lambda\varepsilon^2\|\Delta v\|^2_{L^2(E)}\right\}. \tag{3.28b}$$

By (3.22) we can immediately write the estimate

$$B_s(v_h, v_h) \geqslant \|v_h\|^2_s \quad \forall v_h \in V^k_h. \tag{3.29}$$

Setting in (3.29) $v_h = z_h - u_h \in V^k_h$ and recalling the error equation (3.21), we get

$$\|z_h - u_h\|^2_s \leqslant B_s(z_h - u_h, z_h - u_h) + B_s(u_h - u, z_h - u_h) = \quad B_s(z_h - u, z_h - u_h). \tag{3.30}$$

Lemma 3.2. *Let u the weak solution of* (3.5) *under Assumption* 3.1 *and $u_0 = 0$. Let $u_h \in V^k_h$ be the finite element solution in* (3.17)*. The approximation error estimate*

$$\|u - u_h\|_s \leqslant C_{0,\text{approx}}\|u - z_h\|_{h,s} \tag{3.31}$$

holds for all $z_h \in V^k_h$, where the constant $C_{0,\text{approx}}$ is independent of h.

Proof. Let a $z_h \in V^k_h$. Recalling the terms appearing in (3.18), we have

$$\begin{aligned} B_s(u - z_h, u_h - z_h) := & B(u - z_h, u_h - z_h) + S(u - z_h, u_h - z_h) \\ & + R_s(u - z_h, u_h - z_h) \end{aligned} \tag{3.32}$$

with

$$\begin{aligned} B(u - z_h, u_h - z_h) := & \int_{Q_T}\{\partial_t(u - z_h)(u_h - z_h) + \varepsilon\nabla_x(u - z_h)\cdot\nabla_x(u_h - z_h) \\ & + \boldsymbol{\beta}\cdot\nabla_x(u - z_h)\,(u_h - z_h) + r(u - z_h)\,(u_h - z_h)\}\,\mathrm{d}x\,\mathrm{d}t, \end{aligned} \tag{3.33}$$

$$S(u - z_h, u_h - z_h) := \sum_{E\in\mathcal{T}_h}\int_E \tau^\lambda\big(\boldsymbol{\beta}_{Q_T}\cdot\nabla(u - z_h)\;\boldsymbol{\beta}_{Q_T}\cdot\nabla(u_h - z_h)\big)\,\mathrm{d}x\,\mathrm{d}t, \tag{3.34}$$

$$R_s(u-z_h,u_h-z_h):=\sum_{E\in\mathcal{T}_h}\int_E -\tau^{\lambda}\big(\varepsilon\Delta(u-z_h)\boldsymbol{\beta}_{Q_T}\cdot\nabla(u_h-z_h)\big)\,\mathrm{d}x\,\mathrm{d}t. \tag{3.35}$$

Using appropriately (2.3a), we derive the following bounds for the terms in (3.33):

$$\int_{Q_T}\varepsilon\nabla_x(u-z_h)\cdot\nabla_x(u_h-z_h)\,\mathrm{d}x\,\mathrm{d}t\leqslant\sqrt{\varepsilon}\|\nabla_x(u-z_h)\|_{L2(Q_T)}\|u_h-z_h\|_s, \tag{3.36}$$

$$\begin{aligned}
&\int_{Q_T}\big((u_t-z_{h,t})+\boldsymbol{\beta}\cdot\nabla_x(u-z_h)\big)(u_h-z_h)\,\mathrm{d}x\,\mathrm{d}t=\int_{Q_T}\big(\boldsymbol{\beta}_{Q_T}\cdot\nabla(u-z_h)\big)(u_h-z_h)\,\mathrm{d}x\,\mathrm{d}t\\
&\overset{(2.4)}{=}-\int_{Q_T}\big(\boldsymbol{\beta}_{Q_T}\cdot\nabla(u_h-z_h)\big)(u-z_h)\,\mathrm{d}x\,\mathrm{d}t+\int_{\partial Q_T}\boldsymbol{\beta}_{Q_T}\cdot\mathbf{n}(u-z_h)(u_h-z_h)\,\mathrm{d}s\\
&\leqslant\Big(\tau^{\lambda}\|\boldsymbol{\beta}_{Q_T}\cdot\nabla(u_h-z_h)\|^2_{L^2(Q_T)}\Big)^{1/2}\Big(\tau^{-\lambda}\|u-z_h\|^2_{L^2(Q_T)}\Big)^{1/2}\quad\text{(using BCs)}\\
&\quad+\|u_h-z_h\|_{L^2(\Sigma_T)}\,\|u-z_h\|_{L^2(\Sigma_T)}\\
&\leqslant\left(2\Big(\tau^{-\lambda}\|u-z_h\|^2_{L^2(Q_T)}\Big)^{1/2}+2\|u-z_h\|_{L^2(\Sigma_T)}\right)\|u_h-z_h\|_s,
\end{aligned} \tag{3.37}$$

and also

$$\begin{aligned}
\int_{Q_T}\Big(\frac{r}{2}\Big)^{1/2}(u_h-z_h)(2r)^{1/2}(u-z_h)\,\mathrm{d}x\,\mathrm{d}t&\leqslant\Big(\frac{r}{2}\|u_h-z_h\|^2_{L^2(Q_T)}\Big)^{1/2}\Big(2r\|u-z_h\|^2_{L^2(Q_T)}\Big)^{1/2}\\
&\leqslant C_r\|u-z_h\|_{L^2(Q_T)}\,\|u_h-z_h\|_s.
\end{aligned} \tag{3.38}$$

Treating together the terms in (3.34) and (3.35), we get

$$\begin{aligned}
&\sum_{E\in\mathcal{T}_h}\int_E\tau^{\lambda/2}\big(-\varepsilon\Delta(u-z_h)+\boldsymbol{\beta}_{Q_T}\cdot\nabla(u-z_h)+r(u-z_h)\big)\tau^{\lambda/2}\boldsymbol{\beta}_{Q_T}\cdot\nabla(u_h-z_h)\,\mathrm{d}x\,\mathrm{d}t\Big|\\
&\leqslant\sum_{E\in\mathcal{T}_h}\big\{\tau^{\lambda/2}\big(\varepsilon\|\Delta u-\Delta z_h\|_{L^2(E)}+\|\boldsymbol{\beta}_{Q_T}\cdot\nabla(u-z_h)\|_{L^2(E)}+r\|u-z_h\|_{L^2(E)}\big)\\
&\quad\times\tau^{\lambda/2}\|\boldsymbol{\beta}_{Q_T}\cdot\nabla(u_h-z_h)\|_{L^2(E)}\big\}\\
&\lesssim\sum_{E\in\mathcal{T}_h}\Big\{\tau^{\lambda}\Big(\varepsilon^2\|\Delta u-\Delta z_h\|^2_{L^2(E)}+\|\boldsymbol{\beta}_{Q_T}\cdot\nabla(u-z_h)\|^2_{L^2(E)}+r^2\|u-z_h\|^2_{L^2(E)}\Big)\Big\}^{1/2}\|u_h-z_h\|_s.
\end{aligned} \tag{3.39}$$

Collecting the previous bounds, and using (3.32) we can obtain

$$B_s(u-z_h,u_h-z_h)\leqslant\|u-z_h\|_{h,s}\|u_h-z_h\|_s. \tag{3.40}$$

Combining (3.40) and (3.30) and adding $\|u-z_h\|_s$ on both sides of the inequality we can deduce

$$\|u_h-z_h\|_s+\|u-z_h\|_s\lesssim 2\|u-z_h\|_{h,s}. \tag{3.41}$$

The result (3.31) can be derived by applying triangle inequality. □

Note that the error estimate (3.31) includes bounds for the term $\|\partial_t u-\partial_t u_h\|_{L^2}$, compare with Remarks 3.1 and 3.2.

3.2.3 Quasi-interpolation

Let $V = W^{\ell,p}(Q_T)$, $\ell \geqslant 2$, $p > 1$. We introduce the quasi-interpolant, [29], [33],

$$\Pi_h \colon V \to V_h^k, \qquad \Pi_h f = \sum_{i=1}^{\dim_{V_h^k}} \lambda_i(f)\varphi_{i,h}, \tag{3.42}$$

where λ_i are the linear dual functionals and $\varphi_{i,h}$ the local shape functions of V_h^k. The multivariate quasi-interpolant given in (3.42) can be constructed by the corresponding uni-variate quasi-interpolants $\pi_1, \dots, \pi_d$, i.e., $\Pi_h f = (\pi_1 \otimes \pi_2 \cdots \otimes \pi_d)(f)$, and each dual functional $\lambda_i(f)$ can be similarly defined by the corresponding uni-variate dual functionals, see, e.g., [29], [33]–[35]. For convenience one can consider a dual basis satisfying $\lambda_i(\varphi_{j,h}) = \delta_{ij}$, $i, j = 1, \dots, \dim_{V_h^k}$, [33]. By this property it can be immediately inferred that $\Pi_h(\varphi_h) = \varphi_h$ for all $\varphi_h \in V_h^k$.

Remark 3.7. A discussion for the dual functionals $\lambda_i(f)$ when f belongs to tensor spaces is given in refs. [34], [35]. In ref. [25] an analysis is given for the case of low regularity functions f and anisotropic meshes.

The following stability bounds can be shown, [29], [33], [35].

Proposition 3.2. *Let f be a smooth function. The bounds*

$$\|\Pi_h f\|_{L^p(Q_T)} \leqslant C_0 \|f\|_{L^p(Q_T)}, \tag{3.43a}$$

$$\|\Pi_h f\|_{W^{1,p}(Q_T)} \leqslant C_1 \|f\|_{W^{1,p}(Q_T)} \tag{3.43b}$$

hold, where C_0, C_1 are positive constants.

Suppose $\mathcal{D} \subset \mathbb{R}^d$ is a cuboid domain and $\mathbf{h}_D = (h_1, \dots, h_d)$ is a vector containing the sizes h_i in each dimension x_i, $i = 1, \dots, d$. Define the set of multi-indices $A = \{me_i,\ i = 1, \dots, d,\ m \in \mathbb{N}\}$, where e_i are the orthonormal basis vectors, and denote with $A^0 = \{\mathbf{m}_0 = (m_1, \dots, m_d) \colon m_i < m\}$ the set of multi-indices which corresponds to the tensor product polynomials of degree less than m. Note that A^0 can also be charactirized as $A^0 = \{\mathbf{m}_0 \colon D^{\mathbf{a}}\mathbf{x}^{\mathbf{m}_0} = 0 \ \ \forall\, \mathbf{a} \in A\}$. For a smooth function f, we define the tensor Taylor polynomial of degree less than m evaluated at $\mathbf{y}$:

$$T_{\mathbf{y}}^m f(\mathbf{x}) = \sum_{\mathbf{m}_0 \in A^0} \frac{1}{\mathbf{m}_0!} D^{\mathbf{m}_0} f(\mathbf{y})(\mathbf{x} - \mathbf{y})^{\mathbf{m}_0} \tag{3.44a}$$

and the averaged Taylor expansion over the ball $B \subset \mathcal{D}$:

$$T_{\mathbf{y},\varphi}^A f(\mathbf{x}) = \int_B T_{\mathbf{y}}^m f(\mathbf{x})\varphi(\mathbf{y})\, d\mathbf{y}, \tag{3.44b}$$

where $\varphi \in C_0^\infty(B)$ is a cut-off function with $\int_B \varphi(\mathbf{y})\, d\mathbf{y} = 1$, see [29, Chapt. 4].

In the analysis below we focus mainly on the case $m = 2$. Let the multi-indices $\mathbf{a} \in A$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)$ such that $|\boldsymbol{\beta}| \leqslant 1$.

Proposition 3.3. *Let f be a smooth function. The remainder $R^A f := f - T_{\mathbf{y},\varphi}^A f$ has the form*

$$R^A f(\mathbf{x}) = \sum_{\mathbf{a} \in A} \int_D k_{\mathbf{a}}(\mathbf{x}, \mathbf{y}) D^{\mathbf{a}} f(\mathbf{y})\, d\mathbf{y}, \tag{3.45a}$$

where

$$k_{\mathbf{a}}(\mathbf{x},\mathbf{y}) = C_{\mathbf{a},m}(\mathbf{x}-\mathbf{y})^{\mathbf{a}} k(\mathbf{x},\mathbf{y}) \tag{3.45b}$$

with

$$k(\mathbf{x},\mathbf{y}) \leqslant C|\mathbf{x}-\mathbf{y}|^{-d}, \qquad |D_{\mathbf{x}}^{\boldsymbol{\beta}} k_{\mathbf{a}}(\mathbf{x},\mathbf{y})| \leqslant C|\mathbf{x}-\mathbf{y}|^{|\mathbf{a}|-|\boldsymbol{\beta}|-d}. \tag{3.45c}$$

Proof. The bounds can be shown by combining the results given in refs. [29], [35], [36]. □

We have the commutativity result.

Proposition 3.4. *Let f be a smooth function then*

$$D^{\boldsymbol{\beta}} T_{\mathbf{y},\varphi}^{A} f = T_{\mathbf{y},\varphi}^{A-\boldsymbol{\beta}} D^{\boldsymbol{\beta}} f. \tag{3.46}$$

Proof. Note that for $\boldsymbol{\beta} > \mathbf{m}_0$ holds that $D_{\mathbf{x}}^{\boldsymbol{\beta}} \mathbf{x}^{\mathbf{m}_0} = 0$ and thus

$$\begin{aligned} D_{\mathbf{x}}^{\boldsymbol{\beta}} T_{\mathbf{y},\varphi}^{A} f(\mathbf{x}) &= \sum_{\mathbf{m}_0 \in A^0} \int_B \varphi(\mathbf{y}) D_{\mathbf{y}}^{\mathbf{m}_0} f(\mathbf{y}) \frac{(\mathbf{x}-\mathbf{y})^{(\mathbf{m}_0-\boldsymbol{\beta})}}{(\mathbf{m}_0-\boldsymbol{\beta})!}\, \mathrm{d}\mathbf{y} \\ &= \sum_{\boldsymbol{\gamma} \in (A^0-\boldsymbol{\beta})} \int_B \varphi(\mathbf{y}) D_{\mathbf{y}}^{\boldsymbol{\beta}+\boldsymbol{\gamma}} f(\mathbf{y}) \frac{(\mathbf{x}-\mathbf{y})^{\boldsymbol{\gamma}}}{\boldsymbol{\gamma}!}\, \mathrm{d}\mathbf{y} \\ &= \sum_{\boldsymbol{\gamma} \in (A^0-\boldsymbol{\beta})} \int_B \varphi(\mathbf{y}) D_{\mathbf{y}}^{\boldsymbol{\gamma}} \left(D_{\mathbf{y}}^{\boldsymbol{\beta}} f(\mathbf{y}) \right) \frac{(\mathbf{x}-\mathbf{y})^{\boldsymbol{\gamma}}}{\boldsymbol{\gamma}!}\, \mathrm{d}\mathbf{y} \\ &= T_{\mathbf{y},\varphi}^{A-\boldsymbol{\beta}} (D^{\boldsymbol{\beta}} f)(\mathbf{x}). \end{aligned} \tag{3.47}$$

□

Let the multi-indices $\mathbf{a} \in A$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)$ such that $|\boldsymbol{\beta}| \leqslant 1$. Following similar ideas as in refs. [25], [29], [35], we show the estimates.

Theorem 3.1. *Let $f \in W^{\ell,p}(D)$ with $\ell = m$. Then there exist a tensor Taylor polynomial $T^A f := T_{\mathbf{y},\varphi}^A f$ such that*

$$\|f - T^A f\|_{L^p(D)} \leqslant C_0 \sum_{\mathbf{a} \in A} \mathbf{h}_D^{\mathbf{a}} \|D^{\mathbf{a}} f\|_{L^p(D)}, \tag{3.48a}$$

$$\|D^{\boldsymbol{\beta}}(f - T^A f)\|_{L^p(D)} \leqslant C_1 \sum_{\mathbf{a} \in A, \boldsymbol{\gamma} = (\mathbf{a}-\boldsymbol{\beta})} \mathbf{h}_D^{\boldsymbol{\gamma}} \|D^{\boldsymbol{\gamma}}(D^{\boldsymbol{\beta}} f)\|_{L^p(D)}, \tag{3.48b}$$

where $C_0 > 0$, $C_1 > 0$ are independent of the mesh sizes h_i, $i = 1, \dots, d$.

Proof. Assume initially that $\mathcal{D} \subset \mathbb{R}^d$ is a (reference) cuboid domain with $\mathrm{diam}(\mathcal{D}) \sim 1$, e.g., $\mathcal{D} := \hat{D} = [-1,1]^d$. By Proposition 3.3 we can deduce

$$\left| \int_{\hat{D}} k_{\mathbf{a}}(\mathbf{x},\mathbf{z}) D^{\mathbf{a}} f(\mathbf{z})\, \mathrm{d}\mathbf{z} \right| \leqslant C \|D^{\mathbf{a}} f\|_{L^p(\hat{D})} \quad \forall \mathbf{a} \in A, \tag{3.49}$$

which implies that

$$\|f - T^A f\|_{L^p(\hat{D})} \leqslant C_0 \sum_{\mathbf{a} \in A} \|D^{\mathbf{a}} f\|_{L^p(\hat{D})}. \tag{3.50}$$

Further, combining the last bounds with Proposition 3.4, we can get

$$\|D^{\boldsymbol{\beta}}(f - T^A f)\|_{L^p(\hat{D})} = \|R^{A-\boldsymbol{\beta}} D^{\boldsymbol{\beta}} f(\mathbf{x})\|_{L^p(\hat{D})} \leqslant C \sum_{\mathbf{a} \in A, \boldsymbol{\gamma} = (\mathbf{a}-\boldsymbol{\beta})} \|D^{\boldsymbol{\gamma}}(D^{\boldsymbol{\beta}} f(\mathbf{x}))\|_{L^p(\hat{D})}. \tag{3.51}$$

Utilizing now the tensor character of the mesh, we can apply the change of variables $\mathbf{x} = \mathcal{B}_D\hat{\mathbf{x}} + \mathbf{b}$, see (3.11), and to transform $\hat{D}$ to D. Then we can show that, see Remark 3.4,

$$\begin{aligned} \|D^{\boldsymbol{\beta}}(f - T^A f)\|^p_{L^p(\hat{D})} &= |\mathcal{B}_D|^{-1}\mathbf{h}_D^{p\boldsymbol{\beta}}\|D^{\boldsymbol{\beta}}(f - T^A f)\|^p_{L^p(D)}, \\ \|D^{\mathbf{a}} f\|^p_{L^p(\hat{D})} &= |\mathcal{B}_D|^{-1}\mathbf{h}_D^{p\mathbf{a}}\|D^{\mathbf{a}} f\|_{L^p(D)}. \end{aligned} \tag{3.52}$$

Inserting the results (3.52) into the estimates (3.50), (3.51), we can deduce the required estimates (3.48). □

Having (3.48) we can show bounds on how well the quasi-interpolant $\Pi_h f$ approximates the function f. We follow standard ideas from the finite element methodology. Recall that $\mathbf{a} \in \mathbf{A}$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)$ with $|\boldsymbol{\beta}| \leqslant 1$.

Lemma 3.3. *Let a function $f \in W^{\ell,2}(Q_T)$, $\ell = m$, and Assumption 3.2 for the mesh $\mathcal{T}_h$. Then for the quasi-interpolant (3.42) the interpolation estimates*

$$\|f - \Pi_h f\|^2_{L^2(E)} \leqslant c_{intp,0}\sum_{\mathbf{a}\in A}\mathbf{h}_E^{\mathbf{a}}\|D^{\mathbf{a}} f\|_{L^2(E)}, \quad E \in \mathcal{T}_h, \tag{3.53a}$$

$$\|D^{\boldsymbol{\beta}}(f - \Pi_h f)\|_{L^2(E)} \leqslant c_{intp,1}\sum_{\mathbf{a}\in A,\boldsymbol{\gamma}=(\mathbf{a}-\boldsymbol{\beta})}\mathbf{h}_E^{\boldsymbol{\gamma}}\|D^{\boldsymbol{\gamma}}(D^{\boldsymbol{\beta}} f)\|_{L^2(E)}, \quad E \in \mathcal{T}_h, \tag{3.53b}$$

hold, where the positive constants $c_{\text{intp},0}, c_{\text{intp},1}$ are independent of $\boldsymbol{h}_E$.

Proof. By the results of Theorem 3.1 there exists a tensor polynomial $p_f \in \mathbb{P}^{\mathbf{m}_0}$, $\mathbf{m}_0 \in A_0$ such that

$$\|f - p_f\|_{L^2(E)} \lesssim \sum_{\mathbf{a}\in A}\mathbf{h}_E^{\mathbf{a}}\|D^{\mathbf{a}} f\|_{L^2(E)}, \quad E \in \mathcal{T}_h. \tag{3.54}$$

Using that $\Pi_h(v_h) = v_h \;\; \forall\, v_h \in V_h^k$ and the stability properties (3.48a), (3.43a) we get

$$\begin{aligned} \|f - \Pi_h f\|_{L^2(E)} &\leqslant \|f - p_f\|_{L^2(E)} + \|\Pi_h(f - p_f)\|_{L^2(E)} \\ &\leqslant C\|f - p_f\|_{L^2(E)} \leqslant \sum_{\mathbf{a}\in A}\mathbf{h}_E^{\mathbf{a}}\|D^{\mathbf{a}} f\|_{L^2(E)} \leqslant \sum_{\mathbf{a}\in A}\mathbf{h}^{\mathbf{a}}\|D^{\mathbf{a}} f\|_{L^2(E)}. \end{aligned} \tag{3.55}$$

Following the same steps as above and using the interpolation estimate (3.48b) we can prove (3.53b). □

Remark 3.8. Note that the anisotropic character of the mesh appears in the interpolation estimates given in (3.53). Similar estimates have been shown in ref. [25] for low regularity functions on anisotropic triangular meshes.

Lemma 3.4. *Consider a mesh $\mathcal{T}_h$, with uniform directional mesh widths, i.e., $h_1 \sim h_2 \sim \cdots \sim h_d$ and an element $E \in \mathcal{T}_h$ with* $\operatorname{diam}(E) := h_E$. *Let a function $f \in W^{\ell,p}(\Omega)$, $1 \leqslant \ell$, $1 < p$, and the local intepolation operator $\Pi_h \in \mathbb{Q}^{k=l-1}$. Then the interpolation estimate*

$$|f - \Pi_h f|_{W^{m,p}(E)} \lesssim h_E^{\ell-m}|f|_{W^{\ell,p}(E)} \tag{3.56}$$

holds for $0 \leqslant m \leqslant \ell$.

Proof. See, cf. [29]. □

For isotropic meshes it is known (cf. [29]), that there is a constant $C_{\text{trc}} > 0$, such that

$$\|v\|^2_{L^p(\partial E)} \leqslant C_{\text{trc}}h^{-1}\big(\|v\|_{L^2(E)} + h\|\nabla v\|_{L^2(E)}\big)^2, \quad E \in \mathcal{T}_h. \tag{3.57}$$

Proposition 3.5. *Let a face $e \subset \partial E$ of a mesh element $E \in \mathcal{T}_h$ such that $|E| = |e|h_{\perp,e}$, where $h_{\perp,e}$ is the measure of the edge perpendicular to e. Let a function $f \in H^1(Q_T)$. Then there is a positive constant C_{trc} such that*

$$\|f\|^2_{L^2(e)} \leqslant C_{trc} h^{-1}_{\perp,e}\left(\|f\|^2_{L^2(E)} + \sum_{i=1}^{d} h^2_{i,E}\|\partial_{x_i} f\|^2_{L^2(E)}\right). \tag{3.58}$$

Proof. Let $\hat{f} := f \circ T_E \in H^1(\hat{E})$. For the face $\hat{e} \subset \partial\hat{E}$ the trace ineguality has the form (cf. [29]),

$$\|\hat{f}\|^2_{L^2(\hat{e})} \lesssim \|\hat{f}\|^2_{L^2(\hat{E})} + \|\nabla\hat{f}\|^2_{L^2(\hat{E})}. \tag{3.59}$$

Recalling (3.11) and Remark 3.4, we can transform (3.59) onto E to deduce

$$\|f\|^2_{L^2(e)}|e|^{-1} \lesssim |E|^{-1}\Big(\|f\|^2_{L^2(E)} + \|\mathcal{B}_E\nabla f\|^2_{L^2(E)}\Big). \tag{3.60}$$

Owing to the form of $\mathcal{B}_E$ we have $\|\mathcal{B}_E\nabla f\|^2_{L^2(E)} \lesssim \sum_{i=1}^{d} h^2_{i,E}\|\partial_{x_i} f\|^2_{L^2(E)}$. Using that $h_{\perp,e}|e| = |E|$ we can obtain (3.58). □

Remark 3.9. For $e \subset \Sigma_T$ inequality (3.58) is valid for $h_{\perp,e} \sim h_t$.

Corollary 3.3. *Let the assumptions of Lemma 3.3 and assume $f \in H^2(Q_T)$. Then the following interpolation estimate*

$$\|f - \Pi_h f\|^2_{L^2(\Sigma_T)} \lesssim \sum_{E\in\mathcal{T}_h, \partial E\subset\Sigma_T} \mathbf{h}_E^{-\mathbf{e}_1}\left(\sum_{\mathbf{a}} \mathbf{h}_E^{2\mathbf{a}} + \sum_{i=1}^{d}\left(\mathbf{h}_E^{3\mathbf{e}_i} + h_i\sum_{j\neq i}^{d} h_j^2\right)\right)\|f\|^2_{H^2(E)} \tag{3.61}$$

holds for all multiindices $\boldsymbol{a} \in A$ and $\mathbf{e}_i$ to be the i-th unit normal vector.

Proof. According to (3.58) and Remark 3.9,

$$\begin{aligned}
\|f - \Pi_h f\|^2_{L^2(\Sigma_T)} &= \sum_{e\in\mathcal{E}, e\subset\Sigma_T}\int_e |f - \Pi_h f|^2\,\mathrm{d}s \\
&\leqslant c\sum_{E\in\mathcal{T}_h, \partial E\subset\Sigma_T} h_t^{-1}\left(\|f - \Pi_h f\|^2_{L^2(E)} + \sum_{i=1}^{d} h_i^2\|\partial_{x_i}(f - \Pi_h f)\|^2_{L^2(E)}\right) \\
&\leqslant c\sum_{E\in\mathcal{T}_h, \partial E\subset\Sigma_T} h_t^{-1}\left(\sum_{\mathbf{a}}\mathbf{h}_E^{2\mathbf{a}}\|D^{\mathbf{a}} f\|^2_{L^2(E)} + \sum_{i=1}^{d} h_i^2\sum_{\gamma=\mathbf{a}-\mathbf{e}_i}\mathbf{h}_E^{\gamma}\|D^{\gamma}(\partial_{x_i} f)\|^2_{L^2(E)}\right) \\
&\leqslant c\sum_{E\in\mathcal{T}_h, \partial E\subset\Sigma_T} \mathbf{h}_E^{-\mathbf{e}_1}\left(\sum_{\mathbf{a}}\mathbf{h}_E^{2\mathbf{a}} + \sum_{i=1}^{d}\sum_{\gamma=\mathbf{a}-\mathbf{e}_i}\mathbf{h}_E^{2\mathbf{e}_i}\mathbf{h}_E^{\gamma}\right)\|f\|^2_{H^2(E)} \\
&\leqslant c\sum_{E\in\mathcal{T}_h, \partial E\subset\Sigma_T} \mathbf{h}_E^{-\mathbf{e}_1}\left(\sum_{\mathbf{a}}\mathbf{h}_E^{2\mathbf{a}} + \sum_{i=1}^{d}\sum_{\mathbf{a}}\mathbf{h}_E^{\mathbf{a}+\mathbf{e}_i}\right)\|f\|^2_{H^2(E)}.
\end{aligned} \tag{3.62}$$

Expanding the last sum in (3.62), we get

$$\|f - \Pi_h f\|^2_{L^2(\Sigma_T)} \leqslant c\sum_{E\in\mathcal{T}_h, \partial E\subset\Sigma_T} \mathbf{h}_E^{-\mathbf{e}_1}\left(\sum_{\mathbf{a}}\mathbf{h}_E^{2\mathbf{a}} + \sum_{i=1}^{d}\left(\mathbf{h}_E^{3\mathbf{e}_i} + h_i\sum_{j\neq i}^{d} h_j^2\right)\right)\|f\|^2_{H^2(E)}$$

as reguired. □

Lemma 3.5. *Let $f \in V$ satisfying the assumptions of Lemma 3.3, and let the associated interpolant $\Pi_h f$, see (3.53). Then there exist a constant independent of f and $\boldsymbol{h}_E$ such that the following quasi-interpolation estimate*

$$\begin{aligned}\|f-\Pi_h f\|_{h,s}^2 \lesssim \sum_{E\in\mathcal{T}_h}\Bigg\{\Bigg(\sum_{\mathbf{a}\in A}\mathbf{h}_E^{2\mathbf{a}-\mathbf{e}_1}+\sum_{i=1}^{d}\sum_{\mathbf{a}\in A}\mathbf{h}_E^{\mathbf{a}+\mathbf{e}_i-\mathbf{e}_1}\Bigg)+\sum_{\mathbf{a}\in A}\tau^{-\lambda}\mathbf{h}_E^{2\mathbf{a}}\\ +\sum_{j=2}^{d}\sum_{\mathbf{a}\in A}\Big\{\tau^{\lambda}\varepsilon^2\mathbf{h}_E^{2(\mathbf{a}-2\mathbf{e}_j)}+\frac{\varepsilon}{2}\mathbf{h}_E^{2(\mathbf{a}-\mathbf{e}_j)}\Big\}+\sum_{i=1}^{d}\sum_{\mathbf{a}\in A}\frac{\tau^{\lambda}}{2}\mathbf{h}_E^{2(\mathbf{a}-\mathbf{e}_i)}\Bigg\}\|f\|_{H^2(E)}^2\end{aligned} \tag{3.63}$$

holds true.

Proof. Recall that

$$\begin{aligned}\|f-\Pi_h f\|_{h,s}^2=\frac{1}{2}\|f-\Pi_h f\|_{L^2(\Sigma_T)}^2+\sum_{E\in\mathcal{T}_h}\tau^{-\lambda}\|f-\Pi_h f\|_{L^2(E)}^2+\sum_{E\in\mathcal{T}_h}\Big\{\tau^{\lambda}\varepsilon^2\|\Delta f-\Delta\Pi_h f\|_{L^2(E)}^2\Big\}\\ +\sum_{E\in\mathcal{T}_h}\Big\{\frac{\varepsilon}{2}\|\nabla_x(f-\Pi_h f)\|_{L^2(E)}^2+\frac{r}{2}\|f-\Pi_h f\|_{L^2(E)}^2+\frac{\tau^{\lambda}}{2}\|\boldsymbol{\beta}_{Q_T}\cdot\nabla(f-\Pi_h f)\|_{L^2(E)}^2\Big\}.\end{aligned} \tag{3.64}$$

Under the regularity assumptions for f and utilizing (3.53) and (3.61), we can infer the following estimates

$$\|f-\Pi_h f\|_{L^2(\Sigma_T)}^2\lesssim\sum_{E\in\mathcal{T}_h,\partial E\subset\Sigma_T}\Bigg(\sum_{\mathbf{a}\in A}\mathbf{h}_E^{2\mathbf{a}-\mathbf{e}_1}+\sum_{i=1}^{d}\sum_{\mathbf{a}\in A}\mathbf{h}_E^{\mathbf{a}+\mathbf{e}_i-\mathbf{e}_1}\Bigg)\|f\|_{H^2(E)}^2$$

and also

$$\sum_{E\in\mathcal{T}_h}\tau^{-\lambda}\|f-\Pi_h f\|_{L^2(E)}^2\lesssim\sum_{E\in\mathcal{T}_h}\sum_{\mathbf{a}\in A}\tau^{-\lambda}\mathbf{h}_E^{2\mathbf{a}}\|f\|_{H^2(E)}^2,$$

$$\sum_{E\in\mathcal{T}_h}\Big\{\tau^{\lambda}\varepsilon^2\|\Delta(f-\Pi_h)f\|_{L^2(E)}^2\Big\}\lesssim\sum_{E\in\mathcal{T}_h}\sum_{j=2}^{d}\sum_{\mathbf{a}\in A}\tau^{\lambda}\varepsilon^2\mathbf{h}_E^{2(\mathbf{a}-2\mathbf{e}_j)}\|f\|_{H^2(E)}^2,$$

$$\sum_{E\in\mathcal{T}_h}\Big\{\frac{\varepsilon}{2}\|\nabla_x(f-\Pi_h f)\|_{L^2(E)}^2\Big\}\lesssim\sum_{E\in\mathcal{T}_h}\sum_{j=2}^{d}\sum_{\mathbf{a}\in A}\frac{\varepsilon}{2}\mathbf{h}_E^{2(\mathbf{a}-\mathbf{e}_j)}\|f\|_{H^2(E)}^2,$$

$$\sum_{E\in\mathcal{T}_h}\Big\{\frac{\tau^{\lambda}}{2}\|\boldsymbol{\beta}_{Q_T}\cdot\nabla(f-\Pi_h f)\|_{L^2(E)}^2\Big\}\lesssim\sum_{E\in\mathcal{T}_h}\sum_{j=1}^{d}\sum_{\mathbf{a}\in A}\frac{\tau^{\lambda}}{2}\mathbf{h}_E^{2(\mathbf{a}-\mathbf{e}_j)}\|f\|_{H^2(E)}^2.$$

Inserting the previous estimates in (3.64) we derive (3.63). □

Example 1 (isotropic-uniform mesh). Assume $Q_T=[0,L]\times[0,L]\subset\mathbb{R}^{d=2}$, and assume a $\mathcal{T}_h$ with uniform uni-directional grid sizes, that means $h_t\sim h_x\sim h$. Consider a function $f\in H^2(Q_T)$ and the set of multi-indices $A=\{\mathbf{a}=(a_1,a_2)\colon \mathbf{a}=2e_i, i=1,2\}$. Then by (3.63) we can conclude that Π_h satisfies on isotropic meshes the following estimate

$$\|f-\Pi_h f\|_{h,s}^2\lesssim\Big\{h^2+\tau^{-\lambda}h^4+\tau^{\lambda}\varepsilon^2+\frac{\varepsilon}{2}h^2+\frac{\tau^{\lambda}}{2}h^2\Big\}\sum_{E\in\mathcal{T}_h}\|f\|_{H^2(E)}^2. \tag{3.65}$$

Example 2 (independent directional mesh sizes for $Q_T\subset\mathbb{R}^{d=3}$). Assume that $Q_T=[0,L]^3\subset\mathbb{R}^{d=3}$, and for the uni-directional grid sizes assume $\mathbf{h}_E:=\mathbf{h}=(h_t:=h_1,h_2:=h_x,h_3:=h_y)$. Consider a function $f\in H^2(Q_T)$ and the set of multi-indices $A=\{\mathbf{a}=(a_1,a_2,a_3)\colon \mathbf{a}=2e_i, i=1,2,3\}$. Then by the estimate given in (3.63) we can have

$$
\begin{aligned}
\|f-\Pi_h f\|_{h,s}^2 \lesssim \sum_{\mathbf{a}\in A}\Big\{ & \Big(h_1^{2a_1-1}h_2^{2a_2}h_3^{2a_3}+h_1^{a_1}h_2^{a_2}h_3^{a_3}+h_1^{a_1-1}h_2^{a_2+1}h_3^{a_3}+h_1^{a_1-1}h_2^{a_2}h_3^{a_3+1}\Big) \\
& +\tau^{-\lambda}h_1^{2a_1}h_2^{2a_2}h_3^{2a_3}+\tau^{\lambda}\varepsilon^2\Big(h_1^{2a_1}h_2^{2a_2-4}h_3^{2a_3}+h_1^{2a_1}h_2^{2a_2}h_3^{2a_3-4}\Big) \\
& +\frac{\varepsilon}{2}\Big(h_1^{2a_1}h_2^{2a_2-2}h_3^{2a_3}+h_1^{2a_1}h_2^{2a_2}h_3^{2a_3-2}\Big) \\
& +\frac{\tau^{\lambda}}{2}\Big(h_1^{2a_1-2}h_2^{2a_2}h_3^{2a_3}+h_1^{2a_1}h_2^{2a_2-2}h_3^{2a_3}+h_1^{2a_1}h_2^{2a_2}h_3^{2a_3-2}\Big)\Big\}\sum_{E\in\mathcal{T}_h}\|f\|_{H^2(E)}^2 \\
\lesssim \big\{ & h_1^3+h_1^{-1}\big(h_2^4+h_3^4\big)+h_1^2+h_2^2+h_3^2+h_1h_2+h_1^{-1}h_2^3+h_1^{-1}h_2h_3^2+h_1h_3+h_1^{-1}h_2^2h_3+h_1^{-1}h_3^3 \\
& +\tau^{-\lambda}\big(h_1^4+h_2^4+h_3^4\big)+\tau^{\lambda}\varepsilon^2\big(2+h_1^4h_2^{-4}+h_2^{-4}h_3^4+h_1^4h_3^{-4}+h_2^4h_3^{-4}\big) \\
& +\frac{\varepsilon}{2}\big(h_1^4h_2^{-2}+h_2^2+h_2^{-2}h_3^4+h_1^4h_3^{-2}+h_2^4h_3^{-2}+h_3^2\big) \\
& +\frac{\tau^{\lambda}}{2}\big(h_1^2+h_1^{-2}h_2^4+h_1^{-2}h_3^4+h_1^4h_2^{-2}+h_2^2+h_2^{-2}h_3^4+h_1^4h_3^{-2}+h_2^4h_3^{-2}+h_3^2\big)\Big\}\sum_{E\in\mathcal{T}_h}\|f\|_{H^2(E)}^2.
\end{aligned}
\tag{3.66}
$$

Theorem 3.2. *Let the solutions u and u_h satisfy the assumptions in Lemma 3.2 and let the Π_h satisfy the assumptions in Lemma 3.3. Then the following error convergence result holds*

$$
\begin{aligned}
\|u-u_h\|_h^2 \lesssim \sum_{E\in\mathcal{T}_h}\Bigg\{ & \Bigg(\sum_{\mathbf{a}\in A}\mathbf{h}_E^{2\mathbf{a}-\mathbf{e}_1}+\sum_{i=1}^{d}\sum_{\mathbf{a}\in A}\mathbf{h}_E^{\mathbf{a}+\mathbf{e}_i-\mathbf{e}_1}\Bigg)+\sum_{\mathbf{a}\in A}\tau^{-\lambda}\mathbf{h}_E^{2\mathbf{a}} \\
& +\sum_{j=2}^{d}\sum_{\mathbf{a}\in A}\Big\{\tau^{\lambda}\varepsilon^2\mathbf{h}_E^{2(\mathbf{a}-2\mathbf{e}_j)}+\frac{\varepsilon}{2}\mathbf{h}_E^{2(\mathbf{a}-\mathbf{e}_j)}\Big\}+\sum_{i=1}^{d}\sum_{\mathbf{a}\in A}\frac{\tau^{\lambda}}{2}\mathbf{h}_E^{2(\mathbf{a}-\mathbf{e}_i)}\Bigg\}\|u\|_{H^2(E)}^2.
\end{aligned}
\tag{3.67}
$$

Proof. We combine Lemmas 3.2 and 3.5 and the assertion follows. □

Corollary 3.4. *Let $\mathcal{T}_h$ be a uniform mesh partition, i.e., $h_1\sim\cdots\sim h_{d_x+1}\sim h$. Let further that $u\in H^{\ell}(Q_T)$ with $\ell\geqslant 2$ and $u_h\in V_h^k$, $k=1$. Under the assumptions of Theorem 3.2 holds*

$$
\|u-u_h\|_h \lesssim\ c\ h^{\ell-1}\|u\|_{H^{\ell}(Q_T)} \tag{3.68}
$$

with $c>0$ depending on the problem data.

Proof. The convergence rate (3.68) follows immediately by applying the mesh uniformity properties in (3.63) and in (3.67). □

4 Numerical examples

In order to validate the estimates derived in the previous sections, a series of numerical tests are presented below choosing different values for the parameters of the problem. For the two-dimensional problems we set $\tau^{\lambda}=\vartheta h_x^{\lambda}$ with $\vartheta=0.5$, $\lambda=2$, and the mesh sizes h_x, h_t are chosen independently depending on the purpose of the numerical computation. For the three dimensional problems we set $\tau^{\lambda}=\vartheta h_{E,\min}^2$. First, we start by considering the problem on a space time cylinder $Q_T\subset\mathbb{R}^2$ with a smooth solution and then with a less regular solution. Note that in this case $\boldsymbol{\beta}:=(1,\beta_x)$. Thereafter, we present computations considering the problem on $Q_T\subset\mathbb{R}^3$. For all

the test cases, we mainly use $\mathbb{Q}^k$ with $k = 1$, local polynomial spaces, and in a few examples we provide results using $k = 2$ local polynomial spaces, see (3.12).

The examples have been solved on a series of mesh refinement levels, $s = 0, 1, 2, \ldots$ with $h_s, h_{s+1}, \ldots$, where the asymptotic convergence behavior of the error $\|u - u_h\|_h$ is investigated. For the first problem the behavior of the error $\|u - u_h\|_{t,h} := (\|u - u_h\|^2_{L^2} + \tau^\lambda \|\partial_t u - \partial_t u_h\|^2_{L^2})^{1/2}$ is also studied. Note that a separation of the magnitudes of the two errors is not completely supported by the discretization analysis. We just provide this numerical test which investigates if $\|u - u_h\|_{t,h}$ and $\|u - u_s\|_s$ follow the same asymptotic behavior. Also, we mention that the "expected values" of the rates, which are written in the tables, are the values resulting from the discretization error analysis. In the numerical computations the initial condition u_0 and the boundary data u_Σ are determined by the L^2-projection of the exact solution u onto polynomial space.

The linear system produced by the method (3.18a) is (in general) not symmetric and a direct LU method has been used to solve it. One can also apply GMRES iterative solvers for its solution. However, several efficient methods have been presented in the literature for speeding up the solution process of the system. We refer to refs. [19], [21], [22], [37] for discussions on developing algebraic multigrid methods and on different parallelization approaches for STFEMs.

The 2-d numerical examples have been performed using an in-house code which has been implemented on a Intel(R) Core(TM) i7-8700 CPU with Gentoo Linux optimized system.

The conclusion from the results presented below is that the proposed space-time FE scheme is stable, behaves well and the numerical convergence rates are in agreement with the theoretically predicted rates.

4.1 Two-dimensional space-time cylinders

4.1.1 Smooth solution, uniform meshes

In the first numerical example the domain is $\bar{Q}_T = [0,1] \times [0,1]$ and the exact solution is $u(t,x) = \sin(2\pi t)\sin(2\pi x)$. The problem has been solved for the problem parameters $\{\varepsilon, \beta_x, r\} = \{0.1, 1, 1\}$. In Table 1, the results of the asymptotic convergence rates are displayed. The exact solution is smooth and optimal convergence rates are expected. We observe that the rates r_t and r_x have similar behavior and are a little higher than the expected rates in the first coarse meshes. As the meshes are progressively refined, both r_t and r_x tend to the expected values. In the last columns we can see the rates $r_{x,k=2}$, which correspond to $\mathbb{Q}^{k=2}$ polynomial spaces. Moving to the finer meshes the rates have the expected values.

Table 1: Example 1: smooth test case. The convergence rates r_x and r_t and $r_{x,k=2}$.

u smooth, with $\{\varepsilon, \beta_x, r\} = \{0.1, 1, 1\}$			
Errors $\mathbb{Q}^k$**-space**	$\|u - u_h\|_s$ $k = 1$	$\|u - u_h\|_{t,h}$ $k = 1$	$\|u - u_h\|_s$ $k = 2$
Expected rates	1	1	2
$h_0 = 0.2$	Computed rates		
$h_s = \frac{h_0}{2^s}$	r_x	r_t	$r_{x,k=2}$
$s = 1$	3.20	2.84	4.2
$s = 2$	1.40	1.23	2.4
$s = 3$	1.40	1.08	2.14
$s = 4$	1.14	1.08	2.15
$s = 5$	1.17	1.02	2.01
$s = 6$	1.05	1.01	2.00
$s = 7$	1.00	1.04	2.02

4.1.2 Point singularity test case

We consider the problem on $\bar{Q}_T = [0,1] \times [0,1]$ with exact solution $u(x,t) = ((x-0.2)^2 + (t-0.2)^2)^{\gamma/2}$ with $\gamma \in \{1.001, 0.51\}$. The singular point of the solution is located at the interior of the domain $\bar{Q}_T$. The problem has been solved for two sets of values for the parameters, i.e., $\{\varepsilon, \beta_x, r\} = \{0.1, 1, 1\}$ and $\{\varepsilon, \beta_x, r\} = \{0.1, 10^{-10}, 0.5\}$. In every case the associated solution u belongs to $H^{\ell=\gamma+1}(Q_T)$ and the computational tests have been performed using $\mathbb{Q}^{k=1}$ and $\mathbb{Q}^{k=2}$ local polynomial spaces on uniform mesh partitions $\mathcal{T}_h$. In view of the interpolation results (3.56) and (3.63), see also Corollary 3.4, and the regularity of the solution u, the convergence rates expected to get values close to $\gamma = \ell - 1$. We compute the rates on a sequence of meshes and present the results in Table 2. Note that the cases, where we use $\gamma + 1 < 2$ and $\mathbb{Q}^{k=2}$ spaces, are not covered by the discretization analysis presented in the previous section. The associated results are added in Table 2 for comparison purposes only and for a more comprehensive computational study of the problem. The "expected values" were formed by a simple formalistic interpretation of the interpolation estimates.

Looking at the table, we observe that for the tests with $\gamma = 1.001$, $\beta_x = 0.5$, see first columns, the rates r_x and $r_{x,k=2}$ have similar behavior and are close to the expected values, even for the first refinement steps. For the second test where we set $\beta = 10^{-10}$, the values of $r_{x,k=2}$ are little higher during the first meshes but tend to get the expected values during the last meshes. In the last test we set $\gamma = 0.6$, see last column, the r_x rates are higher than the expected values on the first meshes. Moving to the last refinement steps the rates tend to get the expected values.

4.1.3 Anisotropic meshes

In this test, we investigate the behavior of the on anisotropic meshes. To illustrate this consider a problem where its solution has anisotropic behavior, i.e., its variations are quite different in each different axial direction. We solve the problem using anisotropic meshes, where the directional size of the mesh elements is small in the direction where the solution has strong variations, and the size of the elements is relatively larger in the direction where the solution has less variations. The domain is chosen to be $\bar{Q}_T = [0,1] \times [0,1]$ and the parameters take the values $\{\varepsilon, \beta_x, r\} = \{0.1, 1, 1\}$. The exact solution is $u(x,t) = \sin(\pi x)\sin(4\pi t)$ and varies 'four times more' in the direction of the time axis. The problem is solved in a series of successive refined mesh levels $s = 1, 2, \ldots$, where for each refinement mesh level we set $h_{s,x} = \lambda h_{s,t}$, here the parameter $\lambda \in \{1, 2, 4\}$ forms the anisotropic

Table 2: Example 2: point singularity case. Convergence rates r_x and $r_{x,k=2}$.

$u \in W^{\ell,2}(Q_T)$ with $\ell = \gamma + 1$					
Errors $\mathbb{Q}^k$**-space**	$\|u-u_h\|_s$ $k=1$	$\|u-u_h\|_s$ $k=2$	$\|u-u_h\|_s$ $k=1$	$\|u-u_h\|_s$ $k=2$	$\|u-u_h\|_s$ $k=1$
Parameters	$\gamma = 1.001$		$\gamma = 1.001$		$\gamma = 0.6$
	$\{\varepsilon, \beta_x, r\} = \{0.1, 0.5, 1\}$		$\{\varepsilon, \beta_x, r\} = \{0.1, 10^{-10}, 0.5\}$		$\{\varepsilon, \beta_x, r\} = \{0.1, 0.5, 0.5\}$
Expected rates	1	1	1	1	0.5
$h_0 = 0.2$	Computed rates				
$h_s = \frac{h_0}{2^s}$	r_x	$r_{x,k=2}$	r_x	$r_{x,k=2}$	r_x
$s=1$	0.5	0.98	0.5	1.5	0.44
$s=2$	0.95	1.04	0.92	1.82	0.63
$s=3$	0.90	0.98	0.90	1.45	0.57
$s=4$	0.89	0.94	0.89	1.34	0.55
$s=5$	0.89	0.91	0.89	0.87	0.54
$s=6$	0.91	0.90	0.90	1.25	0.52
$s=7$	0.92	0.89	0.92	1.24	0.51

character of the mesh. For each mesh level s, three different computations have been performed, where each of which corresponds to a different value of the parameter λ. We note that, for a better comparison of the results, the three refinement procedures start from a different initial mesh.

Our goal is to investigate the asymptotic behavior of the rates r_x, how they are affected by the anisotropic variational properties of u and the anisotropic character of the mesh. Also we investigate the computing resources needed to reduce the error to the same levels as those which result when isotropic meshes are used. Table 3 shows the results of the numerical convergence rates. For each λ-case, we can observe that the values of r_x rates are a little higher during the first meshes than the theoretically predicted values. However, moving to the last refinement steps, the rates reduce and get the expected values derived by the error analysis, see also Example 2.

The left part of Figure 1 shows the reduction of the error of the numerical solution relative to the number of elements for each anisotropic mesh. Starting with a mesh with $h_x = h_t$, we move to next refinement steps, $s = 2, 3, \ldots$, such that the directional mesh sizes have the fixed relation $h_{s,x} = \lambda h_{s,t}$, $\lambda = 1, 2, 4$. During the first meshes, the values of the corresponding errors are very close for the same number of elements. However, looking

Table 3: Example 3: anisotropic meshes. The convergence rates r_x for all λ.

$u(x,t)$ smooth, $\{\varepsilon, \beta_x, r\} = \{0.1, 1, 1\}$			
Errors	$\|u - u_h\|_s$	$\|u - u_h\|_s$	$\|u - u_h\|_s$
Parameter	$\lambda = 1$	$\lambda = 2$	$\lambda = 4$
Expected rates	1	1	1
Ref. step s	**Computed rates**		
	r_x	r_x	r_x
$s = 1$	0.5	2.01	1.8
$s = 2$	2.3	1.87	1.62
$s = 3$	1.7	1.61	1.85
$s = 4$	1.31	1.27	1.33
$s = 5$	1.11	1.09	1.11
$s = 6$	1.03	1.02	1.03
$s = 7$	1.01	1.00	1.2
$s = 8$	1.02	1.00	1.00
$s = 9$	1.01	1.02	1.02
$s = 10$	1.00	1.00	1.01

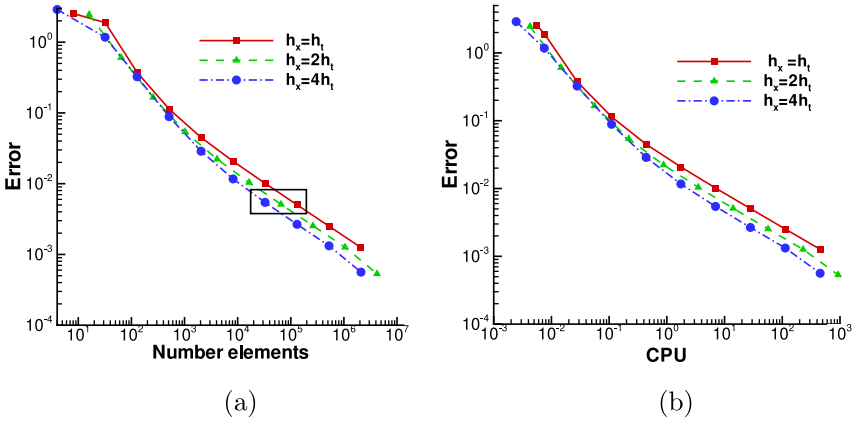

Figure 1: Example 3: anisotropic meshes, (a) the graph on the left shows the reduction of the error against the number of elements for each anisotropic mesh. (b) The discretization error against the CPU time for each λ test case.

the values in the next refinement levels, we observe that the $\lambda = 4$, $\lambda = 2$ meshes require a significantly reduced number of elements compared to the $\lambda = 1$ case, in order to reduce the error to the same level, (see the micro box in the graph). Thus, to reduce the error to a desired level, a uniform refinement with $h_{s,x} \approx h_{s,t}$ (isotropic) is not necessary. This also illustrates the usefulness of the anisotropic meshes when solving such problems. This can be seen more clearly in the right graph in Figure 1, where the error reduction against CPU time is shown. Note that in every case, the corresponding CPU time considers the whole performance-solution of the numerical example and not only the solution of the linear system. As expected the performance of the test with $\lambda = 4$ requires the least computing resources, when we refer to the same mesh refinement step, i.e., the same $h_{s,t}$. It is observed that the error, which corresponds to the numerical solution of $\lambda = 4$ case, decreases at the same level and at the same rate with what corresponds to the isotropic case, i.e., the $\lambda = 1$ case. Figure 1 somehow indicates that an appropriate use of anisotropic meshes can lead to high resolution numerical solutions at low computational cost.

4.2 Examples on three-dimensional space-time cylinders

4.2.1 Smooth solution, isotropic meshes

The purpose of this example is to investigate the convergence behavior of the discretization error $\|u - u_h\|_s$ for the case of having a three-dimensional space-time cylinder. We consider the problem on $\bar{Q}_T = [0,1]^3$ with exact solution $u(x,y,t) = \cos(2\pi x)\cos(2\pi y)\cos(2\pi t)$. We solve the problem using $k = 1$ and $k = 2$ polynomial spaces for two groups of parameter values, the first is $\{\varepsilon, \beta_x, \beta_y, r\} = \{0.1, 1, 1, 1\}$ and the second $\{\varepsilon, \beta_x, \beta_y, r\} = \{0.1, 0, 0, 1\}$. We compare the numerical results with the theoretical findings given in Theorem 3.2, see also (3.66). The numerical results are reported in Table 4. For each $k = 1$-case, we observe that the convergence rates r_x are close to the expected rate value equal to one. The rates $r_{x,k=2}$ are shown in third and the fifth column for the two parameter test cases. The rates are a little lower in the first meshes but moving to the last meshes they have the expected value (with respect to the regularity of the solution) and follow the theoretical convergence rates, compare to the "Smooth solution, uniform meshes" above.

Table 4: Example 4: smooth solution for $Q_T \subset \mathbb{R}^3$. Convergence rates r_x and $r_{x,k=2}$.

	u smooth, $Q_T \subset \mathbb{R}^3$			
Errors	$\|u - u_h\|_s$	$\|u - u_h\|_s$	$\|u - u_h\|_s$	$\|u - u_h\|_s$
Parameters	$\{\varepsilon, \beta_x, \beta_y, r\} = \{0.1, 1, 1, 1\}$		$\{\varepsilon, \beta_x, \beta_y, r\} = \{0.1, 0, 0, 1\}$	
Expected rates	1	2	1	2
$\mathbb{Q}^k$	**Computed rates**			
	$k = 1$	$k = 2$	$k = 1$	$k = 2$
$h_s = \frac{1}{2^s}$	r_x	$r_{x,k=2}$	r_x	$r_{x,k=2}$
$s = 1$	0.554	1.825	0.490	1.873
$s = 2$	1.055	1.985	1.026	1.952
$s = 3$	1.08	1.984	1.049	1.967
$s = 4$	1.058	1.987	1.037	1.976
$s = 5$	1.041	1.989	1.027	1.983
$s = 6$	1.030	1.992	1.020	1.987
$s = 7$	1.023	1.993	1.016	1.990

4.2.2 An example using anisotropic meshes

As it has been illustrated by the examples above that an appropriate construction and use of anisotropic meshes can help on relaxing the high time requirements for the computation of the solution. Also, we have seen the use of the anisotropic meshes preserves the convergence properties of the numerical solution. Similar points are going to be investigated in this numerical example. The problem is considered in $Q_T = [0,1]^3$ with exact solution $u(t,x,y) = \sin(2\pi t)\mathrm{e}^{-(x-0.5)^2/0.01}y$ and the parameters are $\{\epsilon, \beta_x, \beta_y, r\} = \{0.1, 1, 1, 1\}$. The solution has different variational behavior in each direction and this supports the use of anisotropic meshes. The problem is solved using $k = 1$-local spaces applying an anisotropic mesh refinement strategy, which is specified by the relations $h_t = \varkappa h_y, \varkappa \in \{1, 1/2\}$ and $h_x = \lambda h_y, \lambda \in \{1, 1/4\}$. The first starting mesh contains equal directional mesh sizes, i.e., $h_y = h_x = h_t$, and for the next refinement levels the mesh sizes h_x and h_t in x and t directions respectively are determined by the values of $\varkappa$ and λ. The associated convergence rates r_x of the error are shown in Table 5. The first column includes the rates which are computed by applying global uniform refinement (isotropic case). The next two columns include the rates for the anisotropic meshes. It can be seen that the rates are very close to the (theoretical) predicted values for all cases. Figure 2(a) illustrates the error variation with respect to the number

Table 5: Example 5: anisotropic meshes $Q_T \subset \mathbb{R}^3$. Convergence rates.

$u = \sin(2\pi t)\mathrm{e}^{-(x-0.5)^2/0.01}y, \{\epsilon, \beta_x, \beta_y, r\} = \{0.1, 1, 1, 1\}$			
Errors	$\|u - u_h\|_s$		
Parameters	$\lambda = 1, \varkappa = 1$	$\lambda = 1/4, \varkappa = 1$	$\lambda = 1/4, \varkappa = 1/2$
Expected rates	1	1	1
Ref. step s	**Computed rates**		
	r_x	r_x	r_x
$s = 1$	0.927	0.917	0.943
$s = 2$	0.958	0.957	0.973
$s = 3$	0.968	0.967	0.983
$s = 4$	0.975	0.982	0.988
$s = 5$	0.979	0.986	0.991
$s = 6$	0.982	0.988	0.993
$s = 7$	0.984	0.990	0.994
$s = 8$	0.986	0.991	0.995
$s = 9$	0.988	0.992	0.996

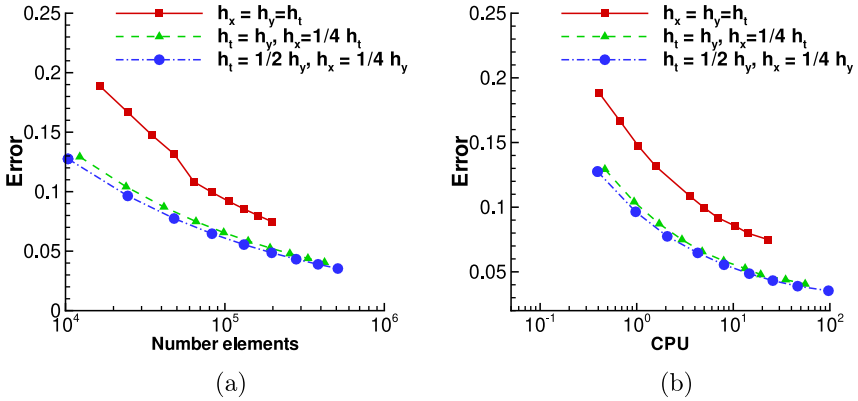

Figure 2: Example 5: anisotropic meshes for $Q_T \subset \mathbb{R}^3$. (a) Left: the variation of the error with respect the number of elements for each mesh, (b) right: the discretization error against the CPU time for each mesh test case.

of mesh elements for the three different meshes. It can be seen that the errors corresponding to the anisotropic meshes have lower values compared to the errors computed using the isotropic meshes (with the same number of elements). Hence, it appears that the anisotropic mesh computations manage more appropriately the number of elements, in the sense that more elements are stacked in the direction where the solution has the most abrupt variation. On the right graph in Figure 2(b) the CPU times are plotted against the error reduction. We observe that the anisotropic meshes bring significant improvements in the time performance of the method. The anisotropic meshes require less computational resources than the isotropic meshes to keep the error at the same low level.

5 Conclusions

In this work space-time FE methods have been developed and analyzed with continuous spaces on anisotropic quadrilateral meshes for solving general linear parabolic problems. The scheme was stabilized following usual upwind streamline methodology. Discretization error estimates were shown in a suitable norm. The proposed method was applied to problems having regular and less regular solutions on isotropic and anisotropic meshes. The numerical convergence rates were in agreement with the theoretical rates.

The proposed scheme can be extended to the case of using discontinuous Galerkin discretizations in time. This can help on solving the problem in a sequential manner, i.e., one space-time slice at a time, [38]. This approach can be further combined with time-Domain Decomposition (DD) iterative solvers materialized in a parallel environment. A incorporation of anisotropic refinement strategies to the proposed method can lead to en efficient method for solving problems with solutions with anisotropic behavior, e.g., boundary layers in fluid problems, re-entrant edges, etc. The development of this type of numerical methods is the subject of a work in progress.

Acknowledgments: The author would like to thank Oliver Koch from the Institute of Computational Mathematics of Johannes Kepler University Linz for his assistance in computing the numerical results.
Research ethics: Not applicable.
Informed consent: Not applicable.
Author contributions: The author has accepted responsibility for the entire content of this manuscript and approved its submission.
Use of Large Language Models, AI and Machine Learning Tools: None declared.
Conflict of interest: The author states no conflict of interest.
Research funding: None declared.
Data availability: Not applicable.

References

[1] L. C. Evans, *Partial Differential Equations*, ser. Graduate Studies in Mathematics, vol. 19, 2nd ed. USA, American Mathematical Society, 2010.
[2] H.-G. Roos, M. Stynes, and L. Tobiska, *Robust Numerical Methods for Singularly Perturbed Differential Equations. Convection—Diffusion—Reaction and Flow Problems*, ser. Springer Series in Computational Mathematics, vol. 24, Berlin, Heidelberg, Springer, 2008.
[3] L. P. Franca, G. Hauke, and A. Masud, "Revisiting stabilized finite element methods for the advective—diffusive equation," *Comput. Methods Appl. Mech. Eng.*, vol. 195, nos. 13—16, pp. 1560—1572, 2006.
[4] A. N. Brooks and T. J. Hughes, "Streamline upwind/Petrov—Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier—Stokes equations," *Comput. Methods Appl. Mech. Eng.*, vol. 32, no. 1, pp. 199—259, 1982.
[5] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson, *Computational Differential Equations*, Cambridge, U.K., Cambridge University Press, 1996.
[6] M. Augustin, *et al*., "An assessment of discretizations for convection-dominated convection—diffusion equations," *Comput. Methods Appl. Mech. Eng.*, vol. 200, no. 47, pp. 3395—3409, 2011.

[7] V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, ser. Springer Series in Computational Mathematics, 2nd ed. Berlin, Heidelberg, Springer-Verlag, vol. 25, 2006.
[8] W. Huang, L. Kamenski, and J. Lang, "Stability of explicit Runge—Kutta methods for high order finite element approximation of linear parabolic equations," *Numer. Math. Adv. Appl.*, vol. 103, pp. 165—173, 2015.
[9] E. Burman and A. Ern, "Implicit-explicit Runge—Kutta schemes and finite elements with symmetric stabilization for advection-diffusion equations," *ESAIM Math. Model. Numer. Anal.*, vol. 46, no. 4, pp. 681—707, 2012.
[10] U. Langer and O. Steinbach, *Space-Time Methods: Applications to Partial Differential Equations*, ser. Radon Series on Computational and Applied Mathematics, vol. 25, U. Langer and O. Steinbach, Eds., Berlin, Walter de Gruyter GmbH and Co KG, 2019.
[11] R. E. Bank, P. S. Vassilevski, and L. T. Zikatanov, "Arbitrary dimension convection—diffusion schemes for space-time discretizations," *J. Comput. Appl. Math.*, vol. 310, pp. 19—31, 2017.
[12] S. Larsson and M. Molteni, "Numerical solution of parabolic problems based on a weak space-time formulation," *Comput. Methods Appl. Math.*, vol. 17, no. 1, pp. 65—84, 2016.
[13] C. Mollet, "Stability of Petrov—Galerkin discretizations: application to the space-time weak formulation for parabolic evolution problems," *Comput. Methods Appl. Math.*, vol. 14, no. 2, pp. 231—255, 2014.
[14] C. Schwab and R. Stevenson, "Space-time adaptive wavelet methods for parabolic evolution problems," *Math. Comput.*, vol. 78, no. 267, pp. 1293—1318, 2009.
[15] R. Stevenson and J. Westerdiep, "Stability of Galerkin discretizations of a mixed space-time variational formulation of parabolic evolution equations," *IMA J. Numer. Anal.*, vol. 41, no. 1, pp. 28—47, 2020.
[16] I. Toulopoulos, "Space-time finite element methods stabilized using bubble function spaces," *Appl. Anal.*, vol. 99, no. 7, pp. 1153—1170, 2018.
[17] I. Toulopoulos, "Numerical solutions of quasilinear parabolic problems by a continuous space-time finite element scheme," *SIAM J. Sci. Comput.*, vol. 44, no. 5, pp. A2944—A2973, 2022.
[18] A. Mantzaflaris, F. Scholz, and I. Toulopoulos, "Low-rank space-time decoupled isogeometric analysis for parabolic problems with varying coefficients," *Comput. Methods Appl. Math.*, vol. 19, no. 1, pp. 123—136, 2019.
[19] C. Hofer, U. Langer, M. Neumüller, and I. Toulopoulos, "Time-multipatch discontinuous Galerkin space-time isogeometric analysis of parabolic evolution problems," *Electron. Trans. Numer. Anal.*, vol. 49, pp. 126—150, 2018.
[20] T. E. Tezduyar and K. Takizawa, "Space-time computations in practical engineering applications: a summary of the 25-year history," *Comput. Mech.*, vol. 63, no. 4, pp. 747—753, 2019.
[21] R. Dyja, B. Ganapathysubramanian, and K. G. van der Zee, "Parallel-in-space-time, adaptive finite element framework for nonlinear parabolic equations," *SIAM J. Sci. Comput.*, vol. 40, no. 3, pp. C283—C304, 2018.
[22] M. Neumüller and I. Smears, "Time-parallel iterative solvers for parabolic evolution equations," *SIAM J. Sci. Comput.*, vol. 41, no. 1, pp. C28—C51, 2018.
[23] C. V. Frontin, G. S. Walters, F. D. Witherden, C. W. Lee, D. M. Williams, and D. L. Darmofal, "Foundations of space-time finite element methods: polytopes, interpolation, and integration," *Appl. Numer. Math.*, vol. 166, pp. 92—113, 2021.
[24] T. G. Lube, "Anisotropic mesh refinement in stabilized Galerkin methods," *Numer. Math.*, vol. 74, no. 3, pp. 261—282, 1996.
[25] T. Apel, "Interpolation of non-smooth functions on anisotropic finite element meshes," *Math. Model. Numer. Anal.*, vol. 33, no. 6, pp. 1149—1185, 1999. https://doi.org/10.1051/m2an:1999139.
[26] S. Micheletti, S. Perotto, and M. Picasso, "Stabilized finite elements on anisotropic meshes: a priori error estimates for the advection-diffusion and the Stokes problems," *SIAM J. Numer. Anal.*, vol. 41, no. 3, pp. 1131—1162, 2003.
[27] L. Formaggia, S. Micheletti, and S. Perotto, "Anisotropic mesh adaptation in computational fluid dynamics: application to the advection—diffusion—reaction and the Stokes problems," *Appl. Numer. Math.*, vol. 51, no. 4, pp. 511—533, 2004.
[28] R. A. Adams and J. J. F. Fournier, *Sobolev Spaces*, ser. Pure and Applied Mathematics, 2nd ed. Oxford U.K., Academic Press-Imprint Elsevier Science, vol. 140, 2003.
[29] S. Brenner and L. Scott, *The Mathematical Theory of Finite Element Methods*, ser. Texts in Applied Mathematics, 2nd ed. New-York, Springer, 2008.
[30] O. Ladyzhenskaya and N. Uraltseva, *Linear and Quasilinear Elliptic Equations*, New York, London, Academic Press, 1968.
[31] A. Koshelev, "*Regularity Problem for Quasilinear Elliptic and Parabolic Systems*, ser. Lecture Notes in Mathematics, Berlin, Heidelberg, Springer-Verlag, 1995, p. 1614.
[32] R. Adreev, "Stability of sparse space-time finite element discretizations of linear parabolic evolution equations," *IMA J. Numer. Anal.*, vol. 33, no. 1, pp. 242—260, 2013.
[33] A. Ern and J. L. Guermond, *One-Dimensional Finite Elements and Tensorization*, ser. Texts in Applied Mathematics, Cham, Switzerland, Springer International Publishing, vol. 72, 2021, no. 1614.
[34] W. Hackbusch and B. N. Khoromskij, "Tensor-product approximation to operators and functions in high dimensions," *J. Complex.*, vol. 23, no. 4, pp. 697—714, 2007.
[35] L. L. Schumaker, *Spline Functions: Basic Theory*, 3rd ed. Cambridge, University Press, 2007.
[36] T. Dupont and R. Scott, "Polynomial approximation of functions in Sobolev spaces," *Math. Comput.*, vol. 34, no. 150, pp. 441—463, 1980.

[37] O. Steinbach and H. Yang, “Comparison of algebraic multigrid methods for an adaptive space-time finite-element discretization of the heat equation in 3D and 4D,” *Numer. Lin. Algebra Appl.*, vol. 25, no. 3, 2018, https://doi.org/10.1002/nla.2143.

[38] I. Toulopoulos, “A unified time discontinuous Galerkin space-time finite element scheme for non-Newtonian power law models,” *Int. J. Numer. Methods Fluid.*, vol. 95, no. 5, pp. 851—868, 2023.