SPACE-TIME VIRTUAL ELEMENTS FOR THE HEAT EQUATION*

SERGIO GOMEZ†, LORENZO MASCOTTO‡, ANDREA MOIOLA§, AND
ILARIA PERUGIA¶

Abstract. We propose and analyze a space-time virtual element method for the discretization of the heat equation in a space-time cylinder, based on a standard Petrov–Galerkin formulation. Local discrete functions are solutions to a heat equation problem with polynomial data. Global virtual element spaces are nonconforming in space, so that the analysis and the design of the method are independent of the spatial dimension. The information between time slabs is transmitted by means of upwind terms involving polynomial projections of the discrete functions. We prove well posedness and optimal error estimates for the scheme, and validate them with several numerical tests.

1. Introduction. The virtual element method (VEM) was introduced in [3] as an extension of the finite element method to general polytopic meshes for the approximation of solutions to the Poisson equation. Trial and test spaces consist of functions that are solutions to local problems related to the PDE problem to be approximated. Moreover, they typically contain polynomials of a given maximum degree, together with nonpolynomial functions allowing for the enforcement of the desired type of conformity in the global spaces. These functions are not required to be explicitly known. Suitable sets of degrees of freedom (DoFs) are chosen so that projections from local virtual element (VE) spaces onto polynomial spaces can be computed out of them. Such polynomial projectors and certain stabilizing bilinear forms are used to define the discrete bilinear forms. A nonconforming version of the VEM was proposed in [2]. Unlike its conforming counterpart, the nonconforming VEM can be presented in a unified framework for any dimension, which significantly simplifies its analysis and implementation.

In the VEM literature, time dependent problems have always been tackled by combining a VE discretization in space with a time-stepping scheme for the solution to the resulting ODE system. The prototypical example is [21], where the heat equation was considered. On the other hand, space–time Galerkin methods are based on

---

†Department of Mathematics and Applications, University of Milano Bicocca, 20125, Milan, Italy (sergio.gomezmacias@unimib.it).

‡Department of Mathematics and Applications, University of Milano Bicocca, 20125, Milan, Italy, and Faculty of Mathematics, University of Vienna, 1090 Vienna, Austria, and IMATI-CNR, 27100, Pavia, Italy (lorenzo.mascotto@unimib.it).

§Department of Mathematics, University of Pavia, 27100, Pavia, Italy (andrea.moiola@unipv.it).

¶Faculty of Mathematics, University of Vienna, 1090 Vienna, Austria (ilaria.perugia@univie.ac.at).

discretizing the space and time variables of a PDE at once. These methods provide a natural framework where high-order accuracy can be obtained in both space and time, and an approximate solution is available on the whole space–time domain.

In this paper, we design and analyze the first space–time VEM for the solution to a time-dependent PDE, namely, the heat equation; we can consider spatial domains in one, two, and three dimensions. We employ prismatic-type elements. This allows us to distinguish two types of mesh facets: *space-like* facets, i.e., facets lying on hyperplanes in space–time that are perpendicular to the time axis; *time-like* facets, i.e., facets whose normals are perpendicular to the time axis. The method we propose is based on a standard space–time variational formulation of the heat equation in the space–time cylinder $Q_T = \Omega \times (0, T)$ with trial space $L^2(0, T; H_0^1(\Omega)) \cap H^1(0, T; H^{-1}(\Omega))$ and test space $L^2(0, T; H_0^1(\Omega))$; see [7, Ch. XVIII, sect. 4.1].

For a recent survey of space–time discretizations of parabolic problems, we refer to [13]. In particular, a continuous finite element discretization of the standard Petrov–Galerkin variational formulation is presented and analyzed in [18]. Additionally, we refer to [1], [17], and [20] for wavelet- or finite element-type discretizations based on a minimal residual Petrov–Galerkin formulation, and to [15] and [6] for discontinuous Galerkin approaches. Motivated by the boundary integral operator analysis, a continuous finite element method based on a fractional-order-in-time variational formulation was studied in [19]. Recently, the space–time first order system least squares (FOSLS) formulation of [4] has been revisited and analyzed; see [8], [9], and [10]. We summarize the main features of the proposed VEM.

- Local VE spaces consist of functions that solve a heat equation with polynomial data on each space–time element; this makes the method particularly suitable for further extensions, e.g., to its Trefftz variant.
- We consider tensor-product in time (prismatic) meshes but the VE spaces are not of tensor-product type. Even for prismatic elements with simplicial bases, the proposed VE spaces do not coincide with their standard tensor-product finite element counterparts.
- Global VE spaces involve approximating continuity constraints across mesh facets. More precisely, we impose nonconformity conditions on time-like facets analogous to those in [2] for the Poisson problem, and allow for discontinuous functions in time. Across space-like facets, we transmit the information between consecutive time slabs by upwinding. In the present VEM context, the upwind terms are defined by means of a polynomial projection.
- To keep the presentation and the analysis of the method as simple as possible, the details are presented for the particular case of space–time tensor-product meshes. However, as discussed in subsection 2.6 below, the method can handle nonmatching time-like or space-like facets, which is desirable for space–time adaptivity.

We summarize the advantages of the proposed space–time VEM over standard space–time conforming finite element methods.

- The nonconforming VEM setting is of arbitrary order and its design is independent of the spatial dimension.
- Nonmatching space-like and time-like facets, which naturally stem from mesh adaptive procedures, can be handled easily.
- As the discrete spaces are discontinuous in time, we can solve the global (expensive) problem as a sequence of local (cheaper) problems on time slabs.
- The definition of the local spaces allows for the construction of space–time discrete Trefftz spaces.

The main advancements of this paper are the following.

- We design a novel space–time VEM for the heat equation in any spatial dimension.
- We prove its well posedness and optimal a priori error estimates.
- We validate numerically the theoretical results on some test cases.

Notation. We denote the first and second partial derivatives with respect to the time variable t by ∂_t and ∂_{tt}, respectively, and the spatial gradient and Laplacian operators by $\nabla_\mathbf{x}$, $\Delta_\mathbf{x}$, respectively. Throughout this paper, standard notation for Sobolev spaces will be employed. For a given bounded Lipschitz domain $D \subset \mathbb{R}^d$ ($d \in \mathbb{N}$), $H^s(D)$ represents the standard Sobolev space of order $s \in \mathbb{N}$, endowed with the standard inner product $(\cdot,\cdot)_{s,D}$, the seminorm $|\cdot|_{s,D}$, and the norm $\|\cdot\|_{s,D}$. In particular, $H^0(D) := L^2(D)$, where $L^2(D)$ is the space of Lebesgue square integrable functions over D and $H_0^1(D)$ is the closure of $C_0^\infty(D)$ in the $H^1(D)$ norm. Whenever s is a fractional or negative number, the Sobolev space $H^s(D)$ is defined by means of interpolation and duality. The Sobolev spaces on ∂D are defined analogously and denoted by $H^s(\partial D)$, $s < 1$.

As is common in space–time variational problems, we shall also use Bochner spaces of functions mapping a time interval (a,b) into a Banach space $(Z, \|\cdot\|_Z)$, which we denote by $H^s(a,b;Z)$, $s \in \mathbb{N}$.

Structure of this paper. In the remainder of this introduction, we introduce the model problem (subsection 1.1), and a regular sequence of meshes (subsection 1.2). The new space–time VEM method is presented in section 2. Section 3 is dedicated to the well-posedness of the method, while in section 4 we present an a priori error analysis and prove quasi-optimal estimates for the h-version of the method. We conclude this work with some numerical experiments in section 5 and some concluding remarks in section 6.

1.1. The model problem and its weak formulation. We are interested in the approximation of solutions to heat equation initial-boundary value problems on the space–time domain $Q_T := \Omega \times (0, T)$, where $\Omega \subset \mathbb{R}^d$ ($d = 1, 2, 3$) and $T > 0$ denote a bounded Lipschitz spatial domain and a final time, respectively.

Let $f : Q_T \to \mathbb{R}$ denote the prescribed right-hand side. We consider a positive constant volumetric heat capacity c_H and a positive constant scalar-valued thermal conductivity ν. The strong formulation of the initial-boundary value problem for the heat equation reads: Find a function $u : Q_T \to \mathbb{R}$ (temperature) such that

$$(1.1) \qquad \begin{cases} c_H \partial_t u - \nu \Delta_\mathbf{x} u = f & \text{in } Q_T; \\ u = 0 \quad \text{on } \Omega \times \{0\}; \qquad u = 0 \quad \text{on } \partial\Omega \times (0, T). \end{cases}$$

See Remark 1.2 below for more general initial and boundary conditions.

Introduce the function spaces
$$(1.2)$$
$$Y := L^2\left(0, T; H_0^1(\Omega)\right), \quad X := \left\{ v \in Y \cap H^1\left(0, T; H^{-1}(\Omega)\right) \mid v = 0 \text{ in } \Omega \times \{0\} \right\},$$

endowed with the norms

$$\|v\|_Y^2 := \left\| \nu^{1/2} \nabla_\mathbf{x} v \right\|_{0,Q_T}^2, \qquad \|v\|_X^2 := \|c_H \partial_t v\|_{L^2(0,T;H^{-1}(\Omega))}^2 + \|v\|_Y^2,$$

respectively. Here, we have used the following definition:

$$\text{for any } \phi \text{ in } L^2(0,T;H^{-1}(\Omega)), \qquad \|\phi\|_{L^2(0,T;H^{-1}(\Omega))} := \sup_{0 \neq v \in Y} \frac{\int_0^T \langle \phi, v \rangle \mathrm{dt}}{\|v\|_Y},$$

where $\langle \cdot, \cdot \rangle$ denotes the duality between $H_0^1(\Omega)$ and $H^{-1}(\Omega)$. Next, we define the space–time bilinear form $b(\cdot, \cdot) \colon X \times Y \to \mathbb{R}$ as

$$(1.3) \qquad b(u, v) := \int_0^T \left(\langle c_H \partial_t u, v \rangle + \int_\Omega \nu \nabla_\mathbf{x} u \cdot \nabla_\mathbf{x} v \ \mathrm{d}\mathbf{x} \right) \mathrm{dt}.$$

The weak formulation of (1.1), see, e.g., [7], reads as follows:

$$(1.4) \qquad \text{Find } u \in X \text{ such that } b(u, v) = \int_0^T \langle f, v \rangle \mathrm{dt} \qquad \forall v \in Y.$$

The following well-posedness result is valid; see e.g., [18, Cor. 2.3].

PROPOSITION 1.1. *If f belongs to $L^2(0, T; H^{-1}(\Omega))$, then the variational formulation (1.4) is well posed with the a priori bound*

$$\|u\|_X \leq 2\sqrt{2}\|f\|_{L^2(0, T; H^{-1}(\Omega))}.$$

Remark 1.2 (inhomogeneous initial and boundary conditions). Given (f, u_0) in $L^2(0, T; H^{-1}(\Omega)) \times L^2(\Omega)$, consider the following problem: find $u \in Y \cap H^1(0, T; H^{-1}(\Omega))$ such that
(1.5)
$$\begin{cases} \displaystyle\int_0^T \left(\langle c_H \partial_t u, v \rangle + \int_\Omega \nu \nabla_\mathbf{x} u \cdot \nabla_\mathbf{x} v \ \mathrm{d}\mathbf{x} \right) \mathrm{dt} = \int_0^T \langle f, v \rangle \mathrm{dt} & \forall v \in L^2(0, T; H^{-1}(\Omega)), \\ \displaystyle\int_\Omega u(\cdot, 0) w \, dx = \int_\Omega u_0 w \, dx & \forall w \in L^2(\Omega). \end{cases}$$

The well-posedness of problem (1.5) is discussed, e.g., in [17, sect. 5].

The case of inhomogeneous Dirichlet boundary conditions $u = g$ on $\partial\Omega \times (0, T)$ can be dealt with by assuming g in $H^1(0, T; H^{1/2}(\partial\Omega))$. Denote by $G \colon Q_T \to \mathbb{R}$ the solution to the family of elliptic problems $-\nu\Delta_\mathbf{x} G(\cdot, t) = 0$ in Ω with $G(\cdot, t) = g(\cdot, t)$ on $\partial\Omega$ for all $0 \leq t \leq T$. The function G belongs to $H^1(0, T; H^1(\Omega))$, since $\partial_t G$ solves a similar family of elliptic problems with boundary data $\partial_t g$ in $L^2(0, T; H^{1/2}(\partial\Omega))$.

For the case of inhomogeneous initial and boundary conditions, denote by w the solution to problem (1.5) with source term $f - c_H \partial_t G$ and initial condition $u_0 - g(\cdot, 0)$. Then, $u = G + w$ solves the inhomogeneous initial-boundary value problem with data (f, u_0, g). In particular, u belongs to $L^2(0, T; H^1(\Omega)) \cap H^1(0, T; H^{-1}(\Omega))$.

1.2. Mesh assumptions. For the sake of presentation, we stick to tensor-product-in-time meshes. We postpone possible generalization to subsection 2.6 below, which are important, e.g., for an adaptive version of the scheme.

We consider a sequence of polytopic meshes $\{\mathcal{T}_h\}_h$ of Q_T. We require that
(G1) the space domain Ω is split into a mesh $\mathcal{T}_h^\mathbf{x}$ of nonoverlapping d-dimensional polytopes with straight facets; the time interval $(0, T)$ is split into N subintervals $I_n := (t_{n-1}, t_n)$ with knots $0 = t_0 < t_1 < \ldots < t_N = T$; each element K in $\mathcal{T}_h$ can be written as $K_\mathbf{x} \times I_n$, for some $K_\mathbf{x}$ in $\mathcal{T}_h^\mathbf{x}$ and $1 \leq n \leq N$.
Essentially, assumption (**G1**) states that (i) each element is the tensor-product of a d-dimensional polytope with a time interval; (ii) each element belongs to a time slab out of the N identified by the partition $\{t_n\}_{n=0}^N$; (iii) each time slab is partitioned by the same space mesh; (iv) all elements within the same time slab have the same extent in time.

Given an element K in $\mathcal{T}_h$, $K = K_\mathbf{x} \times I_n$, we denote its diameter by h_K and the diameter of $K_\mathbf{x}$ by $h_{K_\mathbf{x}}$, and set $h_{I_n} := t_n - t_{n-1}$. We let $h := \max_{K \in \mathcal{T}_h} h_K$ and

$h_{\mathbf{x}} := \max_{K \in \mathcal{T}_h^{\mathbf{x}}} h_{K_{\mathbf{x}}}$. Furthermore, the set of all $(d-1)$-dimensional facets of $K_{\mathbf{x}}$ is denoted by $\mathcal{F}^{K_{\mathbf{x}}}$, and for any $F_{\mathbf{x}} \in \mathcal{F}^{K_{\mathbf{x}}}$ we define

$$h_{F_{\mathbf{x}}} := \begin{cases} \min\{h_{K_{\mathbf{x}}}, h_{\widetilde{K}_{\mathbf{x}}}\} & \text{if } F_{\mathbf{x}} = K_{\mathbf{x}} \cap \widetilde{K}_{\mathbf{x}} \text{ for some } \widetilde{K}_{\mathbf{x}} \in \mathcal{T}_h^{\mathbf{x}}, \\ h_{K_{\mathbf{x}}} & \text{if } F_{\mathbf{x}} \subset \partial\Omega. \end{cases}$$

For $d = 1$, $F_{\mathbf{x}}$ is a point and $\int_{F_{\mathbf{x}}} v(x,t)\mathrm{d}S$ is equal to $v(F_{\mathbf{x}}, t)$. For each spatial facet $F_{\mathbf{x}}$ in $\mathcal{F}^{K_{\mathbf{x}}}$, we introduce the time-like facet $F := F_{\mathbf{x}} \times I_n$; we collect all these time-like facets into the set $\mathcal{F}^K$.

We fix one of the two unit normal d-dimensional vectors associated with $F_{\mathbf{x}}$ and denote it by $\mathbf{n}_{F_{\mathbf{x}}}$. For $d \geq 1$, each time-like facet $F = F_{\mathbf{x}} \times I_n$ lies in a d-dimensional hyperplane with unit normal vector $\mathbf{n}_F := (\mathbf{n}_{F_{\mathbf{x}}}, 0)$.

Next, we require further assumptions on the spatial mesh: there exists $\gamma > 0$ independent of the meshsize such that

(G2) each spatial element $K_{\mathbf{x}}$ in $\mathcal{T}_h^{\mathbf{x}}$ is star-shaped with respect to a ball of radius $\rho_{K_{\mathbf{x}}}$ with $h_{K_{\mathbf{x}}} \leq \gamma \rho_{K_{\mathbf{x}}}$ and the number of $(d-1)$-dimensional facets of $K_{\mathbf{x}}$ is uniformly bounded with respect to the meshsize;

(G3) given two neighboring elements $K_{\mathbf{x}}$ and $\widetilde{K}_{\mathbf{x}}$ of $\mathcal{T}_h^{\mathbf{x}}$, we have that $\gamma^{-1} h_{\widetilde{K}_{\mathbf{x}}} \leq h_{K_{\mathbf{x}}} \leq \gamma h_{\widetilde{K}_{\mathbf{x}}}$.

For a given space–time element $K \subset \mathbb{R}^{d+1}$ and any space-like or time-like facet $F \subset \partial K$, we denote the space of polynomials of total degree at most $p \in \mathbb{N}$ on K and F by $\mathbb{P}_p(K)$ and $\mathbb{P}_p(F)$, respectively. For a given time interval I, $\mathbb{P}_p(I)$ denotes the space of polynomials in I of total degree at most p in $\mathbb{N}$.

Henceforth, for a positive natural number k, we define the spaces of broken H^k functions over $\mathcal{T}_h^{\mathbf{x}}$ and $\mathcal{T}_h$, respectively, by

$$H^k(\mathcal{T}_h^{\mathbf{x}}) := \left\{ v \in L^2(Q_T) \mid v_{|K_{\mathbf{x}}} \in H^k(K_{\mathbf{x}}) \ \forall K_{\mathbf{x}} \in \mathcal{T}_h^{\mathbf{x}} \right\};$$
$$H^k(\mathcal{T}_h) := \left\{ v \in L^2(Q_T) \mid v_{|K} \in H^k(K) \ \forall K \in \mathcal{T}_h \right\}.$$

We denote the broken Sobolev k seminorm on $\mathcal{T}_h$ by $|\cdot|_{k,\mathcal{T}_h}$ and the space of piecewise polynomials of degree at most ℓ in $\mathbb{N}$ on $\mathcal{T}_h$ by $\mathcal{S}_\ell(\mathcal{T}_h)$.

2. The virtual element method. In this section, we introduce a VEM for the discretization of problem (1.4) based on the regular meshes introduced in subsection 1.2. First, local VE spaces are introduced in subsection 2.1 together with their DoFs. Based on the choice of such DoFs, in subsection 2.2, we show that we can compute different polynomial projections of the VE functions. Such polynomial projections are instrumental in the design of the global VE spaces; see subsection 2.3. Likewise, in subsection 2.4, we design computable discrete bilinear forms and require sufficient properties that will allow us to prove the well posedness of the scheme, introduced in subsection 2.5, as well as convergence estimates. Finally, in subsection 2.6, we present more general types of meshes that can be used, e.g., in an adaptive framework.

2.1. Local virtual element spaces. We present a VE discretization of the infinite-dimensional spaces X and Y introduced in (1.2).

Given an approximation degree $p \in \mathbb{N}$ and an element $K = K_{\mathbf{x}} \times I_n$ in $\mathcal{T}_h$, we define the following local VE spaces:

(2.1)
$$V_h(K) := \Big\{ v_h \in L^2(K) \mid \widetilde{c}_H^K \partial_t v_h - \widetilde{\nu}^K \Delta_{\mathbf{x}} v_h \in \mathbb{P}_{p-1}(K), \quad v_{h|K_{\mathbf{x}} \times \{t_{n-1}\}} \in \mathbb{P}_p(K_{\mathbf{x}});$$

$$\mathbf{n}_{F_{\mathbf{x}}} \cdot \nabla_{\mathbf{x}} v_{h|F} \in \mathbb{P}_p(F) \ \forall F := F_{\mathbf{x}} \times I_n \text{ with } F_{\mathbf{x}} \in \mathcal{F}^{K_{\mathbf{x}}} \Big\},$$

where $\widetilde{c}_H^K := h_{I_n}$ and $\widetilde{\nu}^K := h_{K_{\mathbf{x}}}^2$.

The space $V_h(K)$ contains $\mathbb{P}_p(K)$. The degree p in the Neumann boundary conditions is not necessary for this inclusion to be valid, as $p-1$ would be sufficient. Nevertheless, the degree p is crucial in the proof of the Poincaré-type inequality in Proposition 2.8 below.

Remark 2.1. Functions in $V_h(K)$ solve a heat equation problem with polynomial source, initial condition, and Neumann boundary conditions. For this reason, $V_h(K) \subset L^2(I_n; H^1(K_{\mathbf{x}}))$; see [14, Thm. 4.1 and sect. 4.7.2 in Ch. 3] with standard modifications to deal with the inhomogeneous Neumann data.

Remark 2.2. As opposed to the standard VEM setting [3], in definition (2.1) we consider solutions to local problems involving some scaling factors ($\widetilde{c}_H^K$ and $\widetilde{\nu}^K$). The reason is that these local problems involve differential operators of different orders. By using a scaling argument and mapping the element K into a "reference" element $\widehat{K} = \widehat{I} \times \widehat{K}_{\widehat{\mathbf{x}}}$, with $|\widehat{I}_n| = \mathrm{diam}(\widehat{K}_{\widehat{\mathbf{x}}}) = 1$, the resulting reference space consists of solutions to a heat equation with both coefficients equal to 1. This allows us to use equivalence of norms results when proving the stability of the scheme.

Let $\{m_\alpha^K\}_{\alpha=1}^{\dim(\mathbb{P}_{p-1}(K))}$, $\{m_\beta^F\}_{\beta=1}^{\dim(\mathbb{P}_p(F))}$, and $\{m_\gamma^{K_{\mathbf{x}}}\}_{\gamma=1}^{\dim(\mathbb{P}_p(K_{\mathbf{x}}))}$ be any bases of $\mathbb{P}_{p-1}(K)$, $\mathbb{P}_p(F)$, and $\mathbb{P}_p(K_{\mathbf{x}})$. We introduce the following set of linear functionals on $V_h(K)$:

- the *bulk* moments

$$(2.2) \qquad \frac{1}{|K|} \int_{I_n} \int_{K_{\mathbf{x}}} v_h \, m_\alpha^K \mathrm{d}\mathbf{x} \, \mathrm{dt} \qquad \forall \alpha = 1, \ldots, \dim(\mathbb{P}_{p-1}(K));$$

- for all space–time facets $F_{\mathbf{x}} \times I_n = F \in \mathcal{F}^K$, the *time-like* moments

$$(2.3) \qquad \frac{1}{|F|} \int_{I_n} \int_{F_{\mathbf{x}}} v_h \, m_\beta^F \mathrm{d}S \, \mathrm{dt} \qquad \forall \beta = 1, \ldots, \dim(\mathbb{P}_p(F));$$

- the *space-like* moments

$$(2.4) \qquad \frac{1}{|K_{\mathbf{x}}|} \int_{K_{\mathbf{x}}} v_h(\cdot, t_{n-1}) m_\gamma^{K_{\mathbf{x}}} \mathrm{d}\mathbf{x} \qquad \forall \gamma = 1, \ldots, \dim(\mathbb{P}_p(K_{\mathbf{x}})).$$

Since functions $v_h \in V_h(K)$ are polynomials at time t_{n-1}, then the integrals in (2.4) are well defined. Moreover, the inclusion $V_h(K) \subset L^2(I_n; H^1(K_{\mathbf{x}}))$ (see Remark 2.1) implies that the integrals in (2.2) and (2.3) are well defined as well.

We introduce the number of the functionals in (2.2)–(2.4) as

$$\# \mathrm{DoFs} := \dim(\mathbb{P}_{p-1}(K)) + \sum_{F \in \mathcal{F}^K} \dim(\mathbb{P}_p(F)) + \dim(\mathbb{P}_p(K_{\mathbf{x}})).$$

In the following lemma, we prove that the linear functionals (2.2)–(2.4) actually define a set of DoFs for $V_h(K)$. For convenience, we denote the set of these linear functionals by $\{\mathrm{DoF}_i\}_{i=1}^{\# \mathrm{DoFs}}$.

LEMMA 2.3. *The linear functionals* (2.2)–(2.4) *are a set of unisolvent* DoFs *for the space* $V_h(K)$.

Proof. Since the right-hand side and the initial and Neumann boundary conditions in (2.1) are independent of each other, the dimension of $V_h(K)$ is equal to the number of the linear functionals (2.2)–(2.4). Thus, it suffices to prove that the set of these linear functionals is unisolvent. In other words, we prove that, whenever $v_h \in V_h(K)$ satisfies $\mathrm{DoF}_i(v_h) = 0$ for all $i = 1, \ldots, \# \mathrm{DoFs}$, then $v_h = 0$.

Thanks to the definition of the DoFs (2.2) and (2.3), we have

$$0 = \int_{I_n} \int_{K_{\mathbf{x}}} v_h \underbrace{\left(\widetilde{c}_H^K \partial_t v_h - \widetilde{\nu}^K \Delta_{\mathbf{x}} v_h\right)}_{\in \mathbb{P}_{p-1}(K)} \mathrm{d}\mathbf{x}\ \ \mathrm{d}t + \widetilde{\nu}^K \sum_{F_{\mathbf{x}} \in \mathcal{F}^{K_{\mathbf{x}}}} \int_{I_n} \int_{F_{\mathbf{x}}} v_h \underbrace{\mathbf{n}_{F_{\mathbf{x}}} \cdot \nabla_{\mathbf{x}} v_h}_{\in \mathbb{P}_p(F)} \mathrm{d}S\ \mathrm{d}t$$

$$= \frac{\widetilde{c}_H^K}{2} \left(\|v_h(\cdot, t_n)\|_{0,K_{\mathbf{x}}}^2 - \|v_h(\cdot, t_{n-1})\|_{0,K_{\mathbf{x}}}^2\right) + \widetilde{\nu}^K \|\nabla_{\mathbf{x}} v_h\|_{0,K}^2.$$

Furthermore, using the definition of the DoFs (2.4), we have $\|v_h(\cdot, t_{n-1})\|_{0,K_{\mathbf{x}}}^2 = 0$ and deduce that

$$\widetilde{\nu}^K \|\nabla_{\mathbf{x}} v_h\|_{0,K}^2 = 0 \quad \Rightarrow \quad \nabla_{\mathbf{x}} v_h = 0 \text{ in } K \quad \Rightarrow \quad v_h = v_h(t).$$

From the definition of the space $V_h(K)$, this implies that $\partial_t v_h$ belongs to $\mathbb{P}_{p-1}(I_n)$; equivalently, v_h belongs to $\mathbb{P}_p(I_n)$. On the other hand, we know that the moments (2.3) are zero, in particular when they are taken with respect to monomials up to degree p in time only. This implies $v_h = 0$. □

2.2. Polynomial projections. Functions in the local VE space $V_h(K)$ are not known in closed form. However, if we have at our disposal the DoFs of a function v_h in $V_h(K)$, then we can compute projections onto polynomial spaces with given maximum degree.

First, for all $K = K_{\mathbf{x}} \times I_n$ in $\mathcal{T}_h$ and $\varepsilon > 0$, we define the operator $\Pi_p^N : H^{\frac{1}{2}+\varepsilon}(I_n; L^2(K_{\mathbf{x}})) \cap L^2(I_n; H^1(K_{\mathbf{x}})) \to \mathbb{P}_p(K)$ as follows: for any v in $H^{\frac{1}{2}+\varepsilon}(I_n; L^2(K_{\mathbf{x}})) \cap L^2(I_n; H^1(K_{\mathbf{x}}))$,

(2.5a) $$\int_{I_n} \int_{K_{\mathbf{x}}} \nabla_{\mathbf{x}} q_p^K \cdot \nabla_{\mathbf{x}} \left(\Pi_p^N v - v\right)\ \mathrm{d}\mathbf{x}\ \mathrm{d}t = 0 \quad \forall q_p^K \in \mathbb{P}_p(K);$$

(2.5b) $$\int_{I_n} \int_{K_{\mathbf{x}}} q_{p-1}(t) \left(\Pi_p^N v - v\right) \mathrm{d}\mathbf{x}\ \mathrm{d}t = 0 \quad \forall q_{p-1} \in \mathbb{P}_{p-1}(I_n);$$

(2.5c) $$\int_{K_{\mathbf{x}}} \left(\Pi_p^N v(\mathbf{x}, t_{n-1}) - v(\mathbf{x}, t_{n-1})\right) \mathrm{d}\mathbf{x} = 0.$$

We have $V_h(K) \subset L^2(I_n; H^1(K_{\mathbf{x}}))$; see Remark 2.1. This and the fact that functions in $V_h(K)$ restricted to the time t_{n-1} are polynomials entail that we can define $\Pi_p^N v$ also for v in $V_h(K)$.

LEMMA 2.4. *The operator Π_p^N is well defined. Moreover, for any v_h in $V_h(K)$, $\Pi_p^N v_h$ is computable via the* DoFs *(2.2)–(2.4).*

Proof. In order to prove that Π_p^N is well defined, we need to show that the number of (linear) conditions in (2.5a)–(2.5c) is equal to $\dim(\mathbb{P}_p(K))$. As (2.5a) is void for all $q_p^K \in \mathbb{P}_p(I_n)$, we have that the number of conditions in (2.5a)–(2.5c) is equal to $\dim(\mathbb{P}_p(K))$. We need only show that they are linearly independent.

To this aim, assume that $v = 0$. Conditions (2.5a) imply that $\nabla_{\mathbf{x}} \Pi_p^N v = 0$, i.e., $\Pi_p^N v$ belongs to $\mathbb{P}_p(I_n)$. Let $L_p(\cdot)$ be the Legendre polynomial of degree p over $[-1, 1]$. Using conditions (2.5b), we deduce that there exists a constant c such that

$$\Pi_p^N v = c L_p \left(\frac{2t - t_{n-1} - t_n}{t_n - t_{n-1}}\right).$$

Since condition (2.5c) entails $\Pi_p^N v(\cdot, t_{n-1}) = 0$ and $L_p(-1) \neq 0$, we deduce $c = 0$, whence $\Pi_p^N v = 0$. Therefore, the conditions are linearly independent and so Π_p^N is well defined.

As for the computability of $\Pi_p^N v_h$ for v_h in $V_h(K)$, conditions (2.5a) and (2.5b) are available via the bulk moments (2.2) (up to order $p-2$) and the time-like moments (2.3) (up to order $p-1$); condition (2.5c) is available via the lowest-order space-like moment in (2.4). $\qquad\square$

Next, for all K in $\mathcal{T}_h$, we define the operator $\Pi_p^\star : \mathcal{C}^0(I_n; L^2(K_\mathbf{x})) \to \mathbb{P}_p(K)$ as follows: for any v in $\mathcal{C}^0(I_n; L^2(K_\mathbf{x}))$,

$$(2.6a) \qquad \int_{I_n} \int_{K_\mathbf{x}} q_{p-1}^K \left(\Pi_p^\star v - v \right) \mathrm{d}\mathbf{x}\, \mathrm{dt} = 0 \quad \forall q_{p-1}^K \in \mathbb{P}_{p-1}(K);$$

$$(2.6b) \qquad \int_{K_\mathbf{x}} q_p^{K_\mathbf{x}} \left(\Pi_p^\star v(\mathbf{x}, t_{n-1}) - v(\mathbf{x}, t_{n-1}) \right) \mathrm{d}\mathbf{x} = 0 \quad \forall q_p^{K_\mathbf{x}} \in \mathbb{P}_p(K_\mathbf{x}).$$

Again, we have $V_h(K) \subset L^2(I_n; H^1(K_\mathbf{x}))$; see Remark 2.1. This and the fact that functions in $V_h(K)$ restricted to the time t_{n-1} are polynomials entail that we can define $\Pi_p^N v$ also for v in $V_h(K)$.

LEMMA 2.5. *The operator $\Pi_p^\star$ is well defined. Moreover, for any v_h in $V_h(K)$, $\Pi_p^\star v_h$ is computable via the DoFs (2.2)–(2.4).*

Proof. As in the proof of Lemma 2.4, we observe that the number of (linear) conditions in (2.6a)–(2.6b) is equal to $\dim(\mathbb{P}_p(K))$. Thus, it suffices to show that they are linearly independent.

Assume that $v = 0$. Then, taking $q_p^{K_\mathbf{x}} = \Pi_p^\star v(\mathbf{x}, t_{n-1})$ in (2.6b), we get $\Pi_p^\star v(\mathbf{x}, t_{n-1}) = 0$. On the other hand, taking $q_{p-1}^K = \partial_t \Pi_p^\star v$ in (2.6a), we get

$$0 = \frac{1}{2} \left(\left\| \Pi_p^\star v(\cdot, t_n) \right\|_{0,K_\mathbf{x}}^2 - \left\| \Pi_p^\star v(\cdot, t_{n-1}) \right\|_{0,K_\mathbf{x}}^2 \right) = \frac{1}{2} \left\| \Pi_p^\star v(\cdot, t_n) \right\|_{0,K_\mathbf{x}}^2.$$

In addition, we observe that

$$\left\| \partial_t \Pi_p^\star v \right\|_{0,K}^2 = \int_{K_\mathbf{x}} \Pi_p^\star v(\mathbf{x}, t)\, \partial_t \Pi_p^\star v(\mathbf{x}, t) \mathrm{d}\mathbf{x} \Big|_{t=t_{n-1}}^{t_n} - \int_{I_n} \int_{K_\mathbf{x}} \Pi_p^\star v \underbrace{\partial_{tt} \Pi_p^\star v}_{\in \mathbb{P}_{p-2}(K)} \mathrm{d}\mathbf{x}\, \mathrm{dt} = 0.$$

This implies that $\partial_t \Pi_p^\star v = 0$, which, together with $\Pi_p^\star v(\cdot, t_n) = 0$, gives that $\Pi_p^\star v = 0$. Therefore, the conditions are linearly independent and so $\Pi_p^\star$ is well defined.

As for the computability of $\Pi_p^\star v_h$ for $v_h \in V_h(K)$, conditions (2.6a) are available via the bulk DoFs (2.2), and conditions (2.6b) are at disposal via the bottom space-like DoFs (2.4). $\qquad\square$

We introduce other polynomial projectors: for all K in $\mathcal{T}_h$ and v in $L^2(K)$,

$$\Pi_{p-1}^{0,K} : L^2(K) \to \mathbb{P}_{p-1}(K), \qquad (q_{p-1}^K, v - \Pi_{p-1}^{0,K} v)_{0,K} = 0 \quad \forall q_{p-1}^K \in \mathbb{P}_{p-1}(K);$$

for each temporal interval I_n and $v \in L^2(I_n)$,

$$\Pi_{p-1}^{0,I_n} : L^2(I_n) \to \mathbb{P}_{p-1}(I_n), \qquad (q_{p-1}^{I_n}, v - \Pi_{p-1}^{0,I_n} v)_{0,I_n} = 0 \quad \forall q_{p-1}^{I_n} \in \mathbb{P}_{p-1}(I_n);$$

for each spatial element $K_\mathbf{x}$ and $v \in L^2(K_\mathbf{x})$,

$$\Pi_0^{0,K_\mathbf{x}} : L^2(K_\mathbf{x}) \to \mathbb{R}, \qquad (q_0, v - \Pi_0^{0,K_\mathbf{x}} v)_{0,K_\mathbf{x}} = 0 \quad \forall q_0 \in \mathbb{R};$$

for all time-like facet F and v in $L^2(F)$,

$$\Pi_p^{0,F} : L^2(F) \to \mathbb{P}_p(F), \qquad (q_p^F, v - \Pi_p^{0,F} v)_{0,F} = 0 \quad \forall q_p^F \in \mathbb{P}_p(F).$$

Given v_h in $V_h(K)$, the computability of the above projectors applied to v_h follows from the definition of the DoFs (2.2)–(2.4). The projector $\Pi_{p-1}^{0,K}$ induces the global piecewise L^2 projector Π_{p-1}^{0,Q_T} over $\mathcal{T}_h$.

The following polynomial inverse inequalities are valid.

LEMMA 2.6. *For any $p \in \mathbb{N}$, there exist positive constants $c_{\Pi_p^N}$ and $c_{\Pi_p^\star}$ independent of h_{I_n} and $h_{K_{\mathbf{x}}}$ such that, for all q_p in $\mathbb{P}_p(K)$,*

$$
(2.7) \quad \begin{aligned}
&\|q_p\|_{0,K}^2 + h_{K_{\mathbf{x}}}^2 \|\nabla_{\mathbf{x}} q_p\|_{0,K}^2 + h_{I_n}^2 \|\partial_t q_p\|_{0,K}^2 \\
&\leq c_{\Pi_p^N} \left(h_{K_{\mathbf{x}}}^2 \|\nabla_{\mathbf{x}} q_p\|_{0,K}^2 + \left\|\Pi_{p-1}^{0,I_n} q_p\right\|_{0,K}^2 + h_{I_n} \left\|\Pi_0^{0,K_{\mathbf{x}}} q_p(\cdot, t_{n-1})\right\|_{0,K_{\mathbf{x}}}^2 \right)
\end{aligned}
$$

and

$$
(2.8) \quad \begin{aligned}
&\|q_p\|_{0,K}^2 + h_{K_{\mathbf{x}}}^2 \|\nabla_{\mathbf{x}} q_p\|_{0,K}^2 + h_{I_n}^2 \|\partial_t q_p\|_{0,K}^2 \\
&\leq c_{\Pi_p^\star} \left(\left\|\Pi_{p-1}^{0,K} q_p\right\|_{0,K}^2 + h_{I_n} \|q_p(\cdot, t_{n-1})\|_{0,K_{\mathbf{x}}}^2 \right).
\end{aligned}
$$

Proof. The assertion follows from the regularity of the spatial mesh in assumption (**G2**), the fact that the functionals on the right-hand side of (2.7) and (2.8) are norms for $\mathbb{P}_p(K)$, and the equivalence of norms for spaces of polynomials with fixed maximum degree. $\square$

The presence of the subscripts appearing in the inverse estimate constants $c_{\Pi_p^N}$ and $c_{\Pi_p^\star}$ is to remind one that the norms on the right-hand side of (2.7) and (2.8) are induced by the definition of the operators Π_p^N and $\Pi_p^\star$.

2.3. Global virtual element spaces. We construct the global VE spaces in a nonconforming fashion. To this aim, we introduce a jump operator on the time-like facets. Each internal time-like facet F is shared by two elements K_1 and K_2 with outward pointing unit normal vectors $\mathbf{n}_{K_1}$ and $\mathbf{n}_{K_2}$, whereas each boundary time-like facet belongs to the boundary of a single element K_3 with outward pointing unit normal vector $\mathbf{n}_{K_3}$. We denote the d-dimensional vector containing the spatial components of the restriction of $\mathbf{n}_{K_j}$ to the time-like facet F by $\mathbf{n}_{K_j}^F$. Then, the normal jump on each time-like facet F is defined as

$$
(2.9) \quad [\![v]\!]_F := \begin{cases} v_{|K_1} \mathbf{n}_{K_1}^F + v_{|K_2} \mathbf{n}_{K_2}^F & \text{if } F \text{ is an internal face;} \\ v_{|K_3} \mathbf{n}_{K_3}^F & \text{if } F \text{ is a boundary face.} \end{cases}
$$

On each time slab I_n, we introduce the nonconforming Sobolev space of order p associated with the mesh $\mathcal{T}_h^{\mathbf{x}}$:

$$
(2.10) \quad \begin{aligned}
H^{1,nc}(\mathcal{T}_h^{\mathbf{x}}; I_n) := \Big\{ &v \in L_2\left(I_n; H^1(\mathcal{T}_h^{\mathbf{x}})\right) \ \Big| \\
&\int_{I_n} \int_{F_{\mathbf{x}}} q_p^F [\![v]\!]_F \cdot \mathbf{n}_{F_{\mathbf{x}}} \mathrm{d}S \, \mathrm{d}t = 0 \quad \forall q_p^F \in \mathbb{P}_p(F) \Big\}.
\end{aligned}
$$

This allows us to define the VE discretization Y_h of the space Y in (1.2) as the space of functions that are possibly discontinuous in time across space-like facets and nonconforming as above in space:

$$
\begin{aligned}
Y_h := \Big\{ &v_h \in L^2(Q_T) \mid v_{h|K} \in V_h(K) \ \forall K \in \mathcal{T}_h, \\
&v_{h|\mathcal{T}_h^{\mathbf{x}} \times I_n} \in H^{1,nc}(\mathcal{T}_h^{\mathbf{x}}; I_n) \ \forall n = 1, \dots, N \Big\}.
\end{aligned}
$$

The functions in the space X in (1.2) are continuous in time, namely,

$$(2.11) \qquad X \hookrightarrow \mathcal{C}^0([0,T]; L^2(\Omega));$$

see, e.g, [22, Thm. 25.5]. Nevertheless, we discretize it with Y_h as well, and impose the time continuity weakly through upwinding. As functions in the local VE space $V_h(K)$ are not known at the local final time t_n, the upwind fluxes are defined in terms of the traces of their polynomial projections $\Pi_p^\star$; see (2.22) below.

Remark 2.7. Due to the choice of the DoFs, one cannot define a continuous-in-time discretization of X with the local spaces $V_h(K)$. If this were possible, then each VE function on $K = K_\mathbf{x} \times I_n$ would be a polynomial of degree p at the local final time t_n. For general choices of the right-hand side, initial condition, and boundary conditions in (2.1), this cannot be true.

2.4. Discrete bilinear forms. On each element K, define the local continuous bilinear form in $V_h(K) \times V_h(K)$ and seminorm

$$a^K(u_h, v_h) := \nu(\nabla_\mathbf{x} u_h, \nabla_\mathbf{x} v_h)_{0,K}, \qquad |v_h|_{Y(K)}^2 := a^K(v_h, v_h).$$

Next, we prove a local Poincaré-type inequality.

PROPOSITION 2.8. *If v_h belongs to $V_h(K)$, $K = K_\mathbf{x} \times I_n$, then $|v_h|_{Y(K)} = 0$ if and only if $v_h = v_h(t)$ belongs to $\mathbb{P}_p(I_n)$. Moreover, there exists a positive constant C_P^K independent of h_{I_n} and $h_{K_\mathbf{x}}$ such that*

$$(2.12) \qquad \inf_{q_p^t \in \mathbb{P}_p(I_n)} \left\| v_h - q_p^t \right\|_{0,K} \le C_P^K h_{K_\mathbf{x}} \|\nabla_\mathbf{x} v_h\|_{0,K} \qquad \forall v_h \in V_h(K).$$

Proof. If v_h belongs to $V_h(K)$ with $\|\nabla_\mathbf{x} v_h\|_{Y(K)} = 0$, then $v_h = v_h(t)$. The definition of $V_h(K)$ in (2.1) implies that $\partial_t v_h$ belongs to $\mathbb{P}_{p-1}(I_n)$ or, equivalently, that v_h belongs to $\mathbb{P}_p(I_n)$. The converse is obviously true.

Inequality (2.12) follows from the equivalence of seminorms with the same kernel on finite-dimensional spaces and the scaling argument in Remark 2.2. □

We define $Y(\mathcal{T}_h) := L^2(0,T; H^1(\mathcal{T}_h^\mathbf{x}))$ and introduce the global broken seminorms

$$\text{for almost all } t, \quad |v(\cdot,t)|_{1,\mathcal{T}_h^\mathbf{x}}^2 := \sum_{K_\mathbf{x} \in \mathcal{T}_h^\mathbf{x}} \|\nabla_\mathbf{x} v(\cdot,t)\|_{0,K_\mathbf{x}}^2;$$

$$|v|_{Y(\mathcal{T}_h)}^2 := \int_0^T \nu |v(\cdot,t)|_{1,\mathcal{T}_h^\mathbf{x}}^2 \mathrm{d}t = \sum_{K \in \mathcal{T}_h} |v|_{Y(K)}^2.$$

PROPOSITION 2.9. *The seminorm $|\cdot|_{Y(\mathcal{T}_h)}$ is a norm in Y_h.*[1]

Proof. Given v_h in Y_h, we need only prove that $|v_h|_{Y(\mathcal{T}_h)} = 0$ implies $v_h = 0$. The identity $|v_h|_{Y(\mathcal{T}_h)} = 0$ implies that $|v_h|_{Y(K)} = 0$ for all elements $K = K_\mathbf{x} \times I_n$. Using Proposition 2.8, we deduce that $v_{h|K}$ only depends on time and belongs to $\mathbb{P}_p(I_n)$. The assertion follows using the spatial nonconformity of the space Y_h (see (2.10)), which is up to order p. □

On each element K in $\mathcal{T}_h$, $K = K_\mathbf{x} \times I_n$, let

$$S^K : [V_h(K) + L^2(I_n; H^1(K_\mathbf{x})) \cap \mathcal{C}^0(I_n; L^2(K_\mathbf{x}))]^2 \to \mathbb{R}$$

---

[1]In fact, $|\cdot|_{Y(\mathcal{T}_h)}$ is a norm on $Y + Y_h$. So, for arguments in $Y + Y_h$, we shall denote it by $\|\cdot\|_{Y(\mathcal{T}_h)}$.

be any symmetric positive semidefinite bilinear form that is computable via the DoFs and satisfies the following properties:

- for any v_h in $V_h(K) \cap \ker(\Pi_p^N)$, we have that

$$(2.13) \qquad S^K(v_h, v_h) = 0 \qquad \Longrightarrow \qquad v_h = 0;$$

- the following bound is valid with a positive constant $\widetilde{c}^* > 0$ independent of h_{I_n}, $h_{K_\mathbf{x}}$, and K:

$$(2.14)$$
$$S^K(v, v) \leq \widetilde{c}^* \left(h_{K_\mathbf{x}}^{-2} \|v\|_{0,K}^2 + \|\nabla_\mathbf{x} v\|_{0,K}^2 + h_{K_\mathbf{x}}^{-2} h_{I_n}^2 \|\partial_t v\|_{0,K}^2 \right) \; \forall v \in H^1(K).$$

Property (2.13) implies that $S^K(\cdot, \cdot)$ induces a norm in $V_h(K) \cap \ker(\Pi_p^N)$. Another consequence of (2.13) and the scaling argument in Remark 2.2, is that there exist two constants $0 < c_* < c^*$ independent of K such that

$$(2.15) \qquad c_* |v_h|_{Y(K)}^2 \leq \nu S^K(v_h, v_h) \leq c^* |v_h|_{Y(K)}^2 \qquad \forall v_h \in V_h(K) \cap \ker(\Pi_p^N).$$

In fact, the functional $|\cdot|_{Y(K)}$ is a norm on $V_h(K) \cap \ker(\Pi_p^N)$.

We define the discrete counterpart of the local bilinear forms $a^K(\cdot, \cdot)$:

$$(2.16) \qquad a_h^K(u_h, v_h) := a^K(\Pi_p^N u_h, \Pi_p^N v_h) + \nu S^K((I - \Pi_p^N)u_h, (I - \Pi_p^N)v_h).$$

LEMMA 2.10. *Property (2.15) implies that there exist two constants $0 < \alpha_* < \alpha^*$ independent of K such that the following local stability bounds are valid:*

$$(2.17) \qquad \alpha_* |v_h|_{Y(K)}^2 \leq a_h^K(v_h, v_h) \leq \alpha^* |v_h|_{Y(K)}^2 \qquad \forall v_h \in V_h(K).$$

Proof. We only show the upper bound as the lower bound follows analogously leading to $\alpha_* := \min(1, c_*)$. We have

$$\begin{aligned} a_h^K(v_h, v_h) &= \nu \|\nabla_\mathbf{x} \Pi_p^N v_h\|_{0,K}^2 + \nu S^K((I - \Pi_p^N)v_h, (I - \Pi_p^N)v_h) \\ &\leq |\Pi_p^N v_h|_{Y(K)}^2 + c^* |(I - \Pi_p^N)v_h|_{Y(K)}^2 \\ &\leq \max(1, c^*) \left(|\Pi_p^N v_h|_{Y(K)}^2 + |(I - \Pi_p^N)v_h|_{Y(K)}^2 \right). \end{aligned}$$

Pythagoras' theorem implies

$$a_h^K(v_h, v_h) \leq \max(1, c^*) |v_h|_{Y(K)}^2.$$

This proves the upper bound in (2.17) with $\alpha^* = \max(1, c^*)$. $\qquad\square$

The global discrete bilinear form associated with the spatial Laplace operator reads

$$a_h(u_h, v_h) := \sum_{K \in \mathcal{T}_h} a_h^K(u_h, v_h) \qquad \forall u_h, \; v_h \in Y_h.$$

Taking into account Proposition 2.9, an immediate consequence of (2.17) is the global stability bounds

$$(2.18) \qquad \alpha_* \|v_h\|_{Y(\mathcal{T}_h)}^2 \leq a_h(v_h, v_h) \leq \alpha^* \|v_h\|_{Y(\mathcal{T}_h)}^2 \qquad \forall v_h \in Y_h.$$

For sufficiently smooth functions, we have the following upper bounds.

PROPOSITION 2.11. *For all v in $H^1(\mathcal{T}_h)$, the following local and global bounds are valid: for all K in $\mathcal{T}_h$*

$$a_h^K(v,v) \leq 3\max(1,\widetilde{c}^*)\nu\left(1+(1+c_{tr})c_{\Pi_p^N}\right)$$

(2.19)
$$\times\left(h_{K_\mathbf{x}}^{-2}\|v\|_{0,K}^2 + \|\nabla_\mathbf{x}v\|_{0,K}^2 + h_{K_\mathbf{x}}^{-2}h_{I_n}^2\|\partial_t v\|_{0,K}^2\right)$$

and

$$a_h(v,v) \leq 3\max(1,\widetilde{c}^*)\nu\left(1+(1+c_{tr})c_{\Pi_p^N}\right)$$

(2.20)
$$\times\sum_{K\in\mathcal{T}_h}\left(h_{K_\mathbf{x}}^{-2}\|v\|_{0,K}^2 + \|\nabla_\mathbf{x}v\|_{0,K}^2 + h_{K_\mathbf{x}}^{-2}h_{I_n}^2\|\partial_t v\|_{0,K}^2\right),$$

where $\widetilde{c}^*$ is the stability constant in (2.14), ν is the thermal conductivity, c_{tr} is the constant appearing in the elemental trace (in time) inequality, and $c_{\Pi_p^N}$ is the inverse estimate constant in (2.8).

Proof. The stability of the Π_p^N projector entails

$$a^K(\Pi_p^N v, \Pi_p^N v) = \nu\|\nabla_\mathbf{x}\Pi_p^N v\|_{0,K}^2 \leq \nu\|\nabla_\mathbf{x}v\|_{0,K}^2.$$

Using definition (2.16) and bound (2.14), we deduce that

(2.21) $a_h^K(v,v) = a^K(\Pi_p^N v, \Pi_p^N v) + \nu S^K((I-\Pi_p^N)v,(I-\Pi_p^N)v) \leq \max(1,\widetilde{c}^*)\nu$
$$\times\left(\|\nabla_\mathbf{x}v\|_{0,K}^2 + h_{K_\mathbf{x}}^{-2}\|(I-\Pi_p^N)v\|_{0,K}^2 + \|\nabla_\mathbf{x}(I-\Pi_p^N)v\|_{0,K}^2\right.$$
$$\left.+ h_{K_\mathbf{x}}^{-2}h_{I_n}^2\|\partial_t(I-\Pi_p^N)v\|_{0,K}^2\right).$$

Using the polynomial inverse estimate (2.7) with $q_p = \Pi_p^N v$, we can write

$$h_{K_\mathbf{x}}^{-2}\|\Pi_p^N v\|_{0,K}^2 + \|\nabla_\mathbf{x}\Pi_p^N v\|_{0,K}^2 + h_{K_\mathbf{x}}^{-2}h_{I_n}^2\|\partial_t\Pi_p^N v\|_{0,K}^2$$
$$= h_{K_\mathbf{x}}^{-2}\left(\|\Pi_p^N v\|_{0,K}^2 + h_{K_\mathbf{x}}^2\|\nabla_\mathbf{x}\Pi_p^N v\|_{0,K}^2 + h_{I_n}^2\|\partial_t\Pi_p^N v\|_{0,K}^2\right)$$
$$\leq c_{\Pi_p^N}h_{K_\mathbf{x}}^{-2}\left(h_{K_\mathbf{x}}^2\|\nabla_\mathbf{x}\Pi_p^N v\|_{0,K}^2 + \|\Pi_{p-1}^{0,K}\Pi_p^N v\|_{0,K}^2 + h_{I_n}\|\Pi_0^{0,K_\mathbf{x}}\Pi_p^N v(\cdot,t_{n-1})\|_{0,K}^2\right).$$

The definition of Π_p^N, and the stability of the L^2 and Π_p^N projectors entail

$$h_{K_\mathbf{x}}^{-2}\|\Pi_p^N v\|_{0,K}^2 + \|\nabla_\mathbf{x}\Pi_p^N v\|_{0,K}^2 + h_{K_\mathbf{x}}^{-2}h_{I_n}^2\|\partial_t\Pi_p^N v\|_{0,K}^2$$
$$\leq c_{\Pi_p^N}h_{K_\mathbf{x}}^{-2}\left(h_{K_\mathbf{x}}^2\|\nabla_\mathbf{x}v\|_{0,K}^2 + \|v\|_{0,K}^2 + h_{I_n}\|v(\cdot,t_{n-1})\|_{0,K}^2\right).$$

Applying a trace inequality along the time variable (with constant c_{tr}) on the last term yields

$$h_{K_\mathbf{x}}^{-2}\|\Pi_p^N v\|_{0,K}^2 + \|\nabla_\mathbf{x}\Pi_p^N v\|_{0,K}^2 + h_{K_\mathbf{x}}^{-2}h_{I_n}^2\|\partial_t\Pi_p^N v\|_{0,K}^2$$
$$\leq (1+c_{tr})c_{\Pi_p^N}\left(h_{K_\mathbf{x}}^{-2}\|v\|_{0,K}^2 + \|\nabla_\mathbf{x}v\|_{0,K}^2 + h_{K_\mathbf{x}}^{-2}h_{I_n}^2\|\partial_t v\|_{0,K}^2\right).$$

We insert this bound into (2.21) after applying the triangle inequality and obtain (2.19). Adding over all elements gives (2.20).   □

Here and in the following, for a given v in $L^2(Q_T)$, we shall write

$$v^{(n)} := v_{|\Omega\times I_n} \qquad \forall n = 1,\ldots,N.$$

For all u_h, v_h in Y_h and K in $\mathcal{T}_h$, $K = K_{\mathbf{x}} \times I_n$, we set

(2.22)

$$b_h^K(u_h, v_h) := \begin{cases} c_H(\partial_t \Pi_p^\star u_h, v_h)_{0,K} + a_h^K(u_h, v_h) \\ + c_H\Big(\Pi_p^\star u_h^{(1)}(\cdot, 0), v_h^{(1)}(\cdot, 0)\Big)_{0,K_{\mathbf{x}}} & \text{if } n = 1; \\[2mm] c_H(\partial_t \Pi_p^\star u_h, v_h)_{0,K} + a_h^K(u_h, v_h) \\ + c_H\Big(\Pi_p^\star u_h^{(n)}(\cdot, t_{n-1}) - \Pi_p^\star u_h^{(n-1)}(\cdot, t_{n-1}), v_h^{(n)}(\cdot, t_{n-1})\Big)_{0,K_{\mathbf{x}}} & \text{if } 2 \le n \le N. \end{cases}$$

The bilinear form $b_h^K(\cdot, \cdot)$ is computable through the DoFs. Actually, $\Pi_p^\star u_h^{(n)}(\cdot, t_{n-1}) = u_h^{(n)}(\cdot, t_{n-1})$ for $1 \le n \le N$ by the definition of $\Pi_p^\star$ in (2.6) and the definition of the local VE spaces.

We define the discrete counterpart of the global bilinear form $b(\cdot, \cdot)$ introduced in (1.3) as follows:

(2.23) $$b_h(u_h, v_h) := \sum_{K \in \mathcal{T}_h} b_h^K(u_h, v_h) \qquad \forall u_h, \ v_h \in Y_h.$$

The third terms in the definition of $b_h^K(u_h, v_h)$ in (2.22) stand for upwind fluxes for the weak imposition of the zero initial condition for $n = 1$, or of time continuity for $2 \le n \le N$.

2.4.1. An admissible stabilization. Consider the following stabilization for $K = K_{\mathbf{x}} \times I_n$:

(2.24)
$$S^K(u_h, v_h) := h_{K_{\mathbf{x}}}^{-2}(\Pi_{p-1}^{0,K} u_h, \Pi_{p-1}^{0,K} v_h)_{0,K} + h_{K_{\mathbf{x}}}^{-1} \sum_{F \in \mathcal{F}^K} (\Pi_p^{0,F} u_h, \Pi_p^{0,F} v_h)_{0,F}$$
$$+ h_{K_{\mathbf{x}}}^{-2} h_{I_n} (u_h(\cdot, t_{n-1}), v_h(\cdot, t_{n-1}))_{0,K_{\mathbf{x}}}.$$

This bilinear form is computable via the DoFs.

PROPOSITION 2.12. *The stabilization in* (2.24) *satisfies properties* (2.13) *and* (2.14).

Proof. Property (2.13) follows from the fact that $S^K(v_h, v_h)$ involves the squares of *all* the DoFs. Furthermore, property (2.14) follows from the stability of the L^2 projectors and the trace inequality applied to the time-like and space-like facet terms. $\square$

As pointed out in subsection 2.4, property (2.13) and the scaling argument in Remark 2.2 imply that property (2.15) is satisfied as well.

2.5. The method. The VEM that we propose reads as follows:

(2.25) $$\text{Find } u_h \in Y_h \text{ such that } b_h(u_h, v_h) = (f, \Pi_{p-1}^{0,Q_T} v_h)_{0,Q_T} \qquad \forall v_h \in Y_h.$$

The projector Π_{p-1}^{0,Q_T} makes the right-hand side computable and is L^2 stable, which is used in the proof of the well posedness of (2.25) in Theorem 3.3 below.

Under assumption (**G1**), the method can be solved in a time-marching fashion by solving the counterpart of (2.25) restricted to the time-slab I_n, for $n = 1, \ldots, N - 1$, and then transmitting the information to the subsequent time-slab I_{n+1} through upwinding.
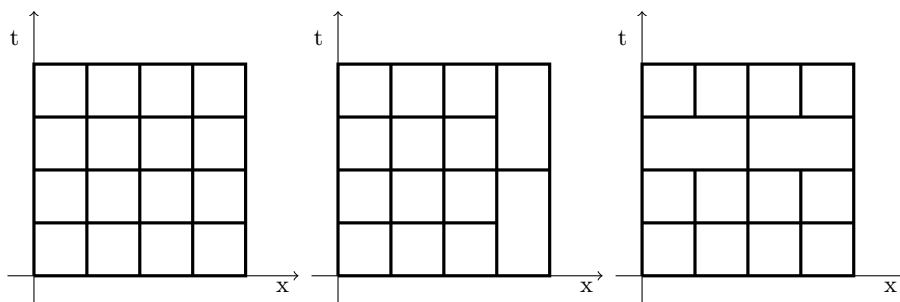
FIG. 1. (a) *Left panel: a mesh satisfying assumption* (**G1**)*.* (b) *Central panel: a mesh with nonmatching time-like facets.* (c) *Right panel: a mesh with nonmatching space-like facets.*

2.6. A glimpse on more general meshes. The reasons why we required assumption (**G1**) is that it is easier to present the construction of the VE spaces. We refer to Figure 1(a) for an example of an admissible mesh in the sense of (**G1**). We can weaken this assumption along two different avenues: we can allow for the following:

- nonmatching time-like facets; see Figure 1(b);
- nonmatching space-like facets; see Figure 1(c).

These generalizations are particularly convenient for space–time adaptivity, where nonmatching time-like and space-like facets typically occur. In the definition of the corresponding VE spaces, a few modifications would take place. In the case of nonmatching time-like facets (as in Figure 1(b)), VE functions have piecewise polynomial Neumann traces. Nonmatching space-like facets (as in Figure 1(c)) have no effect on the definition of the local VE spaces; see [11] for more details.

3. Well posedness of the virtual element method. In this section, we prove well posedness of the method in (2.25). To this aim, we endow the trial space with a suitable norm, which is defined by means of a VE Newton potential; see subsection 3.1. In subsection 3.2, we prove a discrete inf-sup condition. This proof extends that of [18, Theorem 2.1] to our setting, where multiple variational crimes have to be taken into account.

Before that, we prove a global Poincaré-type inequality for functions in the space Y_h.

PROPOSITION 3.1. *Let assumptions* (**G1**)*–*(**G2**) *be valid. Then, there exists a positive constant C_P independent of the mesh size h such that*

$$\|v_h\|_{0,Q_T} \leq C_P \|v_h\|_{Y(\mathcal{T}_h)} \qquad \forall v_h \in Y_h. \tag{3.1}$$

Proof. It suffices to prove the counterpart of (3.1) over each time slab I_n. On any time-like face F, we define the scalar jump $[\![v_h]\!]$ as $[\![v_h]\!]_F \cdot \mathbf{n}_F$.[2] We start from the spatial Poincaré inequality in [5, eq. (1.3) for $d \geq 2$ and sect. 8 for $d = 1$] with constant c_{PB} and integrate it in time over the time slab I_n:

$$\|v_h\|_{0,\Omega \times I_n}^2 \leq c_{PB} \left(\sum_{K \in \mathcal{T}_h, K \subset \Omega \times I_n} \|\nabla_\mathbf{x} v_h\|_{0,K}^2 + \sum_{F_\mathbf{x} \in \mathcal{F}_h^\mathbf{x}} h_{F_\mathbf{x}}^{-1} \int_{I_n} \left(\int_{F_\mathbf{x}} [\![v_h]\!] dS \right)^2 dt \right).$$

For $d = 1$, the integral over the point $F_\mathbf{x}$ is the evaluation at $F_\mathbf{x}$.

To conclude, we have to estimate the second term on the right-hand side. To this aim, we prove estimates on each time-like facet and then collect them together. For

---

[2]We have that $[\![\cdot]\!]$ is a scalar function whereas $[\![\cdot]\!]_F$ defined in (2.9) is a vector field.

simplicity, we further assume that $F = F_{\mathbf{x}} \times I_n$ is an internal (time-like) facet shared by two elements $K_1 = K_{\mathbf{x},1} \times I_n$ and $K_2 = K_{\mathbf{x},2} \times I_n$. The case of a boundary time-like facet can be dealt with similarly. Recall from the nonconformity of the space Y_h (see (2.10)) that $\Pi_p^{0,F}[\![v_h]\!] = 0$.

Using Jensen's inequality, we write

$$h_{F_{\mathbf{x}}}^{-1} \int_{I_n} \Big(\int_{F_{\mathbf{x}}} [\![v_h]\!] \mathrm{d}S \Big)^2 \mathrm{dt} \le \int_{I_n} \int_{F_{\mathbf{x}}} [\![v_h]\!]^2 \mathrm{d}S \ \mathrm{dt} = \big\| [\![v_h]\!] - \Pi_p^{0,F}[\![v_h]\!] \big\|_{0,F}^2 .$$

Denote the L^2 projection onto $\mathbb{P}_p(I_n)$ of the restriction of v_h to K_j by $q_p^{t,j}$, $j = 1, 2$. Let q_p^t be defined on $K_1 \cup K_2$ piecewise as $q_{p|_{K_j}}^t = q_p^{t,j}$, $j = 1, 2$. A standard trace inequality in space with constant c_{tr}, and the local Poincaré inequality (2.12) give

$$h_{F_{\mathbf{x}}}^{-1} \int_{I_n} \Big(\int_{F_{\mathbf{x}}} [\![v_h]\!] \mathrm{d}S \Big)^2 \mathrm{dt} \le c_{tr} \Big(h_{K_{1,\mathbf{x}}}^{-1} \big\| v_h - q_p^t \big\|_{0,K_1}^2 + h_{K_{1,\mathbf{x}}} \big\| \nabla_{\mathbf{x}}(v_h - q_p^t) \big\|_{0,K_1}^2$$
$$+ h_{K_{2,\mathbf{x}}}^{-1} \big\| v_h - q_p^t \big\|_{0,K_2}^2 + h_{K_{2,\mathbf{x}}} \big\| \nabla_{\mathbf{x}}(v_h - q_p^t) \big\|_{0,K_2}^2 \Big)$$
$$\le 2c_{tr} \Big(\max_{K \in \mathcal{T}_h} C_P^K \Big)^2 \Big(h_{K_{1,\mathbf{x}}} \big\| \nabla_{\mathbf{x}}(v_h - q_p^t) \big\|_{0,K_1}^2 + h_{K_{2,\mathbf{x}}} \big\| \nabla_{\mathbf{x}}(v_h - q_p^t) \big\|_{0,K_2}^2 \Big)$$
$$\le 2c_{tr} \Big(\max_{K \in \mathcal{T}_h} C_P^K \Big)^2 \max(h_{K_{1,\mathbf{x}}}, h_{K_{2,\mathbf{x}}}) \sum_{j=1}^2 \big\| \nabla_{\mathbf{x}} v_h \big\|_{0,K_j}^2 .$$

Summing over all the time-like facets of the nth time slab and recalling that the number of $(d-1)$-dimensional facets of each $K_{\mathbf{x}}$ is uniformly bounded with respect to the meshsize (see assumption (**G2**)), we get the assertion.   □

3.1. A virtual element Newton potential. We define a VE Newton potential $\mathfrak{N}_h : \mathcal{S}_p(\mathcal{T}_h) \to Y_h$ as follows: for any ϕ_h in $\mathcal{S}_p(\mathcal{T}_h)$, $\mathfrak{N}_h \phi_h$ in Y_h solves

(3.2)
$$a_h(\mathfrak{N}_h \phi_h, v_h) = b_h(\phi_h, v_h) - a_h(\phi_h, v_h)$$
$$= c_H \Big[(\partial_t \phi_h, v_h)_{0,Q_T} + \Big(\phi_h^{(1)}(\cdot, 0), v_h^{(1)}(\cdot, 0) \Big)_{0,\Omega}$$
$$+ \sum_{n=2}^N \Big(\phi_h^{(n)}(\cdot, t_{n-1}) - \phi_h^{(n-1)}(\cdot, t_{n-1}), v_h^{(n)}(\cdot, t_{n-1}) \Big)_{0,\Omega} \Big] \quad \forall v_h \in Y_h .$$

Thanks to the stability bounds (2.18), the bilinear form $a_h(\cdot, \cdot)$ is continuous and coercive, and the continuity in the $Y(\mathcal{T}_h)$ norm of the functional on the right-hand side of (3.2) follows from Proposition 3.1. Therefore, the VE Newton potential is well defined.

We introduce the following norm on the sum space $X + Y_h$: for all v in $X + Y_h$,

(3.3)
$$\|v\|_{X(\mathcal{T}_h)}^2 := \|v\|_{Y(\mathcal{T}_h)}^2 + \big\| \mathfrak{N}_h(\Pi_p^\star v) \big\|_{Y(\mathcal{T}_h)}^2 + \frac{c_H}{2} \Big(\big\| \Pi_p^\star v^{(1)}(\cdot, 0) \big\|_{0,\Omega}^2$$
$$+ \sum_{n=2}^N \big\| (\Pi_p^\star v^{(n)} - \Pi_p^\star v^{(n-1)})(\cdot, t_{n-1}) \big\|_{0,\Omega}^2 + \big\| \Pi_p^\star v^{(N)}(\cdot, T) \big\|_{0,\Omega}^2 \Big) .$$

Recalling the embedding $X \hookrightarrow \mathcal{C}^0(0, T; L^2(\Omega))$ in (2.11), we have that $\Pi_p^\star$ in (2.6) is well defined for functions in X. In section 4 below, we shall present the convergence analysis of the method with respect to the $\|\cdot\|_{X(\mathcal{T}_h)}$ norm.

3.2. A discrete inf-sup condition and well posedness of the method. In this section, we prove a discrete inf-sup condition in the spaces $(Y_h, \|\cdot\|_{X(\mathcal{T}_h)})$ for the trial functions and $(Y_h, \|\cdot\|_{Y(\mathcal{T}_h)})$ for the test functions.

PROPOSITION 3.2. *There exists a positive constant γ_I independent of $\mathcal{T}_h$ such that*

$$(3.4) \qquad \sup_{0 \neq v_h \in Y_h} \frac{b_h(u_h, v_h)}{\|v_h\|_{Y(\mathcal{T}_h)}} \geq \gamma_I \|u_h\|_{X(\mathcal{T}_h)} \qquad \forall u_h \in Y_h.$$

Proof. For any u_h in Y_h, define $w_h := \mathfrak{N}_h(\Pi_p^\star u_h)$ in Y_h. It suffices to prove that

$$\frac{b_h(u_h, u_h + \delta w_h)}{\|u_h + \delta w_h\|_{Y(\mathcal{T}_h)}} \geq \gamma_I \|u_h\|_{X(\mathcal{T}_h)}$$

for a suitable real parameter $\delta > 0$, which will be fixed below.

The triangle inequality and the definition of the norm $\|\cdot\|_{X(\mathcal{T}_h)}$ in (3.3) imply

$$\|u_h + \delta w_h\|_{Y(\mathcal{T}_h)}^2 \leq 2\left(\|u_h\|_{Y(\mathcal{T}_h)}^2 + \delta^2 \|w_h\|_{Y(\mathcal{T}_h)}^2\right) \leq 2\max(1, \delta^2)\|u_h\|_{X(\mathcal{T}_h)}^2,$$

whence we deduce

$$(3.5) \qquad \|u_h + \delta w_h\|_{Y(\mathcal{T}_h)} \leq \sqrt{2}\max(1, \delta)\|u_h\|_{X(\mathcal{T}_h)}.$$

Next, recalling (2.23) and (2.22), we write

$$(3.6) \quad \begin{aligned} b_h(u_h, u_h) &= \sum_{K \in \mathcal{T}_h} \left(c_H(\partial_t \Pi_p^\star u_h, u_h)_{0,K} + a_h^K(u_h, u_h)\right) + c_H \left\|\Pi_p^\star u_h^{(1)}(\cdot, 0)\right\|_{0,\Omega}^2 \\ &\quad + c_H \sum_{n=2}^N \left(\Pi_p^\star u_h^{(n)}(\cdot, t_{n-1}) - \Pi_p^\star u_h^{(n-1)}(\cdot, t_{n-1}), u_h^{(n)}(\cdot, t_{n-1})\right)_{0,\Omega}. \end{aligned}$$

For $K = K_{\mathbf{x}} \times I_n$, we have

$$\begin{aligned} (\partial_t \Pi_p^\star u_h, u_h)_{0,K} &\overset{(2.6a)}{=} (\partial_t \Pi_p^\star u_h, \Pi_p^\star u_h)_{0,K} \\ &= \frac{1}{2}\left(\left\|\Pi_p^\star u_h^{(n)}(\cdot, t_n)\right\|_{0,K_{\mathbf{x}}}^2 - \left\|\Pi_p^\star u_h^{(n)}(\cdot, t_{n-1})\right\|_{0,K_{\mathbf{x}}}^2\right). \end{aligned}$$

By (2.6b), we have $u_h^{(n)}(\cdot, t_{n-1}) = \Pi_p^\star u_h^{(n)}(\cdot, t_{n-1})$. Simple calculations give

$$\begin{aligned} &\sum_{K \in \mathcal{T}_h} (\partial_t \Pi_p^\star u_h, u_h)_{0,K} + \left\|\Pi_p^\star u_h^{(1)}(\cdot, 0)\right\|_{0,\Omega}^2 \\ &\quad + \sum_{n=2}^N \left(\Pi_p^\star u_h^{(n)}(\cdot, t_{n-1}) - \Pi_p^\star u_h^{(n-1)}(\cdot, t_{n-1}), u_h^{(n)}(\cdot, t_{n-1})\right)_{0,\Omega} \\ &= \sum_{n=1}^N \left(\frac{1}{2}\left\|\Pi_p^\star u_h^{(n)}(\cdot, t_n)\right\|_{0,\Omega}^2 - \frac{1}{2}\left\|\Pi_p^\star u_h^{(n)}(\cdot, t_{n-1})\right\|_{0,\Omega}^2\right) + \left\|\Pi_p^\star u_h^{(1)}(\cdot, 0)\right\|_{0,\Omega}^2 \\ &\quad + \sum_{n=2}^N \left(\left\|\Pi_p^\star u_h^{(n)}(\cdot, t_{n-1})\right\|_{0,\Omega}^2 - \left(\Pi_p^\star u_h^{(n-1)}(\cdot, t_{n-1}), \Pi_p^\star u_h^{(n)}(\cdot, t_{n-1})\right)_{0,\Omega}\right) \\ &= \frac{1}{2}\left\|\Pi_p^\star u_h^{(1)}(\cdot, 0)\right\|_{0,\Omega}^2 + \sum_{n=2}^N \frac{1}{2}\left\|\Pi_p^\star u_h^{(n)}(\cdot, t_{n-1}) - \Pi_p^\star u_h^{(n-1)}(\cdot, t_{n-1})\right\|_{0,\Omega}^2 \\ &\quad + \frac{1}{2}\left\|\Pi_p^\star u_h^{(N)}(\cdot, T)\right\|_{0,\Omega}^2. \end{aligned}$$

Therefore, from (3.6) and (2.18), we get

$$
\begin{aligned}
b_h(u_h, u_h) \geq & \alpha_* \|u_h\|^2_{Y(\mathcal{T}_h)} + \frac{c_H}{2} \Big(\big\| \Pi_p^\star u_h^{(1)}(\cdot, 0) \big\|^2_{0,\Omega} \\
& + \sum_{n=2}^{N} \big\| (\Pi_p^\star u_h^{(n)} - \Pi_p^\star u_h^{(n-1)})(\cdot, t_{n-1}) \big\|^2_{0,\Omega} + \big\| \Pi_p^\star u_h^{(N)}(\cdot, T) \big\|^2_{0,\Omega} \Big).
\end{aligned}
\tag{3.7}
$$

Moreover, the definition of $b_h(\cdot, \cdot)$ in (2.22) and (2.23), and the definition of the VE Newton potential in (3.2) imply

$$
b_h(u_h, \delta w_h) = \delta \left(a_h(w_h, w_h) + a_h(u_h, w_h) \right).
\tag{3.8}
$$

Since (2.18) gives $a_h(w_h, w_h) \geq \alpha_* \|w_h\|^2_{Y(\mathcal{T}_h)}$, then Young's inequality entails, for all positive ε,

$$
\begin{aligned}
a_h(u_h, w_h) &\geq -\left(a_h(u_h, u_h) \right)^{\frac{1}{2}} \left(a_h(w_h, w_h) \right)^{\frac{1}{2}} \geq -\alpha^* \|u_h\|_{Y(\mathcal{T}_h)} \|w_h\|_{Y(\mathcal{T}_h)} \\
&\geq -\frac{\alpha^*}{2\varepsilon} \|u_h\|^2_{Y(\mathcal{T}_h)} - \frac{\alpha^* \varepsilon}{2} \|w_h\|^2_{Y(\mathcal{T}_h)}.
\end{aligned}
$$

Inserting the two above inequalities into (3.8) yields

$$
b_h(u_h, \delta w_h) \geq \delta \left(\alpha_* - \frac{\alpha^* \varepsilon}{2} \right) \|w_h\|^2_{Y(\mathcal{T}_h)} - \frac{\alpha^* \delta}{2\varepsilon} \|u_h\|^2_{Y(\mathcal{T}_h)}.
\tag{3.9}
$$

As a final step, we sum (3.7) and (3.9):

$$
\begin{aligned}
& b_h(u_h, u_h + \delta w_h) \\
& \geq \left(\alpha_* - \frac{\alpha^* \delta}{2\varepsilon} \right) \|u_h\|^2_{Y(\mathcal{T}_h)} + \delta \left(\alpha_* - \frac{\alpha^* \varepsilon}{2} \right) \|w_h\|^2_{Y(\mathcal{T}_h)} \\
& \quad + \frac{c_H}{2} \Big(\big\| \Pi_p^\star u_h^{(1)}(\cdot, 0) \big\|^2_{0,\Omega} + \sum_{n=2}^{N} \big\| \Pi_p^\star u_h^{(n)}(\cdot, t_{n-1}) - \Pi_p^\star u_h^{(n-1)}(\cdot, t_{n-1}) \big\|^2_{0,\Omega} \\
& \quad + \big\| \Pi_p^\star u_h^{(N)}(\cdot, T) \big\|^2_{0,\Omega} \Big).
\end{aligned}
$$

Taking $0 < \varepsilon < (2\alpha_*)/\alpha^*$ and $0 < \delta < (2\varepsilon \alpha_*)/\alpha^*$, defining

$$
\beta := \min \left(\alpha_* - \frac{\alpha^* \delta}{2\varepsilon}, \delta \left(\alpha_* - \frac{\alpha^* \varepsilon}{2} \right) \right) > 0,
$$

and recalling (3.3) and (3.5), we can write

$$
\begin{aligned}
& b_h(u_h, u_h + \delta w_h) \\
& \geq \beta \left(\|u_h\|^2_{Y(\mathcal{T}_h)} + \|w_h\|^2_{Y(\mathcal{T}_h)} \right) + \frac{c_H}{2} \Big(\big\| \Pi_p^\star u_h^{(1)}(\cdot, 0) \big\|^2_{0,\Omega} \\
& \quad + \sum_{n=2}^{N} \big\| \Pi_p^\star u_h^{(n)}(\cdot, t_{n-1}) - \Pi_p^\star u_h^{(n-1)}(\cdot, t_{n-1}) \big\|^2_{0,\Omega} + \big\| \Pi_p^\star u_h^{(N)}(\cdot, T) \big\|^2_{0,\Omega} \Big) \\
& \geq \min(1, \beta) \|u_h\|^2_{X(\mathcal{T}_h)} \geq \frac{\min(1, \beta)}{\sqrt{2} \max(1, \delta)} \|u_h\|_{X(\mathcal{T}_h)} \|u_h + \delta w_h\|_{Y(\mathcal{T}_h)}.
\end{aligned}
$$

The assertion follows with $\gamma_I := \min(1, \beta) / (\sqrt{2} \max(1, \delta))$. □

We are in a position to prove the well posedness of the method in (2.25).

THEOREM 3.3. *There exists a unique solution u_h to the method in* (2.25) *with the following continuous dependence on the data:*

$$\|u_h\|_{X(\mathcal{T}_h)} \le \gamma_I^{-1} C_P \nu^{-1} \|f\|_{0,Q_T},$$

where γ_I is the discrete inf-sup constant in (3.4), *C_P is the Poincaré-type inequality constant in* (3.1), *and ν is the thermal conductivity.*

Proof. The discrete inf-sup condition (3.4) implies uniqueness of the solution. The existence follows from the uniqueness, owing to the finite dimensionality of Y_h. As for the stability bound, we apply again the inf-sup condition (3.4) and recall the definition (2.25) of the method:

$$\|u_h\|_{X(\mathcal{T}_h)} \le \frac{1}{\gamma_I} \sup_{0 \ne v_h \in Y_h} \frac{b_h(u_h, v_h)}{\|v_h\|_{Y(\mathcal{T}_h)}} = \frac{1}{\gamma_I} \sup_{0 \ne v_h \in Y_h} \frac{(f, \Pi_{p-1}^{0,Q_T} v_h)_{0,Q_T}}{\|v_h\|_{Y(\mathcal{T}_h)}}.$$

The Cauchy–Schwarz inequality, the L^2 stability of Π_{p-1}^{0,Q_T}, and the global Poincaré-type inequality (3.1) give the assertion. □

4. Convergence analysis. In this section, we analyze the convergence of the method in (2.25). We start by introducing further technical tools in subsection 4.1, which are typical of the nonconforming framework. Then, in subsections 4.2 and 4.3, we develop an a priori error analysis in two steps: first, we prove a convergence result à la Strang; next, we derive optimal convergence rates, by using interpolation and polynomial approximation results, assuming sufficient regularity on the solution.

4.1. Technical results. Introduce the bilinear form $\mathcal{N}_h : L^2\big(0, T; H^{\frac{3}{2}+\varepsilon}(\Omega)\big) \times Y_h \to \mathbb{R}$ given by

$$(4.1) \qquad \mathcal{N}_h(u, v_h) := \nu \sum_{n=1}^{N} \int_{I_n} \sum_{F_{\mathbf{x}} \in \mathcal{F}_h^{\mathbf{x}}} \int_{F_{\mathbf{x}}} \nabla_{\mathbf{x}} u \cdot [\![v_h]\!]_F \, \mathrm{d}S \, \mathrm{d}t.$$

This bilinear form encodes information on the nonconformity of the space Y_h across time-like facets.

On $K = I_n \times K_{\mathbf{x}}$, define the local bilinear form

$$b^K(w, v) := \int_{I_n} \int_{K_{\mathbf{x}}} (c_H \partial_t w \, v + \nu \nabla_{\mathbf{x}} w \cdot \nabla_{\mathbf{x}} v) \, \mathrm{d}\mathbf{x} \, \mathrm{d}t.$$

LEMMA 4.1. *Assume that the solution u to the continuous problem* (1.4) *belongs to $L^2(0, T; H^{\frac{3}{2}+\varepsilon}(\Omega))$. Then, for all v_h in Y_h,*

$$(4.2) \qquad \sum_{K \in \mathcal{T}_h} b^K(u, v_h) = (f, v_h)_{0,Q_T} + \mathcal{N}_h(u, v_h).$$

Proof. Integrating by parts in space and recalling the definition of $\mathcal{N}_h$ in (4.1), we can write

$$\sum_{K \in \mathcal{T}_h} b^K(u, v_h)$$

$$= \sum_{K \in \mathcal{T}_h} \int_{I_n} \left(\int_{K_{\mathbf{x}}} (c_H \partial_t u - \nu \Delta_{\mathbf{x}} u) v_h \mathrm{d}\mathbf{x} + \nu \sum_{F_{\mathbf{x}} \in \mathcal{F}^{K_{\mathbf{x}}}} \int_{F_{\mathbf{x}}} v_h (\mathbf{n}_{F_{\mathbf{x}}} \cdot \nabla_{\mathbf{x}} u) \mathrm{d}S \right) \mathrm{d}t$$

$$= (f, v_h)_{0,Q_T} + \mathcal{N}_h(u, v_h),$$

which proves (4.2). □

We introduce another preliminary result, which characterizes the polynomial inconsistency of the method in (2.25). To this aim, we define the bilinear form $\mathcal{J}^K : H^1(\mathcal{T}_h) \times Y_h \to \mathbb{R}$ given on each $K = K_\mathbf{x} \times I_n$ in $\mathcal{T}_h$ as follows: for all w in $H^1(\mathcal{T}_h)$, v_h in Y_h,

(4.3)

$$\mathcal{J}^K(w, v_h) := \begin{cases} c_H \left(\Pi_p^\star w^{(1)}(\cdot, 0), v_h^{(1)}(\cdot, 0) \right)_{0, K_\mathbf{x}} & \text{if } n = 1; \\ c_H \left(\left(\Pi_p^\star w^{(n)} - \Pi_p^\star w^{(n-1)} \right)(\cdot, t_{n-1}), v_h^{(n)}(\cdot, t_{n-1}) \right)_{0, K_\mathbf{x}} & \text{if } 2 \leq n \leq N. \end{cases}$$

This bilinear form encodes the polynomial inconsistency of the method at space-like facets, as stated in the following lemma.

LEMMA 4.2. *The local bilinear forms $b_h^K(\cdot, \cdot)$ satisfy*

(4.4) $b_h^K(q_p, v_h) = b^K(q_p, v_h) + \mathcal{J}^K(q_p, v_h)$ $\quad \forall q_p \in \mathcal{S}_p(\mathcal{T}_h), \forall v_h \in V_h(K), \forall K \in \mathcal{T}_h.$

Proof. Thanks to the definition of the bilinear form $a^K(\cdot, \cdot)$, the orthogonality properties of the projector Π_p^N, and the fact that the projectors $\Pi_p^\star$ and Π_p^N preserve polynomials of degree p, we have

$b_h^K(q_p, v_h)$
$= c_H(\partial_t \Pi_p^\star q_p, v_h)_{0,K} + a^K(\Pi_p^N q_p, \Pi_p^N v_h) + \nu S^K((I - \Pi_p^N)q_p, (I - \Pi_p^N)v_h) + \mathcal{J}^K(\Pi_p^\star q_p, v_h)$
$= c_H(\partial_t \Pi_p^\star q_p, v_h)_{0,K} + a^K(q_p, \Pi_p^N v_h) + \mathcal{J}^K(q_p, v_h)$
$= c_H(\partial_t q_p, v_h)_{0,K} + a^K(q_p, v_h) + \mathcal{J}^K(q_p, v_h) = b^K(q_p, v_h) + \mathcal{J}^K(q_p, v_h).$

This completes the proof. $\qquad\square$

4.2. A Strang-type result. We prove an a priori estimate for the method in (2.25).

THEOREM 4.3. *Let u and u_h be the solutions to (1.4) and (2.25), let u belong to $X \cap L^2(0, T, H^{\frac{3}{2}+\varepsilon}(\Omega))$ for some $\varepsilon > 0$, let u_I in Y_h be the DoF interpolant of u in Y_h, and let γ_I be the discrete inf-sup constant appearing in (3.4). Then, we have*

(4.5)
$$\|u - u_h\|_{X(\mathcal{T}_h)} \leq \|u - u_I\|_{Y(\mathcal{T}_h)}$$
$$+ \gamma_I^{-1} \sup_{0 \neq v_h \in Y_h} \left[\frac{|(f - \Pi_{p-1}^{0,Q_T} f, v_h)_{0,Q_T}|}{\|v_h\|_{Y(\mathcal{T}_h)}} + \frac{|\mathcal{N}_h(u, v_h)|}{\|v_h\|_{Y(\mathcal{T}_h)}} \right.$$
$$\left. + \inf_{q_p \in \mathcal{S}_p(\mathcal{T}_h)} \frac{\sum_{K \in \mathcal{T}_h} \left(b_h^K(u - q_p, v_h) - b^K(u - q_p, v_h) + \mathcal{J}^K(q_p, v_h) \right)}{\|v_h\|_{Y(\mathcal{T}_h)}} \right].$$

Proof. By the triangle inequality, we have

$$\|u - u_h\|_{X(\mathcal{T}_h)} \leq \|u - u_I\|_{X(\mathcal{T}_h)} + \|u_I - u_h\|_{X(\mathcal{T}_h)} =: T_1 + T_2.$$

Since $\Pi_p^\star$ is computable from the DoFs, we have that $\Pi_p^\star(u - u_I) = 0$ in each element. Taking into account (3.3), this yields

$$T_1^2 = \|u - u_I\|_{Y(\mathcal{T}_h)}^2 + \left\| \mathfrak{N}_h \left(\Pi_p^\star(u - u_I) \right) \right\|_{Y(\mathcal{T}_h)}^2 + \frac{c_H}{2} \left(\left\| \Pi_p^\star(u - u_I)(\cdot, 0) \right\|_{0,\Omega}^2 \right.$$
$$\left. + \sum_{n=2}^N \left\| \Pi_p^\star(u - u_I^{(n)})(\cdot, t_{n-1}) - \Pi_p^\star(u - u_I^{(n-1)})(\cdot, t_{n-1}) \right\|_{0,\Omega}^2 + \left\| \Pi_p^\star(u - u_I)(\cdot, T) \right\|_{0,\Omega}^2 \right)$$
$$= \|u - u_I\|_{Y(\mathcal{T}_h)}^2.$$

The rest of this proof is devoted to estimate the term T_2. The definition of u_I implies

$$b_h(u_I, v_h) = b_h(u, v_h) \qquad \forall v_h \in Y_h.$$

Using this property and the discrete inf-sup condition (3.4), we get

$$\|u_I - u_h\|_{X(\mathcal{T}_h)} \le \gamma_I^{-1} \sup_{0 \ne v_h \in Y_h} \frac{b_h(u_I - u_h, v_h)}{\|v_h\|_{Y(\mathcal{T}_h)}} = \gamma_I^{-1} \sup_{0 \ne v_h \in Y_h} \frac{b_h(u - u_h, v_h)}{\|v_h\|_{Y(\mathcal{T}_h)}}.$$

We recall (2.25), add and subtract any q_p in $\mathcal{S}_p(\mathcal{T}_h)$, use the inconsistency property (4.4), add and subtract u, recall the property of the nonconformity bilinear form $\mathcal{N}_h(\cdot, \cdot)$ in (4.2), and deduce that

$$b_h(u - u_h, v_h) = \sum_{K \in \mathcal{T}_h} b_h^K(u, v_h) - (f, \Pi_{p-1}^{0,Q_T} v_h)_{0,Q_T}$$

$$= \sum_{K \in \mathcal{T}_h} \left(b_h^K(u - q_p, v_h) + b_h^K(q_p, v_h) \right) - (f, \Pi_{p-1}^{0,Q_T} v_h)_{0,Q_T}$$

$$= \sum_{K \in \mathcal{T}_h} \left(b_h^K(u - q_p, v_h) + b^K(q_p, v_h) + \mathcal{J}^K(q_p, v_h) \right) - (f, \Pi_{p-1}^{0,Q_T} v_h)_{0,Q_T}$$

$$= \sum_{K \in \mathcal{T}_h} \left(b_h^K(u - q_p, v_h) + b^K(q_p - u, v_h) + b^K(u, v_h) + \mathcal{J}^K(q_p, v_h) \right) - (f, \Pi_{p-1}^{0,Q_T} v_h)_{0,Q_T}$$

$$= \sum_{K \in \mathcal{T}_h} \left(b_h^K(u - q_p, v_h) - b^K(u - q_p, v_h) + \mathcal{J}^K(q_p, v_h) \right) + (f - \Pi_{p-1}^{0,Q_T} f, v_h)_{0,Q_T} + \mathcal{N}_h(u, v_h).$$

The assertion follows from taking the infimum over all q_p in $\mathcal{S}_p(\mathcal{T}_h)$ and then the supremum over all v_h in Y_h.   □

4.3. A priori error estimate. The aim of this section is to prove optimal convergence rates for the method in (2.25). So far, we derived all estimates with explicit constants, so as to track the use of different type of inequalities (Poincaré, trace, inverse estimates, ...). Furthermore, we kept separated the contributions of $h_{K_\mathbf{x}}$ and h_{I_n}. In this section, we shall not keep this level of detail. As a matter of notation, we henceforth write $a \lesssim b$ meaning that there exists a positive constant c independent of the meshsize, such that $a \le cb$. We also write $a \simeq b$ if $a \lesssim b$ and $b \lesssim a$ at once.

We prove error estimates under some regularity assumptions on the exact solution and focus on the case of isotropic space–time meshes, i.e., assume that

(4.6) $h_{K_\mathbf{x}} \simeq h_{I_n} \simeq h_K \qquad \forall K = K_\mathbf{x} \times I_n \in \mathcal{T}_h.$

In Theorem 4.3, we proved that the error of the method in (2.25) is bounded by the sum of four terms of different flavour: (i) a VE interpolation error; (ii) a term involving the discretization of the right-hand side f; (iii) a term measuring the spatial nonconformity of the discrete space; (iv) a term involving polynomial error estimates, which appears because of the temporal nonconformity and the polynomial inconsistency of the discrete bilinear form. Based on that result, we prove the following theorem.

THEOREM 4.4. *Let assumptions (**G1**)–(**G3**) be valid, and let $\mathcal{T}_h$ be isotropic in the sense of (4.6). Let u, the solution of (1.4), and f, the right-hand side of (1.4), belong to $H^{p+1}(\mathcal{T}_h)$ and $H^p(\mathcal{T}_h)$, respectively, where $p \ge 1$ denotes the degree of approximation of the method in (2.25). Let u_h be the solution to (2.25). Then,*

(4.7) $\|u - u_h\|_{X(\mathcal{T}_h)} \lesssim h^p(|u|_{p+1, \mathcal{T}_h} + |f|_{p, \mathcal{T}_h}).$

Proof. We estimate the four terms on the right-hand side of (4.5) separately. The assertion then follows by combining the four bounds we provide below.

Part (i) VE interpolation error. For any q_p in $\mathcal{S}_p(\mathcal{T}_h)$, the triangle inequality implies

$$(4.8) \qquad \|u - u_I\|_{Y(\mathcal{T}_h)} \leq |u - q_p|_{Y(\mathcal{T}_h)} + |q_p - u_I|_{Y(\mathcal{T}_h)}.$$

We focus on the second term on the right-hand side. For any K in $\mathcal{T}_h$, $(q_p - u_I)_{|_K}$ belongs to $V_h(K)$. Therefore, the stability bounds (2.17) entail

$$|q_p - u_I|_{Y(K)}^2 \lesssim a_h^K(q_p - u_I, q_p - u_I) \qquad \forall K \in \mathcal{T}_h.$$

Since u_I is the DoFs interpolant of u and $a_h(\cdot, \cdot)$ is computed via the DoFs, the above inequality also implies

$$|q_p - u_I|_{Y(\mathcal{T}_h)}^2 \lesssim a_h(q_p - u_I, q_p - u_I) = a_h(q_p - u, q_p - u).$$

Furthermore, using the discrete continuity property (2.20), we arrive at

$$|q_p - u_I|_{Y(\mathcal{T}_h)}^2 \lesssim \sum_{K \in \mathcal{T}_h} \left(h_K^{-2} \|u - q_p\|_{0,K}^2 + |u - q_p|_{1,K}^2 \right).$$

Inserting this into (4.8) and using standard polynomial approximation results yield

$$\|u - u_I\|_{Y(\mathcal{T}_h)} \lesssim h^p |u|_{p+1, \mathcal{T}_h}.$$

Part (ii) Handling the variational crime on the right-hand side f. Using the definition of Π_{p-1}^{0,Q_T}, standard polynomial approximation estimates, and the global discrete Poincaré inequality (3.1) entail

$$(f - \Pi_{p-1}^{0,Q_T} f, v_h)_{0,Q_T} \leq \sum_{K \in \mathcal{T}_h} \left\| f - \Pi_{p-1}^{0,K} f \right\|_{0,K} \|v_h\|_{0,K} \lesssim h^p |f|_{p,\mathcal{T}_h} \|v_h\|_{0,Q_T}$$

$$\lesssim h^p |f|_{p,\mathcal{T}_h} \|v_h\|_{Y(\mathcal{T}_h)}.$$

Part (iii) Handling the variational crime of the time-like nonconformity. We estimate

$$\sup_{0 \neq v_h \in Y_h} \frac{|\mathcal{N}_h(u, v_h)|}{\|v_h\|_{Y(\mathcal{T}_h)}} = \sup_{0 \neq v_h \in Y_h} \frac{|\nu \sum_{F \in \mathcal{F}_h} \int_{I_n} \int_{F_{\mathbf{x}}} \nabla_{\mathbf{x}} u \cdot [\![v_h]\!]_F \mathrm{d}S \, \mathrm{d}t|}{\|v_h\|_{Y(\mathcal{T}_h)}}.$$

We present estimates on a single facet $F = F_{\mathbf{x}} \times I_n$. For the sake of simplicity, we assume that F is an internal time-like facet shared by the elements K_1 and K_2. Using the definition of the spatial nonconformity of the space Y_h (see (2.10)) and the properties of L^2 projectors, for all $q_p^{t,1}$, $q_p^{t,2}$ in $\mathbb{P}_p(I_n)$, we write[3]

$$\left| \int_{I_n} \int_{F_{\mathbf{x}}} \nabla_{\mathbf{x}} u \cdot [\![v_h]\!]_F \mathrm{d}S \, \mathrm{d}t \right| = \left| \int_{I_n} \int_{F_{\mathbf{x}}} \nabla_{\mathbf{x}} u \cdot \mathbf{n}_{F_{\mathbf{x}}} [\![v_h]\!] \mathrm{d}S \, \mathrm{d}t \right|$$

$$= \left| \int_{I_n} \int_{F_{\mathbf{x}}} \left(\nabla_{\mathbf{x}} u \cdot \mathbf{n}_{F_{\mathbf{x}}} - \Pi_p^{0,F} \left(\nabla_{\mathbf{x}} u \cdot \mathbf{n}_{F_{\mathbf{x}}} \right) \right) [\![v_h]\!] \mathrm{d}S \, \mathrm{d}t \right|$$

$$= \left| \int_{I_n} \int_{F_{\mathbf{x}}} \left(\nabla_{\mathbf{x}} u \cdot \mathbf{n}_{F_{\mathbf{x}}} - \Pi_p^{0,F} \left(\nabla_{\mathbf{x}} u \cdot \mathbf{n}_{F_{\mathbf{x}}} \right) \right) \left((v_{h|_{K_2}} - q_p^{t,2}) - (v_{h|_{K_1}} - q_p^{t,1}) \right) \mathrm{d}S \, \mathrm{d}t \right|.$$

---

[3]Here we use the scalar normal jump $[\![\cdot]\!]$ defined in the proof of Proposition 3.1.

Next, use the Cauchy–Schwarz inequality, the triangle inequality, standard properties of the L^2 projector, a trace inequality, the local quasi-uniformity of the space–time mesh, and arrive at

$$
\left| \int_{I_n} \int_{F_{\mathbf{x}}} \nabla_{\mathbf{x}} u \cdot [\![v_h]\!]_F \mathrm{d}S \, \mathrm{dt} \right|
$$
$$
\leq \left\| \nabla_{\mathbf{x}} u \cdot \mathbf{n}_{F_{\mathbf{x}}} - \Pi_p^{0,F} \left(\nabla_{\mathbf{x}} u \cdot \mathbf{n}_{F_{\mathbf{x}}} \right) \right\|_{L^2(F)} \left(\left\| v_{h|_{K_2}} - q_p^{t,2} \right\|_{L^2(F)} + \left\| v_{h|_{K_1}} - q_p^{t,1} \right\|_{L^2(F)} \right)
$$
$$
\leq \left\| \nabla_{\mathbf{x}} \left(u - \Pi_{p+1}^{0,K_1} u \right) \cdot \mathbf{n}_{F_{\mathbf{x}}} \right\|_{L^2(F)} \left(\left\| v_{h|_{K_2}} - q_p^{t,2} \right\|_{L^2(F)} + \left\| v_{h|_{K_1}} - q_p^{t,1} \right\|_{L^2(F)} \right)
$$
$$
\lesssim \left(h_{K_1}^{-\frac{1}{2}} \left| u - \Pi_{p+1}^{0,K_1} u \right|_{Y(K_1)} + h_{K_1}^{\epsilon} \left| u - \Pi_{p+1}^{0,K_1} u \right|_{\frac{3}{2}+\varepsilon, K_1} \right)
$$
$$
\times \left(\sum_{j=1,2} \left(h_{K_j}^{-\frac{1}{2}} \left\| v_h - q_p^{t,j} \right\|_{0, K_j} + h_{K_j}^{\frac{1}{2}} |v_h|_{Y(K_j)} \right) \right).
$$

An application of (2.12) yields

$$
\left| \int_{I_n} \int_{F_{\mathbf{x}}} \nabla_{\mathbf{x}} u \cdot [\![v_h]\!]_F \mathrm{d}S \, \mathrm{dt} \right|
$$
$$
\lesssim \left(\left| u - \Pi_{p+1}^{0,K_1} u \right|_{Y(K_1)} + h_{K_1}^{\epsilon+\frac{1}{2}} \left| u - \Pi_{p+1}^{0,K_1} u \right|_{\frac{3}{2}+\varepsilon, K_1} \right) \sum_{j=1,2} |v_h|_{Y(K_j)}.
$$

Summing up over all the elements and using approximation properties of the L^2 projector, we eventually get

$$
\sup_{0 \neq v_h \in Y_h} \frac{|\mathcal{N}_h(u, v_h)|}{\|v_h\|_{Y(\mathcal{T}_h)}} \lesssim h^p |u|_{p+1, \mathcal{T}_h}.
$$

Part (iv.a) Polynomial approximation error of $b^K(\cdot, \cdot)$ type. Let q_p be in $\mathcal{S}_p(\mathcal{T}_h)$. Using the Cauchy–Schwarz inequality twice and the definition of the bilinear form $b^K(\cdot, \cdot)$ give

$$
b^K(u - q_p, v_h) = c_H \left(\partial_t (u - q_p), v_h \right)_{0,K} + \nu \left(\nabla_{\mathbf{x}}(u - q_p), \nabla_{\mathbf{x}} v_h \right)_{0,K}
$$
$$
\lesssim |u - q_p|_{1,K} \left(\|v_h\|_{0,K} + |v_h|_{Y(K)} \right).
$$

Summing up over all the elements, using an ℓ^2 Cauchy–Schwarz inequality, and recalling the global Poincaré-type inequality (3.1), we can write

$$
(4.9) \qquad \sum_{K \in \mathcal{T}_h} b^K(u - q_p, v_h) \lesssim |u - q_p|_{1, \mathcal{T}_h} \|v_h\|_{Y(\mathcal{T}_h)}.
$$

Part (iv.b) Polynomial approximation error of $b_h^K(\cdot, \cdot) + \mathcal{J}^K(\cdot, \cdot)$ type. Thanks to definitions (2.22) and (4.3) on each element $K = K_{\mathbf{x}} \times I_n$ in $\mathcal{T}_h$, for all v_h in Y_h, we have

$$
(4.10) \qquad
\begin{aligned}
b_h^K(u - q_p, v_h) + \mathcal{J}^K(q_p, v_h) &= c_H(\partial_t \Pi_p^\star (u - q_p), v_h)_{0,K} \\
&\quad + a_h^K(u - q_p, v_h) + \mathcal{J}^K(u, v_h),
\end{aligned}
$$

where q_p is the same as in Part (iv.a). We first focus on the second term. Using the stability bound (2.17) and the continuity property (2.19) with $h_{K_{\mathbf{x}}} \simeq h_{I_n}$, we arrive at

$$a_h^K(u - q_p, v_h) \le a_h^K(u - q_p, u - q_p)^{\frac{1}{2}} a_h^K(v_h, v_h)^{\frac{1}{2}}$$
$$\lesssim \left(h_K^{-1} \|u - q_p\|_{0,K} + |u - q_p|_{1,K} \right) |v_h|_{Y(K)}.$$

Next, we deal with the first term on the right-hand side of (4.10). The Cauchy–Schwarz inequality yields

$$(\partial_t \Pi_p^\star(u - q_p), v_h)_{0,K} \le \left\| \partial_t \Pi_p^\star(u - q_p) \right\|_{0,K} \|v_h\|_{0,K}.$$

A polynomial inverse inequality gives

(4.11) $$\qquad \left\| \partial_t \Pi_p^\star(u - q_p) \right\|_{0,K} \lesssim h_K^{-1} \left\| \Pi_p^\star(u - q_p) \right\|_{0,K}.$$

By using (2.8), the definition of $\Pi_p^\star$, the stability of the L^2 orthogonal projection, and the trace inequality, we arrive at

(4.12)
$$\left\| \Pi_p^\star(u - q_p) \right\|_{0,K} \lesssim \left\| \Pi_{p-1}^{0,K} \Pi_p^\star(u - q_p) \right\|_{0,K} + h_K^{\frac{1}{2}} \left\| \Pi_p^{0,K_\mathbf{x}} \Pi_p^\star(u - q_p)(\cdot, t_{n-1}) \right\|_{0,K_\mathbf{x}}$$
$$= \left\| \Pi_{p-1}^{0,K}(u - q_p) \right\|_{0,K} + h_K^{\frac{1}{2}} \left\| \Pi_p^{0,K_\mathbf{x}}(u - q_p)(\cdot, t_{n-1}) \right\|_{0,K_\mathbf{x}}$$
$$\le \|u - q_p\|_{0,K} + h_K^{\frac{1}{2}} \|(u - q_p)(\cdot, t_{n-1})\|_{0,K_\mathbf{x}}$$
$$\lesssim \|u - q_p\|_{0,K} + h_K |u - q_p|_{1,K}.$$

Therefore, we obtain

$$(\partial_t \Pi_p^\star(u - q_p), v_h)_{0,K} \lesssim \left(h_K^{-1} \|u - q_p\|_{0,K} + |u - q_p|_{1,K} \right) \|v_h\|_{0,K}.$$

Finally, we estimate the third term on the right-hand side of (4.10). Since the initial condition $u(\cdot, 0)$ is zero, $\mathcal{J}^K(u, v_h) = 0$ if $n = 1$. So, we consider the case $n \ge 2$:

$$\mathcal{J}^K(u, v_h) = c_H(\Pi_p^\star u^{(n)}(\cdot, t_{n-1}) - \Pi_p^\star u^{(n-1)}(\cdot, t_{n-1}), v_h^{(n)}(\cdot, t_{n-1}))_{0,K_\mathbf{x}}$$
$$\lesssim h_K^{\frac{1}{2}} \left(\left\| u(\cdot, t_{n-1}) - \Pi_p^\star u^{(n)}(\cdot, t_{n-1}) \right\|_{0,K_\mathbf{x}} \right.$$
$$\left. + \left\| u(\cdot, t_{n-1}) - \Pi_p^\star u^{(n-1)}(\cdot, t_{n-1}) \right\|_{0,K_\mathbf{x}} \right) h_K^{-\frac{1}{2}} \left\| v_h^{(n)}(\cdot, t_{n-1}) \right\|_{0,K_\mathbf{x}}.$$

Proceeding as in Proposition 2.12, it is possible to show that

$$h_K^{-\frac{1}{2}} \left\| v_h^{(n)}(\cdot, t_{n-1}) \right\|_{0,K_\mathbf{x}} \lesssim |v_h|_{Y(K)}.$$

Thus, we can focus on the two terms involving u. As for the first one, we use a trace inequality along the time direction, add and subtract the same q_p as above, recall that $\Pi_p^\star$ preserves polynomials of degree at most p, use the triangle inequality, apply the polynomial inverse estimate (4.11), and get

$$h_K^{\frac{1}{2}} \left\| u(\cdot, t_{n-1}) - \Pi_p^\star u^{(n)}(\cdot, t_{n-1}) \right\|_{0,K_\mathbf{x}} \lesssim \left\| u - \Pi_p^\star u \right\|_{0,K} + h_K \left\| \partial_t(u - \Pi_p^\star u) \right\|_{0,K}$$
$$\le \|u - q_p\|_{0,K} + h_K \|\partial_t(u - q_p)\|_{0,K} + \left\| \Pi_p^\star(u - q_p) \right\|_{0,K} + h_K \left\| \partial_t(\Pi_p^\star(u - q_p)) \right\|_{0,K}$$
$$\le \|u - q_p\|_{0,K} + h_K \|\partial_t(u - q_p)\|_{0,K} + \left\| \Pi_p^\star(u - q_p) \right\|_{0,K}.$$

Next, we apply estimate (4.12) and get

$$h_K^{\frac{1}{2}} \left\| u(\cdot, t_{n-1}) - \Pi_p^\star u^{(n)}(\cdot, t_{n-1}) \right\|_{0,K_\mathbf{x}} \lesssim \|u - q_p\|_{0,K} + h_K |u - q_p|_{1,K}.$$

For the second term involving u, we proceed analogously. Setting $K' := K_{\mathbf{x}} \times I_{n-1}$ and using the local quasi-uniformity of the space–time mesh, we get

$$h_K^{\frac{1}{2}} \left\| u(\cdot, t_{n-1}) - \Pi_p^\star u^{(n-1)}(\cdot, t_{n-1}) \right\|_{0,K_{\mathbf{x}}} \lesssim \| u - q_p \|_{0,K'} + h_{K'} |u - q_p|_{1,K'}.$$

Summing over all the elements, using standard manipulations (including ℓ^2 Cauchy–Schwarz inequalities), and applying the global Poincaré type inequality (3.1) give

$$\text{(4.13)} \qquad \sum_{K \in \mathcal{T}_h} \left(b_h^K(u - q_p, v_h) + \mathcal{J}^K(q_p, v_h) \right)$$

$$\lesssim \sum_{K \in \mathcal{T}_h} \left(h_K^{-2} \| u - q_p \|_{L^2(K)}^2 + |u - q_p|_{1,K}^2 \right)^{\frac{1}{2}} \| v_h \|_{Y(\mathcal{T}_h)}.$$

Conclusion of Part (iv). From (4.9) and (4.13), which are valid for any $q_p \in V_h(K)$, and standard polynomial approximation results, we obtain

$$\sup_{0 \neq v_h \in Y_h} \inf_{q_p \in \mathcal{S}_p(\mathcal{T}_h)} \frac{\sum_{K \in \mathcal{T}_h} \left(b_h^K(u - q_p, v_h) - b^K(u - q_p, v_h) + \mathcal{J}^K(q_p, v_h) \right)}{\| v_h \|_{Y(\mathcal{T}_h)}} \lesssim h^p |u|_{p+1, \mathcal{T}_h}.$$

This concludes Part (iv) and completes the whole proof. $\qquad\qquad\square$

5. Numerical results. In this section, we assess the error estimates proven in Theorem 4.4. We developed an object-oriented MATLAB implementation to obtain high-order approximations of space–time (1+1)- and (2+1)-dimensional problems. We briefly mention some relevant computational aspects regarding the numerical results below.

- In case of inhomogeneous initial and/or boundary conditions, we set moments at $\Omega \times \{0\}$ and/or at $\partial\Omega \times (0,T)$ accordingly and modify the right-hand side. This corresponds to a standard lifting procedure, where the lifting has all the remaining moments equal to zero. In this way, in the presence of incompatible initial and boundary data, no artificial compatibility condition is enforced on the discrete solutions.

- In Theorem 4.4, error bounds are provided in the $\|\cdot\|_{X(\mathcal{T}_h)}$ norm. Since the virtual element solution u_h to (2.25) is not known in closed form and the error in the $X(\mathcal{T}_h)$ norm is not computable, we report the following associated error quantities:

(5.1a)
$$\mathcal{E}^Y := \left\| u - \Pi_p^N u_h \right\|_{Y(\mathcal{T}_h)}, \quad \mathcal{E}^N := \left\| \Pi_p^N (\mathfrak{N}_h \Pi_p^\star(u - u_h)) \right\|_{Y(\mathcal{T}_h)},$$

$$(\mathcal{E}^U)^2 := \frac{c_H}{2} \left(\left\| \Pi_p^\star(u - u_h)(\cdot, 0) \right\|_{L^2(\Omega)}^2 + \left\| \Pi_p^\star(u - u_h)(\cdot, T) \right\|_{L^2(\Omega)}^2 \right.$$

$$\left. + \sum_{n=2}^N \left\| \Pi_p^\star(u - u_h)^{(n)}(\cdot, t_{n-1}) - \Pi_p^\star(u - u_h)^{(n-1)}(\cdot, t_{n-1}) \right\|_{L^2(\Omega)}^2 \right).$$

The $X(\mathcal{T}_h)$ norm is related to the sum of $\mathcal{E}^Y$, $\mathcal{E}^N$, and $\mathcal{E}^U$. We also show the error in the $L^2(Q_T)$ norm, namely

(5.1b)
$$\mathcal{E}^L := \left\| u - \Pi_p^\star u_h \right\|_{L^2(Q_T)},$$

which is not covered by our theory.

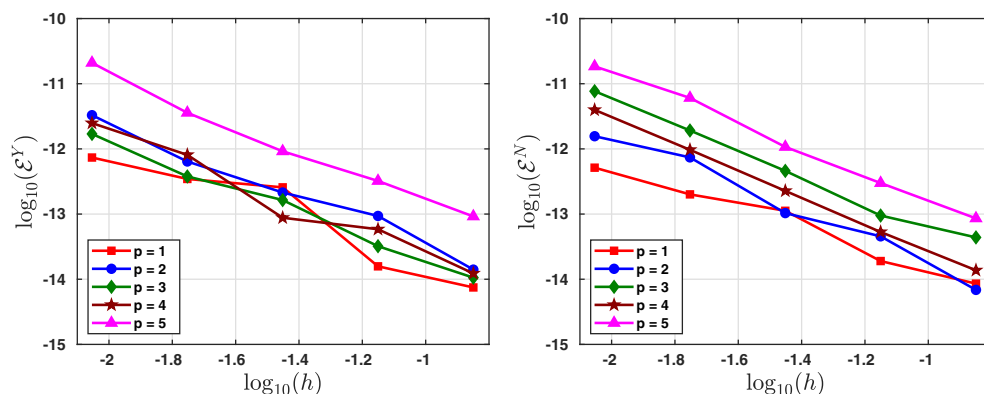- In all experiments, we take $c_H = \nu = 1$ and employ the stabilization in (2.24).

FIG. 2. *h-dependence of the errors in* (5.1) *for the patch tests with solution u_p in* (5.2).

5.1. Results in $(1+1)$-dimension. We use tensor-product meshes and uniform partitions along the space and time directions.

5.1.1. Patch test. The discrete bilinear form $b_h(\cdot,\cdot)$ in (2.23) is polynomial inconsistent; see Lemma 4.2. However, thanks to the error estimates (4.7), the method in (2.25) passes the patch test, i.e., up to round-off errors, polynomial solutions of order p are approximated exactly.

We consider the following family of exact solutions on $Q_T = (0,1) \times (0,1)$:

$$(5.2) \qquad u_p(x,t) = \begin{cases} t^{p/2}x^{p/2} & \text{if } p \text{ is even;} \\ t^{(p-1)/2}x^{(p+1)/2} + t^{(p+1)/2}x^{(p-1)/2} & \text{if } p \text{ is odd.} \end{cases}$$

For any $p \in \mathbb{N}$, u_p belongs to $\mathbb{P}_p(Q_T)$. In Figure 2, for $p = 1,\dots,5$, we show the errors in the approximation of u_p obtained using a sequence of meshes with $h_{K_\mathbf{x}} = h_{I_n} = 5 \times 10^{-2}/2^{i-1}$, $i = 1,\dots,4$, and approximation degree p. The scale of 10^{-10} in the figures validates the patch test. The growth of the error observed while decreasing the mesh size represents the actual effect of the condition number when solving the linear systems stemming from (2.25).

5.1.2. Smooth solution. On the space–time domain $Q_T = (0,1) \times (0,1)$, we consider the problem with exact smooth solution

$$(5.3) \qquad\qquad\qquad u(x,t) = \sin(t)\sin(3\pi x).$$

In Figure 3, we show the rates of convergence of the errors in (5.1) obtained using a sequence of meshes with $h_{K_\mathbf{x}} = h_{I_n} = 0.2 \times 2^{-i}$, for $i = 1,\dots,5$, and different approximation degrees p. We observe convergence of order $\mathcal{O}(h^p)$ for the error $\mathcal{E}^Y$, of order $\mathcal{O}(h^{p+\frac{1}{2}})$ for the error $\mathcal{E}^U$, and of order $\mathcal{O}(h^{p+1})$ for the errors $\mathcal{E}^N$ and $\mathcal{E}^L$. Such rates of convergence are in agreement with estimate (4.7) and the approximation rates that might be expected from the norms in (5.1).

5.1.3. Singular solutions. We assess the convergence of the method for solutions with finite Sobolev regularity. We use the same sequence of meshes as in subsection 5.1.1. For $Q_T = (0,1) \times (0,1)$ and $\alpha > -1/2$, we consider the singular solutions

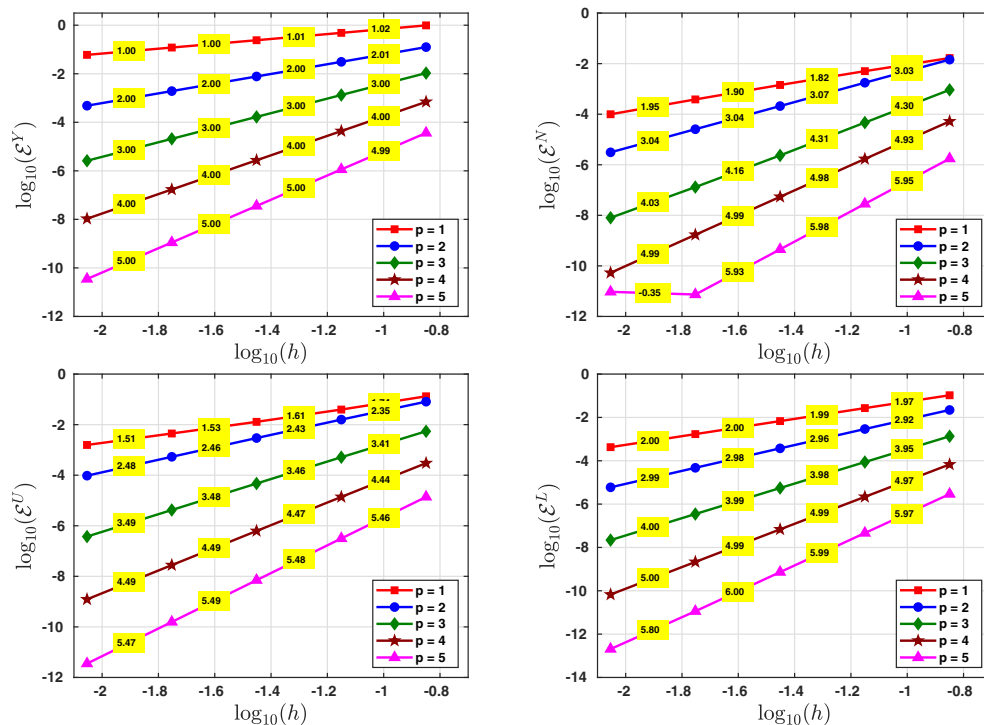$$(5.4) \qquad\qquad\qquad u_\alpha(x,t) = t^\alpha \sin(\pi x).$$

FIG. 3. *h-convergence of the errors in* (5.1) *for the test case with smooth solution* (5.3). *The numbers in the yellow rectangles denote the experimental orders of convergence. (Figure in color online.)*

We have that u_α and $\partial_x u_\alpha$ belong to $H^{\alpha+1/2-\epsilon}(0,1;\mathcal{C}^\infty(0,1))$ for any $\epsilon > 0$. The singularity occurs at the initial time. The errors in (5.1) are depicted in Figures 4 and 5 for $\alpha = 0.55$ and $\alpha = 0.75$. We observe convergence of order $\mathcal{O}(h^{\min\{p,\alpha+1/2\}})$ for the error $\mathcal{E}^Y$, of order $\mathcal{O}(h^{\alpha-\frac{1}{2}})$ for the error $\mathcal{E}^N$, of order $\mathcal{O}(h^\alpha)$ for the error $\mathcal{E}^U$, and of order $\mathcal{O}(h^{\alpha+\frac{1}{2}})$ for the error $\mathcal{E}^L$.

For a continuous finite element discretization of formulation (1.4), lower rates of convergence are obtained; see [12, sect. 7.5.3].

5.1.4. Incompatible initial and boundary conditions. On the space–time domain $Q_T = (0,1) \times (0,1)$, we consider the heat equation problem (1.1) with zero source term ($f = 0$), homogeneous Dirichlet boundary conditions ($u = 0$ on $\partial\Omega \times (0,T)$), and constant initial condition ($u = 1$ on $\Omega \times \{0\}$). The corresponding exact solution is given by the Fourier series

$$(5.5) \qquad u(x,t) = \sum_{n=0}^\infty \frac{4}{(2n+1)\pi} \sin((2n+1)\pi x) \exp\left(-(2n+1)^2\pi^2 t\right).$$

Due to the incompatibility of the initial and boundary conditions, u is discontinuous at $(0,0)$ and $(1,0)$, and does not belong to $H^1(Q_T)$ but belongs to $H^s\left(0,1;H_0^1(0,1)\right)$ for any $s < 1/4$; see [16, sect. 7.1]. Therefore, the rates of convergence obtained cannot be predicted by Theorem 4.4.

In Figure 6, we show the errors obtained with $p = 1, 2$ on a sequence of uniform Cartesian meshes for the proposed VEM and on a sequence of structured triangular
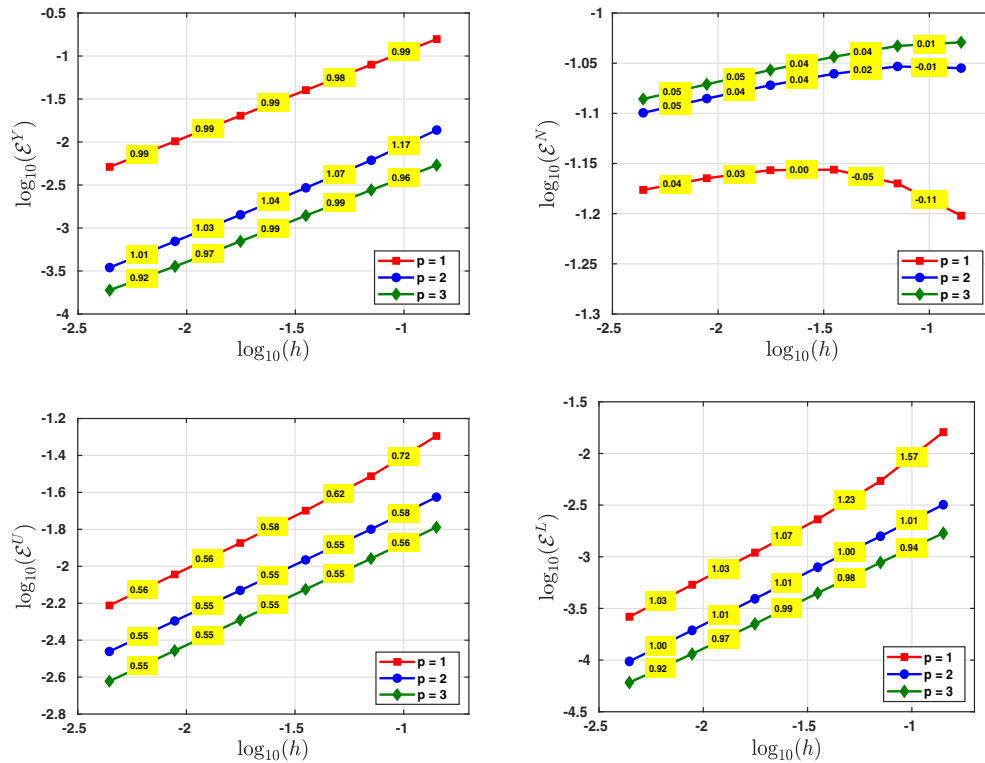
FIG. 4. *h-convergence of the errors in* (5.1) *for the test case with singular solution* u_α (5.4) *with* $\alpha = 0.55$.

meshes for the continuous finite element method in [18]. The continuous finite element method does not converge in the Y-norm, while the error $\mathcal{E}^Y$ of the proposed VEM converges with order $\mathcal{O}(h^{1/4})$. For the computation of the error, we truncate the series (5.5) at $n = 250$.

5.1.5. Increasing the degree of approximation. We are also interested in the performance of the p-version of the method, i.e., we fix a mesh and increase the degree of approximation. This is worth investigating also in view of the design of hp refinements. We consider the smooth solution test case from subsection 5.1.2 with a fixed mesh with $h_{I_n} = h_{K_\mathbf{x}} = 0.1$. The results shown in Figure 7 in *semilogy* scale. We observe the expected exponential convergence in terms of the square root of N_{DoFs} for all the VEM errors.

5.2. Results in (2+1)-dimension. We use tensor-product-in-time meshes and uniform partitions of the time interval $(0, T)$, and discretize the spatial domain Ω with sequences of quadrilateral meshes such as that in Figure 8 (left panel). We checked that the method passes the patch test also in the $(2+1)$-dimensional case. We do not report the results for the sake of brevity.

On $Q_T = (0, 1)^2 \times (0, 1)$, we consider

$$(5.6) \qquad u(\mathbf{x}, t) = \exp(-t) \sin(\pi x_1) \sin(\pi x_2).$$

In Figure 8 (right panel), we display the rates of convergence using different values of p and observe the expected rates of convergence for the error $\mathcal{E}^Y$.
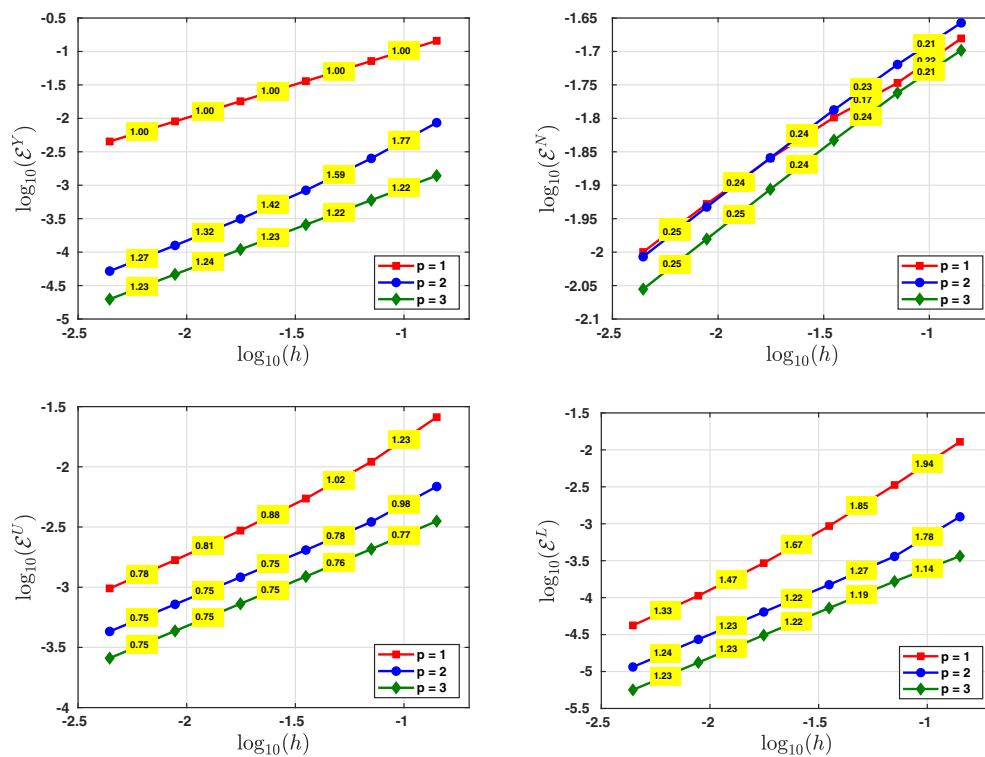
FIG. 5. *h-convergence of the errors in* (5.1) *for the test case with singular solution* u_α (5.4) *with* $\alpha = 0.75$.

(a) Proposed VEM   (b) Continuous finite element method
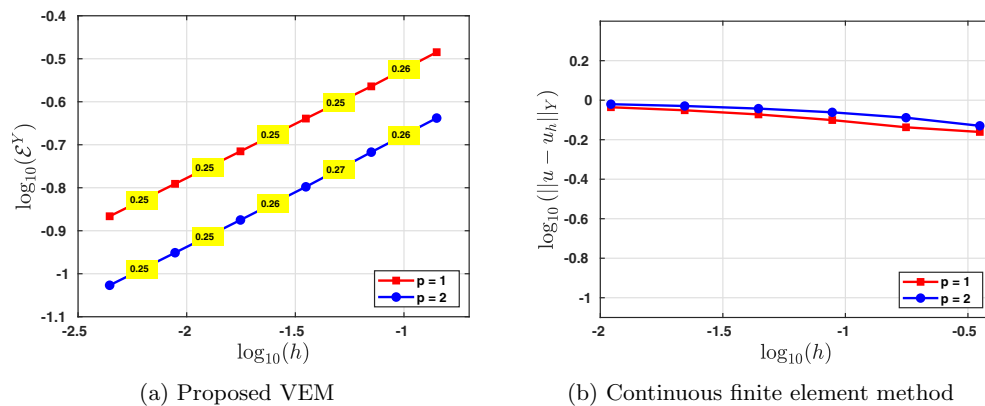
FIG. 6. *h-convergence for the test case with exact solution* (5.5) *with incompatible initial and boundary conditions.*

6. Conclusions. We designed and analyzed a space–time VEM for the heat equation based on a standard Petrov–Galerkin variational formulation. The advantages of using the proposed space–time VEM over standard space–time finite element methods are that it allows for decomposing the linear system stemming from the method into smaller systems associated with different time slabs, can be modified into a Trefftz variant, and permits the treatment of incompatible initial and boundary
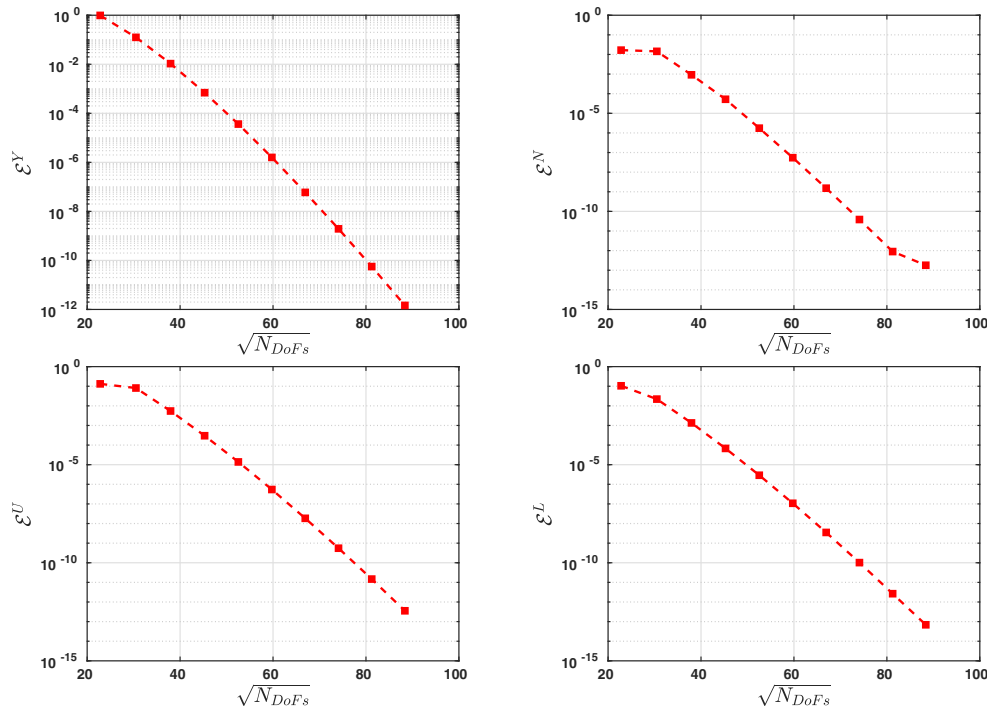
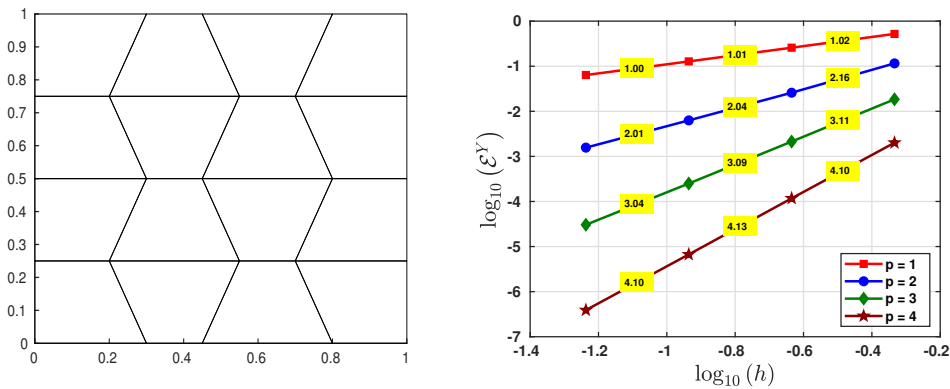FIG. 7. *p-convergence of the errors in* (5.1) *for the test case with smooth solution in* (5.3).

FIG. 8. Left panel*: example of mesh for the* 2*-dimensional spatial domain used in the numerical experiments.* Right panel: *h-convergence for the test case with smooth solution* (5.6).

data. We proved well posedness of the method and optimal a priori error estimates. Numerical results validate the expected rates of convergence.

In [11], the method introduced in this paper has been extended to more general prismatic meshes with hanging facets and variable degrees of accuracy, enabling the implementation of *hp*-adaptive mesh refinements. Tests of an adaptive procedure driven by a residual-type error indicator are also presented there.

Acknowledgments. The authors are very grateful to the anonymous reviewers for their careful reading of our manuscript, and for their valuable comments and suggestions.

REFERENCES

[1] R. ANDREEV, *Stability of sparse space-time finite element discretizations of linear parabolic evolution equations*, IMA J. Numer. Anal., 33 (2013), pp. 242–260.

[2] B. P. AYUSO DE DIOS, K. LIPNIKOV, AND G. MANZINI, *The nonconforming virtual element method*, ESAIM Math. Model. Numer. Anal., 50 (2016), pp. 879–904.

[3] L. BEIRÃO DA VEIGA, F. BREZZI, A. CANGIANI, G. MANZINI, L. MARINI, AND A. RUSSO, *Basic principles of virtual element methods*, Math. Models Methods Appl. Sci., 23 (2013), pp. 199–214.

[4] P. B. BOCHEV AND M. D. GUNZBURGER, *Least-squares Finite Element Methods*, Appl. Math. Sci. 166, Springer, New York, 2009.

[5] S. C. BRENNER, *Poincaré–Friedrichs inequalities for piecewise H^1 functions*, SIAM J. Numer. Anal., 41 (2003), pp. 306–324, https://doi.org/10.1137/S0036142902401311.

[6] A. CANGIANI, Z. DONG, AND E. H. GEORGOULIS, *hp-version space-time discontinuous Galerkin methods for parabolic problems on prismatic meshes*, SIAM J. Sci. Comput., 39 (2017), pp. A1251–A1279, https://doi.org/10.1137/16M1073285.

[7] R. DAUTRAY AND J.-L. LIONS, *Mathematical Analysis and Numerical Methods for Science and Technology, Vol. 5, Evolution Problems* I, Springer-Verlag, 1992.

[8] T. FÜHRER AND M. KARKULIK, *Space-time least-squares finite elements for parabolic equations*, Comput. Math. Appl., 92 (2021), pp. 27–36.

[9] G. GANTNER AND R. STEVENSON, *Further results on a space-time FOSLS formulation of parabolic PDEs*, ESAIM Math. Model. Numer. Anal., 55 (2021), pp. 283–299.

[10] G. GANTNER AND R. STEVENSON, *Improved Rates for a Space-Time FOSLS of Parabolic PDEs*, preprint, https://arxiv.org/abs/2208.10824, 2022.

[11] S. GÓMEZ, L. MASCOTTO, AND I. PERUGIA, *Design and performance of a space-time virtual element method for the heat equation on prismatic meshes*, Comput. Methods Appl. Mech. Eng., 418(A) (2024), 116491.

[12] S. GÓMEZ, *Nonconforming Space–time Methods for evolution PDEs*, Ph.D. thesis, University of Pavia, Pavia, Italy, 2023, in preparation.

[13] U. LANGER AND O. STEINBACH, *Space-Time Methods: Applications to Partial Differential Equations*, Radon Ser. Comput. Appl. Math. 25, de Gruyter, Berlin, 2019.

[14] J.-L. LIONS AND E. MAGENES, *Non-homogeneous Boundary Value Problems and Applications*, Vol. I, Die Grundlehren der mathematischen Wissenschaften, Band 181, Springer-Verlag, New York, Heidelberg, 1972.

[15] M. NEUMÜLLER, *Space-Time Methods: Fast Solvers and Applications*, Ph.D. thesis, TU Graz, Graz, Austria, 2013.

[16] D. SCHÖTZAU AND C. SCHWAB, *Time discretization of parabolic problems by the hp-version of the discontinuous Galerkin finite element method*, SIAM J. Numer. Anal., 38 (2000), pp. 837–875.

[17] C. SCHWAB AND R. STEVENSON, *Space-time adaptive wavelet methods for parabolic evolution problems*, Math. Comp., 78 (2009), pp. 1293–1318.

[18] O. STEINBACH, *Space-time finite element methods for parabolic problems*, Comput. Methods Appl. Math., 15 (2015), pp. 551–566.

[19] O. STEINBACH AND M. ZANK, *Coercive space-time finite element methods for initial boundary value problems*, Electron. Trans. Numer. Anal., 52 (2020), pp. 154–194.

[20] R. STEVENSON AND J. WESTERDIEP, *Minimal residual space-time discretizations of parabolic equations: Asymmetric spatial operators*, Comput. Math. Appl., 101 (2021), pp. 107–118.

[21] G. VACCA AND L. BEIRÃO DA VEIGA, *Virtual element methods for parabolic problems on polygonal meshes*, Numer. Methods Partial Differential Equations, 31 (2015), pp. 2110–2134.

[22] J. WLOKA, *Partial Differential Equations*, Cambridge University Press, Cambridge, 1987.