SIAM J. NUMER. ANAL.
Vol. 61, No. 5, pp. 2534–2557

THE UNIQUE CONTINUATION PROBLEM FOR THE HEAT EQUATION DISCRETIZED WITH A HIGH-ORDER SPACE-TIME NONCONFORMING METHOD*

ERIK BURMAN†, GUILLAUME DELAY‡, AND ALEXANDRE ERN§

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 in a target space-time subdomain 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. Owing to the ill-posed nature of the problem, an additional estimate on the residual dual norm is needed to prove the convergence of the discrete solution to the exact solution in the energy norm in the target space-time subdomain. This is achieved by combining the above results 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.

Key words. unique continuation, data assimilation, heat equation, discontinuous Galerkin, hybridized discontinuous Galerkin, regularization, error estimate

MSC codes. 65M30, 65M60, 65J20, 35R25, 35K05

DOI. 10.1137/22M1508637

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 $u : J \times \Omega \to \mathbb{R}$ that satisfies

$$L(u) := \frac{\partial u}{\partial t} - \Delta u = f \qquad \text{in } J \times \Omega, \tag{1.1}$$

$$u = g \qquad \text{in } J \times \varpi, \tag{1.2}$$

*Received by the editors July 12, 2022; accepted for publication (in revised form) May 30, 2023; published electronically October 25, 2023.
https://doi.org/10.1137/22M1508637
Funding: The work of the first author was supported by EPSRC grants EP/T033126/1 and EP/V050400/1. The work of the second author was supported by the Emergence grant of Sorbonne University.

†Department of Mathematics, University College London, London, WC1E 6BT, UK (e.burman@ucl.ac.uk).
‡Sorbonne Université, CNRS, Université Paris Cité, LJLL, F-75005 Paris, France (guillaume.delay@sorbonne-universite.fr).
§CERMICS, École des Ponts, 77455, Marne-la-Vallée cedex 2, and INRIA, 75589 Paris, France (alexandre.ern@enpc.fr).

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 the restriction to $J \times \varpi$ of a solution to the heat equation in $J \times \Omega$. The model problem (1.1)–(1.2) is ill-posed. Indeed, whenever a solution exists (this is the case whenever g satisfies the heat equation (1.1) 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 1.1 (conditional stability estimate). *Let B be a connected subset of Ω such that $\overline{B} \subset \Omega$. 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*

$$\begin{aligned} \|v\|_{L^2(T_1,T_2;H^1(B))} \le C_{\text{stb}} \big(\|v\|_{L^2(J;L^2(\varpi))} + \|L(v)\|_{L^2(J;H^{-1}(\Omega))}\big)^{\alpha} \\ \times \big(\|v\|_{L^2(J;L^2(\Omega))} + \|L(v)\|_{L^2(J;H^{-1}(\Omega))}\big)^{1-\alpha}. \end{aligned} \tag{1.3}$$

Lemma 1.1 with the additional assumption $\varpi \subset B$ corresponds to Theorem 1 in [8]. This assumption can be removed by using the arguments of the proof of Theorem 1.1 from [23]; see also [24, 17, 1]. The conditional stability estimate from Lemma 1.1 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 $k \ge 1$ (resp., $\ell \ge 0$) 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 3.8). We also mention that Lemma 1.1 admits some interesting extensions. First, in the case of known boundary conditions (i.e., if we add the assumption $v = 0$ on $\partial\Omega$), then Lemma 1.1 remains valid with $T_2 = T_f$, $B = \Omega$, and $\alpha = 1$ (see Theorem 2 from [8]). This setting will be considered in our numerical experiments as well. Moreover, in the case of known initial conditions (i.e., if we add the assumption $v(0, .) = v_0$ in Ω for some $v_0 \in L^2(\Omega)$), then Lemma 1.1 remains valid with $T_1 = 0$ provided $\|v_0\|_\Omega$ is added to $\|L(v)\|_{L^2(J;H^{-1}(\Omega))}$ in (1.3) [13]. Other stability estimates are available in the literature, for instance, using Cauchy data on the boundary instead of interior data [22]. For an overview on analysis techniques of unique continuation for parabolic equations, we refer the reader to [27].

Since the problem (1.1)–(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 [3, 4, 16] or the Tikhonov regularization [20]. These methods have already been applied to solve the data assimilation problem subject to the heat equation in the one-dimensional case; see [26, 2]. Other numerical approaches have also been proposed and analyzed for the one-dimensional heat equation (see [21, 28]), but, to the best of our knowledge, none of the abovementioned 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 [5, 6, 9] for the stationary version of the present problem. The analysis was extended to the data assimilation problem subject to the Helmholtz equation using a high-order discretization in [11]. Data assimilation subject to nonstationary problems was considered in [12] for the wave equation and in [8, 10] for the heat equation with known Dirichlet boundary data for the numerical tests. In particular, [8] considers the semidiscretization in space using lowest-order Lagrange finite elements, discussing several stability situations, whereas [10] 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 [15], 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 3.2), 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 data assimilation 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 $L^2(J; H^1_0(\Omega))$-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. Moreover, in addition to the main analysis arguments used for well-posed problems (inf-sup stability, consistency), we need an additional, nontrivial estimate on a suitable dual residual norm. This estimate is indeed key to invoke the conditional stability estimate from Lemma 1.1 and prove an error estimate in the energy norm.

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, which combines arguments relevant to well-posed problems (sections 3.2 to 3.4) with arguments specific to ill-posed problems (section 3.5). These steps prepare the stage for our main results, which are the error estimate (Theorem 3.6) and the devising of the regularization to achieve optimal error decay rates (Theorem 3.8). Finally, in section 4, we present some numerical results that corroborate our theoretical results. Moreover, we illustrate the benefits of high-order discretization methods.

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 $p \le q$, we use the notation $\{p{:}q\} := \{m \in \mathbb{N},\, p \le m \le q\}$. 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 $0 = t_0 < t_1 < \cdots < t_N = T_f$, 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 the above partition of 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 $[\![\Phi^\tau]\!]^n := \Phi^\tau(t_n^+) - \Phi^\tau(t_n^-)$. For convenience, we also set $[\![\Phi^\tau]\!]^0 := 0$; 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 $\boldsymbol{n}_T$ 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, $\mathcal{F}_h^{\mathrm{int}}$, and the set of the mesh boundary faces, $\mathcal{F}_h^{\partial}$. Any mesh interface $F \in \mathcal{F}_h^{\mathrm{int}}$ is oriented by a fixed unit normal vector $\boldsymbol{n}_F$. Moreover, for a piecewise smooth function v and any mesh interface $F \in \mathcal{F}_h^{\mathrm{int}}$, $[\![v]\!]_F$ denotes the jump of v across F in the direction of $\boldsymbol{n}_F$. We also use the broken gradient and Laplacian operators, $\nabla_{\mathcal{T}}$ and $\Delta_{\mathcal{T}}$, which are defined such that $(\nabla_{\mathcal{T}} v)|_T := \nabla(v|_T)$ and $(\Delta_{\mathcal{T}} v)|_T := \Delta(v|_T)$ 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 $k \ge 1$ be the polynomial degree of the hybridized dG method in space and let $\ell \ge 0$ 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

$$\hat{U}_h^k := U_{\mathcal{T}}^k \times U_{\mathcal{F}}^k, \quad U_{\mathcal{T}}^k := \times_{T \in \mathcal{T}_h} \mathbb{P}^k(T), \quad U_{\mathcal{F}}^k := \times_{F \in \mathcal{F}_h} \mathbb{P}^k(F). \tag{2.1}$$

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 (dofs) attached to the mesh boundary faces are null. For a generic pair $\hat{v}_h \in \hat{U}_h^k$, its dofs associated with a generic mesh cell $T \in \mathcal{T}_h$ are denoted by

$$\hat{v}_T := (v_T, v_{\partial T} := (v_F)_{F\in\mathcal{F}_{\partial T}}) \in \hat{U}_T^k := \mathbb{P}^k(T) \times \mathbb{P}^k(\mathcal{F}_{\partial T}), \tag{2.2}$$

where $\mathbb{P}^k(\mathcal{F}_{\partial T}) := \times_{F\in\mathcal{F}_{\partial T}} \mathbb{P}^k(F)$ 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

$$\widetilde{U}_h^\tau := \{\hat{v}_h^\tau \in L^2(J;\hat{U}_h^k) \mid \hat{v}_h^\tau|_{I_n} \in \mathbb{P}^\ell(I_n;\hat{U}_h^k)\ \forall n \in \{1{:}N\}\}, \tag{2.3}$$

$$\widetilde{U}_{h0}^\tau := \{\hat{v}_h^\tau \in \widetilde{U}_h^\tau \mid \hat{v}_h^\tau|_{I_n} \in \mathbb{P}^\ell(I_n;\hat{U}_{h0}^k)\ \forall n \in \{1{:}N\}\}. \tag{2.4}$$

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

$$U_{\mathcal{T}}^\tau := \{v_{\mathcal{T}}^\tau \in L^2(J;U_{\mathcal{T}}^k) \mid v_{\mathcal{T}}^n := v_{\mathcal{T}}^\tau|_{I_n} \in \mathbb{P}^\ell(I_n;U_{\mathcal{T}}^k)\ \forall n \in \{1{:}N\}\}, \tag{2.5}$$

$$U_{\mathcal{F}}^\tau := \{v_{\mathcal{F}}^\tau \in L^2(J;U_{\mathcal{F}}^k) \mid v_{\mathcal{F}}^n := v_{\mathcal{F}}^\tau|_{I_n} \in \mathbb{P}^\ell(I_n;U_{\mathcal{F}}^k)\ \forall n \in \{1{:}N\}\}. \tag{2.6}$$

For a generic function $\hat{v}_h^\tau \in \widetilde{U}_h^\tau$, 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 $\widetilde{U}_h^\tau$ (primal variables) and let $\hat{\zeta}_h^\tau$, $\hat{\eta}_h^\tau$ be generic functions in $\widetilde{U}_{h0}^\tau$ (dual variables). We use the subscript $_h$ to indicate the bilinear forms related to the space discretization, and the superscript $^\tau$ to indicate those related to the time discretization. The two bilinear forms associated with the discretization of the heat equation are

$$a_h(\hat{v}_h^\tau, \hat{\eta}_h^\tau) := \sum_{T\in\mathcal{T}_h} \big\{(\nabla v_T^\tau, \nabla \eta_T^\tau)_{J\times T} - (\nabla v_T^\tau{\cdot}\boldsymbol{n}_T, \eta_T^\tau - \eta_{\partial T}^\tau)_{J\times\partial T}$$
$$- (v_T^\tau - v_{\partial T}^\tau, \nabla\eta_T^\tau{\cdot}\boldsymbol{n}_T)_{J\times\partial T}\big\}, \tag{2.7}$$

$$b^\tau(v_{\mathcal{T}}^\tau, \eta_{\mathcal{T}}^\tau) := \sum_{n\in\{1:N\}} \big\{(\partial_t v_{\mathcal{T}}^n, \eta_{\mathcal{T}}^n)_{I_n\times\Omega} + ([\![v_{\mathcal{T}}^\tau]\!]^{n-1}, \eta_{\mathcal{T}}^n(t_{n-1}^+))_\Omega\big\}. \tag{2.8}$$

The stabilization bilinear forms read as follows:

$$s_h^\tau(\hat{v}_h^\tau, \hat{w}_h^\tau) := d_h(\hat{v}_h^\tau, \hat{w}_h^\tau) + d^\tau(v_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau), \tag{2.9}$$

$$\sigma_h(\hat{\zeta}_h^\tau, \hat{\eta}_h^\tau) := (\nabla_{\mathcal{T}}\zeta_{\mathcal{T}}^\tau, \nabla_{\mathcal{T}}\eta_{\mathcal{T}}^\tau)_{J\times\Omega} + d_h(\hat{\zeta}_h^\tau, \hat{\eta}_h^\tau), \tag{2.10}$$

with

$$d_h(\hat{v}_h^\tau, \hat{w}_h^\tau) := \sum_{T\in\mathcal{T}_h} h^{-1}(v_T^\tau - v_{\partial T}^\tau, w_T^\tau - w_{\partial T}^\tau)_{J\times\partial T}, \tag{2.11}$$

$$d^\tau(v_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau) := \sum_{n\in\{1:N-1\}} \big\{([\![v_{\mathcal{T}}^\tau]\!]^n, [\![w_{\mathcal{T}}^\tau]\!]^n)_\Omega + ([\![\nabla_{\mathcal{T}} v_{\mathcal{T}}^\tau]\!]^n, [\![\nabla_{\mathcal{T}} w_{\mathcal{T}}^\tau]\!]^n)_\Omega\big\}. \tag{2.12}$$

(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

$$t_h^\tau(v_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau) := \gamma c(\tau,h)^2 (v_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau)_{J\times\Omega}, \tag{2.13}$$

$$m_\varpi(v_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau) := (v_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau)_{J\times\varpi}, \tag{2.14}$$

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 (3.34)). For later use, we define

$$\|w_{\mathcal{T}}^\tau\|_{J\times\varpi}^2 := m_\varpi(w_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau). \tag{2.15}$$

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

$$\begin{aligned}\mathcal{L}_h^\tau(\hat{v}_h^\tau, \hat{\zeta}_h^\tau) := \frac{1}{2}\|v_{\mathcal{T}}^\tau - g_\delta\|_{J\times\varpi}^2 + \frac{1}{2}t_h^\tau(v_{\mathcal{T}}^\tau, v_{\mathcal{T}}^\tau) + \frac{1}{2}s_h^\tau(\hat{v}_h^\tau, \hat{v}_h^\tau) - \frac{1}{2}\sigma_h(\hat{\zeta}_h^\tau, \hat{\zeta}_h^\tau)\\ + a_h(\hat{v}_h^\tau, \hat{\zeta}_h^\tau) + b^\tau(v_{\mathcal{T}}^\tau, \zeta_{\mathcal{T}}^\tau) - (f, \zeta_{\mathcal{T}}^\tau)_{J\times\Omega},\end{aligned} \tag{2.16}$$

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 motivation for looking for a saddle-point of the discrete Lagrangian is to minimize the discrepancy with respect to data under the constraint of the partial differential equation. This idea is classical, but the key point here is that the Lagrangian is defined at the discrete level. This makes it simpler to enhance the convexity/concavity of the discrete Lagrangian by suitable consistent stabilization terms; see [9] for further discussion.

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

$$m_\varpi(u_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau) + t_h^\tau(u_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau) + s_h^\tau(\hat{u}_h^\tau, \hat{w}_h^\tau) + a_h(\hat{w}_h^\tau, \hat{\xi}_h^\tau) + b^\tau(w_{\mathcal{T}}^\tau, \xi_{\mathcal{T}}^\tau) = m_\varpi(g_\delta, w_{\mathcal{T}}^\tau), \tag{2.17}$$

$$a_h(\hat{u}_h^\tau, \hat{\eta}_h^\tau) + b^\tau(u_{\mathcal{T}}^\tau, \eta_{\mathcal{T}}^\tau) - \sigma_h(\hat{\xi}_h^\tau, \hat{\eta}_h^\tau) = (f, \eta_{\mathcal{T}}^\tau)_{J\times\Omega}, \tag{2.18}$$

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

$$\begin{aligned}A_h^\tau((\hat{v}_h^\tau, \hat{\zeta}_h^\tau), (\hat{w}_h^\tau, \hat{\eta}_h^\tau)) := m_\varpi(v_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau) + t_h^\tau(v_{\mathcal{T}}^\tau, w_{\mathcal{T}}^\tau) + s_h^\tau(\hat{v}_h^\tau, \hat{w}_h^\tau) + a_h(\hat{w}_h^\tau, \hat{\zeta}_h^\tau)\\ + b^\tau(w_{\mathcal{T}}^\tau, \zeta_{\mathcal{T}}^\tau) + b^\tau(v_{\mathcal{T}}^\tau, \eta_{\mathcal{T}}^\tau) + a_h(\hat{v}_h^\tau, \hat{\eta}_h^\tau) - \sigma_h(\hat{\zeta}_h^\tau, \hat{\eta}_h^\tau).\end{aligned} \tag{2.19}$$

The discrete problem (2.17)–(2.18) can be rewritten as follows: Find $(\hat{u}_h^\tau, \hat{\xi}_h^\tau) \in \widetilde{U}_h^\tau \times \widetilde{U}_{h0}^\tau$ such that

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

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., [19, section 69.2.3] or [25, section 2.3] and the references therein. Then, we study the stablity properties of A_h^τ in a suitable residual norm, we introduce interpolation operators in space and in time, and we bound the consistency error. The second important step, specific to ill-posed problems, is to bound a suitable dual residual norm. We combine these bounds with the abstract conditional estimate from Lemma 1.1 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 $A \le CB$ 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

$$(\partial_t R^n(v_{\mathcal{T}}^\tau), q_{\mathcal{T}}^n)_{I_n\times\Omega} := (\partial_t v_{\mathcal{T}}^n, q_{\mathcal{T}}^n)_{I_n\times\Omega} + ([\![v_{\mathcal{T}}^\tau]\!]^{n-1}, q_{\mathcal{T}}^n(t_{n-1}^+))_\Omega \tag{3.1}$$

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 in $L^2(J; U_{\mathcal{T}}^k)$. The main consequence of (3.1) is that the bilinear form b^τ can be rewritten as

$$b^\tau(v_{\mathcal{T}}^\tau, \eta_{\mathcal{T}}^\tau) = (\partial_t R^\tau(v_{\mathcal{T}}^\tau), \eta_{\mathcal{T}}^\tau)_{J\times\Omega} \quad \forall (v_{\mathcal{T}}^\tau, \eta_{\mathcal{T}}^\tau) \in U_{\mathcal{T}}^\tau \times U_{\mathcal{T}}^\tau. \tag{3.2}$$

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

$$R^n(v_{\mathcal{T}}^\tau)(t,\boldsymbol{x}) = v_{\mathcal{T}}^n(t,\boldsymbol{x}) - [\![v_{\mathcal{T}}]\!]^{n-1}(\boldsymbol{x}) \frac{(-1)^\ell}{2}(L_\ell - L_{\ell+1}) \circ T_n^{-1}(t) \tag{3.3}$$

for all $(t,\boldsymbol{x}) \in I_n \times \Omega$, 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 (3.3) is that

$$R^n(v_{\mathcal{T}}^\tau)(t_n^-) = v_{\mathcal{T}}^n(t_n^-) \quad \forall n \in \{1{:}N\}. \tag{3.4}$$

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

LEMMA 3.1 (stability properties of R^τ). *For all $v_{\mathcal{T}}^\tau \in U_{\mathcal{T}}^\tau$, we have*

$$\|\partial_t^\tau(R^\tau(v_{\mathcal{T}}^\tau) - v_{\mathcal{T}}^\tau)\|_{J\times\Omega} \lesssim \tau^{-\frac{1}{2}} d^\tau(v_{\mathcal{T}}^\tau, v_{\mathcal{T}}^\tau)^{\frac{1}{2}}, \tag{3.5}$$

$$\|R^\tau(v_{\mathcal{T}}^\tau) - v_{\mathcal{T}}^\tau\|_{J\times\Omega} + \|\nabla_{\mathcal{T}}(R^\tau(v_{\mathcal{T}}^\tau) - v_{\mathcal{T}}^\tau)\|_{J\times\Omega} \lesssim \tau^{\frac{1}{2}} d^\tau(v_{\mathcal{T}}^\tau, v_{\mathcal{T}}^\tau)^{\frac{1}{2}}. \tag{3.6}$$

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

$$\|h^{-\frac{1}{2}}[\![R^\tau(v_{\mathcal{T}}^\tau)]\!]_{\mathcal{F}_h^{int}}\|_{J\times\mathcal{F}_h^{int}} := \left(\sum_{F\in\mathcal{F}_h^{int}} h^{-1}\|[\![R^\tau(v_{\mathcal{T}}^\tau)]\!]_F\|_{J\times F}^2\right)^{\frac{1}{2}} \lesssim d_h(\hat{v}_h^\tau, \hat{v}_h^\tau)^{\frac{1}{2}}. \tag{3.7}$$

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

$$|\!|\!|\hat{v}_h^\tau, \hat{\zeta}_h^\tau|\!|\!|^2 := \|v_{\mathcal{T}}^\tau\|_{\mathrm{R}}^2 + \|v_{\mathcal{T}}^\tau\|_{J\times\varpi}^2 + t_h^\tau(v_{\mathcal{T}}^\tau, v_{\mathcal{T}}^\tau) + s_h^\tau(\hat{v}_h^\tau, \hat{v}_h^\tau) + \sigma_h(\hat{\zeta}_h^\tau, \hat{\zeta}_h^\tau), \tag{3.8}$$

with

$$\|v_{\mathcal{T}}^\tau\|_{\mathrm{R}}^2 := \|h(\partial_t R^\tau(v_{\mathcal{T}}^\tau) - \Delta_{\mathcal{T}} v_{\mathcal{T}}^\tau)\|_{J\times\Omega}^2 + \|h^{\frac{1}{2}}[\![\nabla v_{\mathcal{T}}^\tau]\!]_{\mathcal{F}_h^{\mathrm{int}}}\cdot\boldsymbol{n}_{\mathcal{F}_h^{\mathrm{int}}}\|_{J\times\mathcal{F}_h^{\mathrm{int}}}^2, \tag{3.9}$$

and $\|h^{\frac{1}{2}}[\![\nabla v_{\mathcal{T}}^\tau]\!]_{\mathcal{F}_h^{\mathrm{int}}}\cdot\boldsymbol{n}_{\mathcal{F}_h^{\mathrm{int}}}\|_{J\times\mathcal{F}_h^{\mathrm{int}}}^2 := \sum_{F\in\mathcal{F}_h^{\mathrm{int}}} h\|[\![\nabla v_{\mathcal{T}}^\tau]\!]_F\cdot\boldsymbol{n}_F\|_{J\times F}^2$. We observe that $|\!|\!|\cdot|\!|\!|$ defines a norm on $\widetilde{U}_h^\tau \times \widetilde{U}_{h0}^\tau$. Our key stability result is the following inf-sup condition. Notice that this result implies the existence and uniqueness of the solution to the discrete problem.

LEMMA 3.2 (inf-sup condition). *The following holds for all $(\hat{v}_h^\tau, \hat{\zeta}_h^\tau) \in \widetilde{U}_h^\tau \times \widetilde{U}_{h0}^\tau$:*

$$|\!|\!|\hat{v}_h^\tau, \hat{\zeta}_h^\tau|\!|\!| \lesssim \sup_{(\hat{w}_h^\tau, \hat{\eta}_h^\tau)\in\widetilde{U}_h^\tau\times\widetilde{U}_{h0}^\tau\setminus\{(0,0)\}} \frac{A_h^\tau((\hat{v}_h^\tau, \hat{\zeta}_h^\tau), (\hat{w}_h^\tau, \hat{\eta}_h^\tau))}{|\!|\!|\hat{w}_h^\tau, \hat{\eta}_h^\tau|\!|\!|}. \tag{3.10}$$

Proof. Let us denote by S the right-hand side of (3.10).

(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

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

(ii) Let $\hat{\eta}_h^\tau := (0, (\eta_F^\tau)_{F\in\mathcal{F}_h})$ with $\eta_F^\tau := h[\![\nabla v_\mathcal{T}^\tau]\!]_F \cdot \boldsymbol{n}_F$ 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

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

Moreover, we also have

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

This implies that

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

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

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

Using the definition of $\hat{\eta}_h^\tau$ and recalling the rewriting (3.2) of b^τ, we have

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

since a discrete trace inverse inequality implies that

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

Furthermore, invoking again inverse inequalities, we have

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

This implies that

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

(iv) Gathering the previous estimates leads to

$$\begin{aligned}\|v_{\mathcal{T}}^\tau\|_{J\times\varpi}^2 &+ t_h^\tau(v_{\mathcal{T}}^\tau, v_{\mathcal{T}}^\tau) + s_h^\tau(\hat{v}_h^\tau, \hat{v}_h^\tau) + \sigma_h(\hat{\zeta}_h^\tau, \hat{\zeta}_h^\tau) + \|h^{\frac{1}{2}}[\![\nabla v_{\mathcal{T}}^\tau]\!]_{\mathcal{F}_h^{\text{int}}}\cdot \boldsymbol{n}_{\mathcal{F}_h^{\text{int}}}\|_{J\times\mathcal{F}_h^{\text{int}}}^2 \\ &+ \|h(\partial_t R^\tau(v_{\mathcal{T}}^\tau) - \Delta_{\mathcal{T}} v_{\mathcal{T}}^\tau)\|_{J\times\Omega}^2 \lesssim S |\!|\!| \hat{v}_h^\tau, \hat{\zeta}_h^\tau |\!|\!| + S^2.\end{aligned}$$

Recalling the definition (3.8) of the triple norm and using Young's inequality gives the expected inf-sup condition. ∎

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

$$\check{I}_n^\ell(v)(t_n^-) := v(t_n), \tag{3.11}$$

$$\int_{I_n} (\check{I}_n^\ell(v) - v, q^n)_\Omega := 0 \qquad \forall q^n \in \mathbb{P}^{\ell-1}(I_n; H^1(\Omega)). \tag{3.12}$$

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 $v \in H^1(J; H^1(\Omega))$,

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

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 $F \in \mathcal{F}_h$,

$$I_T^n(v) := I_T^\tau(v)|_{I_n} := \Pi_T^k(\check{I}_n^\ell(v)), \qquad I_F^n(v) := I_F^\tau(v)|_{I_n} := \Pi_F^k(\check{I}_n^\ell(v)). \tag{3.13}$$

We notice that the approximation operators in space and in time commute. Moreover, proceeding as in [19, Lemma 69.16], we derive the following useful orthogonality property: For all $v \in H^1(J; H^1(\Omega))$ and all $w_{\mathcal{T}}^\tau \in \mathbb{P}^\ell(J; U_{\mathcal{T}}^k)$,

$$(\partial_t v - \partial_t R^\tau(I_{\mathcal{T}}^\tau(v)), w_{\mathcal{T}}^\tau)_{J\times\Omega} = 0. \tag{3.14}$$

Let us finally define the discrete and interpolation errors on the primal unknown. Recall that u denotes the solution to the exact problem (1.1)–(1.2) and that $(\hat{u}_h^\tau, \hat{\xi}_h^\tau)$ denotes the solution to the discrete problem (2.17)–(2.18). The discrete and interpolation errors on the primal unknown are then defined as

$$\hat{e}_h^\tau := \hat{u}_h^\tau - \hat{I}_h^\tau(u), \qquad \hat{\theta}_h^\tau := (u, u|_{\mathcal{F}_h}) - \hat{I}_h^\tau(u), \tag{3.15}$$

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

$$
\begin{aligned}
\|\hat\theta_h^\tau\|_\#^2 := {} & \|\nabla_\mathcal{T}\theta_\mathcal{T}^\tau\|_{J\times\Omega}^2 + \|h\Delta_\mathcal{T}\theta_\mathcal{T}^\tau\|_{J\times\Omega}^2 + \|\theta_\mathcal{T}^\tau\|_{J\times\Omega}^2 \\
& + \sum_{T\in\mathcal{T}_h}\left\{\|h^{\frac12}\nabla\theta_T^\tau\|_{J\times\partial T}^2 + \|h^{-\frac12}\theta_T^\tau\|_{J\times\partial T}^2 + \|h^{-\frac12}\theta_{\partial T}^\tau\|_{J\times\partial T}^2\right\} \\
& + \sum_{n\in\{1:N-1\}}\left\{\|[\![\theta_\mathcal{T}^\tau]\!]^n\|_\Omega^2 + \|[\![\nabla_\mathcal{T}\theta_\mathcal{T}^\tau]\!]^n\|_\Omega^2\right\} + \tau\|\partial_t^\tau\theta_\mathcal{T}^\tau\|_{J\times\Omega}^2. \qquad (3.16)
\end{aligned}
$$

Recall that the stability norm $|\!|\!|\cdot|\!|\!|$ is defined in (3.8)–(3.9). In what follows, we assume that

$$
u \in H^1(J;H^1(\Omega)) \cap L^2(J;H^2(\Omega)). \qquad (3.17)
$$

Lemma 3.3 (consistency and boundedness). *Let $(\hat u_h^\tau, \hat\xi_h^\tau)$ denote the solution to the discrete problem* (2.17)–(2.18). *Let $\hat e_h^\tau$ and $\hat\theta_h^\tau$ be the discrete and interpolation errors on the primal unknown defined in* (3.15). *Under the regularity assumption* (3.17), *we have, for all $(\hat w_h^\tau, \hat\eta_h^\tau) \in \widetilde U_h^\tau \times \widetilde U_{h0}^\tau$,*

$$
|A_h^\tau((\hat e_h^\tau, \hat\xi_h^\tau), (\hat w_h^\tau, 0))| \lesssim \left(\|\hat\theta_h^\tau\|_\# + \|\delta\|_{J\times\varpi} + t_h^\tau(I_\mathcal{T}^\tau(u), I_\mathcal{T}^\tau(u))^{\frac12}\right)|\!|\!|\hat w_h^\tau, 0|\!|\!|, \qquad (3.18)
$$

$$
|A_h^\tau((\hat e_h^\tau, \hat\xi_h^\tau), (0, \hat\eta_h^\tau))| \lesssim \|\hat\theta_h^\tau\|_\# |\!|\!|0, \hat\eta_h^\tau|\!|\!|. \qquad (3.19)
$$

Proof. (i) Proof of (3.18). Using $(\hat e_h^\tau, \hat\xi_h^\tau) = (\hat u_h^\tau, \hat\xi_h^\tau) - (\hat I_h^\tau(u), 0)$, (2.17) in the discrete problem, the definition (2.19) of A_h^τ, and $g_\delta = g + \delta$, we have, for all $\hat w_h^\tau \in \widetilde U_h^\tau$,

$$
\begin{aligned}
& A_h^\tau((\hat e_h^\tau, \hat\xi_h^\tau), (\hat w_h^\tau, 0)) \\
& \quad = m_\varpi(g_\delta, w_\mathcal{T}^\tau) - m_\varpi(I_\mathcal{T}^\tau(u), w_\mathcal{T}^\tau) - t_h^\tau(I_\mathcal{T}^\tau(u), w_\mathcal{T}^\tau) - s_h^\tau(\hat I_h^\tau(u), \hat w_h^\tau) \\
& \quad = m_\varpi(\delta, w_\mathcal{T}^\tau) + m_\varpi(\theta_\mathcal{T}^\tau, w_\mathcal{T}^\tau) - t_h^\tau(I_\mathcal{T}^\tau(u), w_\mathcal{T}^\tau) - s_h^\tau(\hat I_h^\tau(u), \hat w_h^\tau).
\end{aligned}
$$

The Cauchy–Schwarz inequality yields

$$
\begin{aligned}
|m_\varpi(\delta, w_\mathcal{T}^\tau)| + |m_\varpi(\theta_\mathcal{T}^\tau, w_\mathcal{T}^\tau)| &\lesssim (\|\delta\|_{J\times\varpi} + \|\theta_\mathcal{T}^\tau\|_{J\times\varpi})\|w_\mathcal{T}^\tau\|_{J\times\varpi}, \\
|t_h^\tau(I_\mathcal{T}^\tau(u), w_\mathcal{T}^\tau)| &\lesssim t_h^\tau(I_\mathcal{T}^\tau(u), I_\mathcal{T}^\tau(u))^{\frac12} t_h^\tau(w_\mathcal{T}^\tau, w_\mathcal{T}^\tau)^{\frac12}, \\
|s_h^\tau(\hat I_h^\tau(u), \hat w_h^\tau)| &\lesssim s_h^\tau(\hat I_h^\tau(u), \hat I_h^\tau(u))^{\frac12} s_h^\tau(\hat w_h^\tau, \hat w_h^\tau)^{\frac12}.
\end{aligned}
$$

Moreover, we have

$$
\begin{aligned}
s_h^\tau(\hat I_h^\tau(u), \hat I_h^\tau(u)) = {} & \sum_{T\in\mathcal{T}_h} h^{-1}\|I_T^\tau(u) - I_{\partial T}^\tau(u)\|_{J\times\partial T}^2 \\
& + \sum_{n\in\{1:N-1\}}\left\{\|[\![\theta_\mathcal{T}^\tau]\!]^n\|_\Omega^2 + \|[\![\nabla_\mathcal{T}\theta_\mathcal{T}^\tau]\!]^n\|_\Omega^2\right\} \lesssim \|\hat\theta_h^\tau\|_\#^2.
\end{aligned}
$$

The bound (3.18) follows by gathering the above estimates.

(ii) Proof of (3.19). Proceeding as above and using now (2.18) in the discrete problem, we have, for all $\hat\eta_h^\tau \in \widetilde U_{h0}^\tau$,

$$
\begin{aligned}
A_h^\tau((\hat e_h^\tau, \hat\xi_h^\tau), (0, \hat\eta_h^\tau)) &= (f, \eta_\mathcal{T}^\tau)_{J\times\Omega} - b^\tau(I_\mathcal{T}^\tau(u), \eta_\mathcal{T}^\tau) - a_h(\hat I_h^\tau(u), \hat\eta_h^\tau) \\
&= -a_h(\hat I_h^\tau(u), \hat\eta_h^\tau) - (\Delta u, \eta_\mathcal{T}^\tau)_{J\times\Omega},
\end{aligned}
$$

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 (3.14) and (3.2). Recalling the definition (2.7) of a_h, we infer that

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

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 $\sum_{T\in\mathcal{T}_h}(\nabla u\cdot\boldsymbol{n}_T, \eta_{\partial T}^\tau)_{J\times\partial T} = 0$. Using Cauchy–Schwarz and inverse inequalities, we readily obtain (3.19). ◻

Lemma 3.4 (a priori residual bound). *Under the regularity assumption* (3.17)*, we have*

$$\|\!|\hat{e}_h^\tau, \hat{\xi}_h^\tau|\!\|+ t_h^\tau(u_{\mathcal{T}}^\tau, u_{\mathcal{T}}^\tau)^{\frac12} + s_h^\tau(\hat{u}_h^\tau, \hat{u}_h^\tau)^{\frac12} \lesssim \|\hat{\theta}_h^\tau\|_\# + \|\delta\|_{J\times\varpi} + t_h^\tau(I_{\mathcal{T}}^\tau(u), I_{\mathcal{T}}^\tau(u))^{\frac12}.$$

Proof. Using inf-sup stability (Lemma 3.2) yields

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

Moreover, owing to consistency (Lemma 3.3), we have, for all $(\hat{w}_h^\tau, \hat{\eta}_h^\tau) \in \widetilde{U}_h^\tau \times \widetilde{U}_{h0}^\tau$,

$$|A_h^\tau((\hat{e}_h^\tau, \hat{\xi}_h^\tau), (\hat{w}_h^\tau, \hat{\eta}_h^\tau))| \lesssim (\|\hat{\theta}_h^\tau\|_\# + \|\delta\|_{J\times\varpi} + t_h^\tau(I_{\mathcal{T}}^\tau(u), I_{\mathcal{T}}^\tau(u))^{\frac12})\|\!|\hat{w}_h^\tau, \hat{\eta}_h^\tau|\!\|.$$

Combining the two bounds gives

$$\|\!|\hat{e}_h^\tau, \hat{\xi}_h^\tau|\!\| \lesssim \|\hat{\theta}_h^\tau\|_\# + \|\delta\|_{J\times\varpi} + t_h^\tau(I_{\mathcal{T}}^\tau(u), I_{\mathcal{T}}^\tau(u))^{\frac12}.$$

Finally, observing that

$$
\begin{aligned}
&t_h^\tau(u_{\mathcal{T}}^\tau, u_{\mathcal{T}}^\tau)^{\frac12} \le t_h^\tau(I_{\mathcal{T}}^\tau(u), I_{\mathcal{T}}^\tau(u))^{\frac12} + t_h^\tau(e_{\mathcal{T}}^\tau, e_{\mathcal{T}}^\tau)^{\frac12} \le t_h^\tau(I_{\mathcal{T}}^\tau(u), I_{\mathcal{T}}^\tau(u))^{\frac12} + \|\!|\hat{e}_h^\tau, \hat{\xi}_h^\tau|\!\|,\\
&s_h^\tau(\hat{u}_h^\tau, \hat{u}_h^\tau)^{\frac12} \le s_h^\tau(\hat{I}_h^\tau(u), \hat{I}_h^\tau(u))^{\frac12} + s_h^\tau(\hat{e}_h^\tau, \hat{e}_h^\tau)^{\frac12} \le \|\hat{\theta}_h^\tau\|_\# + \|\!|\hat{e}_h^\tau, \hat{\xi}_h^\tau|\!\|
\end{aligned}
$$

concludes the proof. ◻

Remark 3.1 (Lemma 3.4). 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 space-discrete functions $v_{\mathcal{T}}^\tau \in H^1(J; U_{\mathcal{T}}^k)$ by setting

$$\langle L_{\mathcal{T}}(v_{\mathcal{T}}^\tau), \eta\rangle_{L^2(H^{-1}),L^2(H_0^1)} := (\partial_t v_{\mathcal{T}}^\tau, \eta)_{J\times\Omega} + (\nabla_{\mathcal{T}} v_{\mathcal{T}}^\tau, \nabla\eta)_{J\times\Omega}. \tag{3.20}$$

The corresponding dual norm is

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

It is also useful to introduce the data oscillation term

$$\Theta_f := f - \Pi^\tau(f), \tag{3.21}$$

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 3.5 (bound on dual error norm). *Under the regularity assumption* (3.17), *we have*

$$\begin{aligned}&\|L_{\mathcal{T}}(R^\tau(u_{\mathcal{T}}^\tau)-u)\|_{L^2(J;H^{-1}(\Omega))}\\ &\quad\lesssim (1+\tau^{-\frac12}h)\|\hat\theta_h^\tau\|_\# + \|\delta\|_{J\times\varpi} + t_h^\tau(I_{\mathcal{T}}^\tau(u),I_{\mathcal{T}}^\tau(u))^{\frac12} + \|\Theta_f\|_{J\times\Omega}.\end{aligned}$$

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

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

where we used the L^2-orthogonality properties of Π^τ. Let $\hat\eta_h^\tau\in\widetilde U_{h0}^\tau$ be such that

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

for all $n\in\{1{:}N\}$. Invoking the second equation (2.18) in the discrete problem and recalling the rewriting (3.2) 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 rearranging 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 (2.7) of the bilinear form a_h shows that

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

since we have

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

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 $(\nabla u_T^{\tau}\cdot \boldsymbol{n}_T)|_{I_n\times\partial T} \in \mathbb{P}^{\ell}(I_n; \mathbb{P}^k(\mathcal{F}_{\partial T}))$ 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 (3.9) of the $\|\cdot\|_{\mathrm{R}}$-norm, we have

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

where we used that

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

Owing to Lemma 3.4, we have $\|e_{\mathcal{T}}^{\tau}\|_{\mathrm{R}} \lesssim \|\hat{\theta}_h^{\tau}\|_{\#} + \|\delta\|_{J\times\varpi} + t_h^{\tau}(I_{\mathcal{T}}^{\tau}(u), I_{\mathcal{T}}^{\tau}(u))^{\frac12}$. Moreover, the triangle inequality, the estimate (3.5), and the above bound on $\|\Pi^{\tau}(\eta) - \eta_{\mathcal{T}}^{\tau}\|_{J\times\Omega}$ give

$$
\begin{aligned}
&|(\partial_t(R^{\tau}(I_{\mathcal{T}}^{\tau}(u)) - u), \Pi^{\tau}(\eta) - \eta_{\mathcal{T}}^{\tau})_{J\times\Omega}| \\
&\quad \lesssim \big(\|\partial_t^{\tau}(R^{\tau}(I_{\mathcal{T}}^{\tau}(u)) - I_{\mathcal{T}}^{\tau}(u))\|_{J\times\Omega} + \|\partial_t^{\tau}(u - I_{\mathcal{T}}^{\tau}(u))\|_{J\times\Omega} \big) h \|\nabla \eta\|_{J\times\Omega} \\
&\quad \lesssim \big(d^{\tau}(I_{\mathcal{T}}^{\tau}(u), I_{\mathcal{T}}^{\tau}(u))^{\frac12} + \tau^{\frac12} \|\partial_t^{\tau} \theta_{\mathcal{T}}^{\tau}\|_{J\times\Omega} \big) \tau^{-\frac12} h \|\nabla \eta\|_{J\times\Omega}.
\end{aligned}
$$

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

$$
|A_1| \lesssim \big((1 + \tau^{-\frac12} h) \|\hat{\theta}_h^{\tau}\|_{\#} + \|\delta\|_{J\times\varpi} + t_h^{\tau}(I_{\mathcal{T}}^{\tau}(u), I_{\mathcal{T}}^{\tau}(u))^{\frac12} \big) \|\nabla \eta\|_{J\times\Omega}.
$$

(ii) Bound on A_2. Owing to the Cauchy–Schwarz inequality and (3.6), we have

$$
|(\nabla_{\mathcal{T}}(R^{\tau}(u_{\mathcal{T}}^{\tau}) - u_{\mathcal{T}}^{\tau}), \nabla \eta)_{J\times\Omega}| \lesssim \tau^{\frac12} d^{\tau}(u_{\mathcal{T}}^{\tau}, u_{\mathcal{T}}^{\tau})^{\frac12} \|\nabla \eta\|_{J\times\Omega}.
$$

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

$$
\left| \sum_{T\in\mathcal{T}_h} (u_T^{\tau} - u_{\partial T}^{\tau}, \nabla \eta_T^{\tau}\cdot \boldsymbol{n}_T)_{J\times\partial T} \right| \lesssim d_h(\hat{u}_h^{\tau}, \hat{u}_h^{\tau})^{\frac12} \|\nabla \eta\|_{J\times\Omega}.
$$

Thus, we have

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

Using $\tau \lesssim 1$ and Lemma 3.4 yields

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

(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 3.4. 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

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

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

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 1.1. We use the target (semi)norm

$$\|v\|_{\mathrm{trg}} := \|\nabla_{\mathcal{T}} v\|_{L^2(T_1,T_2;L^2(B))},$$

where T_1, T_2, and B are defined in Lemma 1.1. Recall the quantities $C_{\mathrm{stb}} > 0$ and $\alpha \in (0,1]$ introduced in Lemma 1.1. Recall that the $\|\cdot\|_\#$-norm is defined in (3.16) and the data oscillation term Θ_f in (3.21), that $\hat{\theta}_h^\tau$ is the interpolation error defined in (3.15), and that the coefficient $c(\tau,h)$ is used to weight the regularization bilinear form t_h^τ defined in (2.13).

THEOREM 3.6 (error estimate in target subdomain). *Under the regularity assumption* (3.17)*, we have*

$$\begin{aligned}\|u - u_{\mathcal{T}}^\tau\|_{\mathrm{trg}} \lesssim{}& C_{\mathrm{stb}}(1+\tau^{-1}h)^\alpha(1+\tau^{-1}h+\gamma^{-\frac{1}{2}}c(\tau,h)^{-1})^{1-\alpha}\\ &\times \big(\|\hat{\theta}_h^\tau\|_\# + \|\delta\|_{J\times\varpi} + t_h^\tau(I_{\mathcal{T}}^\tau(u), I_{\mathcal{T}}^\tau(u))^{\frac{1}{2}} + \|\Theta_f\|_{J\times\Omega}\big).\end{aligned} \tag{3.22}$$

Proof. (i) Using an averaging operator in space (see [7, Lemmas 3.2 and 5.3] or [18, Chapter 22] and the references therein), we can build a piecewise polynomial function $\widetilde{u}_{\mathcal{T}}^\tau \in \{v_{\mathcal{T}}^\tau \in H^1(J;H^1(\Omega)) \mid v_{\mathcal{T}}^\tau|_{I_n\times T} \in \mathbb{P}^{\ell+1}(I_n;\mathbb{P}^k(T)), n\in\{1{:}N\}, T\in\mathcal{T}_h\}$ such that

$$\begin{aligned}&\|h^{-1}(\widetilde{u}_{\mathcal{T}}^\tau - R^\tau(u_{\mathcal{T}}^\tau))\|_{J\times\Omega} + \|\nabla_{\mathcal{T}}(\widetilde{u}_{\mathcal{T}}^\tau - R^\tau(u_{\mathcal{T}}^\tau))\|_{J\times\Omega}\\ &\quad\lesssim \|h^{-\frac{1}{2}}[\![R^\tau(u_{\mathcal{T}}^\tau)]\!]_{\mathcal{F}_h^{\mathrm{int}}}\|_{J\times\mathcal{F}_h^{\mathrm{int}}}\\ &\quad\lesssim d_h(\hat{u}_h^\tau, \hat{u}_h^\tau)^{\frac{1}{2}},\end{aligned} \tag{3.23}$$

where the first bound results from the approximation properties in space of the averaging operator and the second bound from (3.7) (see Lemma 3.1). Invoking the triangle inequality yields

$$\|u - u_{\mathcal{T}}^\tau\|_{\mathrm{trg}} \le \|u_{\mathcal{T}}^\tau - \widetilde{u}_{\mathcal{T}}^\tau\|_{\mathrm{trg}} + \|u - \widetilde{u}_{\mathcal{T}}^\tau\|_{\mathrm{trg}},$$

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

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

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

The first term on the right-hand side is estimated by invoking (3.6) (see Lemma 3.1) and the second term by invoking (3.23). Owing to Lemma 3.4 and since $\tau \lesssim 1$, we infer that

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

(iii) Bound on $\|u-\widetilde{u}_{\mathcal{T}}^{\tau}\|_{\mathrm{trg}}$. Since $\widetilde{u}_{\mathcal{T}}^{\tau}-u \in H^1(J;H^1(\Omega)) \subset H^1(J;H^{-1}(\Omega)) \cap L^2(J;H^1(\Omega))$, we can invoke the conditional stability estimate from Lemma 1.1. This gives

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

It remains to bound $\|\widetilde{u}_{\mathcal{T}}^{\tau}-u\|_{J\times\varpi}$, $\|\widetilde{u}_{\mathcal{T}}^{\tau}-u\|_{J\times\Omega}$, and $\|L(\widetilde{u}_{\mathcal{T}}^{\tau}-u)\|_{L^2(J;H^{-1}(\Omega))}$.

(iii.a) Bound on $\|\widetilde{u}_{\mathcal{T}}^{\tau}-u\|_{J\times\varpi}$. We have

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

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 (3.6) (see Lemma 3.1) and the bound (3.23) imply that

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

where we used the definition (2.9) of s_h^{τ} and $\tau \lesssim 1$, $h \lesssim 1$. Gathering the above estimates and using Lemma 3.4 gives

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

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

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

The terms $\|u-I_{\mathcal{T}}^{\tau}(u)\|_{J\times\Omega}+\|u_{\mathcal{T}}^{\tau}-\widetilde{u}_{\mathcal{T}}^{\tau}\|_{J\times\Omega}$ 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. Recalling the definition (2.13) of t_h^{τ}, we have

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

Invoking Lemma 3.4 gives

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

(iii.c) Bound on $\|L(\widetilde{u}_{\mathcal{T}}^{\tau}-u)\|_{L^2(J;H^{-1}(\Omega))}$. Recalling that $L_{\mathcal{T}}$ denotes the extension of L to $H^1(J;U_{\mathcal{T}}^k)$, the triangle inequality gives

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

For the second term on the right-hand side, we invoke Lemma 3.5 to infer

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

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

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

where we used an inverse inequality in time, the estimate (3.23), and a global Poincaré inequality in Ω for η. On the other hand, using again (3.23) gives

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

Combining the last two bounds and invoking Lemma 3.4, we infer that

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

Altogether, this gives

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

(iv) Combining the bounds from the above steps proves (3.22). ∎

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

$$\|I_T^n(v)-v\|_{I_n\times T}\lesssim\tau^{\ell+1}\|v\|_{H^{\ell+1}(I_n;L^2(T))}+h^{k+1}\|v\|_{H^1(I_n;H^{k+1}(T))},\tag{3.24}$$

$$\|\nabla(I_T^n(v)-v)\|_{I_n\times T}\lesssim\tau^{\ell+1}\|v\|_{H^{\ell+1}(I_n;H^1(T))}+h^{k}\|v\|_{H^1(I_n;H^{k+1}(T))},\tag{3.25}$$

$$h^{\frac{1}{2}}\|\nabla(I_T^n(v)-v)\|_{I_n\times\partial T}\lesssim h^{\frac{1}{2}}\tau^{\ell+1}\|v\|_{H^{\ell+1}(I_n;H^2(T))}+h^{k}\|v\|_{H^1(I_n;H^{k+1}(T))},\tag{3.26}$$

$$h\|\Delta(I_T^n(v)-v)\|_{I_n\times T}\lesssim h\tau^{\ell+1}\|v\|_{H^{\ell+1}(I_n;H^2(T))}+h^{k}\|v\|_{H^1(I_n;H^{k+1}(T))},\tag{3.27}$$

$$\|\partial_t(I_T^n(v)-v)\|_{I_n\times T}\lesssim\tau^{\ell}\|v\|_{H^{\ell+1}(I_n;L^2(T))}+h^{k+1}\|v\|_{H^1(I_n;H^{k+1}(T))}.\tag{3.28}$$

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

$$\|(I_T^n(v) - v)(s)\|_T \lesssim \tau^{\ell+\frac{1}{2}} \|v\|_{H^{\ell+1}(I_n;L^2(T))} + h^{k+1}\|v\|_{H^1(I_n;H^{k+1}(T))}, \tag{3.29}$$

$$\|\nabla(I_T^n(v) - v)(s)\|_T \lesssim \tau^{\ell+\frac{1}{2}} \|v\|_{H^{\ell+1}(I_n;H^1(T))} + h^{k}\|v\|_{H^1(I_n;H^{k+1}(T))}. \tag{3.30}$$

Finally, we have

$$\begin{aligned} h^{-\frac{1}{2}}\|I_T^n(v) - v\|_{I_n\times\partial T} &+ h^{-\frac{1}{2}}\|I_{\partial T}^n(v) - v\|_{I_n\times\partial T} \\ &\lesssim h^{-\frac{1}{2}}\tau^{\ell+1}\|v\|_{H^{\ell+1}(I_n;H^1(T))} + h^k\|v\|_{H^1(I_n;H^{k+1}(T))}. \end{aligned} \tag{3.31}$$

We consider the functional space

$$U_* := H^{\ell+2}(J;L^2(\Omega)) \cap H^{\ell+1}(J;H^2(\Omega)) \cap H^1(J;H^{k+1}(\Omega)), \tag{3.32}$$

equipped with its natural norm denoted by $\|\cdot\|_*$. An important consequence of Lemma 3.7 is that, under the assumption $u \in U_*$, we have

$$\|\hat{\theta}_h^\tau\|_\# + t_h^\tau(I_{\mathcal{T}}^\tau(u), I_{\mathcal{T}}^\tau(u))^{\frac{1}{2}} + \|\Theta_f\|_{J\times\Omega} \lesssim \big((1+h^{-1}\tau)^{\frac{1}{2}}\tau^{\ell+\frac{1}{2}} + h^k + \gamma^{\frac{1}{2}}c(\tau,h)\big)\|u\|_*, \tag{3.33}$$

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 3.8 (decay rates). *Assume that* $u \in U_*$ *and that* $\tau \lesssim h$, $h \lesssim \tau$. *Set*

$$c(\tau,h) := \tau^{\ell+\frac{1}{2}} + h^k. \tag{3.34}$$

We have the following decay rates on the errors:

$$|\!|\!|\hat{e}_h^\tau, \hat{\xi}_h^\tau|\!|\!| + t_h^\tau(u_{\mathcal{T}}^\tau, u_{\mathcal{T}}^\tau)^{\frac{1}{2}} + s_h^\tau(\hat{u}_h^\tau, \hat{u}_h^\tau)^{\frac{1}{2}} \lesssim (\tau^{\ell+\frac{1}{2}} + h^k)\|u\|_* + \|\delta\|_{J\times\varpi}, \tag{3.35}$$

$$\|u - u_{\mathcal{T}}^\tau\|_{trg} \lesssim C_{stb}(\tau^{\ell+\frac{1}{2}} + h^k)^\alpha(\|u\|_* + (\tau^{\ell+\frac{1}{2}} + h^k)^{-1}\|\delta\|_{J\times\varpi}), \tag{3.36}$$

where the hidden constant scales as $\max(\gamma,1)^{\frac{1}{2}}$ *in* (3.35) *and* $\max(\gamma,\gamma^{\alpha-1})^{\frac{1}{2}}$ *in* (3.36).

Proof. Plug (3.34) into (3.33) and invoke Lemma 3.4 for (3.35) and Theorem 3.6 for (3.36). ∎

Remark 3.2 (choosing discretization parameters). The estimates of Theorem 3.8 indicate that space and time refinements have to be stopped when $(\tau^{\ell+\frac{1}{2}} + h^k)\|u\|_* \simeq \|\delta\|_{J\times\varpi}$. Hence, as the noise level diminishes, finer discretizations can be employed.

3.8. Technical proofs. In this section, we outline the proofs of Lemmas 3.1 and 3.7. The proofs use standard arguments from finite element approximation theory.

3.8.1. Proof of Lemma 3.1. Let $\hat{v}_h^\tau := (v_{\mathcal{T}}^\tau, v_{\mathcal{F}}^\tau) \in \widetilde{U}_h^\tau$.

(i) Proof of (3.5). For all $n \in \{1{:}N\}$ and all $q_{\mathcal{T}}^n \in \mathbb{P}^\ell(I_n; U_{\mathcal{T}}^k)$, we have

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

Invoking a discrete trace inequality in time implies

$$\|\partial_t(R^n(v_{\mathcal{T}}^\tau) - v_{\mathcal{T}}^n)\|_{I_n\times\Omega} \lesssim \tau^{-\frac{1}{2}}\|[\![v_{\mathcal{T}}^\tau]\!]^{n-1}\|_\Omega. \tag{3.37}$$

Summing over $n \in \{1{:}N\}$ proves (3.5) since $\sum_{n\in\{1:N\}}\|[\![v_{\mathcal{T}}^\tau]\!]^{n-1}\|_\Omega^2 \le d^\tau(v_{\mathcal{T}}^\tau, v_{\mathcal{T}}^\tau)$ (recall that $[\![\cdot]\!]^0 = 0$ by convention).

(ii) Proof of (3.6). Since we have $(R^n(v_\mathcal{T}^\tau) - v_\mathcal{T}^n)(t_n^-) = 0$ (see (3.4)), a Poincaré inequality in time over I_n and the bound (3.37) yield

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

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

$$\|R^\tau(v_\mathcal{T}^\tau) - v_\mathcal{T}^\tau\|_{J\times\Omega} \lesssim \tau^{\frac{1}{2}} d^\tau(v_\mathcal{T}^\tau, v_\mathcal{T}^\tau)^{\frac{1}{2}}. \tag{3.38}$$

Moreover, (3.3) shows that the operators R^τ and $\nabla_\mathcal{T}$ commute. Hence, using the same arguments as above shows that

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

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

$$\|\nabla_\mathcal{T}(R^\tau(v_\mathcal{T}^\tau) - v_\mathcal{T}^\tau)\|_{J\times\Omega}^2 \lesssim \tau d^\tau(v_\mathcal{T}^\tau, v_\mathcal{T}^\tau). \tag{3.39}$$

Summing (3.38) and (3.39) proves (3.6).

(iii) Proof of (3.7). For all $F\in\mathcal{F}_h^{\mathrm{int}}$, the operators R^τ and $[\![\cdot]\!]_F$ commute. In particular, we have, for all $n\in\{1{:}N\}$ and all $q_F^n\in\mathbb{P}^\ell(I_n;\mathbb{P}^k(F))$,

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

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

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

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

$$\|h^{-\frac{1}{2}}[\![R^\tau(v_\mathcal{T}^\tau) - v_\mathcal{T}^\tau]\!]_{\mathcal{F}_h^{\mathrm{int}}}\|_{J\times\mathcal{F}_h^{\mathrm{int}}} \lesssim \|h^{-\frac{1}{2}}[\![v_\mathcal{T}^\tau]\!]_{\mathcal{F}_h^{\mathrm{int}}}\|_{J\times\mathcal{F}_h^{\mathrm{int}}}.$$

Finally, (3.7) follows from

$$\begin{aligned}&\|h^{-\frac{1}{2}}[\![R^\tau(v_\mathcal{T}^\tau)]\!]_{\mathcal{F}_h^{\mathrm{int}}}\|_{J\times\mathcal{F}_h^{\mathrm{int}}}\\ &\quad\le \|h^{-\frac{1}{2}}[\![R^\tau(v_\mathcal{T}^\tau) - v_\mathcal{T}^\tau]\!]_{\mathcal{F}_h^{\mathrm{int}}}\|_{J\times\mathcal{F}_h^{\mathrm{int}}} + \|h^{-\frac{1}{2}}[\![v_\mathcal{T}^\tau]\!]_{\mathcal{F}_h^{\mathrm{int}}}\|_{J\times\mathcal{F}_h^{\mathrm{int}}}\\ &\quad\lesssim \|h^{-\frac{1}{2}}[\![v_\mathcal{T}^\tau]\!]_{\mathcal{F}_h^{\mathrm{int}}}\|_{J\times\mathcal{F}_h^{\mathrm{int}}} \lesssim d_h(\hat{v}_h^\tau, \hat{v}_h^\tau)^{\frac{1}{2}}.\end{aligned}$$

3.8.2. Proof of Lemma 3.7. 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 (3.13)), where Π_T^k is the L^2-orthogonal projection onto $\mathbb{P}^k(T)$ and $\check{I}_n^\ell$ is the approximation operator defined in (3.11)–(3.12). The stability and approximation properties of Π_T^k are classical (see, e.g., [18, section 11.5.3]); those of $\check{I}_n^\ell$ are discussed in [19, section 69.3.2 and example 69.7].

(i) Proof of (3.24). The triangle inequality gives

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

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

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

where we used the approximation properties in space of Π_T^k in the first estimate and the H^1-stability in time of $\check{I}_n^\ell$ 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 $\check{I}_n^\ell$. Combining these two estimates proves (3.24).

(ii) Proof of (3.25), (3.26), (3.27), and (3.28). We can use arguments similar to those invoked in step (i) since we have

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

because the operators $\check{I}_n^\ell$ 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 (3.29), (3.30), and (3.31) employ similar arguments.

4. Numerical results. In this section, we present numerical experiments to check the convergence rates established in Theorem 3.8. We also study the influence of the noise in the measurements, and we illustrate the benefits of using a high-order discretization.

The space domain is $\Omega := (0,1)^d$ with $d \in \{1,2\}$. 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, with the subset $(T_1,T_2)\times B$ specified below for each test case. Some noise is added to the measurements in the following way: (i) the time-space domain is divided into 10^{d+1} subdomains; (ii) a random noise level $\delta := a * \text{rand}()$ is assigned to each subdomain, with $a \ge 0$ the noise amplitude and rand() is a `C++` function returning a random number in $[-1,1]$. Thus, every subdomain has the same noise during the whole mesh-refinement process. Moreover, the Tikhonov regularization coefficient is set to $\gamma := 10^{-3}$.

All the tests are run with the `DiSk++` library [14], 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, we address 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 dofs by searching $\hat{u}_h^\tau$ in $\widetilde{U}_{h0}^\tau$ and not in $\widetilde{U}_h^\tau$. Notice that in this case, we have $\alpha = 1$ in Lemma 1.1 (see Theorem 2 from [8]). We then expect the same convergence rates as for a well-posed problem, i.e., k in 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 4.1. 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 3.8. (Recall that the errors are computed as the difference between the numerical solution and a projection of the exact solution.) The errors in Figure 4.1 are reported as a function of the total number of space-time dofs. This gives a perspective on the efficiency of the method with respect to its cost. In particular, we see that, for a given number of dofs, we can reach a better precision by employing a high-order method. We also notice that the errors with known boundary data are always smaller than those with unknown boundary data. For completeness, the number of dofs is reported in Table 4.1. The numbers correspond to the case of known boundary data; those for unknown boundary data are slightly higher due to the use of additional boundary unknowns.

Let us now consider some noise in the measurements. In this situation, we obtain the same results as above until the discrete error becomes small and then the convergence stops. To illustrate this point, let us focus on the highest polynomial degree (which generates the lowest error). The results are reported in Figure 4.2 for $\ell = 2$ and $k = 3$. As expected, the higher the noise, the larger the value at which the error

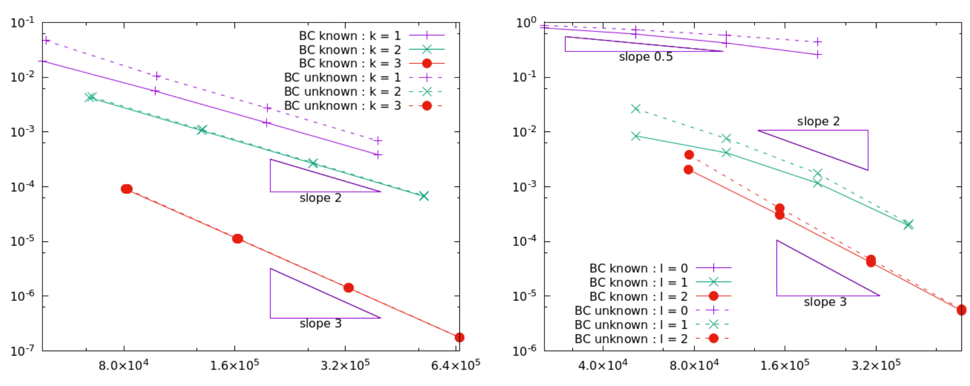

FIG. 4.1. *$L^2(T_1, T_2; H^1(B))$-seminorm errors with respect to the number of dofs for the one-dimensional test case without any noise. Both cases with known or unknown boundary data are considered. Left: space convergence ($M \in \{16, 32, 64, 128\}$, $N = 128$, $\ell = 3$). Right: time convergence ($N \in \{10, 20, 40, 80\}$, $M = 256$, $k = 3$).*

Table 4.1

Number of dofs (one-dimensional test case with known boundary data). Left: space refinement for $\ell = 3$ and $N = 128$. Right: time refinement for $M = 256$ and $k = 3$.

M	$k=1$	$k=2$	$k=3$	N	$\ell=0$	$\ell=1$	$\ell=2$
16	$4.8{\cdot}10^4$	$6.5{\cdot}10^4$	$8.1{\cdot}10^4$	10	$2.6{\cdot}10^4$	$5.1{\cdot}10^4$	$7.7{\cdot}10^4$
32	$9.7{\cdot}10^4$	$1.3{\cdot}10^5$	$1.6{\cdot}10^5$	20	$5.1{\cdot}10^4$	$1.0{\cdot}10^5$	$1.5{\cdot}10^5$
64	$2.0{\cdot}10^5$	$2.6{\cdot}10^5$	$3.3{\cdot}10^5$	40	$1.0{\cdot}10^5$	$2.0{\cdot}10^5$	$3.1{\cdot}10^5$
128	$3.9{\cdot}10^5$	$5.2{\cdot}10^5$	$6.5{\cdot}10^5$	80	$2.0{\cdot}10^5$	$4.1{\cdot}10^5$	$6.1{\cdot}10^5$

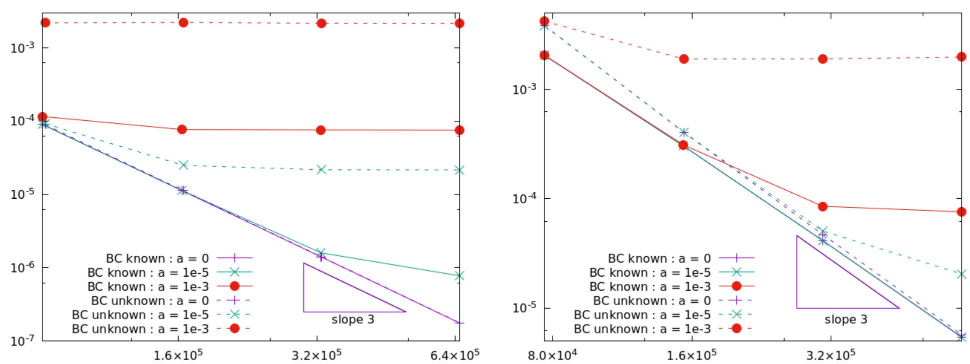

Fig. 4.2. *$L^2(T_1, T_2; H^1(B))$-seminorm errors with respect to the number of dofs for the one-dimensional test case and various noise levels. Left: space convergence ($M \in \{16, 32, 64, 128\}$, $k = 3$, $N = 128$, $\ell = 3$). Right: time convergence ($N \in \{10, 20, 40, 80\}$, $\ell = 2$, $M = 256$, $k = 3$).*

stagnates. More precisely, in the case of unknown boundary data, the error stagnates at about twice the level of noise (2.10^{-5} for $a = 10^{-5}$ and 2.10^{-3} for $a = 10^{-3}$), whereas it stagnates at 8.10^{-5} for $a = 10^{-3}$ in the case of known boundary data. This means that the test case with unknown boundary data is more sensitive to the presence of noise.

4.2. Two-dimensional test case. In this section, we consider the following two-dimensional setting:

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

Notice that $\overline{B} \cap \partial\Omega = \emptyset$ and that $\varpi \not\subset B$. Both cases with and without boundary data are considered.

The space convergence is studied on a sequence of seven triangulations of increasing refinement ($h \in \{0.25, 0.18, 0.125, 0.09, 0.0625, 0.045, 0.03125\}$) and $k \in \{1, 2\}$. The first, third, fifth, and seventh triangulations of the sequence are shown in Figure 4.3. Notice that all the triangulations are fitted to the sets ϖ and B. We use $N = 40$ time intervals and the time degree $\ell = 2$. The time convergence is studied using $N \in \{5, 10, 15, 20, 25, 30\}$, $\ell \in \{0, 1\}$, and the space discretization using the finest mesh and $k = 2$.

The $L^2(T_1, T_2; H^1(B))$-seminorm errors obtained without any noise are reported in Figure 4.4 as a function of the total number of space-time dofs. The convergence in space (left panel) indicates that the scheme has a convergence rate of k with and without the knowledge of boundary data. (Since the number of dofs scales as h^{-d},

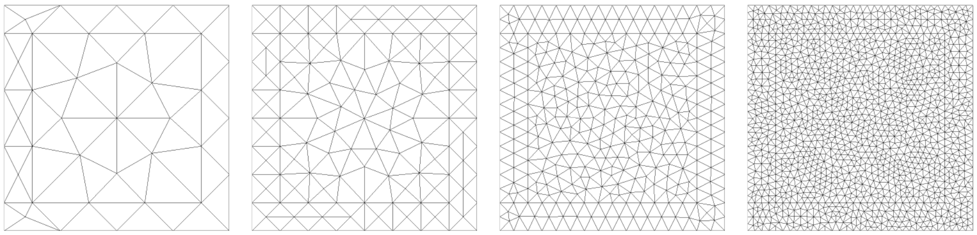

Fig. 4.3. *First, third, fifth, and seventh triangulations used for the two-dimensional test case; all the triangulations are fitted to ϖ and B.*

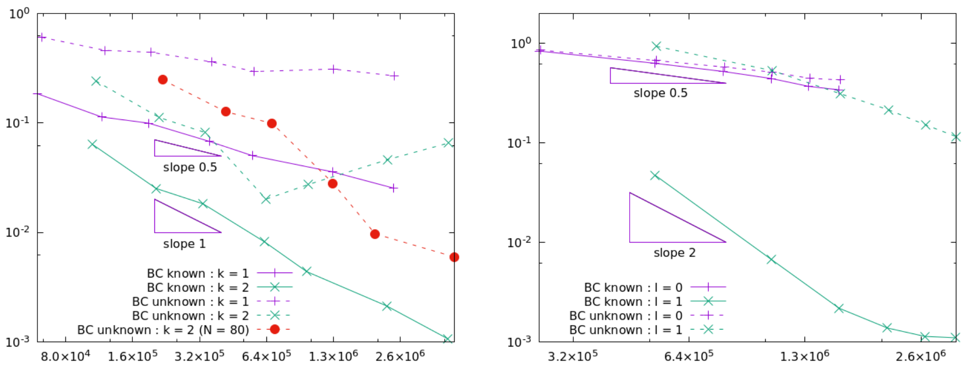

Fig. 4.4. *$L^2(T_1, T_2; H^1(B))$-seminorm errors for the two-dimensional test case without any noise. Left: space convergence ($N = 40$ except the curve with red bullets for which $N = 80$, $\ell = 1$). Right: time convergence (finest mesh, $k = 2$).*

the slopes in the left panel of Figure 4.4 have to be multiplied by two to get the convergence rate.) This convergence stops in the case of unknown boundary data when we reach the lowest error enabled by the time refinement (this error can be extrapolated using the curve corresponding to $\ell = 1$ on the right panel). When this lowest error is reached, continuing to refine the mesh increases the error. To confirm that the obstruction comes from the time discretization error, we also report in the left panel of Figure 4.4 the space convergence error obtained with a finer time discretization ($N = 80$). We also notice that, since the convergence rate for $k = 1$ is indeed one in the case of known boundary data, we conjecture that the superconvergence observed in the one-dimensional test case was due to the fact that the domain was one-dimensional. The convergence in time (right panel of Figure 4.4) seems to correspond to a convergence at rate $\ell + \frac{1}{2}$, even if the boundary data are not known. (Recall that for the one-dimensional case, we got $\ell + 1$ for $\ell \in \{1, 2\}$.) An interesting observation is that in the case of known boundary data, space refinement has to be finer since the error stops converging on the right panel for $\ell = 1$ (which indicates a lack of space refinement), whereas for unknown boundary data, convergence stops on the left panel for $k = 2$ (which indicates a lack of time refinement). For completeness, the number of dofs is reported in Table 4.2. We see that we reach several millions of dofs, which is high for the use of a direct solver like Pardiso. In fact, these numerical simulations use more than 100 GB of RAM.

Finally, we consider some noise at level $a \in \{0, 0.05, 0.1\}$. The results are reported in Figure 4.5 along with the solution without any noise for $k = 2$ and $\ell = 1$. In the

Table 4.2

Number of dofs (two-dimensional test case with known boundary value). Left: space refinement for $\ell = 1$ and $N = 40$. Right: time refinement for $h = 0.03125$ and $k = 2$.

h	$k=1$	$k=2$	N	$\ell=0$	$\ell=1$
$2.5{\cdot}10^{-1}$	$6.0{\cdot}10^{4}$	$1.1{\cdot}10^{5}$	5	$2.6{\cdot}10^{5}$	$5.2{\cdot}10^{5}$
$1.8{\cdot}10^{-1}$	$1.2{\cdot}10^{5}$	$2.0{\cdot}10^{5}$	10	$5.2{\cdot}10^{5}$	$1.0{\cdot}10^{6}$
$1.3{\cdot}10^{-1}$	$1.9{\cdot}10^{5}$	$3.3{\cdot}10^{5}$	15	$7.8{\cdot}10^{5}$	$1.6{\cdot}10^{6}$
$9.0{\cdot}10^{-2}$	$3.6{\cdot}10^{5}$	$6.2{\cdot}10^{5}$	20	$1.0{\cdot}10^{6}$	$2.1{\cdot}10^{6}$
$6.3{\cdot}10^{-2}$	$5.5{\cdot}10^{5}$	$9.7{\cdot}10^{5}$	25	$1.3{\cdot}10^{6}$	$2.6{\cdot}10^{6}$
$4.5{\cdot}10^{-2}$	$1.3{\cdot}10^{6}$	$2.2{\cdot}10^{6}$	30	$1.6{\cdot}10^{6}$	$3.1{\cdot}10^{6}$
$3.1{\cdot}10^{-2}$	$2.4{\cdot}10^{6}$	$4.2{\cdot}10^{6}$	.	.	.

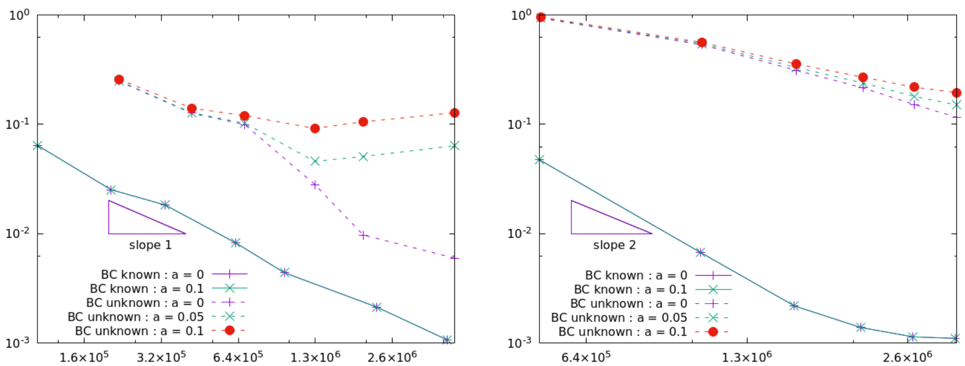

Fig. 4.5. *$L^2(T_1, T_2; H^1(B))$-seminorm errors for the two-dimensional test case with noised data ($a \in \{0, 0.05, 0.1\}$). Left: space convergence ($k = 2$ and $\ell = 1$). Right: time convergence (finest mesh from Figure* 4.3*, $k = 2$ and $\ell = 1$).*

left panel, $N = 40$ (resp., $N = 80$) is considered for known (resp., unknown) boundary data. We observe once more that the test case without the knowledge of boundary data is much more sensitive to the presence of noise. Indeed, the presence of noise changes the result only in this case.

REFERENCES

[1] G. Alessandrini, L. Rondi, E. Rosset, and S. Vessella, *The stability for the Cauchy problem for elliptic equations*, Inverse Problems, 25 (2009), 123004.

[2] 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 (2015), pp. 971–1002.

[3] L. Bourgeois, *A mixed formulation of quasi-reversibility to solve the Cauchy problem for Laplace's equation*, Inverse Problems, 21 (2005), pp. 1087–1104.

[4] L. Bourgeois, *Convergence rates for the quasi-reversibility method to solve the Cauchy problem for Laplace's equation*, Inverse Problems, 22 (2006), pp. 413–430.

[5] E. Burman, *Stabilized finite element methods for nonsymmetric, noncoercive, and ill-posed problems. Part I: Elliptic equations*, SIAM J. Sci. Comput., 35 (2013), pp. A2752–A2780.

[6] E. Burman, *Error estimates for stabilized finite element methods applied to ill-posed problems*, C. R. Math. Acad. Sci. Paris, 352 (2014), pp. 655–659.

[7] E. Burman and A. Ern, *Continuous interior penalty hp-finite element methods for advection and advection-diffusion equations*, Math. Comp., 76 (2007), pp. 1119–1140.

[8] E. Burman and L. Oksanen, *Data assimilation for the heat equation using stabilized finite element methods*, Numer. Math., 139 (2018), pp. 505–528.

[9] 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 (2018), 035004.

[10] 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 (2018), pp. 2065–2082.

[11] 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 (2021), pp. 2368–2392.

[12] 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 (2021), pp. S969–S991.

[13] E. Burman, G. Delay, A. Ern, and L. Oksanen, *A stability estimate for data assimilation subject to the heat equation with initial datum*, to appear in Comptes Rendus Math. (CRAS).

[14] 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., 344 (2018), pp. 852–874.

[15] W. Dahmen, R. Stevenson, and J. Westerdiep, *Accuracy controlled data assimilation for parabolic problems*, Math. Comp., 91 (2022), pp. 557–595.

[16] 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 (2013), pp. 2123–2148.

[17] O. Y. Èmanuilov, *Controllability of parabolic equations*, Mat. Sb., 186 (1995), pp. 109–132.

[18] A. Ern and J.-L. Guermond, *Finite Elements* I: *Approximation and Interpolation*, Texts in Appl. Math. 72, Springer, Cham, 2021.

[19] A. Ern and J.-L. Guermond, *Finite Elements* III*: First-order and time-dependent PDEs*, Texts in Appl. Math. 74, Springer, Cham, 2021.

[20] K. Ito and B. Jin, *Inverse Problems. Tikhonov Theory and Algorithms*, Ser. Appl. Math. 22, World Scientific, Hackensack, NJ, 2015.

[21] 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 (2015), pp. 567–588.

[22] M. V. Klibanov, *Estimates of initial conditions of parabolic equations and inequalities via lateral Cauchy data*, Inverse Problems, 22 (2006), pp. 495–514.

[23] J. Korevaar and J. L. H. Meyers, *Logarithmic convexity for supremum norms of harmonic functions*, Bull. London Math. Soc., 26 (1994), pp. 353–362.

[24] L. Robbiano, *Théorème d'unicité adapté au contrôle des solutions des problèmes hyperboliques*, Comm. Partial Differential Equations, 16 (1991), pp. 789–800.

[25] I. Smears, *Robust and efficient preconditioners for the discontinuous Galerkin time-stepping method*, IMA J. Numer. Anal., 37 (2017), pp. 1961–1985.

[26] M. Tadi, M. V. Klibanov, and W. Cai, *An inversion method for parabolic equations based on quasireversibility*, Comput. Math. Appl., 43 (2002), pp. 927–941.

[27] S. Vessella, *Unique continuation properties and quantitative estimates of unique continuation for parabolic equations*, in Handbook of Differential Equations: Evolutionary Equations, Handb. Differ. Equ. 5, Elsevier/North-Holland, Amsterdam, 2009, pp. 421–500.

[28] 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 (2010), pp. 655–671.