



HAL
open science

The unique continuation problem for the heat equation discretized with a high-order space-time nonconforming method

Erik Burman, Guillaume Delay, Alexandre Ern

► **To cite this version:**

Erik Burman, Guillaume Delay, Alexandre Ern. The unique continuation problem for the heat equation discretized with a high-order space-time nonconforming method. 2022. hal-03720960v1

HAL Id: hal-03720960

<https://hal.science/hal-03720960v1>

Preprint submitted on 12 Jul 2022 (v1), last revised 16 Oct 2023 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The unique continuation problem for the heat equation discretized with a high-order space-time nonconforming method*

Erik Burman[†] Guillaume Delay[‡] Alexandre Ern[§]

July 12, 2022

Abstract

We are interested in solving the unique continuation problem for the heat equation, i.e., we want to reconstruct the solution of the heat equation given its (noised) value in a subset of the computational domain. Both initial and boundary data can be unknown. We discretize this problem using a space-time discontinuous Galerkin method (including hybrid variables in space) and look for the solution that minimizes a discrete Lagrangian. We establish discrete inf-sup stability and bound the consistency error, leading to a priori estimates on the residual. Our main result, proving the convergence of the discrete solution to the exact solution, combines this a priori bound with a conditional stability estimate at the continuous level. The rate of convergence depends on the conditional stability, the approximation order in space and in time, and the size of the perturbations in data. Quite importantly, the weight of the regularization term depends on the time-step and the mesh-size, and we show how to choose it to preserve the best possible decay rates on the error. Finally, we run numerical simulations to assess the performance of the method in practice.

1 Introduction

In the present work, we are interested in solving numerically a data assimilation problem subject to the heat equation. In this problem, neither the boundary conditions nor the initial data are known. In order to compensate for the lack of initial and boundary data, we use the knowledge of the solution in a subdomain. We also investigate the influence of noise on this additional datum. Specifically, we consider a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$, a subset $\varpi \subset \Omega$, and a time interval $J := (0, T_f)$ with final time $T_f > 0$. Our goal is to approximate the function $u : J \times \Omega \rightarrow \mathbb{R}$ that satisfies

$$L(u) := \frac{\partial u}{\partial t} - \Delta u = f \quad \text{in } J \times \Omega, \tag{1}$$

$$u = g \quad \text{in } J \times \varpi, \tag{2}$$

*Draft version, July 12, 2022

[†]Department of Mathematics, University College London, London, UK-WC1E 6BT, UK.

[‡]Sorbonne Université, CNRS, Université de Paris, LJLL, F-75005 Paris, France.

[§]CERMICS, École des Ponts, 77455 Marne-la-Vallée cedex 2, and INRIA, Paris, France.

where $f \in L^2(J \times \Omega)$ is a given source term and $g \in H^1(J; H^1(\varpi)') \cap L^2(J; H^1(\varpi))$ is some datum resulting from measurements. The model problem (1)-(2) is ill-posed. Indeed, whenever a solution exists (this is the case whenever g satisfies the heat equation (1) in $J \times \varpi$), it is unique, but there is no a priori estimate on the solution in the usual Hadamard sense. Another difficulty is that we want to consider perturbed data g_δ instead of g to account for some noise in the measurements.

Although we do not have usual stability estimates, so-called conditional stability estimates are available. This type of estimate will play a key role in our error analysis. Conditional stability estimates essentially allow one to control the norm of a function in a target subdomain by means of weaker norms in a larger domain. Our analysis hinges on the following result.

Lemma 1 (Theorem 1 from [7]). *Let B be a connected subset of Ω such that $\varpi \subset B$ and $\overline{B} \subset \Omega$. Let $0 < T_1 < T_2 < T_f$. There exist $C_{\text{stb}} > 0$ and $\alpha \in (0, 1]$ such that for all $v \in H^1(J; H^{-1}(\Omega)) \cap L^2(J; H^1(\Omega))$, we have*

$$\begin{aligned} \|v\|_{L^2(T_1, T_2; H^1(B))} &\leq C_{\text{stb}} \left(\|v\|_{L^2(J; L^2(\varpi))} + \|L(v)\|_{L^2(J; H^{-1}(\Omega))} \right)^\alpha \\ &\quad \times \left(\|v\|_{L^2(J; L^2(\Omega))} + \|L(v)\|_{L^2(J; H^{-1}(\Omega))} \right)^{1-\alpha}. \end{aligned} \quad (3)$$

The conditional stability estimate from Lemma 1 will be used to prove the convergence of our approximation method. The constant α therein has an influence on the convergence rate. For instance, if h is the mesh size and τ the time step, if polynomials of degree $k \geq 1$ (resp., $\ell \geq 0$) are used for the space (resp., time) discretization, and provided the noise in the measurements is small enough, our main result establishes an error bound with decay rate $O((h^k + \tau^{\ell+\frac{1}{2}})^\alpha)$ in the target subdomain B (see Theorem 9). We also mention that Lemma 1 admits some interesting extensions. First, Lemma 1 still holds true if $\varpi \not\subset B$ as can be verified by using the arguments from [22]. Moreover, in the case of known boundary conditions (i.e., if we add the additional constraint $v = 0$ on $\partial\Omega$), then Lemma 1 remains valid with $T_2 = T_f$, $B = \Omega$, and $\alpha = 1$ (see Theorem 2 from [7]). This setting will be considered in our numerical experiments as well. Moreover, in the case of known initial conditions (i.e., if we add $v(0, \cdot) = v_0$ in Ω for some $v_0 \in L^2(\Omega)$), then Lemma 1 remains valid with $T_1 = 0$ provided $\|v_0\|_\Omega$ is added to $\|L(v)\|_{L^2(J; H^{-1}(\Omega))}$ in (3) [21]. Other stability estimates are available in the literature, for instance using Cauchy data on the boundary instead of interior data [19]. For an overview on analysis techniques of unique continuation for parabolic equations, we refer the reader to [25].

Since the problem (1)-(2) is ill-posed, a regularization must be considered to devise a reasonable approximation method. The usual approach is to regularize the continuous problem before embarking on any discretization method. Several regularization methods are available, for instance the quasi-reversibility method [2, 3, 14] or the Tikhonov regularization [17]. These methods have already been applied to solve the data assimilation problem subject to the heat equation in the one-dimensional case; see [24, 1]. Other numerical approaches have also been proposed and analyzed for the one-dimensional heat equation, see [18, 26], but, to the best of our knowledge, none of the above-mentioned references provides an error analysis balancing the stability of the regularization method, the conditional stability estimate, and the approximation order of the discretization in space and in time.

An alternative approach is to first discretize the ill-posed problem and then regularize it at the discrete level. The main advantage of discretizing first and then regularizing is that it makes it possible to design regularization terms that allow for a rigorous numerical analysis, leading to error estimates with rates that match the best possible rates deduced from the conditional stability estimates. The discretize-then-regularize approach has already been considered for instance in [4, 5, 8] for the stationary version of the present problem. The authors have also extended this method to the data assimilation problem subject to the Helmholtz equation using a high-order discretization [10]. Data assimilation subject to non-stationary problems was considered in [11] for the wave equation and in [7, 9] for the heat equation with known Dirichlet boundary data for the numerical tests. In particular, [7] considers the semi-discretization in space using lowest-order Lagrange finite elements, discussing several different stability situations, whereas [9] deals with the full discretization of the problem using the backward Euler scheme in time and still lowest-order Lagrange finite elements in space. In the latter reference, only the case of unknown initial data, but known boundary data, is considered. In [13], a method minimizing the dual norm of the PDE residual augmented by the least-squares error on data fitting was proposed, together with error estimates for the reconstruction problem. Also in this latter work, the boundary data are assumed to be known.

In the present work, we use higher-order methods for the discretization in space and in time. Specifically, we employ a discontinuous Galerkin (dG) method in time and a hybridized dG method in space (recall that such methods attach discrete unknowns to the mesh cells and to the mesh faces). The use of a dG method in time is the natural way to extend the backward Euler scheme to higher order. Furthermore, the use of a hybridized dG method in space reduces the number of stabilization terms that are needed with respect to standard finite elements (irrespective of the order of approximation). Indeed, as we shall see below (see the inf-sup condition from Lemma 3), the hybridized dG method naturally gives a control on the residuals without the need of any stabilization (which is needed for standard finite elements). Notice that, at the algebraic level, the discrete unknowns are globally coupled in time, as is anyway the case for all the methods to solve such problems.

To sum up, the added value of the present work is that we consider the data assimilation problem with unknown boundary conditions (and initial conditions) and that we devise a higher-order method in space and in time to discretize the above ill-posed problem. The salient feature of our approximation method is that the regularization term matches the decay rates of the approximation method. We also emphasize that we rigorously derive an a priori error estimate that decays at the best possible rate in view of the approximation capacities of the discretization and of the conditional stability of the continuous problem. Finally, we mention that the analysis entails various subtleties because some of the usual arguments in the analysis of time dG methods cannot be applied here. The reason for this is that, because of the ill-posed nature of the problem, we cannot invoke a plain coercivity argument that gives control in the $L^2(J; H_0^1(\Omega))$ -norm. As a consequence, we need to add a stabilization term controlling the time jumps of the discrete solution (and of its gradient) at all the discrete time nodes.

The rest of this work is organized as follows. In Section 2, we present the high-order discretization method in space and in time. In Section 3, we perform the error analysis through the following steps: inf-sup stability of the discrete problem, consistency error

bound, a priori bounds on the residual and some dual norm on the error. These steps prepare the stage for our main results, which are the error estimate (Theorem 7) and the devising of the regularization to achieve optimal error decay rates (Theorem 9). Finally, in Section 4, we present some numerical results that corroborate our theoretical results.

2 Problem discretization

In this section, we describe the space and time discretization, and we present the numerical scheme studied in the present work.

2.1 Time discretization

In what follows, for two integer numbers $p \leq q$, we use the notation $\{p:q\} := \{m \in \mathbb{N}, p \leq m \leq q\}$. We discretize the time interval J using a uniform time step $\tau := T_f/N$ (for simplicity), where $N \in \mathbb{N}^*$ is the number of time steps. We then have $0 = t_0 < t_1 < \dots < t_N = T_f$ where $t_n := n\tau$ for all $n \in \{0:N\}$. We define $I_n := (t_{n-1}, t_n)$ for all $n \in \{1:N\}$.

For a piecewise smooth function Φ^τ defined on J , we use the shorthand notation $\Phi^n := \Phi^\tau|_{I_n}$ for all $n \in \{1:N\}$. We define $\Phi^\tau(t_n^-) := \Phi^n(t_n)$, $\Phi^\tau(t_n^+) := \Phi^{n+1}(t_n)$ and $\llbracket \Phi^\tau \rrbracket^n := \Phi^\tau(t_n^+) - \Phi^\tau(t_n^-)$. For convenience, we also set $\llbracket \Phi^\tau \rrbracket^0 := 0$; this convention is motivated by the fact that there is no initial condition to enforce in the present problem. Finally, $\partial_t^\tau \phi^\tau$ denotes the broken time derivative of ϕ^τ such that $(\partial_t^\tau \phi^\tau)|_{I_n} := \partial_t \phi^n$ for all $n \in \{1:N\}$.

2.2 Space discretization

Let $(\mathcal{T}_h)_{h>0}$ be a family of matching meshes of Ω . In principle, the meshes can have cells that are polyhedra with planar faces in \mathbb{R}^d , and hanging nodes are also possible. However, the analysis below requires the mesh to be such that the underlying discontinuous polynomial approximation space has a global H^1 -conforming subspace with optimal approximation properties. For simplicity, we will therefore restrict the discussion to meshes composed of simplices (one can also readily consider meshes composed of cuboids). The mesh cells are conventionally taken to be open subsets of \mathbb{R}^d , and \mathbf{n}_T denotes the unit outward normal to the generic mesh cell $T \in \mathcal{T}_h$. For a subset $S \subset \mathbb{R}^d$, h_S denotes the diameter of S , and for a mesh \mathcal{T}_h , the index h refers to the maximal diameter of the mesh cells.

The mesh faces are collected in the set \mathcal{F}_h which is split into the set of the mesh interfaces, $\mathcal{F}_h^{\text{int}}$, and the set of the mesh boundary faces, \mathcal{F}_h^∂ . Any mesh interface $F \in \mathcal{F}_h^{\text{int}}$ is oriented by a fixed unit normal vector \mathbf{n}_F . Moreover, for a piecewise smooth function v and any mesh interface $F \in \mathcal{F}_h^{\text{int}}$, $\llbracket v \rrbracket_F$ denotes the jump of v across F in the direction of \mathbf{n}_F . We also use the broken gradient and Laplacian operators, $\nabla_{\mathcal{T}}$ and $\Delta_{\mathcal{T}}$, which are defined such that $(\nabla_{\mathcal{T}}\phi)|_T := \nabla(\phi|_T)$ and $(\Delta_{\mathcal{T}}\phi)|_T := \Delta(\phi|_T)$ for all $T \in \mathcal{T}_h$.

To avoid technicalities, we assume henceforth that the mesh family $(\mathcal{T}_h)_{h>0}$ is quasi-uniform. Therefore, we will use h to measure the diameter of any mesh cell or any mesh face. Moreover, we assume that all the meshes are fitted to the subset ϖ .

2.3 Discrete spaces and bilinear forms

Let $k \geq 1$ be the polynomial degree of the hybridized dG method in space and let $\ell \geq 0$ be the polynomial degree of the dG method in time. We denote by $\mathbb{P}^k(S)$ the set of polynomials of total degree at most k on the subset $S \subseteq \Omega$. Moreover, for a linear space U composed of functions defined on Ω , we denote by $\mathbb{P}^\ell(I; U)$ the set of U -valued polynomials of degree at most ℓ on $I \subseteq \bar{J} = [0, T_f]$.

The discrete unknowns in space are piecewise polynomials of degree k attached to the mesh cells and of the same degree k attached to the mesh faces. We define the discrete spaces

$$\hat{U}_h^k := U_{\mathcal{T}}^k \times U_{\mathcal{F}}^k, \quad U_{\mathcal{T}}^k := \chi_{T \in \mathcal{T}_h} \mathbb{P}^k(T), \quad U_{\mathcal{F}}^k := \chi_{F \in \mathcal{F}_h} \mathbb{P}^k(F). \quad (4)$$

For a generic pair $\hat{v}_h \in \hat{U}_h^k$, we write $\hat{v}_h := (v_{\mathcal{T}}, v_{\mathcal{F}})$ with $v_{\mathcal{T}} := (v_T)_{T \in \mathcal{T}_h} \in U_{\mathcal{T}}^k$ and $v_{\mathcal{F}} := (v_F)_{F \in \mathcal{F}_h} \in U_{\mathcal{F}}^k$. We denote by \hat{U}_{h0}^k the linear subspace of \hat{U}_h^k in which all the degrees of freedom attached to the mesh boundary faces are null. For a generic pair $\hat{v}_h \in \hat{U}_h^k$, its degrees of freedom associated with a generic mesh cell $T \in \mathcal{T}_h$ are denoted by

$$\hat{v}_T := (v_T, v_{\partial T} := (v_F)_{F \in \mathcal{F}_{\partial T}}) \in \hat{U}_T^k := \mathbb{P}^k(T) \times \mathbb{P}^k(\mathcal{F}_{\partial T}), \quad (5)$$

where $\mathbb{P}^k(\mathcal{F}_{\partial T}) := \chi_{F \in \mathcal{F}_{\partial T}} \mathbb{P}^k(F)$ and $\mathcal{F}_{\partial T} := \{F \in \mathcal{F}_h \mid F \subset \partial T\}$ collects the mesh faces composing the boundary of T . We introduce the space-time discrete spaces

$$\tilde{U}_h^\tau := \{\hat{v}_h^\tau \in L^2(J; \hat{U}_h^k) \mid \hat{v}_h^\tau|_{I_n} \in \mathbb{P}^\ell(I_n; \hat{U}_h^k), \forall n \in \{1:N\}\}, \quad (6)$$

$$\tilde{U}_{h0}^\tau := \{\hat{v}_h^\tau \in \tilde{U}_h^\tau \mid \hat{v}_h^\tau|_{I_n} \in \mathbb{P}^\ell(I_n; \hat{U}_{h0}^k), \forall n \in \{1:N\}\}. \quad (7)$$

Notice that we have $\tilde{U}_h^\tau = U_{\mathcal{T}}^\tau \times U_{\mathcal{F}}^\tau$ with

$$U_{\mathcal{T}}^\tau := \{v_{\mathcal{T}}^\tau \in L^2(J; U_{\mathcal{T}}^k) \mid v_{\mathcal{T}}^\tau := v_{\mathcal{T}}^\tau|_{I_n} \in \mathbb{P}^\ell(I_n; U_{\mathcal{T}}^k), \forall n \in \{1:N\}\}, \quad (8)$$

$$U_{\mathcal{F}}^\tau := \{v_{\mathcal{F}}^\tau \in L^2(J; U_{\mathcal{F}}^k) \mid v_{\mathcal{F}}^\tau := v_{\mathcal{F}}^\tau|_{I_n} \in \mathbb{P}^\ell(I_n; U_{\mathcal{F}}^k), \forall n \in \{1:N\}\}. \quad (9)$$

For a generic function $\hat{v}_h^\tau \in \tilde{U}_h^\tau$, we write $\hat{v}_h^\tau := (v_{\mathcal{T}}^\tau, v_{\mathcal{F}}^\tau)$ with $v_{\mathcal{T}}^\tau \in U_{\mathcal{T}}^\tau$ and $v_{\mathcal{F}}^\tau \in U_{\mathcal{F}}^\tau$.

We can now introduce the various bilinear forms needed to formulate the discrete problem. Let $\hat{v}_h^\tau, \hat{w}_h^\tau$ be generic functions in \tilde{U}_h^τ (primal variables) and let $\hat{\zeta}_h^\tau, \hat{\eta}_h^\tau$ be generic functions in \tilde{U}_{h0}^τ (dual variables). We use the subscript h to indicate the bilinear forms related to the space discretization, and the superscript τ to indicate those related to the time discretization. The two bilinear forms associated with the discretization of the heat equation are

$$a_h(\hat{v}_h^\tau, \hat{\eta}_h^\tau) := \sum_{T \in \mathcal{T}_h} \{(\nabla v_T^\tau, \nabla \eta_T^\tau)_{J \times T} - (\nabla v_T^\tau \cdot \mathbf{n}_T, \eta_T^\tau - \eta_{\partial T}^\tau)_{J \times \partial T} - (v_T^\tau - v_{\partial T}^\tau, \nabla \eta_T^\tau \cdot \mathbf{n}_T)_{J \times \partial T}\}, \quad (10)$$

$$b^\tau(v_{\mathcal{T}}^\tau, \eta_{\mathcal{T}}^\tau) := \sum_{n \in \{1:N\}} \{(\partial_t v_{\mathcal{T}}^n, \eta_{\mathcal{T}}^n)_{I_n \times \Omega} + (\llbracket v_{\mathcal{T}}^\tau \rrbracket^{n-1}, \eta_{\mathcal{T}}^n(t_{n-1}^+))_\Omega\}. \quad (11)$$

The stabilization bilinear forms read as follows:

$$s_h^\tau(\hat{v}_h^\tau, \hat{w}_h^\tau) := d_h(\hat{v}_h^\tau, \hat{w}_h^\tau) + d^\tau(v_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau), \quad (12)$$

$$\sigma_h(\hat{\zeta}_h^\tau, \hat{\eta}_h^\tau) := (\nabla_{\mathcal{T}} \zeta_{\mathcal{T}}^\tau, \nabla_{\mathcal{T}} \eta_{\mathcal{T}}^\tau)_{J \times \Omega} + d_h(\hat{\zeta}_h^\tau, \hat{\eta}_h^\tau), \quad (13)$$

with

$$d_h(\hat{v}_h^\tau, \hat{w}_h^\tau) := \sum_{T \in \mathcal{T}_h} h^{-1} (v_T - v_{\partial T}, w_T - w_{\partial T})_{J \times \partial T}, \quad (14)$$

$$d^\tau(v_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau) := \sum_{n \in \{1:N-1\}} \{ (\llbracket v_{\mathcal{T}}^\tau \rrbracket^n, \llbracket w_{\mathcal{T}}^\tau \rrbracket^n)_\Omega + (\llbracket \nabla_{\mathcal{T}} v_{\mathcal{T}}^\tau \rrbracket^n, \llbracket \nabla_{\mathcal{T}} w_{\mathcal{T}}^\tau \rrbracket^n)_\Omega \}. \quad (15)$$

(Notice that the stabilization bilinear form for the dual variable is only related to the space discretization.) Finally, the bilinear forms associated with the regularization and the measurements are

$$t_h^\tau(v_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau) := \gamma c(\tau, h)^2 (v_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau)_{J \times \Omega}, \quad (16)$$

$$m_\varpi(v_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau) := (v_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau)_{J \times \varpi}, \quad (17)$$

where $\gamma > 0$ and the value of $c(\tau, h)$ will result from the error analysis (we shall obtain $c(\tau, h) = h^k + \tau^{\ell + \frac{1}{2}}$, see (56)). Finally, for later use, we define

$$\|w_{\mathcal{T}}^\tau\|_{J \times \varpi}^2 := m_\varpi(w_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau), \quad \ell(w_{\mathcal{T}}^\tau) := (f, w_{\mathcal{T}}^\tau)_{J \times \Omega}. \quad (18)$$

2.4 Lagrangian and discrete problem

We want to find the saddle-point of the Lagrangian defined for all $(\hat{v}_h^\tau, \hat{\zeta}_h^\tau) \in \tilde{U}_h^\tau \times \tilde{U}_{h0}^\tau$ by

$$\begin{aligned} \mathcal{L}_h^\tau(\hat{v}_h^\tau, \hat{\zeta}_h^\tau) := & \frac{1}{2} \|v_{\mathcal{T}}^\tau - g_\delta\|_{J \times \varpi}^2 + \frac{1}{2} t_h^\tau(v_{\mathcal{T}}^\tau, v_{\mathcal{T}}^\tau) + \frac{1}{2} s_h^\tau(\hat{v}_h^\tau, \hat{v}_h^\tau) - \frac{1}{2} \sigma_h(\hat{\zeta}_h^\tau, \hat{\zeta}_h^\tau) \\ & + a_h(\hat{v}_h^\tau, \hat{\zeta}_h^\tau) + b^\tau(v_{\mathcal{T}}^\tau, \zeta_{\mathcal{T}}^\tau) - \ell(\zeta_{\mathcal{T}}^\tau), \end{aligned} \quad (19)$$

where $g_\delta := g + \delta$ denotes the available perturbed measurement of g . Notice that there is no boundary condition on the primal variable, whereas there is a Dirichlet boundary condition on the dual variable.

The discrete problem is derived by seeking a critical point of the Lagrangian and reads as follows: Find $(\hat{u}_h^\tau, \hat{\xi}_h^\tau) \in \tilde{U}_h^\tau \times \tilde{U}_{h0}^\tau$ such that

$$m_\varpi(u_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau) + t_h^\tau(u_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau) + s_h^\tau(\hat{u}_h^\tau, \hat{w}_h^\tau) + a_h(\hat{w}_h^\tau, \hat{\xi}_h^\tau) + b^\tau(w_{\mathcal{T}}^\tau, \xi_{\mathcal{T}}^\tau) = m_\varpi(g_\delta, w_{\mathcal{T}}^\tau), \quad (20)$$

$$a_h(\hat{u}_h^\tau, \hat{\eta}_h^\tau) + b^\tau(u_{\mathcal{T}}^\tau, \eta_{\mathcal{T}}^\tau) - \sigma_h(\hat{\xi}_h^\tau, \hat{\eta}_h^\tau) = \ell(\eta_{\mathcal{T}}^\tau), \quad (21)$$

where the first equation holds for all $\hat{w}_h^\tau \in \tilde{U}_h^\tau$ and the second for all $\hat{\eta}_h^\tau \in \tilde{U}_{h0}^\tau$. For all $(\hat{v}_h^\tau, \hat{\zeta}_h^\tau)$ and $(\hat{w}_h^\tau, \hat{\eta}_h^\tau)$ in $\tilde{U}_h^\tau \times \tilde{U}_{h0}^\tau$, we define the bilinear form

$$\begin{aligned} A_h^\tau((\hat{v}_h^\tau, \hat{\zeta}_h^\tau), (\hat{w}_h^\tau, \hat{\eta}_h^\tau)) := & m_\varpi(v_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau) + t_h^\tau(v_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau) + s_h^\tau(\hat{v}_h^\tau, \hat{w}_h^\tau) + a_h(\hat{w}_h^\tau, \hat{\zeta}_h^\tau) \\ & + b^\tau(w_{\mathcal{T}}^\tau, \zeta_{\mathcal{T}}^\tau) + b^\tau(v_{\mathcal{T}}^\tau, \eta_{\mathcal{T}}^\tau) + a_h(\hat{v}_h^\tau, \hat{\eta}_h^\tau) - \sigma_h(\hat{\zeta}_h^\tau, \hat{\eta}_h^\tau). \end{aligned} \quad (22)$$

The discrete problem (20)-(21) can be rewritten as follows: Find $(\hat{u}_h^\tau, \hat{\xi}_h^\tau) \in \tilde{U}_h^\tau \times \tilde{U}_{h0}^\tau$ such that

$$A_h^\tau((\hat{u}_h^\tau, \hat{\xi}_h^\tau), (\hat{w}_h^\tau, \hat{\eta}_h^\tau)) = m_\varpi(g_\delta, w_{\mathcal{T}}^\tau) + \ell(\eta_{\mathcal{T}}^\tau), \quad \forall (\hat{w}_h^\tau, \hat{\eta}_h^\tau) \in \tilde{U}_h^\tau \times \tilde{U}_{h0}^\tau.$$

3 Analysis

This section is organized as follows. We first introduce a time reconstruction operator to rewrite the bilinear form b^τ . This operator is classical in the context of dG methods in time; see, e.g., [16, Section 69.2.3] or [23, Section 2.3] and the references therein. Then, we study the stability properties of A_h^τ in a suitable residual norm. Next, we introduce interpolation operators in space and in time and we bound the consistency error. This allows us to bound the residual in various norms. We combine these bounds with the abstract conditional estimate from Lemma 1 to derive error estimates in the target subdomain B introduced therein. Finally, we state the approximation properties of the interpolation operator, we tune the size of the stabilization parameter and we establish the convergence rates for the method.

In what follows, we use the convention $A \lesssim B$ to abbreviate the inequality $A \leq CB$ for positive real numbers A and B , where the constant $C > 0$ does not depend on h, τ , the solution of the continuous and discrete problems. Unless explicitly specified, the constant C is also independent of the parameter γ .

3.1 Time reconstruction operator

For all $v_\mathcal{T}^\tau \in U_\mathcal{T}^\tau$, its time reconstruction $R^\tau(v_\mathcal{T}^\tau) \in C^0(J; U_\mathcal{T}^k)$ is defined such that $R^\tau(v_\mathcal{T}^\tau)(t_0^+) := v_\mathcal{T}^\tau(t_0^+)$ and such that, for all $n \in \{1:N\}$, $R^n(v_\mathcal{T}^\tau) := R^\tau(v_\mathcal{T}^\tau)|_{I_n} \in \mathbb{P}^{\ell+1}(I_n; U_\mathcal{T}^k)$ satisfies

$$(\partial_t R^n(v_\mathcal{T}^\tau), q_\mathcal{T}^n)_{I_n \times \Omega} := (\partial_t v_\mathcal{T}^n, q_\mathcal{T}^n)_{I_n \times \Omega} + (\llbracket v_\mathcal{T}^\tau \rrbracket^{n-1}, q_\mathcal{T}^n(t_{n-1}^+))_\Omega, \quad (23)$$

for all $q_\mathcal{T}^n \in \mathbb{P}^\ell(I_n; U_\mathcal{T}^k)$. Since the function $R^\tau(v_\mathcal{T}^\tau)$ is continuous in time, its time derivative is well defined over the whole time interval J . The main consequence of (23) is that the bilinear form b^τ can be rewritten as

$$b^\tau(v_\mathcal{T}^\tau, \eta_\mathcal{T}^\tau) = (\partial_t R^\tau(v_\mathcal{T}^\tau), \eta_\mathcal{T}^\tau)_{J \times \Omega}, \quad \forall (\hat{v}_\mathcal{T}^\tau, \hat{\eta}_\mathcal{T}^\tau) \in \tilde{U}_h^\tau \times \tilde{U}_{h0}^\tau. \quad (24)$$

Moreover, it is well-known that the reconstruction operator can be rewritten as

$$R^n(v_\mathcal{T}^\tau)(t, \mathbf{x}) = v_\mathcal{T}^n(t, \mathbf{x}) - \llbracket v_\mathcal{T}^\tau \rrbracket^{n-1}(\mathbf{x}) \frac{(-1)^\ell}{2} (L_\ell - L_{\ell+1}) \circ T_n^{-1}(t), \quad (25)$$

for all $(t, \mathbf{x}) \in I_n \times \Omega$, where L_ℓ is the Legendre polynomial of degree ℓ defined on $(-1, 1)$ and T_n is the affine mapping from $(-1, 1)$ to I_n . A consequence of (25) is that

$$R^n(v_\mathcal{T}^\tau)(t_n^-) = v_\mathcal{T}^n(t_n^-), \quad \forall n \in \{1:N\}. \quad (26)$$

The following stability properties are useful for the present analysis (the proof is outlined in Section 3.8.1).

Lemma 2 (Stability properties of R^τ). *For all $v_\mathcal{T}^\tau \in U_\mathcal{T}^\tau$, we have*

$$\|\partial_t^\tau (R^\tau(v_\mathcal{T}^\tau) - v_\mathcal{T}^\tau)\|_{J \times \Omega} \lesssim \tau^{-\frac{1}{2}} d^\tau(v_\mathcal{T}^\tau, v_\mathcal{T}^\tau)^{\frac{1}{2}}, \quad (27)$$

$$\|R^\tau(v_\mathcal{T}^\tau) - v_\mathcal{T}^\tau\|_{J \times \Omega} + \|\nabla_\mathcal{T} (R^\tau(v_\mathcal{T}^\tau) - v_\mathcal{T}^\tau)\|_{J \times \Omega} \lesssim \tau^{\frac{1}{2}} d^\tau(v_\mathcal{T}^\tau, v_\mathcal{T}^\tau)^{\frac{1}{2}}. \quad (28)$$

Moreover, for all $\hat{v}_h^\tau = (v_{\mathcal{T}}^\tau, v_{\mathcal{F}}^\tau) \in \tilde{U}_h^\tau$, we have

$$\|h^{-\frac{1}{2}} \llbracket R^\tau(v_{\mathcal{T}}^\tau) \rrbracket_{\mathcal{F}_h^{\text{int}}} \|_{J \times \mathcal{F}_h^{\text{int}}} := \left(\sum_{F \in \mathcal{F}_h^{\text{int}}} h^{-1} \|\llbracket R^\tau(v_{\mathcal{T}}^\tau) \rrbracket_F\|_{J \times F}^2 \right)^{\frac{1}{2}} \lesssim d_h(\hat{v}_h^\tau, \hat{v}_h^\tau)^{\frac{1}{2}}. \quad (29)$$

3.2 Residual stability

For all $(\hat{v}_h^\tau, \hat{\zeta}_h^\tau) \in \tilde{U}_h^\tau \times \tilde{U}_{h0}^\tau$, we define the residual norm

$$\|\hat{v}_h^\tau, \hat{\zeta}_h^\tau\|^2 := \|v_{\mathcal{T}}^\tau\|_{\mathbb{R}}^2 + \|v_{\mathcal{T}}^\tau\|_{J \times \varpi}^2 + t_h^\tau(v_{\mathcal{T}}^\tau, v_{\mathcal{T}}^\tau) + s_h^\tau(\hat{v}_h^\tau, \hat{v}_h^\tau) + \sigma_h(\hat{\zeta}_h^\tau, \hat{\zeta}_h^\tau), \quad (30)$$

with

$$\|v_{\mathcal{T}}^\tau\|_{\mathbb{R}}^2 := \|h(\partial_t R^\tau(v_{\mathcal{T}}^\tau) - \Delta_{\mathcal{T}} v_{\mathcal{T}}^\tau)\|_{J \times \Omega}^2 + \|h^{\frac{1}{2}} \llbracket \nabla v_{\mathcal{T}}^\tau \rrbracket_{\mathcal{F}_h^{\text{int}}} \cdot \mathbf{n}_{\mathcal{F}_h^{\text{int}}}\|_{J \times \mathcal{F}_h^{\text{int}}}^2, \quad (31)$$

and $\|h^{\frac{1}{2}} \llbracket \nabla v_{\mathcal{T}}^\tau \rrbracket_{\mathcal{F}_h^{\text{int}}} \cdot \mathbf{n}_{\mathcal{F}_h^{\text{int}}}\|_{J \times \mathcal{F}_h^{\text{int}}}^2 := \sum_{F \in \mathcal{F}_h^{\text{int}}} h \|\llbracket \nabla v_{\mathcal{T}}^\tau \rrbracket_F \cdot \mathbf{n}_F\|_{J \times F}^2$. Our key stability result is the following inf-sup condition.

Lemma 3 (Inf-sup condition). *The following holds for all $(\hat{v}_h^\tau, \hat{\zeta}_h^\tau) \in \tilde{U}_h^\tau \times \tilde{U}_{h0}^\tau$,*

$$\|\hat{v}_h^\tau, \hat{\zeta}_h^\tau\| \lesssim \sup_{(\hat{w}_h^\tau, \hat{\eta}_h^\tau) \in \tilde{U}_h^\tau \times \tilde{U}_{h0}^\tau \setminus \{(0,0)\}} \frac{A_h^\tau((\hat{v}_h^\tau, \hat{\zeta}_h^\tau), (\hat{w}_h^\tau, \hat{\eta}_h^\tau))}{\|\hat{w}_h^\tau, \hat{\eta}_h^\tau\|}. \quad (32)$$

Proof. Let us denote by S the right-hand side of (32).

(i) We first use the test functions $\hat{w}_h^\tau := \hat{v}_h^\tau$ and $\hat{\eta}_h^\tau := -\hat{\zeta}_h^\tau$ to get

$$\begin{aligned} \|v_{\mathcal{T}}^\tau\|_{J \times \varpi}^2 + t_h^\tau(v_{\mathcal{T}}^\tau, v_{\mathcal{T}}^\tau) + s_h^\tau(\hat{v}_h^\tau, \hat{v}_h^\tau) + \sigma_h(\hat{\zeta}_h^\tau, \hat{\zeta}_h^\tau) &= A_h^\tau((\hat{v}_h^\tau, \hat{\zeta}_h^\tau), (\hat{v}_h^\tau, -\hat{\zeta}_h^\tau)) \\ &\leq S \|\hat{v}_h^\tau, \hat{\zeta}_h^\tau\|. \end{aligned}$$

(ii) Let $\hat{\eta}_h^\tau := (0, (\eta_F^\tau)_{F \in \mathcal{F}_h})$ with $\eta_F^\tau := h \llbracket \nabla v_{\mathcal{T}}^\tau \rrbracket_F \cdot \mathbf{n}_F$ for all $F \in \mathcal{F}_h^{\text{int}}$ and $\eta_F^\tau := 0$ for all $F \in \mathcal{F}_h^\partial$. Since $\|0, \hat{\eta}_h^\tau\| = \sigma_h(\hat{\eta}_h^\tau, \hat{\eta}_h^\tau)^{\frac{1}{2}}$, we have

$$\begin{aligned} \|h^{\frac{1}{2}} \llbracket \nabla v_{\mathcal{T}}^\tau \rrbracket_{\mathcal{F}_h^{\text{int}}} \cdot \mathbf{n}_{\mathcal{F}_h^{\text{int}}}\|_{J \times \mathcal{F}_h^{\text{int}}}^2 &= a_h(\hat{v}_h^\tau, \hat{\eta}_h^\tau) \\ &= A_h^\tau((\hat{v}_h^\tau, \hat{\zeta}_h^\tau), (0, \hat{\eta}_h^\tau)) + \sigma_h(\hat{\zeta}_h^\tau, \hat{\eta}_h^\tau) \\ &\leq S \sigma_h(\hat{\eta}_h^\tau, \hat{\eta}_h^\tau)^{\frac{1}{2}} + \sigma_h(\hat{\zeta}_h^\tau, \hat{\eta}_h^\tau) \\ &\leq (S + \sigma_h(\hat{\zeta}_h^\tau, \hat{\zeta}_h^\tau)^{\frac{1}{2}}) \sigma_h(\hat{\eta}_h^\tau, \hat{\eta}_h^\tau)^{\frac{1}{2}}. \end{aligned}$$

Moreover, we also have

$$\sigma_h(\hat{\eta}_h^\tau, \hat{\eta}_h^\tau) = d_h(\hat{\eta}_h^\tau, \hat{\eta}_h^\tau) = \sum_{T \in \mathcal{T}_h} h^{-1} \|\eta_{\partial T}^\tau\|_{J \times \partial T}^2 \lesssim \|h^{\frac{1}{2}} \llbracket \nabla v_{\mathcal{T}}^\tau \rrbracket_{\mathcal{F}_h^{\text{int}}} \cdot \mathbf{n}_{\mathcal{F}_h^{\text{int}}}\|_{J \times \mathcal{F}_h^{\text{int}}}^2.$$

This implies that

$$\|h^{\frac{1}{2}} \llbracket \nabla v_{\mathcal{T}}^\tau \rrbracket_{\mathcal{F}_h^{\text{int}}} \cdot \mathbf{n}_{\mathcal{F}_h^{\text{int}}}\|_{J \times \mathcal{F}_h^{\text{int}}}^2 \lesssim S^2 + \sigma_h(\hat{\zeta}_h^\tau, \hat{\zeta}_h^\tau).$$

(iii) We now consider $\hat{\eta}_h^\tau := (\eta_\tau^\tau, 0)$ with $\eta_\tau^\tau := h^2(\partial_t R^\tau(v_\tau^\tau) - \Delta_{\mathcal{T}} v_\tau^\tau)$. We have

$$\begin{aligned} & (\partial_t R^\tau(v_\tau^\tau), \eta_\tau^\tau)_{J \times \Omega} + a_h^\tau(\hat{v}_h^\tau, \hat{\eta}_h^\tau) \\ &= (\partial_t R^\tau(v_\tau^\tau), \eta_\tau^\tau)_{J \times \Omega} + \sum_{T \in \mathcal{T}_h} \{ (\nabla v_T^\tau, \nabla \eta_T^\tau)_{J \times T} - (\nabla v_T^\tau \cdot \mathbf{n}_T, \eta_T^\tau)_{J \times \partial T} \\ & \quad - (v_T^\tau - v_{\partial T}^\tau, \nabla \eta_T^\tau \cdot \mathbf{n}_T)_{J \times \partial T} \} \\ &= (\partial_t R^\tau(v_\tau^\tau) - \Delta_{\mathcal{T}} v_\tau^\tau, \eta_\tau^\tau)_{J \times \Omega} - \sum_{T \in \mathcal{T}_h} (v_T^\tau - v_{\partial T}^\tau, \nabla \eta_T^\tau \cdot \mathbf{n}_T)_{J \times \partial T}. \end{aligned}$$

Using the definition of $\hat{\eta}_h^\tau$ and recalling the rewriting (24) of b^τ , we have

$$\begin{aligned} & \|h(\partial_t R^\tau(v_\tau^\tau) - \Delta_{\mathcal{T}} v_\tau^\tau)\|_{J \times \Omega}^2 \\ &= a_h(\hat{v}_h^\tau, \hat{\eta}_h^\tau) + b^\tau(v_\tau^\tau, \eta_\tau^\tau) + \sum_{T \in \mathcal{T}_h} (v_T^\tau - v_{\partial T}^\tau, \nabla \eta_T^\tau \cdot \mathbf{n}_T)_{J \times \partial T} \\ &= A_h^\tau((\hat{v}_h^\tau, \hat{\zeta}_h^\tau), (0, \hat{\eta}_h^\tau)) + \sigma_h(\hat{\zeta}_h^\tau, \hat{\eta}_h^\tau) + \sum_{T \in \mathcal{T}_h} (v_T^\tau - v_{\partial T}^\tau, \nabla \eta_T^\tau \cdot \mathbf{n}_T)_{J \times \partial T} \\ &\lesssim S \|0, \hat{\eta}_h^\tau\| + (\sigma_h(\hat{\zeta}_h^\tau, \hat{\zeta}_h^\tau) + d_h(\hat{v}_h^\tau, \hat{v}_h^\tau))^{\frac{1}{2}} \sigma_h(\hat{\eta}_h^\tau, \hat{\eta}_h^\tau)^{\frac{1}{2}}, \end{aligned}$$

since a discrete trace inverse inequality implies that

$$\sum_{T \in \mathcal{T}_h} h \|\nabla \eta_T^\tau \cdot \mathbf{n}_T\|_{J \times \partial T}^2 \lesssim \|\nabla \eta_\tau^\tau\|_{J \times \Omega}^2 \leq \sigma_h(\hat{\eta}_h^\tau, \hat{\eta}_h^\tau).$$

Furthermore, invoking again inverse inequalities, we have

$$\begin{aligned} \sigma_h(\hat{\eta}_h^\tau, \hat{\eta}_h^\tau) &= \sum_{T \in \mathcal{T}_h} \left(\|\nabla \eta_T^\tau\|_{J \times T}^2 + h^{-1} \|\eta_T^\tau\|_{J \times \partial T}^2 \right) \\ &\lesssim \|h^{-1} \eta_\tau^\tau\|_{J \times \Omega}^2 = \|h(\partial_t R^\tau(v_\tau^\tau) - \Delta_{\mathcal{T}} v_\tau^\tau)\|_{J \times \Omega}^2. \end{aligned}$$

This implies that

$$\|h(\partial_t R^\tau(v_\tau^\tau) - \Delta_{\mathcal{T}} v_\tau^\tau)\|_{J \times \Omega}^2 \lesssim S^2 + d_h(\hat{v}_h^\tau, \hat{v}_h^\tau) + \sigma_h(\hat{\zeta}_h^\tau, \hat{\zeta}_h^\tau).$$

Gathering the previous estimates leads to the expected inf-sup condition. \square

3.3 Interpolation operator and error decomposition

In this section, we define the space-time interpolation operator used in the error analysis. Its definition is motivated by orthogonality properties. To facilitate the reading, the approximation properties of this interpolation operator are discussed later in Section 3.7.

Let us first consider the approximation in time. For all $v \in H^1(J; H^1(\Omega))$ and all $n \in \{1:N\}$, we define $\check{I}_n^\ell(v) \in \mathbb{P}^\ell(I_n; H^1(\Omega))$ by

$$\check{I}_n^\ell(v)(t_n^-) := v(t_n), \quad (33)$$

$$\int_{I_n} (\check{I}_n^\ell(v) - v, q^n)_\Omega := 0, \quad \forall q^n \in \mathbb{P}^{\ell-1}(I_n; H^1(\Omega)). \quad (34)$$

Furthermore, the approximation in space is realized by using L^2 -orthogonal projections. Let $y \in H^1(\Omega)$. For all $T \in \mathcal{T}_h$, we define $\Pi_T^k(y)$ as the L^2 -orthogonal projection of $y|_T$ onto $\mathbb{P}^k(T)$, i.e., $(\Pi_T^k(y), q_T)_T := (y, q_T)_T$ for all $q_T \in \mathbb{P}^k(T)$. For all $F \in \mathcal{F}_h$, we define $\Pi_F^k(y)$ as the L^2 -orthogonal projection of $y|_F$ onto $\mathbb{P}^k(F)$, i.e., $(\Pi_F^k(y), q_F)_F := (y, q_F)_F$ for all $q_F \in \mathbb{P}^k(F)$.

We can now define our space-time interpolation operator by setting, for all $v \in H^1(J; H^1(\Omega))$,

$$\hat{I}_h^\tau(v) := (I_{\mathcal{T}}^\tau(v), I_{\mathcal{F}}^\tau(v)) \in \tilde{U}_h^\tau,$$

where $I_{\mathcal{T}}^\tau(v) \in U_{\mathcal{T}}^\tau$ and $I_{\mathcal{F}}^\tau(v) \in U_{\mathcal{F}}^\tau$ are such that, for all $n \in \{1:N\}$, all $T \in \mathcal{T}_h$, and all $F \in \mathcal{F}_h$,

$$I_T^n(v) := I_T^\tau(v)|_{I_n} := \Pi_T^k(\check{I}_n^\ell(v)), \quad I_F^n(v) := I_F^\tau(v)|_{I_n} := \Pi_F^k(\check{I}_n^\ell(v)). \quad (35)$$

We notice that the approximation operators in space and in time commute. Moreover, proceeding as in [16, Lemma 69.16], we derive the following useful orthogonality property: For all $v \in H^1(J; H^1(\Omega))$ and all $w_{\mathcal{T}}^\tau \in \mathbb{P}^\ell(J; U_{\mathcal{T}}^k)$,

$$(\partial_t v - \partial_t R^\tau(I_{\mathcal{T}}^\tau(v)), w_{\mathcal{T}}^\tau)_{J \times \Omega} = 0. \quad (36)$$

Let us finally define the discrete and interpolation errors on the primal unknown. Recall that u denotes the solution to the exact problem (1)-(2) and that $(\hat{u}_h^\tau, \hat{\xi}_h^\tau)$ denotes the solution to the discrete problem (20)-(21). The discrete and interpolation errors on the primal unknown are then defined as

$$\hat{e}_h^\tau := \hat{u}_h^\tau - \hat{I}_h^\tau(u), \quad \hat{\theta}_h^\tau := (u, u|_{\mathcal{F}_h}) - \hat{I}_h^\tau(u), \quad (37)$$

so that we have $\hat{\theta}_h^\tau := ((\theta_T^\tau)_{T \in \mathcal{T}_h}, (\theta_F^\tau)_{F \in \mathcal{F}_h})$ with $\theta_T^\tau := u|_T - I_T^\tau(u)$ for all $T \in \mathcal{T}_h$, and $\theta_F^\tau := u|_F - I_F^\tau(u)$ for all $F \in \mathcal{F}_h$.

3.4 Consistency and a priori residual bound

We now bound the consistency error in the discrete formulation. To this purpose, we consider the norm

$$\begin{aligned} \|\hat{\theta}_h^\tau\|_{\#}^2 &:= \|\nabla_{\mathcal{T}} \theta_{\mathcal{T}}^\tau\|_{J \times \Omega}^2 + \|h \Delta_{\mathcal{T}} \theta_{\mathcal{T}}^\tau\|_{J \times \Omega}^2 + \|\theta_{\mathcal{T}}^\tau\|_{J \times \Omega}^2 \\ &\quad + \sum_{T \in \mathcal{T}_h} \{ \|h^{\frac{1}{2}} \nabla \theta_T^\tau\|_{J \times \partial T}^2 + \|h^{-\frac{1}{2}} \theta_T^\tau\|_{J \times \partial T}^2 + \|h^{-\frac{1}{2}} \theta_{\partial T}^\tau\|_{J \times \partial T}^2 \} \\ &\quad + \sum_{n \in \{1:N-1\}} \{ \|[\![\theta_{\mathcal{T}}^\tau]\!]^n\|_{\Omega}^2 + \|[\![\nabla_{\mathcal{T}} \theta_{\mathcal{T}}^\tau]\!]^n\|_{\Omega}^2 \} + \tau \|\partial_t \theta_{\mathcal{T}}^\tau\|_{J \times \Omega}^2. \end{aligned} \quad (38)$$

Recall that the stability norm $\|\cdot\|$ is defined in (30)-(31). In what follows, we assume that

$$u \in H^1(J; H^1(\Omega)) \cap L^2(J; H^2(\Omega)). \quad (39)$$

Lemma 4 (Consistency). *Let $(\hat{u}_h^\tau, \hat{\xi}_h^\tau)$ denote the solution to the discrete problem (20)-(21). Let \hat{e}_h^τ and $\hat{\theta}_h^\tau$ be the discrete and interpolation errors on the primal unknown defined in (37). Under the regularity assumption (39), we have, for all $(\hat{w}_h^\tau, \hat{\eta}_h^\tau) \in \tilde{U}_h^\tau \times \tilde{U}_{h0}^\tau$,*

$$|A_h^\tau((\hat{e}_h^\tau, \hat{\xi}_h^\tau), (\hat{w}_h^\tau, 0))| \lesssim (\|\hat{\theta}_h^\tau\|_{\#} + \|\delta\|_{J \times \infty} + t_h^\tau (I_{\mathcal{T}}^\tau(u), I_{\mathcal{T}}^\tau(u))^{\frac{1}{2}}) \|\hat{w}_h^\tau, 0\|, \quad (40)$$

$$|A_h^\tau((\hat{e}_h^\tau, \hat{\xi}_h^\tau), (0, \hat{\eta}_h^\tau))| \lesssim \|\hat{\theta}_h^\tau\|_{\#} \|0, \hat{\eta}_h^\tau\|. \quad (41)$$

Proof. (i) Proof of (40). Using $(\hat{e}_h^\tau, \hat{\xi}_h^\tau) = (\hat{u}_h^\tau, \hat{\xi}_h^\tau) - (\hat{I}_h^\tau(u), 0)$, the equation (20) in the discrete problem, the definition (22) of A_h^τ , and $g_\delta = g + \delta$, we have, for all $\hat{w}_h^\tau \in \tilde{U}_h^\tau$,

$$\begin{aligned} & A_h^\tau((\hat{e}_h^\tau, \hat{\xi}_h^\tau), (\hat{w}_h^\tau, 0)) \\ &= m_\varpi(g_\delta, w_\mathcal{T}^\tau) - m_\varpi(I_\mathcal{T}^\tau(u), w_\mathcal{T}^\tau) - t_h^\tau(I_\mathcal{T}^\tau(u), w_\mathcal{T}^\tau) - s_h^\tau(\hat{I}_h^\tau(u), \hat{w}_h^\tau) \\ &= m_\varpi(\delta, w_\mathcal{T}^\tau) + m_\varpi(\theta_\mathcal{T}^\tau, w_\mathcal{T}^\tau) - t_h^\tau(I_\mathcal{T}^\tau(u), w_\mathcal{T}^\tau) - s_h^\tau(\hat{I}_h^\tau(u), \hat{w}_h^\tau). \end{aligned}$$

The Cauchy–Schwarz inequality yields

$$\begin{aligned} |m_\varpi(\delta, w_\mathcal{T}^\tau)| + |m_\varpi(\theta_\mathcal{T}^\tau, w_\mathcal{T}^\tau)| &\lesssim (\|\delta\|_{J \times \varpi} + \|\theta_\mathcal{T}^\tau\|_{J \times \varpi}) \|w_\mathcal{T}^\tau\|_{J \times \varpi}, \\ |t_h^\tau(I_\mathcal{T}^\tau(u), w_\mathcal{T}^\tau)| &\lesssim t_h^\tau(I_\mathcal{T}^\tau(u), I_\mathcal{T}^\tau(u))^{\frac{1}{2}} t_h^\tau(w_\mathcal{T}^\tau, w_\mathcal{T}^\tau)^{\frac{1}{2}}, \\ |s_h^\tau(\hat{I}_h^\tau(u), \hat{w}_h^\tau)| &\lesssim s_h^\tau(\hat{I}_h^\tau(u), \hat{I}_h^\tau(u))^{\frac{1}{2}} s_h^\tau(\hat{w}_h^\tau, \hat{w}_h^\tau)^{\frac{1}{2}}. \end{aligned}$$

Moreover, we have

$$\begin{aligned} s_h^\tau(\hat{I}_h^\tau(u), \hat{I}_h^\tau(u)) &= \sum_{T \in \mathcal{T}_h} h^{-1} \|I_T^\tau(u) - I_{\partial T}^\tau(u)\|_{J \times \partial T}^2 \\ &\quad + \sum_{n \in \{1:N-1\}} \{ \|\llbracket \theta_\mathcal{T}^\tau \rrbracket^n\|_\Omega^2 + \|\llbracket \nabla_\mathcal{T} \theta_\mathcal{T}^\tau \rrbracket^n\|_\Omega^2 \} \lesssim \|\hat{\theta}_h^\tau\|_\#^2. \end{aligned}$$

The bound (40) follows by gathering the above estimates.

(ii) Proof of (41). Proceeding as above and using now the equation (21) in the discrete problem, we have, for all $\hat{\eta}_h^\tau \in U_{h0}^\tau$,

$$\begin{aligned} A_h^\tau((\hat{e}_h^\tau, \hat{\xi}_h^\tau), (0, \hat{\eta}_h^\tau)) &= (f, \eta_\mathcal{T}^\tau)_{J \times \Omega} - b^\tau(I_\mathcal{T}^\tau(u), \eta_\mathcal{T}^\tau) - a_h(\hat{I}_h^\tau(u), \hat{\eta}_h^\tau) \\ &= -a_h(\hat{I}_h^\tau(u), \hat{\eta}_h^\tau) - (\Delta u, \eta_\mathcal{T}^\tau)_{J \times \Omega}, \end{aligned}$$

where we used that $f = \partial_t u - \Delta u$ in $J \times \Omega$ and $(\partial_t u, \eta_\mathcal{T}^\tau)_{J \times \Omega} = b^\tau(I_\mathcal{T}^\tau(u), \eta_\mathcal{T}^\tau)$ owing to (36) and (24). Recalling the definition (10) of a_h , we infer that

$$\begin{aligned} A_h^\tau((\hat{e}_h^\tau, \hat{\xi}_h^\tau), (0, \hat{\eta}_h^\tau)) &= \sum_{T \in \mathcal{T}_h} \{ -(\nabla I_T^\tau(u), \nabla \eta_T^\tau)_{J \times T} + (\nabla I_T^\tau(u) \cdot \mathbf{n}_T, \eta_T^\tau - \eta_{\partial T}^\tau)_{J \times \partial T} \\ &\quad + (I_T^\tau(u) - I_{\partial T}^\tau(u), \nabla \eta_T^\tau \cdot \mathbf{n}_T)_{J \times \partial T} - (\Delta u, \eta_T^\tau)_{J \times T} \} \\ &= \sum_{T \in \mathcal{T}_h} \{ (\nabla \theta_T^\tau, \nabla \eta_T^\tau)_{J \times T} - (\nabla \theta_T^\tau \cdot \mathbf{n}_T, \eta_T^\tau - \eta_{\partial T}^\tau)_{J \times \partial T} \\ &\quad + (\theta_{\partial T}^\tau - \theta_T^\tau, \nabla \eta_T^\tau \cdot \mathbf{n}_T)_{J \times \partial T} \}, \end{aligned}$$

where we used the assumed regularity of u and the fact that $\eta_F^\tau = 0$ for all $F \in \mathcal{F}_h^\partial$ to infer that $\sum_{T \in \mathcal{T}_h} (\nabla u \cdot \mathbf{n}_T, \eta_{\partial T}^\tau)_{J \times \partial T} = 0$. Using Cauchy–Schwarz and inverse inequalities, we readily obtain (41). \square

Lemma 5 (A priori residual bound). *Under the regularity assumption (39), we have*

$$\|\hat{e}_h^\tau, \hat{\xi}_h^\tau\| + t_h^\tau(u_\mathcal{T}^\tau, u_\mathcal{T}^\tau)^{\frac{1}{2}} + s_h^\tau(\hat{u}_h^\tau, \hat{u}_h^\tau)^{\frac{1}{2}} \lesssim \|\hat{\theta}_h^\tau\|_\# + \|\delta\|_{J \times \varpi} + t_h^\tau(I_\mathcal{T}^\tau(u), I_\mathcal{T}^\tau(u))^{\frac{1}{2}}.$$

Proof. Using inf-sup stability (Lemma 3) yields

$$\|\hat{e}_h^\tau, \hat{\xi}_h^\tau\| \lesssim \sup_{(\hat{w}_h^\tau, \hat{\eta}_h^\tau) \in \tilde{U}_h^\tau \times \tilde{U}_{h0}^\tau \setminus \{(0,0)\}} \frac{A_h^\tau((\hat{e}_h^\tau, \hat{\xi}_h^\tau), (\hat{w}_h^\tau, \hat{\eta}_h^\tau))}{\|\hat{w}_h^\tau, \hat{\eta}_h^\tau\|}.$$

Moreover, owing to consistency (Lemma 4), we have, for all $(\hat{w}_h^\tau, \hat{\eta}_h^\tau) \in \tilde{U}_h^\tau \times \tilde{U}_{h0}^\tau$,

$$|A_h^\tau((\hat{e}_h^\tau, \hat{\xi}_h^\tau), (\hat{w}_h^\tau, \hat{\eta}_h^\tau))| \lesssim (\|\hat{\theta}_h^\tau\|_\# + \|\delta\|_{J \times \varpi} + t_h^\tau(I_{\mathcal{T}}^\tau(u), I_{\mathcal{T}}^\tau(u))^{\frac{1}{2}}) \|\hat{w}_h^\tau, \hat{\eta}_h^\tau\|.$$

Combining the two bounds gives

$$\|\hat{e}_h^\tau, \hat{\xi}_h^\tau\| \lesssim \|\hat{\theta}_h^\tau\|_\# + \|\delta\|_{J \times \varpi} + t_h^\tau(I_{\mathcal{T}}^\tau(u), I_{\mathcal{T}}^\tau(u))^{\frac{1}{2}}.$$

Finally, observing that

$$\begin{aligned} t_h^\tau(u_{\mathcal{T}}^\tau, u_{\mathcal{T}}^\tau)^{\frac{1}{2}} &\leq t_h^\tau(I_{\mathcal{T}}^\tau(u), I_{\mathcal{T}}^\tau(u))^{\frac{1}{2}} + t_h^\tau(e_{\mathcal{T}}^\tau, e_{\mathcal{T}}^\tau)^{\frac{1}{2}} \leq t_h^\tau(I_{\mathcal{T}}^\tau(u), I_{\mathcal{T}}^\tau(u))^{\frac{1}{2}} + \|\hat{e}_h^\tau, \hat{\xi}_h^\tau\|, \\ s_h^\tau(\hat{u}_h^\tau, \hat{u}_h^\tau)^{\frac{1}{2}} &\leq s_h^\tau(\hat{I}_h^\tau(u), \hat{I}_h^\tau(u))^{\frac{1}{2}} + s_h^\tau(\hat{e}_h^\tau, \hat{e}_h^\tau)^{\frac{1}{2}} \leq \|\hat{\theta}_h^\tau\|_\# + \|\hat{e}_h^\tau, \hat{\xi}_h^\tau\|, \end{aligned}$$

concludes the proof. \square

Remark 3.1. (Lemma 5) The a priori residual bound involves three terms, one depending on the interpolation error of the exact solution measured in the $\|\cdot\|_\#$ -norm, one resulting from the presence of noise in the measurements, and one resulting from the regularization.

3.5 Bound on dual error norm

In this section, we prove another important result bounding some dual norm of the error. Recall the operator $L := \partial_t - \Delta$ so that, for all $v \in H^1(J; L^2(\Omega)) \cap L^2(J; H^1(\Omega))$ and all $\eta \in L^2(J; H_0^1(\Omega))$, we have $\langle L(v), \eta \rangle_{L^2(H^{-1}), L^2(H_0^1)} = (\partial_t v, \eta)_{J \times \Omega} + (\nabla v, \nabla \eta)_{J \times \Omega}$, where $\langle \cdot, \cdot \rangle_{L^2(H^{-1}), L^2(H_0^1)}$ stands for the duality product between $L^2(J; H^{-1}(\Omega))$ and $L^2(J; H_0^1(\Omega))$. The operator L can be extended to discrete functions $v_{\mathcal{T}}^\tau \in H^1(J; U_{\mathcal{T}}^k)$ by setting

$$\langle L_{\mathcal{T}}(v_{\mathcal{T}}^\tau), \eta \rangle_{L^2(H^{-1}), L^2(H_0^1)} := (\partial_t v_{\mathcal{T}}^\tau, \eta)_{J \times \Omega} + (\nabla_{\mathcal{T}} v_{\mathcal{T}}^\tau, \nabla \eta)_{J \times \Omega}. \quad (42)$$

The corresponding dual norm is

$$\|L_{\mathcal{T}}(v)\|_{L^2(J; H^{-1}(\Omega))} = \sup_{\substack{\eta \in L^2(J; H_0^1(\Omega)) \\ \|\nabla \eta\|_{J \times \Omega} = 1}} \{(\partial_t v, \eta)_{J \times \Omega} + (\nabla_{\mathcal{T}} v, \nabla \eta)_{J \times \Omega}\}.$$

It is also useful to introduce the data oscillation term

$$\Theta_f := f - \Pi^\tau(f), \quad (43)$$

where $\Pi^\tau(f)|_{I_n} := \Pi_n^\ell(f|_{I_n})$ for all $n \in \{1:N\}$, and Π_n^ℓ denotes the L^2 -orthogonal projection from $L^2(I_n; L^2(\Omega))$ onto $\mathbb{P}^\ell(I_n; L^2(\Omega))$.

Lemma 6 (Bound on dual error norm). *Under the regularity assumption (39), we have*

$$\begin{aligned} & \|L_{\mathcal{T}}(R^{\tau}(u_{\mathcal{T}}^{\tau}) - u)\|_{L^2(J; H^{-1}(\Omega))} \\ & \lesssim (1 + \tau^{-\frac{1}{2}}h)\|\hat{\theta}_h^{\tau}\|_{\#} + \|\delta\|_{J \times \varpi} + t_h^{\tau}(I_{\mathcal{T}}^{\tau}(u), I_{\mathcal{T}}^{\tau}(u))^{\frac{1}{2}} + \|\Theta_f\|_{J \times \Omega}. \end{aligned}$$

Proof. Let $\eta \in L^2(J; H_0^1(\Omega))$ with $\|\nabla \eta\|_{J \times \Omega} = 1$. Recalling the definition of Θ_f , we have

$$\begin{aligned} & \langle L_{\mathcal{T}}(R^{\tau}(u_{\mathcal{T}}^{\tau}) - u), \eta \rangle_{L^2(H^{-1}), L^2(H_0^1)} \\ & = (\partial_t R^{\tau}(u_{\mathcal{T}}^{\tau}), \eta)_{J \times \Omega} + (\nabla_{\mathcal{T}} R^{\tau}(u_{\mathcal{T}}^{\tau}), \nabla \eta)_{J \times \Omega} - (\Pi^{\tau}(f), \eta)_{J \times \Omega} - (\Theta_f, \eta)_{J \times \Omega} \\ & = (\partial_t R^{\tau}(u_{\mathcal{T}}^{\tau}), \Pi^{\tau}(\eta))_{J \times \Omega} + (\nabla_{\mathcal{T}} R^{\tau}(u_{\mathcal{T}}^{\tau}), \nabla \eta)_{J \times \Omega} - (f, \Pi^{\tau}(\eta))_{J \times \Omega} - (\Theta_f, \eta)_{J \times \Omega}, \end{aligned}$$

where we used the L^2 -orthogonality properties of Π^{τ} . Let $\hat{\eta}_h^{\tau} \in \tilde{U}_{h0}^{\tau}$ be such that

$$\hat{\eta}_h^{\tau}|_{I_n} := ((\Pi_n^{\ell} \Pi_T^k(\eta|_T))_{T \in \mathcal{T}_h}, (\Pi_n^{\ell} \Pi_F^k(\eta|_F))_{F \in \mathcal{F}_h}) \in \mathbb{P}^{\ell}(I_n; \hat{U}_{h0}^k),$$

for all $n \in \{1:N\}$. Invoking the second equation (21) in the discrete problem and recalling the rewriting (24) of b^{τ} gives

$$(\partial_t R^{\tau}(u_{\mathcal{T}}^{\tau}), \eta_{\mathcal{T}}^{\tau})_{J \times \Omega} + a_h(\hat{u}_h^{\tau}, \hat{\eta}_h^{\tau}) - \sigma_h(\hat{\xi}_h^{\tau}, \hat{\eta}_h^{\tau}) = (f, \eta_{\mathcal{T}}^{\tau})_{J \times \Omega}.$$

Subtracting this equation from the above expression, recalling that $f = \partial_t u - \Delta u$, and re-arranging the terms leads to

$$\langle L_{\mathcal{T}}(R^{\tau}(u_{\mathcal{T}}^{\tau}) - u), \eta \rangle_{L^2(H^{-1}), L^2(H_0^1)} = A_1 + A_2 + A_3,$$

with

$$\begin{aligned} A_1 & := (\partial_t (R^{\tau}(u_{\mathcal{T}}^{\tau}) - u) - \Delta_{\mathcal{T}}(u_{\mathcal{T}}^{\tau} - u), \Pi^{\tau}(\eta) - \eta_{\mathcal{T}}^{\tau})_{J \times \Omega}, \\ A_2 & := (\nabla_{\mathcal{T}} R^{\tau}(u_{\mathcal{T}}^{\tau}), \nabla \eta)_{J \times \Omega} - a_h(\hat{u}_h^{\tau}, \hat{\eta}_h^{\tau}) + (\Delta_{\mathcal{T}} u_{\mathcal{T}}^{\tau}, \Pi^{\tau}(\eta) - \eta_{\mathcal{T}}^{\tau})_{J \times \Omega}, \\ A_3 & := \sigma_h(\hat{\xi}_h^{\tau}, \hat{\eta}_h^{\tau}) - (\Theta_f, \eta)_{J \times \Omega}. \end{aligned}$$

A straightforward calculation using the definition (10) of the bilinear form a_h shows that

$$A_2 = (\nabla_{\mathcal{T}}(R^{\tau}(u_{\mathcal{T}}^{\tau}) - u_{\mathcal{T}}^{\tau}), \nabla \eta)_{J \times \Omega} + \sum_{T \in \mathcal{T}_h} (u_T^{\tau} - u_{\partial T}^{\tau}, \nabla \eta_T^{\tau} \cdot \mathbf{n}_T)_{J \times \partial T},$$

since we have

$$\begin{aligned} & (\nabla_{\mathcal{T}} u_{\mathcal{T}}^{\tau}, \nabla \eta)_{J \times \Omega} - a_h(\hat{u}_h^{\tau}, \hat{\eta}_h^{\tau}) \\ & = (\nabla_{\mathcal{T}} u_{\mathcal{T}}^{\tau}, \nabla_{\mathcal{T}}(\Pi^{\tau}(\eta) - \eta_{\mathcal{T}}^{\tau}))_{J \times \Omega} \\ & \quad + \sum_{T \in \mathcal{T}_h} \{(\nabla u_T^{\tau} \cdot \mathbf{n}_T, \eta_T^{\tau} - \eta_{\partial T}^{\tau})_{J \times \partial T} + (u_T^{\tau} - u_{\partial T}^{\tau}, \nabla \eta_T^{\tau} \cdot \mathbf{n}_T)_{J \times \partial T}\} \\ & = (-\Delta_{\mathcal{T}} u_{\mathcal{T}}^{\tau}, \Pi^{\tau}(\eta) - \eta_{\mathcal{T}}^{\tau})_{J \times \Omega} \\ & \quad + \sum_{T \in \mathcal{T}_h} \{(\nabla u_T^{\tau} \cdot \mathbf{n}_T, \Pi^{\tau}(\eta) - \eta_{\partial T}^{\tau})_{J \times \partial T} + (u_T^{\tau} - u_{\partial T}^{\tau}, \nabla \eta_T^{\tau} \cdot \mathbf{n}_T)_{J \times \partial T}\} \\ & = (-\Delta_{\mathcal{T}} u_{\mathcal{T}}^{\tau}, \Pi^{\tau}(\eta) - \eta_{\mathcal{T}}^{\tau})_{J \times \Omega} + \sum_{T \in \mathcal{T}_h} (u_T^{\tau} - u_{\partial T}^{\tau}, \nabla \eta_T^{\tau} \cdot \mathbf{n}_T)_{J \times \partial T}, \end{aligned}$$

where we used that $(\nabla_{\mathcal{T}} u_{\mathcal{T}}^{\tau}, \nabla(\eta - \Pi^{\tau}(\eta)))_{J \times \Omega} = 0$ in the first equality and the fact that $(\nabla u_{\mathcal{T}}^{\tau} \cdot \mathbf{n}_T)|_{I_n \times \partial T} \in \mathbb{P}^{\ell}(I_n; \mathbb{P}^k(\mathcal{F}_{\partial T}))$ and the definition of $\eta_{\partial T}^{\tau}$ in the third equality. It remains to bound A_1, A_2, A_3 .

(i) Bound on A_1 . This is the most delicate term. We observe that

$$\begin{aligned} \partial_t(R^{\tau}(u_{\mathcal{T}}^{\tau}) - u) - \Delta_{\mathcal{T}}(u_{\mathcal{T}}^{\tau} - u) &= \partial_t R^{\tau}(u_{\mathcal{T}}^{\tau} - I_{\mathcal{T}}^{\tau}(u)) - \Delta_{\mathcal{T}}(u_{\mathcal{T}}^{\tau} - I_{\mathcal{T}}^{\tau}(u)) \\ &\quad + \partial_t(R^{\tau}(I_{\mathcal{T}}^{\tau}(u)) - u) - \Delta_{\mathcal{T}}(I_{\mathcal{T}}^{\tau}(u) - u). \end{aligned}$$

Recalling the definition (31) of the $\|\cdot\|_{\mathbb{R}}$ -norm, we have

$$|(\partial_t R^{\tau}(u_{\mathcal{T}}^{\tau} - I_{\mathcal{T}}^{\tau}(u)) - \Delta_{\mathcal{T}}(u_{\mathcal{T}}^{\tau} - I_{\mathcal{T}}^{\tau}(u)), \Pi^{\tau}(\eta) - \eta_{\mathcal{T}}^{\tau})_{J \times \Omega}| \lesssim \|e_{\mathcal{T}}^{\tau}\|_{\mathbb{R}} \|\nabla \eta\|_{J \times \Omega},$$

where we used that

$$\|\Pi^{\tau}(\eta) - \eta_{\mathcal{T}}^{\tau}\|_{J \times \Omega} \lesssim h \|\nabla \eta\|_{J \times \Omega}.$$

Owing to Lemma 5, we have $\|e_{\mathcal{T}}^{\tau}\|_{\mathbb{R}} \lesssim \|\hat{\theta}_h^{\tau}\|_{\#} + \|\delta\|_{J \times \varpi} + t_h^{\tau}(I_{\mathcal{T}}^{\tau}(u), I_{\mathcal{T}}^{\tau}(u))^{\frac{1}{2}}$. Moreover, the triangle inequality, the estimate (27), and the above bound on $\|\Pi^{\tau}(\eta) - \eta_{\mathcal{T}}^{\tau}\|_{J \times \Omega}$ give

$$\begin{aligned} &|(\partial_t(R^{\tau}(I_{\mathcal{T}}^{\tau}(u)) - u), \Pi^{\tau}(\eta) - \eta_{\mathcal{T}}^{\tau})_{J \times \Omega}| \\ &\lesssim (\|\partial_t^{\tau}(R^{\tau}(I_{\mathcal{T}}^{\tau}(u)) - I_{\mathcal{T}}^{\tau}(u))\|_{J \times \Omega} + \|\partial_t^{\tau}(u - I_{\mathcal{T}}^{\tau}(u))\|_{J \times \Omega}) h \|\nabla \eta\|_{J \times \Omega} \\ &\lesssim (d^{\tau}(I_{\mathcal{T}}^{\tau}(u), I_{\mathcal{T}}^{\tau}(u))^{\frac{1}{2}} + \tau^{\frac{1}{2}} \|\partial_t^{\tau} \theta_{\mathcal{T}}^{\tau}\|_{J \times \Omega}) \tau^{-\frac{1}{2}} h \|\nabla \eta\|_{J \times \Omega}. \end{aligned}$$

Finally, we have

$$|(\Delta_{\mathcal{T}}(I_{\mathcal{T}}^{\tau}(u) - u), \Pi^{\tau}(\eta) - \eta_{\mathcal{T}}^{\tau})_{J \times \Omega}| \lesssim h \|\Delta_{\mathcal{T}} \theta_{\mathcal{T}}^{\tau}\|_{J \times \Omega} \|\nabla \eta\|_{J \times \Omega}.$$

Gathering the above estimates gives

$$|A_1| \lesssim ((1 + \tau^{-\frac{1}{2}} h) \|\hat{\theta}_h^{\tau}\|_{\#} + \|\delta\|_{J \times \varpi} + t_h^{\tau}(I_{\mathcal{T}}^{\tau}(u), I_{\mathcal{T}}^{\tau}(u))^{\frac{1}{2}}) \|\nabla \eta\|_{J \times \Omega}.$$

(ii) Bound on A_2 . Owing to the Cauchy–Schwarz inequality and (28), we have

$$|(\nabla_{\mathcal{T}}(R^{\tau}(u_{\mathcal{T}}^{\tau}) - u_{\mathcal{T}}^{\tau}), \nabla \eta)_{J \times \Omega}| \lesssim \tau^{\frac{1}{2}} d^{\tau}(u_{\mathcal{T}}^{\tau}, u_{\mathcal{T}}^{\tau})^{\frac{1}{2}} \|\nabla \eta\|_{J \times \Omega}.$$

Moreover, using Cauchy–Schwarz and inverse trace inequalities, we have

$$\left| \sum_{T \in \mathcal{T}_h} (u_{\mathcal{T}}^{\tau} - u_{\partial T}^{\tau}, \nabla \eta_{\mathcal{T}}^{\tau} \cdot \mathbf{n}_T)_{J \times \partial T} \right| \lesssim d_h(\hat{u}_h^{\tau}, \hat{u}_h^{\tau})^{\frac{1}{2}} \|\nabla \eta\|_{J \times \Omega}.$$

Thus, we have

$$|A_2| \lesssim (\tau^{\frac{1}{2}} d^{\tau}(u_{\mathcal{T}}^{\tau}, u_{\mathcal{T}}^{\tau})^{\frac{1}{2}} + d_h(\hat{u}_h^{\tau}, \hat{u}_h^{\tau})^{\frac{1}{2}}) \|\nabla \eta\|_{J \times \Omega}.$$

Using $\tau \lesssim 1$ and Lemma 5 yields

$$|A_2| \lesssim s_h^{\tau}(\hat{u}_h^{\tau}, \hat{u}_h^{\tau})^{\frac{1}{2}} \|\nabla \eta\|_{J \times \Omega} \lesssim (\|\hat{\theta}_h^{\tau}\|_{\#} + \|\delta\|_{J \times \varpi} + t_h^{\tau}(I_{\mathcal{T}}^{\tau}(u), I_{\mathcal{T}}^{\tau}(u))^{\frac{1}{2}}) \|\nabla \eta\|_{J \times \Omega}.$$

(iii) Bound on A_3 . Since $\sigma_h(\hat{\eta}_h^{\tau}, \hat{\eta}_h^{\tau}) = \|\nabla_{\mathcal{T}} \eta_{\mathcal{T}}^{\tau}\|_{J \times \Omega}^2 + d_h(\hat{\eta}_h^{\tau}, \hat{\eta}_h^{\tau})$, we infer that $\sigma_h(\hat{\eta}_h^{\tau}, \hat{\eta}_h^{\tau})^{\frac{1}{2}} \lesssim \|\nabla \eta\|_{J \times \Omega}$. This implies that $|\sigma_h(\hat{\xi}_h^{\tau}, \hat{\eta}_h^{\tau})| \lesssim \sigma_h(\hat{\xi}_h^{\tau}, \hat{\xi}_h^{\tau})^{\frac{1}{2}} \sigma_h(\hat{\eta}_h^{\tau}, \hat{\eta}_h^{\tau})^{\frac{1}{2}} \lesssim \sigma_h(\hat{\xi}_h^{\tau}, \hat{\xi}_h^{\tau})^{\frac{1}{2}} \|\nabla \eta\|_{J \times \Omega}$ and $\sigma_h(\hat{\xi}_h^{\tau}, \hat{\xi}_h^{\tau})^{\frac{1}{2}}$ is bounded in Lemma 5. Moreover, the bound $|(\Theta_f, \eta)_{J \times \Omega}| \lesssim \|\Theta_f\|_{J \times \Omega} \|\nabla \eta\|_{J \times \Omega}$ results from the Cauchy–Schwarz inequality combined and the global Poincaré inequality in Ω . Hence, we have

$$|A_3| \lesssim (\|\hat{\theta}_h^{\tau}\|_{\#} + \|\delta\|_{J \times \varpi} + t_h^{\tau}(I_{\mathcal{T}}^{\tau}(u), I_{\mathcal{T}}^{\tau}(u))^{\frac{1}{2}} + \|\Theta_f\|_{J \times \Omega}) \|\nabla \eta\|_{J \times \Omega}.$$

(iv) Combining the above estimates and recalling that $\|\nabla \eta\|_{J \times \Omega} = 1$ gives the expected bound. \square

3.6 Main result: error estimate in the target subdomain

We are now ready to derive our main error estimate. The idea of the proof is to combine the results of Sections 3.4 and 3.5 with the conditional stability estimate from Lemma 1. We use the target (semi)norm

$$\|v\|_{\text{trg}} := \|\nabla v\|_{L^2(T_1, T_2; L^2(B))},$$

where T_1, T_2 and B are defined in Lemma 1. Recall the quantities $C_{\text{stb}} > 0$ and $\alpha \in (0, 1]$ introduced in Lemma 1. Recall that the $\|\cdot\|_{\#}$ -norm is defined in (38), the data oscillation term Θ_f in (43), that $\hat{\theta}_h^\tau$ is the interpolation error defined in (37), and that the coefficient $c(\tau, h)$ is used to weight the regularization bilinear form t_h^τ defined in (16).

Theorem 7 (Error estimate in target subdomain). *Under the regularity assumption (39), we have*

$$\begin{aligned} \|u - u_{\mathcal{T}}^\tau\|_{\text{trg}} &\lesssim C_{\text{stb}}(1 + \tau^{-1}h)^\alpha(1 + \tau^{-1}h + \gamma^{-\frac{1}{2}}c(\tau, h)^{-1})^{1-\alpha} \\ &\quad \times (\|\hat{\theta}_h^\tau\|_{\#} + \|\delta\|_{J \times \varpi} + t_h^\tau(I_{\mathcal{T}}^\tau(u), I_{\mathcal{T}}^\tau(u))^{\frac{1}{2}} + \|\Theta_f\|_{J \times \Omega}). \end{aligned} \quad (44)$$

Proof. (i) Using an averaging operator in space (see [6, Lemmas 3.2 and 5.3] or [15, Chap. 22] and the references therein), we can build a piecewise polynomial function $\tilde{u}_{\mathcal{T}}^\tau \in \{v_{\mathcal{T}}^\tau \in H^1(J; H^1(\Omega)) \mid v_{\mathcal{T}}^\tau|_{I_n \times T} \in \mathbb{P}^{\ell+1}(I_n; \mathbb{P}^k(T)), n \in \{1:N\}, T \in \mathcal{T}_h\}$ such that

$$\begin{aligned} \|h^{-1}(\tilde{u}_{\mathcal{T}}^\tau - R^\tau(u_{\mathcal{T}}^\tau))\|_{J \times \Omega} + \|\nabla_{\mathcal{T}}(\tilde{u}_{\mathcal{T}}^\tau - R^\tau(u_{\mathcal{T}}^\tau))\|_{J \times \Omega} &\lesssim \|h^{-\frac{1}{2}}\llbracket R^\tau(u_{\mathcal{T}}^\tau) \rrbracket_{\mathcal{F}_h^{\text{int}}}\|_{J \times \mathcal{F}_h^{\text{int}}} \\ &\lesssim d_h(\hat{u}_h^\tau, \hat{u}_h^\tau)^{\frac{1}{2}}, \end{aligned} \quad (45)$$

where the first bound results from the approximation properties in space of the averaging operator and the second bound from (29) (see Lemma 2). Invoking the triangle inequality yields

$$\|u - u_{\mathcal{T}}^\tau\|_{\text{trg}} \leq \|u_{\mathcal{T}}^\tau - \tilde{u}_{\mathcal{T}}^\tau\|_{\text{trg}} + \|u - \tilde{u}_{\mathcal{T}}^\tau\|_{\text{trg}},$$

and we are left with bounding the two terms on the right-hand side.

(ii) Bound on $\|u_{\mathcal{T}}^\tau - \tilde{u}_{\mathcal{T}}^\tau\|_{\text{trg}}$. We observe that

$$\|u_{\mathcal{T}}^\tau - \tilde{u}_{\mathcal{T}}^\tau\|_{\text{trg}} \leq \|\nabla_{\mathcal{T}}(u_{\mathcal{T}}^\tau - R^\tau(u_{\mathcal{T}}^\tau))\|_{J \times \Omega} + \|\nabla_{\mathcal{T}}(\tilde{u}_{\mathcal{T}}^\tau - R^\tau(u_{\mathcal{T}}^\tau))\|_{J \times \Omega}.$$

The first term on the right-hand side is estimated by invoking (28) (see Lemma 2) and the second term by invoking (45). Owing to Lemma 5 and since $\tau \lesssim 1$, we infer that

$$\begin{aligned} \|u_{\mathcal{T}}^\tau - \tilde{u}_{\mathcal{T}}^\tau\|_{\text{trg}} &\leq d^\tau(u_{\mathcal{T}}^\tau, u_{\mathcal{T}}^\tau)^{\frac{1}{2}} + d_h(\hat{u}_h^\tau, \hat{u}_h^\tau)^{\frac{1}{2}} \lesssim s_h^\tau(\hat{u}_h^\tau, \hat{u}_h^\tau)^{\frac{1}{2}} \\ &\lesssim \|\hat{\theta}_h^\tau\|_{\#} + \|\delta\|_{J \times \varpi} + t_h^\tau(I_{\mathcal{T}}^\tau(u), I_{\mathcal{T}}^\tau(u))^{\frac{1}{2}}. \end{aligned}$$

(iii) Bound on $\|u - \tilde{u}_{\mathcal{T}}^\tau\|_{\text{trg}}$. Since $\tilde{u}_{\mathcal{T}}^\tau - u \in H^1(J; H^1(\Omega)) \subset H^1(J; H^{-1}(\Omega)) \cap L^2(J; H^1(\Omega))$, we can invoke the conditional stability estimate from Lemma 1. This gives

$$\begin{aligned} \|\tilde{u}_{\mathcal{T}}^\tau - u\|_{\text{trg}} &\lesssim C_{\text{stb}} \left(\|\tilde{u}_{\mathcal{T}}^\tau - u\|_{J \times \varpi} + \|L(\tilde{u}_{\mathcal{T}}^\tau - u)\|_{L^2(J; H^{-1}(\Omega))} \right)^\alpha \\ &\quad \times \left(\|\tilde{u}_{\mathcal{T}}^\tau - u\|_{J \times \Omega} + \|L(\tilde{u}_{\mathcal{T}}^\tau - u)\|_{L^2(J; H^{-1}(\Omega))} \right)^{1-\alpha}. \end{aligned}$$

It remains to bound $\|\tilde{u}_{\mathcal{T}}^\tau - u\|_{J \times \varpi}$, $\|\tilde{u}_{\mathcal{T}}^\tau - u\|_{J \times \Omega}$, and $\|L(\tilde{u}_{\mathcal{T}}^\tau - u)\|_{L^2(J; H^{-1}(\Omega))}$.
 (iii.a) Bound on $\|\tilde{u}_{\mathcal{T}}^\tau - u\|_{J \times \varpi}$. We have

$$\|\tilde{u}_{\mathcal{T}}^\tau - u\|_{J \times \varpi} \leq \|u - I_{\mathcal{T}}^\tau(u)\|_{J \times \varpi} + \|u_{\mathcal{T}}^\tau - I_{\mathcal{T}}^\tau(u)\|_{J \times \varpi} + \|u_{\mathcal{T}}^\tau - \tilde{u}_{\mathcal{T}}^\tau\|_{J \times \varpi},$$

where $I_{\mathcal{T}}^\tau(u)$ is defined in Section 3.3. We have

$$\begin{aligned} \|u - I_{\mathcal{T}}^\tau(u)\|_{J \times \varpi} &= \|\theta_{\mathcal{T}}^\tau\|_{J \times \varpi} \leq \|\theta_{\mathcal{T}}^\tau\|_{J \times \Omega}, \\ \|u_{\mathcal{T}}^\tau - I_{\mathcal{T}}^\tau(u)\|_{J \times \varpi} &= \|e_{\mathcal{T}}^\tau\|_{J \times \varpi} \leq \|\hat{e}_h^\tau, \hat{\xi}_h^\tau\|. \end{aligned}$$

Furthermore, the bound (28) (see Lemma 2) and the bound (45) imply that

$$\begin{aligned} \|u_{\mathcal{T}}^\tau - \tilde{u}_{\mathcal{T}}^\tau\|_{J \times \varpi} &\leq \|u_{\mathcal{T}}^\tau - R^\tau(u_{\mathcal{T}}^\tau)\|_{J \times \varpi} + \|R^\tau(u_{\mathcal{T}}^\tau) - \tilde{u}_{\mathcal{T}}^\tau\|_{J \times \varpi} \\ &\lesssim \tau^{\frac{1}{2}} d^\tau(u_{\mathcal{T}}^\tau, u_{\mathcal{T}}^\tau)^{\frac{1}{2}} + h d_h(\hat{u}_h^\tau, \hat{u}_h^\tau)^{\frac{1}{2}}. \end{aligned}$$

Gathering the above estimates and using Lemma 5 (and $\tau \lesssim 1$, $h \lesssim 1$) gives

$$\|\tilde{u}_{\mathcal{T}}^\tau - u\|_{J \times \varpi} \lesssim \|\hat{\theta}_h^\tau\|_{\#} + \|\delta\|_{J \times \varpi} + t_h^\tau(I_{\mathcal{T}}^\tau(u), I_{\mathcal{T}}^\tau(u))^{\frac{1}{2}}.$$

(iii.b) Bound on $\|\tilde{u}_{\mathcal{T}}^\tau - u\|_{J \times \Omega}$. Using the same decomposition, we have

$$\|\tilde{u}_{\mathcal{T}}^\tau - u\|_{J \times \Omega} \leq \|u - I_{\mathcal{T}}^\tau(u)\|_{J \times \Omega} + \|u_{\mathcal{T}}^\tau - I_{\mathcal{T}}^\tau(u)\|_{J \times \Omega} + \|u_{\mathcal{T}}^\tau - \tilde{u}_{\mathcal{T}}^\tau\|_{J \times \Omega}.$$

The terms $\|u - I_{\mathcal{T}}^\tau(u)\|_{J \times \Omega} + \|u_{\mathcal{T}}^\tau - \tilde{u}_{\mathcal{T}}^\tau\|_{J \times \Omega}$ are estimated by using the same arguments as in Step (iii.a). The bound on $\|u_{\mathcal{T}}^\tau - I_{\mathcal{T}}^\tau(u)\|_{J \times \Omega}$ uses, however, a different argument since we can only invoke here the regularization. This gives

$$\|u_{\mathcal{T}}^\tau - I_{\mathcal{T}}^\tau(u)\|_{J \times \Omega} = \|e_{\mathcal{T}}^\tau\|_{J \times \Omega} = \gamma^{-\frac{1}{2}} c(\tau, h)^{-1} t_h^\tau(e_{\mathcal{T}}^\tau, e_{\mathcal{T}}^\tau)^{\frac{1}{2}}.$$

Invoking Lemma 5 gives

$$\|\tilde{u}_{\mathcal{T}}^\tau - u\|_{J \times \Omega} \lesssim (1 + \gamma^{-\frac{1}{2}} c(\tau, h)^{-1}) (\|\hat{\theta}_h^\tau\|_{\#} + \|\delta\|_{J \times \varpi} + t_h^\tau(I_{\mathcal{T}}^\tau(u), I_{\mathcal{T}}^\tau(u))^{\frac{1}{2}}).$$

(iii.c) Bound on $\|L(\tilde{u}_{\mathcal{T}}^\tau - u)\|_{L^2(J; H^{-1}(\Omega))}$. We have

$$\begin{aligned} \|L(\tilde{u}_{\mathcal{T}}^\tau - u)\|_{L^2(J; H^{-1}(\Omega))} &\leq \|L_{\mathcal{T}}(\tilde{u}_{\mathcal{T}}^\tau - R^\tau(u_{\mathcal{T}}^\tau))\|_{L^2(J; H^{-1}(\Omega))} \\ &\quad + \|L_{\mathcal{T}}(R^\tau(u_{\mathcal{T}}^\tau) - u)\|_{L^2(J; H^{-1}(\Omega))}. \end{aligned}$$

For the second term on the right-hand side, we invoke Lemma 6 to infer

$$\begin{aligned} \|L_{\mathcal{T}}(R^\tau(u_{\mathcal{T}}^\tau) - u)\|_{L^2(J; H^{-1}(\Omega))} &\lesssim (1 + \tau^{-\frac{1}{2}} h) \|\hat{\theta}_h^\tau\|_{\#} \\ &\quad + \|\delta\|_{J \times \varpi} + t_h^\tau(I_{\mathcal{T}}^\tau(u), I_{\mathcal{T}}^\tau(u))^{\frac{1}{2}} + \|\Theta_f\|_{J \times \Omega}. \end{aligned}$$

To estimate the first term on the right-hand side, we bound the dual norm by considering an arbitrary test function $\eta \in L^2(J; H_0^1(\Omega))$ with $\|\nabla \eta\|_{J \times \Omega} = 1$. On the one hand, we have

$$\begin{aligned} |(\partial_t(\tilde{u}_{\mathcal{T}}^\tau - R^\tau(u_{\mathcal{T}}^\tau)), \eta)_{J \times \Omega}| &\lesssim \|\partial_t(\tilde{u}_{\mathcal{T}}^\tau - R^\tau(u_{\mathcal{T}}^\tau))\|_{J \times \Omega} \|\eta\|_{J \times \Omega} \\ &\lesssim \tau^{-1} h d_h(\hat{u}_h^\tau, \hat{u}_h^\tau)^{\frac{1}{2}} \|\nabla \eta\|_{J \times \Omega}, \end{aligned}$$

where we used an inverse inequality in time, the estimate (45), and a global Poincaré inequality in Ω . On the other hand, using again (45) gives

$$|(\nabla_{\mathcal{T}}(\tilde{u}_{\mathcal{T}}^{\tau} - R^{\tau}(u_{\mathcal{T}}^{\tau})), \nabla\eta)_{J \times \Omega}| \lesssim d_h(\hat{u}_h^{\tau}, \hat{u}_h^{\tau})^{\frac{1}{2}} \|\nabla\eta\|_{J \times \Omega}.$$

Combining the last two bounds and invoking Lemma 5, we infer that

$$\begin{aligned} \|L_{\mathcal{T}}(\tilde{u}_{\mathcal{T}}^{\tau} - R^{\tau}(u_{\mathcal{T}}^{\tau}))\|_{L^2(J; H^{-1}(\Omega))} &\lesssim (1 + \tau^{-1}h) d_h(\hat{u}_h^{\tau}, \hat{u}_h^{\tau})^{\frac{1}{2}} \\ &\lesssim (1 + \tau^{-1}h) (\|\hat{\theta}_h^{\tau}\|_{\#} + \|\delta\|_{J \times \varpi} + t_h^{\tau}(I_{\mathcal{T}}^{\tau}(u), I_{\mathcal{T}}^{\tau}(u))^{\frac{1}{2}}). \end{aligned}$$

Altogether, this gives

$$\begin{aligned} \|L(\tilde{u}_{\mathcal{T}}^{\tau} - u)\|_{L^2(J; H^{-1}(\Omega))} \\ \lesssim (1 + \tau^{-1}h) (\|\hat{\theta}_h^{\tau}\|_{\#} + \|\delta\|_{J \times \varpi} + t_h^{\tau}(I_{\mathcal{T}}^{\tau}(u), I_{\mathcal{T}}^{\tau}(u))^{\frac{1}{2}}) + \|\Theta_f\|_{J \times \Omega}. \end{aligned}$$

(iv) Combining the bounds from the above steps proves (44). \square

3.7 Weighting the regularization

The last step in our error analysis is to identify the weighting coefficient $c(\tau, h)$ in the regularization bilinear form. The interpolation operator defined in Section 3.3 fulfills the following convergence properties (the proof is outlined in Section 3.8.2).

Lemma 8 (Approximation). *The following holds for all $n \in \{1:N\}$, all $T \in \mathcal{T}_h$, and all $v \in H^{\ell+1}(I_n; H^2(T)) \cap H^1(I_n; H^{k+1}(T))$,*

$$\|I_T^n(v) - v\|_{I_n \times T} \lesssim \tau^{\ell+1} \|v\|_{H^{\ell+1}(I_n; L^2(T))} + h^{k+1} \|v\|_{H^1(I_n; H^{k+1}(T))}, \quad (46)$$

$$\|\nabla(I_T^n(v) - v)\|_{I_n \times T} \lesssim \tau^{\ell+1} \|v\|_{H^{\ell+1}(I_n; H^1(T))} + h^k \|v\|_{H^1(I_n; H^{k+1}(T))}, \quad (47)$$

$$h^{\frac{1}{2}} \|\nabla(I_T^n(v) - v)\|_{I_n \times \partial T} \lesssim h^{\frac{1}{2}} \tau^{\ell+1} \|v\|_{H^{\ell+1}(I_n; H^2(T))} + h^k \|v\|_{H^1(I_n; H^{k+1}(T))}, \quad (48)$$

$$h \|\Delta(I_T^n(v) - v)\|_{I_n \times T} \lesssim h \tau^{\ell+1} \|v\|_{H^{\ell+1}(I_n; H^2(T))} + h^k \|v\|_{H^1(I_n; H^{k+1}(T))}, \quad (49)$$

$$\|\partial_t(I_T^n(v) - v)\|_{I_n \times T} \lesssim \tau^{\ell} \|v\|_{H^{\ell+1}(I_n; L^2(T))} + h^{k+1} \|v\|_{H^1(I_n; H^{k+1}(T))}. \quad (50)$$

Moreover, we also have for all $s \in \overline{I_n}$,

$$\|(I_T^n(v) - v)(s)\|_T \lesssim \tau^{\ell+\frac{1}{2}} \|v\|_{H^{\ell+1}(I_n; L^2(T))} + h^{k+1} \|v\|_{H^1(I_n; H^{k+1}(T))}, \quad (51)$$

$$\|\nabla(I_T^n(v) - v)(s)\|_T \lesssim \tau^{\ell+\frac{1}{2}} \|v\|_{H^{\ell+1}(I_n; H^1(T))} + h^k \|v\|_{H^1(I_n; H^{k+1}(T))}. \quad (52)$$

Finally, we have

$$\begin{aligned} h^{-\frac{1}{2}} \|I_T^n(v) - v\|_{I_n \times \partial T} + h^{-\frac{1}{2}} \|I_{\partial T}^n(v) - v\|_{I_n \times \partial T} \\ \lesssim h^{-\frac{1}{2}} \tau^{\ell+1} \|v\|_{H^{\ell+1}(I_n; H^1(T))} + h^k \|v\|_{H^1(I_n; H^{k+1}(T))}. \end{aligned} \quad (53)$$

We consider the functional space

$$U_* := H^{\ell+2}(J; L^2(\Omega)) \cap H^{\ell+1}(J; H^2(\Omega)) \cap H^1(J; H^{k+1}(\Omega)), \quad (54)$$

equipped with its natural norm $\|\cdot\|_*$. An important consequence of Lemma 8 is that, under the assumption $u \in U_*$, we have

$$\|\hat{\theta}_h^\tau\|_{\#} + t_h^\tau (I_{\mathcal{T}}^\tau(u), I_{\mathcal{T}}^\tau(u))^{\frac{1}{2}} + \|\Theta_f\|_{J \times \Omega} \lesssim ((1 + h^{-1}\tau)^{\frac{1}{2}} \tau^{\ell + \frac{1}{2}} + h^k + \gamma^{\frac{1}{2}} c(\tau, h)) \|u\|_*, \quad (55)$$

where we used that $\tau \lesssim 1$ and $h \lesssim 1$ to simplify the expression. Notice also that $u \in U_*$ implies that $f \in H^{\ell+1}(J; L^2(\Omega))$.

Theorem 9 (Decay rates). *Assume that $u \in U_*$ and that $\tau \lesssim h$, $h \lesssim \tau$. Set*

$$c(\tau, h) := \tau^{\ell + \frac{1}{2}} + h^k. \quad (56)$$

We have the following decay rates on the errors:

$$\|\hat{e}_h^\tau, \hat{\xi}_h^\tau\| + t_h^\tau (u_{\mathcal{T}}^\tau, u_{\mathcal{T}}^\tau)^{\frac{1}{2}} + s_h^\tau (\hat{u}_h^\tau, \hat{u}_h^\tau)^{\frac{1}{2}} \lesssim (\tau^{\ell + \frac{1}{2}} + h^k) \|u\|_* + \|\delta\|_{J \times \varpi}, \quad (57)$$

$$\|u - u_{\mathcal{T}}^\tau\|_{\text{trg}} \lesssim C_{\text{stb}} (\tau^{\ell + \frac{1}{2}} + h^k)^\alpha (\|u\|_* + (\tau^{\ell + \frac{1}{2}} + h^k)^{-1} \|\delta\|_{J \times \varpi}). \quad (58)$$

where the hidden constant scales as $\max(\gamma, 1)^{\frac{1}{2}}$ in (57) and $\max(\gamma, \gamma^{\alpha-1})^{\frac{1}{2}}$ in (58).

Proof. Plug (56) into Lemma 5 and Theorem 7 and use (55). \square

3.8 Technical proofs

In this section, we outline the proofs of Lemma 2 and Lemma 8. The proofs use standard arguments from finite element approximation theory.

3.8.1 Proof of Lemma 2

Let $\hat{v}_h^\tau := (v_{\mathcal{T}}^\tau, v_{\mathcal{F}}^\tau) \in \tilde{U}_h^\tau$.

(i) Proof of (27). For all $n \in \{1:N\}$ and all $q_{\mathcal{T}}^n \in \mathbb{P}^\ell(I_n; U_{\mathcal{T}}^k)$, we have

$$(\partial_t (R^n(v_{\mathcal{T}}^\tau) - v_{\mathcal{T}}^n), q_{\mathcal{T}}^n)_{I_n \times \Omega} = (\llbracket v_{\mathcal{T}}^\tau \rrbracket^{n-1}, q_{\mathcal{T}}^n(t_{n-1}^+))_{\Omega}.$$

Invoking a discrete trace inequality in time implies

$$\|\partial_t (R^n(v_{\mathcal{T}}^\tau) - v_{\mathcal{T}}^n)\|_{I_n \times \Omega} \lesssim \tau^{-\frac{1}{2}} \|\llbracket v_{\mathcal{T}}^\tau \rrbracket^{n-1}\|_{\Omega}. \quad (59)$$

Summing over $n \in \{1:N\}$ proves (27) since $\sum_{n \in \{1:N\}} \|\llbracket v_{\mathcal{T}}^\tau \rrbracket^{n-1}\|_{\Omega}^2 \leq d^\tau (v_{\mathcal{T}}^\tau, v_{\mathcal{T}}^\tau)$ (recall that $\llbracket \cdot \rrbracket^0 = 0$ by convention).

(ii) Proof of (28). Since we have $(R^n(v_{\mathcal{T}}^\tau) - v_{\mathcal{T}}^n)(t_n^-) = 0$ (see (26)), a Poincaré inequality in time over I_n and the bound (59) yield

$$\|R^n(v_{\mathcal{T}}^\tau) - v_{\mathcal{T}}^n\|_{I_n \times \Omega} \lesssim \tau \|\partial_t (R^n(v_{\mathcal{T}}^\tau) - v_{\mathcal{T}}^n)\|_{I_n \times \Omega} \lesssim \tau^{\frac{1}{2}} \|\llbracket v_{\mathcal{T}}^\tau \rrbracket^{n-1}\|_{\Omega}.$$

Summing over $n \in \{1:N\}$ proves as above that

$$\|R^\tau(v_{\mathcal{T}}^\tau) - v_{\mathcal{T}}^\tau\|_{J \times \Omega} \lesssim \tau^{\frac{1}{2}} d^\tau (v_{\mathcal{T}}^\tau, v_{\mathcal{T}}^\tau)^{\frac{1}{2}}. \quad (60)$$

Moreover, (25) shows that the operators R^τ and $\nabla_{\mathcal{T}}$ commute. Hence, using the same arguments as above shows that

$$\|\nabla_{\mathcal{T}}(R^n(v_{\mathcal{T}}^\tau) - v_{\mathcal{T}}^n)\|_{I_n \times \Omega} \lesssim \tau^{\frac{1}{2}} \|\llbracket \nabla_{\mathcal{T}} v_{\mathcal{T}}^\tau \rrbracket^{n-1}\|_{\Omega}.$$

Summing over $n \in \{1:N\}$ and since $\sum_{n \in \{1:N\}} \|\llbracket \nabla_{\mathcal{T}} v_{\mathcal{T}}^\tau \rrbracket^{n-1}\|_{\Omega}^2 \leq d^\tau(v_{\mathcal{T}}^\tau, v_{\mathcal{T}}^\tau)$ gives

$$\|\nabla_{\mathcal{T}}(R^\tau(v_{\mathcal{T}}^\tau) - v_{\mathcal{T}}^\tau)\|_{J \times \Omega}^2 \lesssim \tau d^\tau(v_{\mathcal{T}}^\tau, v_{\mathcal{T}}^\tau). \quad (61)$$

Summing (60) and (61) proves (28).

(iii) Proof of (29). For all $F \in \mathcal{F}_h^{\text{int}}$, the operators R^τ and $\llbracket \cdot \rrbracket_F$ commute. In particular, we have, for all $n \in \{1:N\}$ and all $q_F^n \in \mathbb{P}^\ell(I_n; \mathbb{P}^k(F))$,

$$(\partial_t(\llbracket R^n(v_{\mathcal{T}}^\tau) - v_{\mathcal{T}}^n \rrbracket_F), q_F^n)_{I_n \times F} = (\llbracket \llbracket v_{\mathcal{T}}^\tau \rrbracket_F \rrbracket^{n-1}, q_F^n(t_{n-1}^+))_F,$$

and $\llbracket R^n(v_{\mathcal{T}}^\tau) - v_{\mathcal{T}}^n \rrbracket_F(t_n^-) = 0$. Hence, using the same arguments as above shows that for all $n \in \{2:N\}$,

$$\begin{aligned} \|\llbracket R^n(v_{\mathcal{T}}^\tau) - v_{\mathcal{T}}^n \rrbracket_F\|_{I_n \times F} &\lesssim \tau \|\partial_t(\llbracket R^n(v_{\mathcal{T}}^\tau) - v_{\mathcal{T}}^n \rrbracket_F)\|_{I_n \times F} \\ &\lesssim \tau^{\frac{1}{2}} \|\llbracket \llbracket v_{\mathcal{T}}^\tau \rrbracket_F \rrbracket^{n-1}\|_F \lesssim \|\llbracket v_{\mathcal{T}}^\tau \rrbracket_F\|_{(I_{n-1} \cup I_n) \times F}. \end{aligned}$$

Summing this relation over $n \in \{2:N\}$, we get

$$\|h^{-\frac{1}{2}} \llbracket R^\tau(v_{\mathcal{T}}^\tau) - v_{\mathcal{T}}^\tau \rrbracket_{\mathcal{F}_h^{\text{int}}}\|_{J \times \mathcal{F}_h^{\text{int}}} \lesssim \|h^{-\frac{1}{2}} \llbracket v_{\mathcal{T}}^\tau \rrbracket_{\mathcal{F}_h^{\text{int}}}\|_{J \times \mathcal{F}_h^{\text{int}}}.$$

Finally, (29) follows from

$$\begin{aligned} \|h^{-\frac{1}{2}} \llbracket R^\tau(v_{\mathcal{T}}^\tau) \rrbracket_{\mathcal{F}_h^{\text{int}}}\|_{J \times \mathcal{F}_h^{\text{int}}} &\leq \|h^{-\frac{1}{2}} \llbracket R^\tau(v_{\mathcal{T}}^\tau) - v_{\mathcal{T}}^\tau \rrbracket_{\mathcal{F}_h^{\text{int}}}\|_{J \times \mathcal{F}_h^{\text{int}}} + \|h^{-\frac{1}{2}} \llbracket v_{\mathcal{T}}^\tau \rrbracket_{\mathcal{F}_h^{\text{int}}}\|_{J \times \mathcal{F}_h^{\text{int}}} \\ &\lesssim \|h^{-\frac{1}{2}} \llbracket v_{\mathcal{T}}^\tau \rrbracket_{\mathcal{F}_h^{\text{int}}}\|_{J \times \mathcal{F}_h^{\text{int}}} \lesssim d_h(\hat{v}_h^\tau, \hat{v}_h^\tau)^{\frac{1}{2}}. \end{aligned}$$

3.8.2 Proof of Lemma 8

Let $v \in H^{\ell+1}(I_n; H^2(T)) \cap H^1(I_n; H^{k+1}(T))$. Recall that $I_T^n(v) := \Pi_T^k(\check{I}_n^\ell(v))$ for all $n \in \{1:N\}$ and all $T \in \mathcal{T}_h$ (see (35)), where Π_T^k is the L^2 -orthogonal projection onto $\mathbb{P}^k(T)$ and \check{I}_n^ℓ is the approximation operator defined in (33)-(34). The stability and approximation properties of Π_T^k are classical (see, e.g., [15, Sec. 11.5.3]); those of \check{I}_n^ℓ are discussed in [16, Sec. 69.3.2 & Ex. 69.7].

(i) Proof of (46). The triangle inequality gives

$$\|I_T^n(v) - v\|_{I_n \times T} \leq \|\Pi_T^k(\check{I}_n^\ell(v)) - \check{I}_n^\ell(v)\|_{I_n \times T} + \|\check{I}_n^\ell(v) - v\|_{I_n \times T}.$$

The first term on the right-hand side is bounded as

$$\|\Pi_T^k(\check{I}_n^\ell(v)) - \check{I}_n^\ell(v)\|_{I_n \times T} \lesssim h^{k+1} \|\check{I}_n^\ell(v)\|_{L^2(I_n; H^{k+1}(T))} \lesssim h^{k+1} \|v\|_{H^1(I_n; H^{k+1}(T))},$$

where we used the approximation properties in space of Π_T^k in the first estimate and the H^1 -stability in time of \check{I}_n^ℓ in the second estimate. Moreover, the second term on the right-hand side is bounded as

$$\|\check{I}_n^\ell(v) - v\|_{I_n \times T} \lesssim \tau^{\ell+1} \|v\|_{H^{\ell+1}(I_n; L^2(T))},$$

where we used the approximation properties in time of \check{I}_n^ℓ . Combining these two estimates proves (46).

(ii) Proof of (47), (48), (49), and (50). We can use similar arguments to those invoked in Step (i) since we have

$$\begin{aligned} \|\nabla(I_T^n(v) - v)\|_{I_n \times T} &\leq \|\nabla(\Pi_T^k(\check{I}_n^\ell(v)) - \check{I}_n^\ell(v))\|_{I_n \times T} + \|\check{I}_n^\ell(\nabla v) - \nabla v\|_{I_n \times T}, \\ \|\partial_t(I_T^n(v) - v)\|_{I_n \times T} &\leq \|\Pi_T^k(\partial_t \check{I}_n^\ell(v)) - \partial_t \check{I}_n^\ell(v)\|_{I_n \times T} + \|\partial_t(\check{I}_n^\ell(v) - v)\|_{I_n \times T}, \end{aligned}$$

because the operators \check{I}_n^ℓ and ∇ commute, as well as the operators Π_T^k and ∂_t . A similar argument can also be used when replacing ∇ by Δ .

(iii) The proofs of (51), (52), and (53) employ similar arguments.

4 Numerical results

In this section, we present numerical experiments to check the convergence rates established in Theorem 9.

All the errors are computed as the difference between the numerical solution and the $L^2(J; L^2(\Omega))$ -orthogonal projection of the exact solution. These errors are measured in the $L^2(T_1, T_2; H^1(B))$ -seminorm, where $(T_1, T_2) \times B$ is specified below for each test case. A noise is added to the measurements in the following way: at each evaluation of the exact solution u to compute g (i.e., at each space-time Gauss node that belongs to $J \times \varpi$), the random noise added to the exact solution is computed as $\delta(t, \mathbf{x}) := a * \text{rand}()$, where $\text{rand}()$ is a C++ function returning a random number in $[-1, 1]$. The noise amplitude is chosen as $a \in \{0, 10^{-5}, 10^{-3}\}$. Moreover, the Tikhonov regularization coefficient is always set to $\gamma := 10^{-3}$.

All the tests are run with the `DiSk++` library [12], and all the linear systems are solved using the Pardiso solver from the MKL library. During the matrix assembly process, we compute the value of the coefficients for the first time interval I_1 and for its coupling with the next time interval I_2 . Then, since the time step is constant, the coefficients associated with the following time intervals are the same and therefore do not need to be recomputed.

4.1 One-dimensional test cases

In this section, we consider the following one-dimensional setting:

$$\Omega := (0, 1), \quad \varpi := (0.25, 0.75), \quad T_f := 2, \quad u(t, x) := \cos(\pi t) \sin(\pi x).$$

All the errors are estimated in the subdomain

$$(T_1, T_2) \times B := (0.2, 1.8) \times (0.125, 0.875).$$

We consider two settings. First, the case with unknown initial data but with known Dirichlet boundary data on $J \times \partial\Omega$. In this situation, the boundary conditions can be enforced in a strong way on the boundary face degrees of freedom by searching \hat{u}_h^τ in \tilde{U}_{h0}^τ and not in \tilde{U}_h^τ . Notice that in this case, we have $\alpha = 1$ in Lemma 1 (see Theorem 2 from [7]). We then expect the same convergence rates as for a well-posed problem, i.e., k in

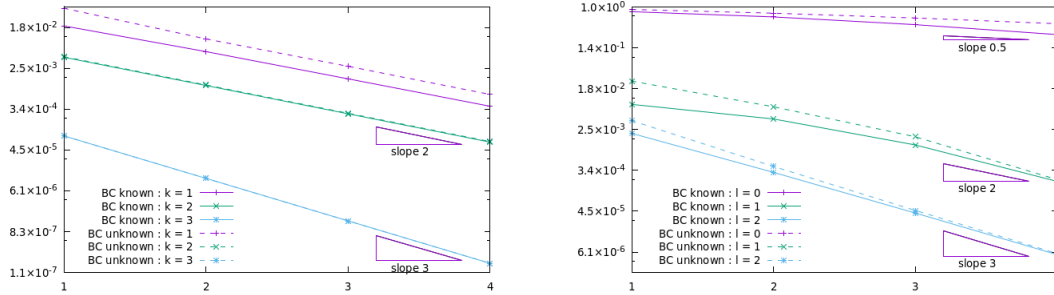


Figure 1: $L^2(T_1, T_2; H^1(B))$ -seminorm errors for the 1d test case without any noise. Both cases with known or unknown boundary data are considered. Left: space convergence ($M \in \{16, 32, 64, 128\}$, $N = 128$, $\ell = 3$). Right: time convergence ($N \in \{10, 20, 40, 80\}$, $M = 256$, $k = 3$).

space and $\ell + \frac{1}{2}$ in time. Then, we consider the case where both initial and boundary data are unknown.

We use a uniform mesh in space and in time (N cells in time and M cells in space). Four levels of refinement are considered in space ($M \in \{16, 32, 64, 128\}$) and in time ($N \in \{10, 20, 40, 80\}$). We run two convergence studies. At first, for a good precision in time ($N = 128$, $\ell = 3$), we study the error for several successive space mesh sizes $M \in \{16, 32, 64, 128\}$ and several polynomial orders $k \in \{1, 2, 3\}$. Then, for a good precision in space ($M = 256$, $k = 3$), we study the error for several successive time mesh sizes $N \in \{10, 20, 40, 80\}$ and several polynomial orders $\ell \in \{0, 1, 2\}$. The results without any noise are reported in Figure 1. We observe optimal space convergence at rate k for $k \in \{2, 3\}$. Instead, we obtain superconvergence at rate 2 for $k = 1$. Moreover, we observe time convergence at rate $\ell + 1$, for $\ell \in \{1, 2\}$, which is slightly better than the expected rate $\ell + \frac{1}{2}$ from Theorem 9. (Recall that the errors are computed as the difference between the numerical solution and a projection of the exact solution.) Another relevant observation is that a higher-order scheme is more efficient in terms of degrees of freedom. Indeed, the scheme with $\ell = 1$ and $N = 40$ has the same number of degrees of freedom as the backward Euler scheme with $\ell = 0$ and $N = 80$, but delivers better results. The same remark can be made for $\ell = 2$.

Let us now consider some noise in the observations. In this situation, we obtain the same results as above until the error becomes small and then the convergence stops. To illustrate this point, let us focus on the case with unknown initial and boundary data. The results are reported in Figure 2 for $\ell = 2$, $M = 256$, $k = 3$, and various mesh refinements in time. As expected, the higher the noise, the larger the value at which the error stagnates. Notice that in the case of known boundary data, the sensitivity to noise is much smaller (results not shown for brevity).

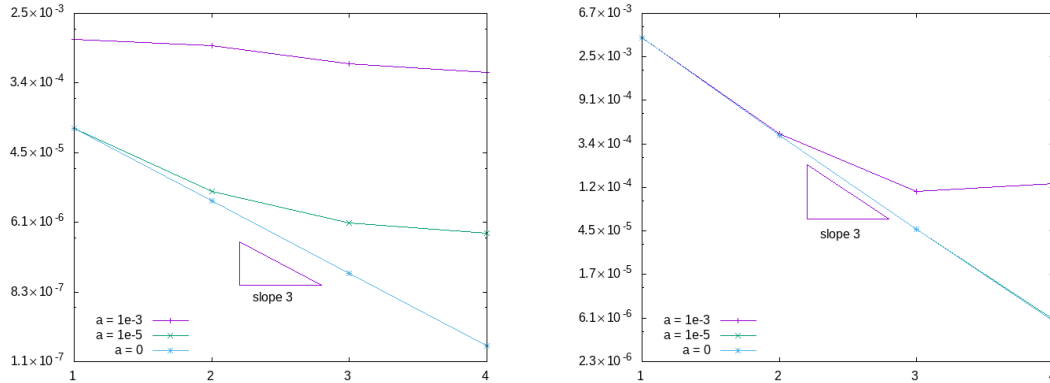


Figure 2: $L^2(T_1, T_2; H^1(B))$ -seminorm errors for the 1d test case. Initial and unknown boundary data with three levels of noise. Left: space convergence ($M \in \{16, 32, 64, 128\}$, $k = 3$, $N = 128$, $\ell = 3$). Right: time convergence ($N \in \{10, 20, 40, 80\}$, $\ell = 2$, $M = 256$, $k = 3$).

4.2 Two-dimensional test case

In this section, we consider the following two-dimensional setting:

$$\begin{aligned} \Omega &:= (0, 1)^2, \quad \varpi := \Omega \setminus (0, 0.875) \times (0.125, 0.875), \quad T_f := 2, \\ (T_1, T_2) \times B &:= (0.2, 1.8) \times (0.125, 0.875)^2, \quad u(t, x, y) := \cos(\pi t) \sin(\pi x) \sin(\pi y). \end{aligned}$$

We show in Figure 3 the domains ϖ and B . Notice that $\overline{B} \cap \partial\Omega = \emptyset$ and that $\varpi \not\subset B$. Both cases with and without boundary data are considered.

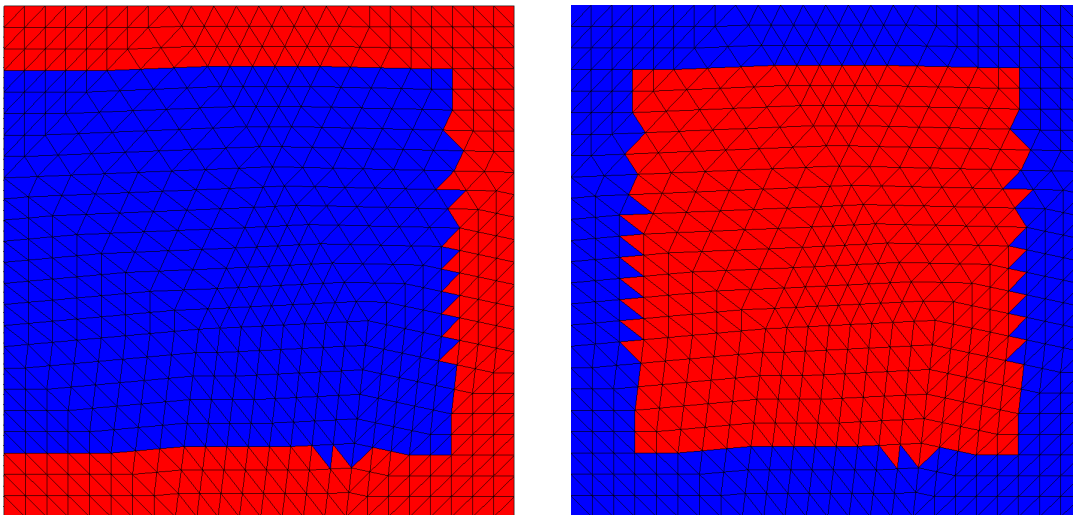


Figure 3: The domains ϖ (left) and B (right) for the 2d test case (in red).

The space convergence is performed with the three triangulations shown in Figure 4 and $k \in \{1, 2\}$. Notice that these meshes are not fitted to ϖ or B . In practice, we consider

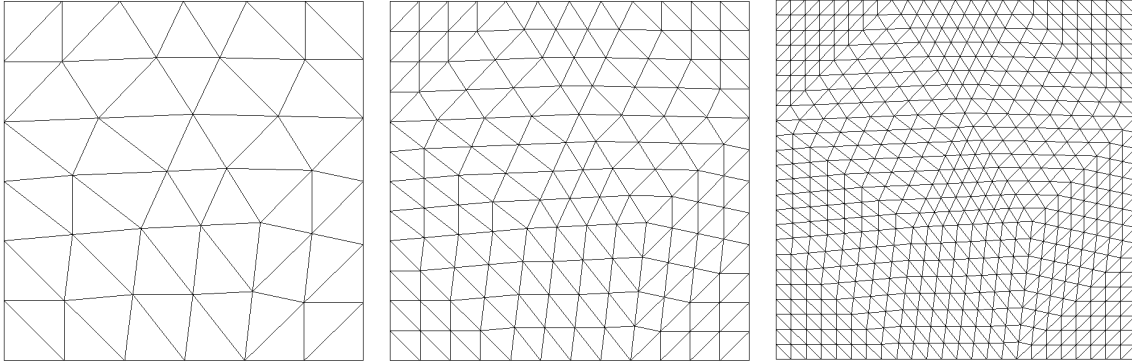


Figure 4: The three meshes used for the 2d test case.

that a triangle is in ϖ (respectively, in B) if its barycenter belongs to ϖ (respectively, to B); see Figure 4. Moreover, we used $N = 10$ time intervals and the time degree $\ell = 2$. The time convergence is performed with $N \in \{5, 10, 20\}$, $\ell \in \{0, 1\}$ and $k = 2$ with the rightmost mesh from Figure 4. Notice that the linear system is generally ill-conditioned, thus precluding the use of effective iterative solvers. The study of adapted preconditioners could help solving this problem, see for instance [20], but we postpone further studies of this point to future work.

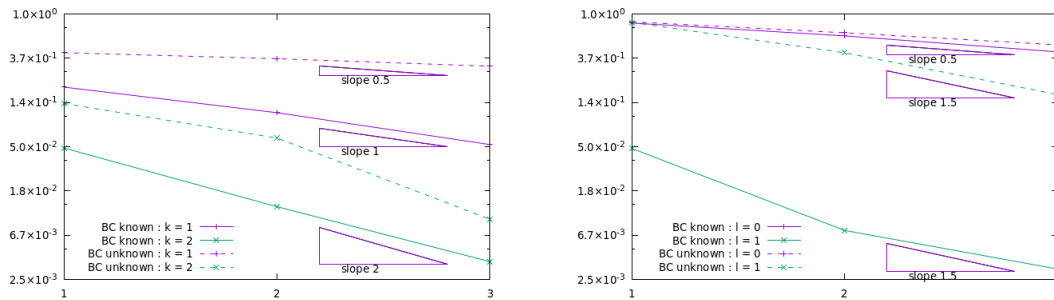


Figure 5: $L^2(T_1, T_2; H^1(B))$ -seminorm errors for the 2d test case with noised data ($a = 10^{-3}$). Left: space convergence ($N = 20$, $\ell = 1$). Right: time convergence (third mesh from Figure 4, $k = 2$).

The $L^2(T_1, T_2; H^1(B))$ -seminorm errors obtained with noised data are reported in Figure 5. The convergence in space (left panel) indicates that a better convergence rate is obtained in the case of known boundary data since the convergence rate is k in this situation. Moreover, this convergence rate is reduced in the case of unknown boundary data: we get a convergence at rate less than 0.5 for $k = 1$. For $k = 2$, it seems that we still get a convergence rate around 2, but it is not as clearcut as when boundary data are known. Also notice that, since the convergence rate for $k = 1$ is 1 in the case of known boundary data, we can conjecture that the superconvergence observed in the 1d test case was due to the fact that the domain was one-dimensional.

The convergence in time (right panel of Figure 5) seems to correspond to a convergence at rate $\ell + \frac{1}{2}$, even if the boundary conditions are not known. Notice that for the one-

dimensional case, we got $\ell + 1$ for $\ell \in \{1, 2\}$. However, according to Theorem 9, we should get a convergence rate of $(\ell + \frac{1}{2})\alpha$. Thus, it seems that, in this numerical experiment, the reduction of convergence due to the lack of boundary data affects only the space convergence.

Acknowledgement: GD was supported by the Emergence grant of Sorbonne University. EB was supported by EPSRC grants EP/T033126/1 and EP/V050400/1.

References

- [1] E. Bécache, L. Bourgeois, L. Franceschini, and J. Dardé. Application of mixed formulations of quasi-reversibility to solve ill-posed problems for heat and wave equations: the 1D case. *Inverse Probl. Imaging*, 9(4):971–1002, 2015.
- [2] L. Bourgeois. A mixed formulation of quasi-reversibility to solve the Cauchy problem for Laplace’s equation. *Inverse Problems*, 21(3):1087–1104, 2005.
- [3] L. Bourgeois. Convergence rates for the quasi-reversibility method to solve the Cauchy problem for Laplace’s equation. *Inverse Problems*, 22(2):413–430, 2006.
- [4] E. Burman. Stabilized finite element methods for nonsymmetric, noncoercive, and ill-posed problems. Part I: Elliptic equations. *SIAM J. Sci. Comput.*, 35(6):A2752–A2780, 2013.
- [5] E. Burman. Error estimates for stabilized finite element methods applied to ill-posed problems. *C. R. Math. Acad. Sci. Paris*, 352(7-8):655–659, 2014.
- [6] E. Burman and A. Ern. Continuous interior penalty *hp*-finite element methods for advection and advection-diffusion equations. *Math. Comp.*, 76(259):1119–1140, 2007.
- [7] E. Burman and L. Oksanen. Data assimilation for the heat equation using stabilized finite element methods. *Numer. Math.*, 139(3):505–528, 2018.
- [8] E. Burman, P. Hansbo, and M. G. Larson. Solving ill-posed control problems by stabilized finite element methods: an alternative to Tikhonov regularization. *Inverse Problems*, 34(3):035004, 36, 2018.
- [9] E. Burman, J. Ish-Horowicz, and L. Oksanen. Fully discrete finite element data assimilation method for the heat equation. *ESAIM Math. Model. Numer. Anal.*, 52(5):2065–2082, 2018.
- [10] E. Burman, G. Delay, and A. Ern. A hybridized high-order method for unique continuation subject to the Helmholtz equation. *SIAM J. Numer. Anal.*, 59(5):2368–2392, 2021.
- [11] E. Burman, A. Feizmohammadi, A. Münch, and L. Oksanen. Space time stabilized finite element methods for a unique continuation problem subject to the wave equation. *ESAIM Math. Model. Numer. Anal.*, 55(suppl.):S969–S991, 2021.

- [12] M. Cicuttin, D. A. Di Pietro, and A. Ern. Implementation of Discontinuous Skeletal methods on arbitrary-dimensional, polytopal meshes using generic programming. *J. Comput. Appl. Math.*, 2017.
- [13] W. Dahmen, R. Stevenson, and J. Westerdiep. Accuracy controlled data assimilation for parabolic problems. *Math. Comp.*, 91(334):557–595, 2022.
- [14] J. Dardé, A. Hannukainen, and N. Hyvönen. An H_{div} -based mixed quasi-reversibility method for solving elliptic Cauchy problems. *SIAM J. Numer. Anal.*, 51(4):2123–2148, 2013.
- [15] A. Ern and J.-L. Guermond. *Finite elements. I Approximation and Interpolation*, volume 72 of *Texts in Applied Mathematics*. Springer, Cham, 2021.
- [16] A. Ern and J.-L. Guermond. *Finite elements III. First-order and time-dependent PDEs*, volume 74 of *Texts in Applied Mathematics*. Springer, Cham, 2021.
- [17] K. Ito and B. Jin. *Inverse problems. Tikhonov theory and algorithms*, volume 22 of *Series on Applied Mathematics*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2015. ISBN 978-981-4596-19-0.
- [18] S. E. Jenkins, C. J. Budd, M. A. Freitag, and N. D. Smith. The effect of numerical model error on data assimilation. *J. Comput. Appl. Math.*, 290:567–588, 2015.
- [19] M. V. Klibanov. Estimates of initial conditions of parabolic equations and inequalities via lateral Cauchy data. *Inverse Problems*, 22(2):495–514, 2006.
- [20] K.-A. Mardal, B. F. Nielsen, and M. Nordaas. Robust preconditioners for PDE-constrained optimization with limited observations. *BIT Numer Math*, 57:405–431, 2017.
- [21] L. Oksanen. Private communication, 2021.
- [22] L. Robbiano. Théorème d’unicité adapté au contrôle des solutions des problèmes hyperboliques. *Comm. Partial Differential Equations*, 16(4-5):789–800, 1991.
- [23] I. Smears. Robust and efficient preconditioners for the discontinuous Galerkin time-stepping method. *IMA J. Numer. Anal.*, 37(4):1961–1985, 2017.
- [24] M. Tadi, M. V. Klibanov, and W. Cai. An inversion method for parabolic equations based on quasireversibility. *Comput. Math. Appl.*, 43(8-9):927–941, 2002.
- [25] S. Vessella. Unique continuation properties and quantitative estimates of unique continuation for parabolic equations. In *Handbook of differential equations: evolutionary equations. Vol. V*, Handb. Differ. Equ., pages 421–500. Elsevier/North-Holland, Amsterdam, 2009. doi: 10.1016/S1874-5717(08)00212-0. URL [https://doi.org/10.1016/S1874-5717\(08\)00212-0](https://doi.org/10.1016/S1874-5717(08)00212-0).
- [26] Y. B. Wang, J. Cheng, J. Nakagawa, and M. Yamamoto. A numerical method for solving the inverse heat conduction problem without initial value. *Inverse Probl. Sci. Eng.*, 18(5):655–671, 2010.