ESAIM: M2AN 59 (2025) 2837–2861
https://doi.org/10.1051/m2an/2025077

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

hp-ERROR ANALYSIS OF A CONTINUOUS PETROV–GALERKIN TIME-STEPPING SCHEME FOR LINEAR PARABOLIC EQUATIONS

Zhe Li and Lijun Yi*

Abstract. We present a fully discrete *hp*-version numerical method for second-order linear parabolic equations, combining a continuous Petrov–Galerkin (CPG) time-stepping scheme with a standard continuous Galerkin (CG) finite element method in space. Our analysis provides *a priori* error estimates in the $L^2(L^2)$- and $L^\infty(L^2)$-norms, where all constants are fully robust – independent of temporal and spatial discretization parameters. For solutions with initial singularities at $t = 0$, exponential temporal convergence is achieved through geometrically graded time meshes and linearly increasing polynomial degrees. Numerical experiments confirm the theoretical convergence rates.

Mathematics Subject Classification. 65M60, 65M12, 65M15.

Received June 28, 2025. Accepted September 9, 2025.

1. Introduction

We consider the second-order linear parabolic equations of the form

$$\begin{cases} \partial_t u + \mathcal{A}u = f(\boldsymbol{x}, t) & \text{in } \Omega \times (0, T], \\ u(\boldsymbol{x}, t) = 0 & \text{on } \partial\Omega \times (0, T], \\ u(\boldsymbol{x}, 0) = u_0(\boldsymbol{x}) & \text{in } \Omega, \end{cases} \tag{1.1}$$

where $\Omega \subset \mathbb{R}^d$ is an open, bounded domain with boundary $\partial\Omega$, and $d \in \{1, 2, 3\}$. Here, $f(\boldsymbol{x}, t)$ represents the source term, and $u_0(\boldsymbol{x})$ is the initial condition. The operator $\mathcal{A}$ is the elliptic spatial operator, defined as

$$\mathcal{A}u = -\sum_{i,j=1}^{d} \partial_{x_j}(a_{ij}(\boldsymbol{x})\partial_{x_i} u) + b(\boldsymbol{x})u,$$

where the coefficients a_{ij} and b are time-independent functions.

Over the past few decades, parabolic equations have attracted significant attention due to their broad range of physical applications. These equations are used to model various phenomena, including heat transfer in draining films [30], diffusion and catalytic population growth [17, 37], dispersion in tide-driven flows [13], contaminant

Keywords and phrases. Parabolic equations, *hp*-version continuous Petrov–Galerkin method, time-stepping scheme, exponential convergence, initial singularity.

Department of Mathematics, Shanghai Normal University, Shanghai 200234, P.R. China.
*Corresponding author: ylj5152@shnu.edu.cn

transport in shallow lakes [43], phase separation in binary alloys [4,39], long-range air pollutant transport [57], and myosin distribution in cells using a two-phase flow framework [33].

Numerous numerical methods have been developed for solving second-order parabolic equations. Classical approaches, such as finite difference schemes in both time and space, have been extensively studied. Additionally, many methods combine finite difference schemes for time discretization with finite element methods (FEMs) or spectral Galerkin/collocation methods for spatial discretization [10,16,20,22,23,34,40,41,56]. While these hybrid approaches often achieve high accuracy in the spatial direction, their temporal accuracy remains relatively low, leading to suboptimal overall approximation precision. Consequently, the development of methods that achieve high accuracy in both time and space is of considerable theoretical and practical significance.

In recent years, significant progress has been made in developing high accuracy space-time methods for solving parabolic equations, particularly using FEMs in both space and time. FEMs can be classified into the h-, p-, and hp-versions. The h-version achieves error reduction by refining the mesh size while keeping the polynomial degree fixed (typically low). In contrast, the p-version reduces the error by increasing the polynomial degree while maintaining a fixed mesh size. The hp-version combines both approaches, enabling simultaneous adjustments of local mesh sizes and polynomial degrees. This flexibility allows the hp-version FEM to adapt efficiently to solution characteristics, achieving exponential convergence rates even for singular solutions, as demonstrated in [8,48,55].

It is well known that Galerkin discretization methods are widely used in the spatial direction, while for time discretization, they are generally categorized into continuous Galerkin (CG) and discontinuous Galerkin (DG) methods. The CG method produces piecewise continuous polynomial solutions, whereas the DG method yields piecewise discontinuous polynomial approximations. Despite this difference, both methods employ test function spaces composed of piecewise polynomials that are discontinuous at time nodes. This structure allows a global scheme defined on $(0, T)$ to be decomposed into local problems on each time subinterval, enabling both methods to be viewed as implicit one-step schemes. The h-version CG time-stepping method for nonlinear first-order initial value problems (IVPs) dates back to the 1970s, pioneered by Hulme [25,26]. The h-version DG time-stepping method was later introduced by Delfour *et al.* [15]. Since 2000, hp-version CG and DG time-stepping methods have been extensively investigated for solving first- and second-order IVPs, as documented in [5,45,50,51,53,54].

In the context of parabolic equations, the h-version DG time-stepping method was first proposed by Jamet [31] and further analyzed by various researchers, including those in [2,3,18,36,42]. For a comprehensive overview of the DG time-stepping method, readers are referred to the monograph [49]. The h-version CG method was initially introduced by Aziz and Monk [6] for the time discretization of the heat equation and was subsequently generalized to broader classes of parabolic equations; see, *e.g.*, [1, 3, 9, 35] and the references therein. The hp-version Galerkin method, known for its flexibility and high accuracy, has also been applied to the time discretization of parabolic equations. One of the earliest attempts in this direction was made by Babuška and Janik [7], who analyzed the hp-version discontinuous Petrov–Galerkin method in time. However, their approach imposed severe restrictions on the spatial discretization, and exponential convergence was not established. More recently, studies such as [46,52] demonstrated that the hp-version DG time-stepping method can achieve exponential convergence for parabolic problems with initial singularities. For *a posteriori* error estimation of the hp-version DG time-stepping method applied to parabolic problems, we refer readers to [24, 38, 47] and references therein.

To the best of our knowledge, there has been no error analysis of the hp-version CG time-stepping method for parabolic equations. The CG method uses distinct spaces for trial and test functions, comprising piecewise continuous and discontinuous polynomials, respectively, making it essentially a continuous Petrov–Galerkin (CPG) scheme. We will adopt this terminology throughout. Since the trial functions in the CPG method are piecewise continuous polynomials with a degree one order higher than the test functions, the discrete solution cannot be used as the test function (as in DG methods) to establish stability. This presents significant challenges for stability and convergence analysis of the CPG time-stepping method for general linear and nonlinear parabolic equations. A feasible approach is to define a special mesh-dependent norm and perform the analysis in terms of this new norm, as shown in [1, 3, 19]. However, the error analysis for the h-version Galerkin method cannot be directly extended to the *hp*-version.

The present work employs the CPG method for the temporal discretization of parabolic equations, motivated by the following considerations:

- The trial space of the CPG method consists of piecewise continuous polynomials, which avoids introducing jump terms into the discrete formulation – commonly required in DG methods. This leads to a certain simplification in both practical implementation and error analysis.
- For regular solutions, both the CPG and DG methods yield an optimal convergence rate of order $q+1$ in the L^2-norm when using polynomials of degree q. However, the CPG method achieves this with fewer degrees of freedom per time step.
- The CPG method is established as A-stable and, as demonstrated in [44] within an abstract Hilbert space setting for general evolution equations, it exhibits an energy-decreasing property when applied to gradient flow equations. This makes it a promising candidate for optimization-related problems.

It is worth emphasizing that both the CPG and DG methods are high-order methods suitable for time discretization, and each has its own advantages. It is difficult to make a straightforward comparison between them, and indeed, such a comparison is not the aim of this paper. For detailed numerical comparisons of the CPG and DG time-stepping applied to IVPs for ordinary differential equations, heat equations, and transient Stokes problems, we refer to [27, 28, 32].

This paper aims to provide a rigorous error analysis of the *hp*-version CPG time discretization scheme combined with the *hp*-version CG spatial discretization for solving the parabolic equation (1.1). The analysis hinges on the stability of the fully discrete Galerkin approximation (see Lem. 2.3), which serves as the foundation for establishing optimal *a priori* error estimates in standard Sobolev norms. The main contributions of this paper are summarized as follows:

- We establish the stability of the *hp*-version fully discrete Galerkin solution. Based on this, we prove the existence and uniqueness of the fully discrete Galerkin approximation.
- We derive *a priori* error estimates in $L^2(0,T;L^2(\Omega))$- and $L^\infty(0,T;L^2(\Omega))$-norms, which are fully explicit in terms of mesh sizes and polynomial approximation degrees in both time and space. Notably, for smooth solutions, our analysis demonstrates spectral accuracy in both temporal and spatial directions.
- For solutions with initial singularities at $t=0$, we show that exponential convergence rates in time can be achieved by employing geometric time partitions combined with linearly increasing polynomial approximation degrees.

The paper is organized as follows. In Section 2, we introduce the space-time *hp*-version Galerkin fully discrete scheme for solving the parabolic equation (1.1), along with proofs of the stability and well-posedness of the fully discrete solutions. In Section 3, we derive error estimates for the fully discrete approximation in the $L^2(0,T;L^2(\Omega))$- and $L^\infty(0,T;L^2(\Omega))$-norms. Specifically, for solutions exhibiting initial singularities at $t=0$, we prove that the fully discrete Galerkin approximation errors exhibit exponential decay in the temporal direction under a geometric time mesh refinement. Section 4 presents numerical experiments that validate the theoretical results. Finally, we conclude the paper with a summary and discuss directions for future research.

2. The *hp*-version CPG-in-time and CG-in-space fully discrete scheme

In Section 2.1, we present the commonly used Sobolev spaces and their associated norm notations. In Section 2.2, we introduce the *hp*-version fully discrete Galerkin scheme for solving the second-order parabolic equation (1.1). Finally, in Section 2.3, we establish the stability of the fully discrete Galerkin solution.

2.1. Notations

Throughout this paper, we adopt standard notation and conventions. Let Ω be a bounded open domain. We denote by $L^2(\Omega)$ the space of square-integrable functions on Ω, equipped with the norm $\|\cdot\|_{L^2(\Omega)}$ and the inner product $(\cdot,\cdot)$. The Sobolev spaces $H^m(\Omega)$ and $H_0^m(\Omega)$ are defined in the standard way, where m is a non-negative integer. These spaces are equipped with the standard norm $\|\cdot\|_{H^m(\Omega)}$ and semi-norm $|\cdot|_{H^m(\Omega)}$. In particular, we have $H^0(\Omega) = L^2(\Omega)$.

For a Banach space B with norm $\|\cdot\|_B$ defined over the spatial domain Ω, and for a time interval $(0,T)$, we define the space-time Sobolev spaces $H^r(0,T;B)$ as follows:

$$H^r(0,T;B) = \big\{u(\boldsymbol{x},t) : (0,T) \to B \text{ and } \|u\|_{H^r(0,T;B)} < \infty\big\},$$

where $r \geq 0$ is an integer, and

$$\|u\|_{H^r(0,T;B)} := \left(\sum_{j=0}^{r} \int_0^T \left\|\partial_t^j u\right\|_B^2 \mathrm{d}t\right)^{\frac{1}{2}}.$$

Similarly, the space-time Sobolev spaces $W^{r,\infty}(0,T;B)$ are defined as

$$W^{r,\infty}(0,T;B) = \big\{u(\boldsymbol{x},t) : (0,T) \to B \text{ and } \|u\|_{W^{r,\infty}(0,T;B)} < \infty\big\},$$

where

$$\|u\|_{W^{r,\infty}(0,T;B)} := \sum_{j=0}^{r} \left(\operatorname*{ess\,sup}_{0\leq t\leq T} \left\|\partial_t^j u\right\|_B\right).$$

For convenience, we typically use $L^2(0,T;B)$ and $L^\infty(0,T;B)$ to represent $H^0(0,T;B)$ and $W^{0,\infty}(0,T;B)$, respectively.

Additionally, we denote by C generic positive constants that are not necessarily identical at different locations, but are always independent of discretization parameters such as the mesh sizes and the approximation degrees.

2.2. Formulation of the *hp*-vesion fully discrete Galerkin scheme

2.2.1. Weak solution

Assume that the coefficients a_{ij} and b of the elliptic spatial operator $\mathcal{A}$ satisfying $a_{ij}, b \in L^\infty(\Omega)$, and $a_{ij} = a_{ji}$ for $1 \leq i,j \leq d$. The bilinear form associated with $\mathcal{A}$ is given by

$$a(u,v) = \int_\Omega \left(\sum_{i,j=1}^{d} a_{ij}(\boldsymbol{x})\partial_{x_i} u \partial_{x_j} v + buv\right) \mathrm{d}\boldsymbol{x}, \tag{2.1}$$

and assume that this form is symmetric, continuous, and coercive. Specifically, for any $u,v \in H_0^1(\Omega)$, there exist positive constants C_1 and C_2, depending only on a_{ij} and b, such that

$$a(u,v) = a(v,u), \tag{2.2a}$$
$$a(u,u) \geq C_1\|u\|_{H^1(\Omega)}^2, \tag{2.2b}$$
$$|a(u,v)| \leq C_2\|u\|_{H^1(\Omega)}\|v\|_{H^1(\Omega)}. \tag{2.2c}$$

Following [21], we define a weak solution of the parabolic equation (1.1) as a function $u \in L^2(0,T;H_0^1(\Omega))$ with $\partial_t u \in L^2(0,T;H^{-1}(\Omega))$ that satisfies the following conditions:

(i) $\langle \partial_t u, v\rangle + a(u,v) = (f,v)$ for each $v \in H_0^1(\Omega)$, almost everywhere in $t \in [0,T]$,
(ii) $u(\boldsymbol{x},0) = u_0(\boldsymbol{x})$.

Here, $\langle\cdot,\cdot\rangle$ denotes the pairing between $H^{-1}(\Omega)$ and $H_0^1(\Omega)$.

As stated in Theorem 5 of Chapter 7 in [21], the problem (1.1) has a unique weak solution u that satisfies

$$u \in L^2\big(0,T;H^2(\Omega)\big) \cap L^\infty\big(0,T;H_0^1(\Omega)\big), \quad \partial_t u \in L^2\big(0,T;L^2(\Omega)\big),$$

provided that $f \in L^2(0,T;L^2(\Omega))$ and $u_0 \in H_0^1(\Omega)$.

2.2.2. Spatial discretization

For the conforming finite element discretization of (1.1), let $\{\mathcal{M}_h\}$ represent a family of shape-regular partitions of Ω into elements $\{\Omega_i\}_{i=1}^{N_{\boldsymbol{x}}}$, where each Ω_i is a d-simplex, quadrilateral, or hexahedron. The diameter of each element Ω_i is denoted by h_i, and the maximum mesh size is defined as $h := \max\{h_i\}_{i=1}^{N_{\boldsymbol{x}}}$. Each element Ω_i is assigned an approximation degree $p_i \geq 1$, which are collectively represented by the vector $\boldsymbol{p} = \{p_i\}_{i=1}^{N_{\boldsymbol{x}}}$. For simplicity, we assume that $(\mathcal{M}_h, \boldsymbol{p})$ forms a quasi-uniform hp-discretization of Ω. That is, there exist positive constants C_h and C_p such that

$$1 \leq \frac{h}{h_i} \leq C_h \quad \text{and} \quad 1 \leq \frac{\bar{p}}{p_i} \leq C_p, \quad \text{for } 1 \leq i \leq N_{\boldsymbol{x}},$$

where $\bar{p} := \max\{p_i\}_{i=1}^{N_{\boldsymbol{x}}}$. In addition, we set $p := \min\{p_i\}_{i=1}^{N_{\boldsymbol{x}}}$.

As usual, the hp-version finite element space, consisting of piecewise continuous pull-back polynomials, is defined as

$$V_0^{\boldsymbol{p}}(\mathcal{M}_h) = V^{\boldsymbol{p}}(\mathcal{M}_h) \cap H_0^1(\Omega),$$

where

$$V^{\boldsymbol{p}}(\mathcal{M}_h) = \big\{u \in H^1(\Omega) : u|_{\Omega_i} \in P_{p_i}(\Omega_i), 1 \leq i \leq N_{\boldsymbol{x}}\big\}.$$

Here, $P_{p_i}(\Omega_i)$ denotes the space of polynomials on Ω_i where the total degree (sum of degrees in all variables) does not exceed p_i if Ω_i is a simplex. Otherwise, it represents the space of polynomials on Ω_i where the separate degree does not exceed p_i in each variable.

2.2.3. Temporal discretization

To introduce the hp-version CPG time-stepping scheme for (1.1), consider an arbitrary partition $\mathcal{J}_\tau$ of the time interval $I := (0, T)$, with time nodes defined as

$$0 = t_0 < t_1 < \cdots < t_{N_t - 1} < t_{N_t} = T.$$

Let $I_n := (t_{n-1}, t_n)$ represent each time subinterval with step size $\tau_n = t_n - t_{n-1}$, and denote the maximum step size by $\tau := \max\{\tau_n\}_{n=1}^{N_t}$.

The approximation degree vector $\boldsymbol{q} = \{q_n\}_{n=1}^{N_t}$ assigns an approximation degree $q_n \geq 1$ to each time interval I_n. We further define the minimum approximation degree $q := \min\{q_n\}_{n=1}^{N_t}$. The discretization $(\mathcal{J}_\tau, \boldsymbol{q})$ is called a quasi-uniform hp-discretization of $(0, T)$ if there exist positive constants C_τ and C_q such that for all $1 \leq n \leq N_t$,

$$1 \leq \frac{\tau}{\tau_n} \leq C_\tau \quad \text{and} \quad 1 \leq \frac{\bar{q}}{q_n} \leq C_q, \quad \text{for } 1 \leq n \leq N_t,$$

where $\bar{q} := \max\{q_n\}_{n=1}^{N_t}$.

When addressing problems with initial singularities at $t = 0$, a specially local refined time mesh and a corresponding tailored approximation degree vector will be employed (see Sect. 3.4).

The trial and test spaces for the hp-version fully discrete Galerkin method are defined as

$$S^{\boldsymbol{p},\boldsymbol{q},1}(\mathcal{M}_h, \mathcal{J}_\tau) = \big\{u \in H^1\big(I; H_0^1(\Omega)\big) : u|_{I_n} \in P_{q_n}(I_n; V_0^{\boldsymbol{p}}(\mathcal{M}_h)), 1 \leq n \leq N_t\big\}$$

and

$$S^{\boldsymbol{p},\boldsymbol{q}-1,0}(\mathcal{M}_h, \mathcal{J}_\tau) = \big\{u \in L^2\big(I; H_0^1(\Omega)\big) : u|_{I_n} \in P_{q_n-1}(I_n; V_0^{\boldsymbol{p}}(\mathcal{M}_h)), 1 \leq n \leq N_t\big\},$$

respectively. Here, $P_{q_n}(I_n; V_0^{\boldsymbol{P}}(\mathcal{M}_h))$, for an integer $q_n \geq 1$, represents the space of $V_0^{\boldsymbol{p}}(\mathcal{M}_h)$-valued polynomials of degree at most q_n over the time interval I_n. The space $P_{q_n-1}(I_n; V_0^{\boldsymbol{p}}(\mathcal{M}_h))$ is defined analogously for polynomials of degree at most $q_n - 1$.

2.2.4. Fully discrete Galerkin Scheme

The *hp*-version fully discrete CPG-in-time and CG-in-space scheme for solving the parabolic equation (1.1) is formulated as follows: find $u_{h\tau} \in S^{\boldsymbol{p},\boldsymbol{q},1}(\mathcal{M}_h, \mathcal{J}_\tau)$ such that

$$\begin{cases} \displaystyle\int_I \big\{(\partial_t u_{h\tau}, v_{h\tau}) + a(u_{h\tau}, v_{h\tau})\big\}\,\mathrm{d}t = \int_I (f, v_{h\tau})\,\mathrm{d}t, \\ u_{h\tau}(\boldsymbol{x}, 0) = R^{\boldsymbol{p}} u_0 \end{cases} \tag{2.3}$$

for any $v_{h\tau} \in S^{\boldsymbol{p},\boldsymbol{q}-1,0}(\mathcal{M}_h, \mathcal{J}_\tau)$. Here, $R^{\boldsymbol{p}} : H_0^1(\Omega) \to V_0^{\boldsymbol{p}}(\mathcal{M}_h)$ is the Ritz–Galerkin projection operator defined by

$$a(u - R^{\boldsymbol{p}} u, v) = 0, \quad \forall v \in V_0^{\boldsymbol{p}}(\mathcal{M}_h). \tag{2.4}$$

Selecting $v = R^{\boldsymbol{p}} u$ in (2.4), and applying the coercivity and continuity of $a(\cdot,\cdot)$ as given in (2.2b) and (2.2c), we obtain the H^1-stability of $R^{\boldsymbol{p}}$:

$$\|R^{\boldsymbol{p}} u\|_{H^1(\Omega)} \le C\|u\|_{H^1(\Omega)}. \tag{2.5}$$

Due to the discontinuity of the test space $S^{\boldsymbol{p},\boldsymbol{q}-1,0}(\mathcal{M}_h, \mathcal{J}_\tau)$ at time nodes, the scheme (2.3) can be interpreted as a time-stepping method. Specifically, if $u_{h\tau}$ is known on I_m for $1 \le m \le n-1$, the solution $u_{h\tau} \in P_{q_n}(I_n; V_0^{\boldsymbol{p}}(\mathcal{M}_h))$ on the next time step I_n can be determined by solving

$$\begin{cases} \displaystyle\int_{I_n} \{(\partial_t u_{h\tau}, v_{h\tau}) + a(u_{h\tau}, v_{h\tau})\}\,\mathrm{d}t = \int_{I_n} (f, v_{h\tau})\,\mathrm{d}t, \\ u_{h\tau}|_{I_n}(\boldsymbol{x}, t_{n-1}) = u_{h\tau}|_{I_{n-1}}(\boldsymbol{x}, t_{n-1}) \end{cases} \tag{2.6}$$

for any $v_{h\tau} \in P_{q_n-1}(I_n; V_0^{\boldsymbol{p}}(\mathcal{M}_h))$. The initial condition is given by $u_{h\tau}|_{I_1}(\boldsymbol{x}, t_0) = R^{\boldsymbol{p}} u_0$.

Remark 2.1. The direct implementation of scheme (2.6) results in a large-scale linear system, particularly in the case of high-dimensional spatial problems. Recently, [14] proposed efficient parallel algorithms for the high-order CPG method in dissipative and wave propagation problems. Notably, their time-parallel algorithms can also be effectively applied to solve (2.6).

2.3. Stability

By applying the well-known standard Poincaré–Friedrichs inequality (see *e.g.*, [11]), we obtain the following result.

Lemma 2.2. *Let $u \in H^1(I; L^2(\Omega))$ with $u(\boldsymbol{x}, 0) = 0$. Then, the following inequality holds*

$$\|u\|_{L^2(I;L^2(\Omega))} \le T\|\partial_t u\|_{L^2(I;L^2(\Omega))}. \tag{2.7}$$

The following lemma establishes the stability of the *hp*-version fully discrete Galerkin approximation $u_{h\tau}$.

Lemma 2.3. *Suppose the right-hand side $f \in L^2(I; L^2(\Omega))$ and the initial data $u_0 \in H_0^1(\Omega)$. Then, there exists a unique Galerkin approximation $u_{h\tau}$ defined by (2.3), satisfying the following inequality:*

$$\|u_{h\tau}\|_{L^2(I;L^2(\Omega))} + \|\partial_t u_{h\tau}\|_{L^2(I;L^2(\Omega))} \le C\|f\|_{L^2(I;L^2(\Omega))} + C\|u_0\|_{H^1(\Omega)}, \tag{2.8}$$

where the constant C depends on T but is independent of any discretization parameters.

Proof. We first establish the stability estimate (2.8). Choosing $v_{h\tau} = \partial_t u_{h\tau}$ in (2.3) gives

$$\int_I \big\{(\partial_t u_{h\tau}, \partial_t u_{h\tau}) + a(u_{h\tau}, \partial_t u_{h\tau})\big\}\,\mathrm{d}t = \int_I (f, \partial_t u_{h\tau})\,\mathrm{d}t. \tag{2.9}$$

Using the symmetry and bilinearity of $a(\cdot,\cdot)$, along with integration by parts in t, we obtain

$$\begin{aligned}\int_I a(u_{h\tau},\partial_t u_{h\tau})\,\mathrm{d}t &= \frac{1}{2}\int_I \big\{a(\partial_t u_{h\tau},u_{h\tau})+a(u_{h\tau},\partial_t u_{h\tau})\big\}\,\mathrm{d}t\\ &= \frac{1}{2}\int_I \partial_t a(u_{h\tau},u_{h\tau})\,\mathrm{d}t\\ &= \frac{1}{2}\big\{a(u_{h\tau}(\boldsymbol{x},T),u_{h\tau}(\boldsymbol{x},T))-a(u_{h\tau}(\boldsymbol{x},0),u_{h\tau}(\boldsymbol{x},0))\big\}.\end{aligned} \tag{2.10}$$

Substituting (2.10) into (2.9) and applying the Cauchy–Schwarz inequality, (2.2b), and (2.2c), we deduce

$$\begin{aligned}\|\partial_t u_{h\tau}\|^2_{L^2(I;L^2(\Omega))} &\le \int_I \|f\|_{L^2(\Omega)}\|\partial_t u_{h\tau}\|_{L^2(\Omega)}\,\mathrm{d}t + C\|u_{h\tau}(\cdot,0)\|^2_{H^1(\Omega)}\\ &\le \|f\|_{L^2(I;L^2(\Omega))}\|\partial_t u_{h\tau}\|_{L^2(I;L^2(\Omega))} + C\|u_{h\tau}(\cdot,0)\|^2_{H^1(\Omega)}\\ &\le \frac{1}{2}\|f\|^2_{L^2(I;L^2(\Omega))} + \frac{1}{2}\|\partial_t u_{h\tau}\|^2_{L^2(I;L^2(\Omega))} + C\|u_{h\tau}(\cdot,0)\|^2_{H^1(\Omega)}.\end{aligned}$$

Rearranging gives

$$\|\partial_t u_{h\tau}\|_{L^2(I;L^2(\Omega))} \le \|f\|_{L^2(I;L^2(\Omega))} + C\|u_{h\tau}(\cdot,0)\|_{H^1(\Omega)}. \tag{2.11}$$

Next, using the triangle inequality and the Poincaré–Friedrichs inequality (2.7), we have

$$\begin{aligned}\|u_{h\tau}\|_{L^2(I;L^2(\Omega))} &\le \|u_{h\tau}-u_{h\tau}(\cdot,0)\|_{L^2(I;L^2(\Omega))} + \|u_{h\tau}(\cdot,0)\|_{L^2(I;L^2(\Omega))}\\ &\le T\|\partial_t(u_{h\tau}-u_{h\tau}(\cdot,0))\|_{L^2(I;L^2(\Omega))} + T^{\frac{1}{2}}\|u_{h\tau}(\cdot,0)\|_{L^2(\Omega)}\\ &= T\|\partial_t u_{h\tau}\|_{L^2(I;L^2(\Omega))} + T^{\frac{1}{2}}\|u_{h\tau}(\cdot,0)\|_{L^2(\Omega)}\\ &\le C\|f\|_{L^2(I;L^2(\Omega))} + C\|u_{h\tau}(\cdot,0)\|_{H^1(\Omega)}.\end{aligned} \tag{2.12}$$

Combining (2.11) and (2.12), we obtain

$$\|u_{h\tau}\|_{L^2(I;L^2(\Omega))} + \|\partial_t u_{h\tau}\|_{L^2(I;L^2(\Omega))} \le C\|f\|_{L^2(I;L^2(\Omega))} + C\|u_{h\tau}(\cdot,0)\|_{H^1(\Omega)}. \tag{2.13}$$

Since $u_{h\tau}(\boldsymbol{x},0)=R^{\boldsymbol{p}}u_0$ by (2.3) and $R^{\boldsymbol{p}}$ is H^1-stable by (2.5), we conclude

$$\|u_{h\tau}(\cdot,0)\|_{H^1(\Omega)} = \|R^{\boldsymbol{p}}u_0\|_{H^1(\Omega)} \le C\|u_0\|_{H^1(\Omega)}.$$

Substituting this into (2.13) gives (2.8).

We now prove the existence and uniqueness of $u_{h\tau}$. Let U, V be two hp-version fully discrete Galerkin approximations satisfying (2.3). Then $W := U - V$ satisfies

$$\begin{cases}\displaystyle\int_I \big\{(\partial_t W, v_{h\tau}) + a(W, v_{h\tau})\big\}\,\mathrm{d}t = 0,\\ W(\boldsymbol{x},0) = 0\end{cases}$$

for any $v_{h\tau}\in S^{\boldsymbol{p},\boldsymbol{q}-1,0}(\mathcal{M}_h,\mathcal{J}_\tau)$. By the stability estimate (2.8), we have

$$\|W\|_{L^2(I;L^2(\Omega))} + \|\partial_t W\|_{L^2(I;L^2(\Omega))} \le 0,$$

implying that $W=0$ on $\Omega\times I$. Thus, $u_{h\tau}$ is unique. Since (2.3) is linear and finite-dimensional, existence follows from the uniqueness result. This completes the proof. $\square$

3. Error analysis

In this section, we derive *a priori* error estimates for the hp-version fully discrete Galerkin approximation defined by (2.3).

3.1. Ritz–Galerkin projection, temporal projection, and their approximation properties

We recall that the Ritz–Galerkin projection operator $R^{\boldsymbol{p}}$ defined in (2.4) applies only to functions with respect to the spatial variable $\boldsymbol{x}$. In the following, we extend $R^{\boldsymbol{p}}$ to functions depending on both $\boldsymbol{x}$ and t in the L^2 sense.

Definition 3.1 (Ritz–Galerkin projection)**.** For any function $u \in L^2(I_n; H_0^1(\Omega))$, we define the extended Ritz–Galerkin projection $R_n^{\boldsymbol{p}} u \in L^2(I_n; V_0^{\boldsymbol{p}}(\mathcal{M}_h))$, as follows

$$\int_{I_n} a(u - R_n^{\boldsymbol{p}} u, v)\,\mathrm{d}t = 0, \quad \forall v \in L^2(I_n; V_0^{\boldsymbol{p}}(\mathcal{M}_h)). \tag{3.1}$$

For any $u \in L^2(I; H_0^1(\Omega))$, we further define the global Ritz–Galerkin projection element-wise by

$$R^{\boldsymbol{p}} u|_{I_n} = R_n^{\boldsymbol{p}} u, \quad 1 \le n \le N_t.$$

Clearly, $R^{\boldsymbol{p}} u \in L^2(I; V_0^{\boldsymbol{p}}(\mathcal{M}_h))$ satisfies

$$\int_{I} a(u - R^{\boldsymbol{p}} u, v)\,\mathrm{d}t = 0, \quad \forall v \in L^2(I; V_0^{\boldsymbol{p}}(\mathcal{M}_h)). \tag{3.2}$$

By choosing $v = R_n^{\boldsymbol{p}} u$ in (3.1) and applying the coercivity and continuity of the bilinear functional $a(\cdot,\cdot)$, we obtain

$$\|R_n^{\boldsymbol{p}} u\|_{L^2(I_n;H^1(\Omega))} \le C\|u\|_{L^2(I_n;H^1(\Omega))}. \tag{3.3}$$

Summing over all elements I_n for $1 \le n \le N_t$, we get

$$\|R^{\boldsymbol{p}} u\|_{L^2(I;H^1(\Omega))} \le C\|u\|_{L^2(I;H^1(\Omega))}. \tag{3.4}$$

Thus, we conclude that the projection operators $R_n^{\boldsymbol{p}}$ and $R^{\boldsymbol{p}}$ are H^1-stable.

Building on the error estimates for the hp-version CG method applied to second-order elliptic boundary value problems, as discussed in [8,48], we conclude that the Ritz–Galerkin projection $R^{\boldsymbol{p}}$ satisfies the following hp-version approximation property.

Lemma 3.2. *Let the tuple $(\mathcal{M}_h, \boldsymbol{p})$ be a quasi-uniform hp-discretization of the spatial domain Ω. Assume that $u \in L^2(I; H_0^1(\Omega) \cap H^r(\Omega))$ with $r \ge 1$. Then, we have*

$$\|u - R^{\boldsymbol{p}} u\|_{L^2(I;L^2(\Omega))} \le C\frac{h^{\min\{p+1,r\}}}{p^r}\|u\|_{L^2(I;H^r(\Omega))}.$$

Next, we introduce the temporal projection operator $\Pi^{\boldsymbol{q}}$ following the frameworks in [48,49,55].

Definition 3.3 (Temporal projection)**.** For any $u \in H^1(I_n; L^2(\Omega))$, the projection $\Pi_{I_n}^{q_n} u \in P_{q_n}(I_n; L^2(\Omega))$, with integer $q_n \ge 1$, is defined as

$$\begin{cases} \displaystyle\int_{I_n} \big(\partial_t(u - \Pi_{I_n}^{q_n} u), v\big)\,\mathrm{d}t = 0, \quad \forall v \in P_{q_n-1}(I_n; L^2(\Omega)), \\ \Pi_{I_n}^{q_n} u(\boldsymbol{x}, t_{n-1}) = u(\boldsymbol{x}, t_{n-1}). \end{cases} \tag{3.5}$$

Here, $P_{q_n}(I_n; L^2(\Omega))$ represents the space of $L^2(\Omega)$-valued polynomials of degree at most q_n over the interval I_n, and $P_{q_n-1}(I_n; L^2(\Omega))$ is defined analogously.

Notably, the following condition also holds

$$\Pi_{I_n}^{q_n} u(\boldsymbol{x}, t_n) = u(\boldsymbol{x}, t_n), \tag{3.6}$$

which can be derived by setting $v = 1$ in (3.5) and applying integration by parts with respect to t.

Furthermore, for any function $u \in H^1(I; L^2(\Omega))$, we define the global projection element-wise as

$$\Pi^{\boldsymbol{q}} u|_{I_n} = \Pi_{I_n}^{q_n} u, \quad 1 \le n \le N_t. \tag{3.7}$$

Using the approximation results provided in Lemma 3.2 of [55] (see also [48]), we can derive the approximation results of $\Pi^{\boldsymbol{q}}u$, as detailed below.

Lemma 3.4. *Let the tuple $(\mathcal{J}_\tau, \boldsymbol{q})$ be an arbitrary hp-discretization of the time interval I. Suppose that $u \in H^1(I; L^2(\Omega))$ satisfies $u|_{\Omega\times I_n} \in H^{k_{0,n}}(I_n; L^2(\Omega))$ with $k_{0,n} \geq 1$. Then, we have the following error estimates*

$$\|u - \Pi^{\boldsymbol{q}}u\|^2_{L^2(I_n;L^2(\Omega))} \leq \left(\frac{\tau_n}{2}\right)^{2k_n} \frac{\Gamma(q_n + 2 - k_n)}{q_n(q_n+1)\Gamma(q_n + k_n)} \|u\|^2_{H^{k_n}(I_n;L^2(\Omega))}, \tag{3.8}$$

$$\|\partial_t(u - \Pi^{\boldsymbol{q}}u)\|^2_{L^2(I_n;L^2(\Omega))} \leq \left(\frac{\tau_n}{2}\right)^{2k_n-2} \frac{\Gamma(q_n + 2 - k_n)}{\Gamma(q_n + k_n)} \|u\|^2_{H^{k_n}(I_n;L^2(\Omega))} \tag{3.9}$$

for $1 \leq n \leq N_t$. Here, $1 \leq k_n \leq \min\{q_n + 1, k_{0,n}\}$.

If the tuple $(\mathcal{J}_\tau, \boldsymbol{q})$ is quasi-uniform and $u \in H^k(I; L^2(\Omega))$ with $k \geq 1$, then we have the following global error estimates

$$\|u - \Pi^{\boldsymbol{q}}u\|_{L^2(I;L^2(\Omega))} \leq C\frac{\tau^{\min\{q+1,k\}}}{q^k} \|u\|_{H^k(I;L^2(\Omega))},$$

$$\|\partial_t(u - \Pi^{\boldsymbol{q}}u)\|^2_{L^2(I;L^2(\Omega))} \leq C\frac{\tau^{\min\{q,k-1\}}}{q^{k-1}} \|u\|_{H^k(I;L^2(\Omega))}.$$

3.2. Abstract error bounds

Let u be the exact solution of the parabolic equation (1.1), and let $u_{h\tau}$ be the *hp*-version fully discrete Galerkin approximation defined by (2.3). We decompose the error $e := u - u_{h\tau}$ into three components as follows:

$$e = (u - R^{\boldsymbol{p}}u) + (R^{\boldsymbol{p}}u - \Pi^{\boldsymbol{q}}R^{\boldsymbol{p}}u) + (\Pi^{\boldsymbol{q}}R^{\boldsymbol{p}}u - u_{h\tau}) := \eta + R^{\boldsymbol{p}}\theta + \psi. \tag{3.10}$$

The terms are defined as follows:

- *Ritz–Galerkin projection error*: $\eta := u - R^{\boldsymbol{p}}u$, which represents the error due to the projection by the Ritz–Galerkin projection $R^{\boldsymbol{p}}$ in the spatial direction.
- *Temporal projection error*: $R^{\boldsymbol{p}}\theta := R^{\boldsymbol{p}}(u - \Pi^{\boldsymbol{q}}u)$, where $\theta := u - \Pi^{\boldsymbol{q}}u$ is the error introduced by the temporal projection operator $\Pi^{\boldsymbol{q}}$. Here, we utilize the commutativity property of the projection operators: $\Pi^{\boldsymbol{q}}R^{\boldsymbol{p}}u = R^{\boldsymbol{p}}\Pi^{\boldsymbol{q}}u$.
- *Galerkin approximation error*: $\psi := \Pi^{\boldsymbol{q}}R^{\boldsymbol{p}}u - u_{h\tau}$, which represents the remaining error due to the space-time Galerkin approximation.

We begin by deriving the $L^2(H^2)$-estimate for the term $\theta = u - \Pi^{\boldsymbol{q}}u$.

Lemma 3.5. *Let the tuple $(\mathcal{J}_\tau, \boldsymbol{q})$ be an arbitrary hp-discretization of the time interval I. Suppose that $u \in H^1(I; H^2(\Omega))$ satisfies $u|_{\Omega\times I_n} \in H^{k_{0,n}}(I_n; H^2(\Omega))$ with $k_{0,n} \geq 1$. Then, we have the following error estimate*

$$\|\theta\|^2_{L^2(I_n;H^2(\Omega))} \leq \left(\frac{\tau_n}{2}\right)^{2k_n} \frac{\Gamma(q_n + 2 - k_n)}{q_n(q_n+1)\Gamma(q_n + k_n)} \|u\|^2_{H^{k_n}(I_n;H^2(\Omega))} \tag{3.11}$$

for $1 \leq n \leq N_t$. Here, $1 \leq k_n \leq \min\{q_n + 1, k_{0,n}\}$.

If the tuple $(\mathcal{J}_\tau, \boldsymbol{q})$ is quasi-uniform and $u \in H^k(I; H^2(\Omega))$ with $k \geq 1$, then we have the following global error estimate

$$\|\theta\|_{L^2(I;H^2(\Omega))} \leq C\frac{\tau^{\min\{q+1,k\}}}{q^k} \|u\|_{H^k(I;H^2(\Omega))}. \tag{3.12}$$

Proof. Using (3.8) from Lemma 3.4, we obtain

$$\|\theta\|^2_{L^2(I_n;H^2(\Omega))} = \|u - \Pi^{\boldsymbol{q}} u\|^2_{L^2(I_n;H^2(\Omega))} = \sum_{|\alpha|=0}^{2} \|D^\alpha u - \Pi^{\boldsymbol{q}}(D^\alpha u)\|^2_{L^2(I_n;L^2(\Omega))}$$
$$\leq \sum_{|\alpha|=0}^{2} \left(\frac{\tau_n}{2}\right)^{2k_n} \frac{\Gamma(q_n+2-k_n)}{q_n(q_n+1)\Gamma(q_n+k_n)} \|D^\alpha u\|^2_{H^{k_n}(I_n;L^2(\Omega))}$$
$$\leq \left(\frac{\tau_n}{2}\right)^{2k_n} \frac{\Gamma(q_n+2-k_n)}{q_n(q_n+1)\Gamma(q_n+k_n)} \|u\|^2_{H^{k_n}(I_n;H^2(\Omega))}.$$

Here, $\alpha = (\alpha_1, \dots, \alpha_d)$ is a multi-index of nonnegative integers α_j, and $D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}$ denotes the differential operator of order $|\alpha| = \sum_{j=1}^{d} \alpha_j$. This completes the proof of the estimate (3.11).

Furthermore, by applying Stirling's formula, we obtain the estimate (3.12). ☐

Next, we establish the $H^1(L^2)$-estimate for the term $\eta = u - R^{\boldsymbol{p}} u$.

Lemma 3.6. *Let the tuple $(\mathcal{M}_h, \boldsymbol{p})$ be a quasi-uniform hp-discretization of the spatial domain Ω. Assume that $u \in H^1(I; H_0^1(\Omega) \cap H^r(\Omega))$ with $r \geq 1$. Then, the following estimate holds*

$$\|\eta\|_{H^1(I;L^2(\Omega))} \leq C \frac{h^{\min\{p+1,r\}}}{p^r} \|u\|_{H^1(I;H^r(\Omega))}. \tag{3.13}$$

Proof. By applying Lemma 3.2, we have

$$\|\eta\|^2_{H^1(I;L^2(\Omega))} = \sum_{j=0}^{1} \left\| \partial_t^j u - R^{\boldsymbol{p}}\left(\partial_t^j u\right) \right\|^2_{L^2(I;L^2(\Omega))} \leq C \sum_{j=0}^{1} \frac{h^{2\min\{p+1,r\}}}{p^{2r}} \|\partial_t^j u\|^2_{L^2(I;H^r(\Omega))}$$
$$\leq C \frac{h^{2\min\{p+1,r\}}}{p^{2r}} \|u\|^2_{H^1(I;H^r(\Omega))}.$$

This completes the proof. ☐

Then, we consider the $L^\infty(L^2)$-estimate for the term $R^{\boldsymbol{p}}\theta = R^{\boldsymbol{p}}(u - \Pi^{\boldsymbol{q}} u)$.

Lemma 3.7. *Let the tuple $(\mathcal{J}_\tau, \boldsymbol{q})$ represent an arbitrary hp-discretization of the time interval I. Suppose that $u \in H^1(I; H_0^1(\Omega))$ satisfies $u|_{\Omega \times I_n} \in W^{k_{0,n},\infty}(I_n; H^1(\Omega))$ with $k_{0,n} \geq 1$. Then, we have the following error estimate*

$$\|R^{\boldsymbol{p}}\theta\|^2_{L^\infty(I_n;L^2(\Omega))} \leq C\left(\frac{\tau_n}{2}\right)^{2k_n} \frac{\Gamma(q_n+2-k_n)}{q_n\Gamma(q_n+k_n)} \|u\|^2_{W^{k_n,\infty}(I_n;H^1(\Omega))} \tag{3.14}$$

for $1 \leq n \leq N_t$. Here, $1 \leq k_n \leq \min\{q_n+1, k_{0,n}\}$.

In particular, if the tuple $(\mathcal{J}_\tau, \boldsymbol{q})$ is quasi-uniform and $u \in W^{k,\infty}(I; H_0^1(\Omega))$ with $k \geq 1$, then we have the following global error estimate

$$\|R^{\boldsymbol{p}}\theta\|_{L^\infty(I;L^2(\Omega))} \leq C \frac{\tau^{\min\{q+1,k\}}}{q^{k-\frac{1}{2}}} \|u\|_{W^{k,\infty}(I;H^1(\Omega))}. \tag{3.15}$$

Proof. By applying the definition of $\Pi^{\boldsymbol{q}}$ from (3.7), (3.5), and (3.6), we have

$$R^{\boldsymbol{p}}\theta(\boldsymbol{x}, t_{n-1}) = \left(R^{\boldsymbol{p}} u - \Pi^{\boldsymbol{q}} R^{\boldsymbol{p}} u\right)(\boldsymbol{x}, t_{n-1}) = 0$$

and

$$R^{\boldsymbol{p}}\theta(\boldsymbol{x}, t_n) = \left(R^{\boldsymbol{p}} u - \Pi^{\boldsymbol{q}} R^{\boldsymbol{p}} u\right)(\boldsymbol{x}, t_n) = 0,$$

which implies that $R^{\boldsymbol{p}}\theta \in H_0^1(I_n)$. Then, by applying the Sobolev inequality (cf. [29])

$$\|v\|_{L^\infty(a,b)}^2 \le \|v\|_{L^2(a,b)}|v|_{H^1(a,b)}, \quad \forall v \in H_0^1(a,b),$$

and utilizing the H^1-stability of the operator $R^{\boldsymbol{p}}$ from (3.3), we get

$$\begin{aligned}
\|R^{\boldsymbol{p}}\theta\|_{L^\infty(I_n;L^2(\Omega))}^2 &\le \left\|R^{\boldsymbol{p}}\theta\right\|_{L^2(I_n;L^2(\Omega))}\left\|\partial_t R^{\boldsymbol{p}}\theta\right\|_{L^2(I_n;L^2(\Omega))} \\
&\le \left\|R^{\boldsymbol{p}}\theta\right\|_{L^2(I_n;H^1(\Omega))}\left\|R^{\boldsymbol{p}}(\partial_t\theta)\right\|_{L^2(I_n;H^1(\Omega))} \\
&\le C\|\theta\|_{L^2(I_n;H^1(\Omega))}\left\|\partial_t\theta\right\|_{L^2(I_n;H^1(\Omega))}.
\end{aligned} \tag{3.16}$$

Similar to the process of proving (3.11), we can easily derive

$$\|\theta\|_{L^2(I_n;H^1(\Omega))}^2 \le \left(\frac{\tau_n}{2}\right)^{2k_n} \frac{\Gamma(q_n+2-k_n)}{q_n(q_n+1)\Gamma(q_n+k_n)}\|u\|_{H^{k_n}(I_n;H^1(\Omega))}^2. \tag{3.17}$$

On the other hand, from the estimate (3.9), we obtain

$$\begin{aligned}
\left\|\partial_t\theta\right\|_{L^2(I_n;H^1(\Omega))}^2 &= \left\|\partial_t(u-\Pi^{\boldsymbol{q}}u)\right\|_{L^2(I_n;H^1(\Omega))}^2 = \sum_{|\alpha|=0}^{1}\left\|D^\alpha\partial_t(u-\Pi^{\boldsymbol{q}}u)\right\|_{L^2(I_n;L^2(\Omega))}^2 \\
&= \sum_{|\alpha|=0}^{1}\left\|\partial_t(D^\alpha u-\Pi^{\boldsymbol{q}}(D^\alpha u))\right\|_{L^2(I_n;L^2(\Omega))}^2 \\
&\le \sum_{|\alpha|=0}^{1}\left(\frac{\tau_n}{2}\right)^{2k_n-2}\frac{\Gamma(q_n+2-k_n)}{\Gamma(q_n+k_n)}\left\|D^\alpha u\right\|_{H^{k_n}(I_n;L^2(\Omega))}^2 \\
&\le \left(\frac{\tau_n}{2}\right)^{2k_n-2}\frac{\Gamma(q_n+2-k_n)}{\Gamma(q_n+k_n)}\|u\|_{H^{k_n}(I_n;H^1(\Omega))}^2.
\end{aligned} \tag{3.18}$$

Taking into account the fact that

$$\|u\|_{H^{k_n}(I_n;H^1(\Omega))}^2 \le C\tau_n\|u\|_{W^{k_n,\infty}(I_n;H^1(\Omega))}^2,$$

and combining (3.16)–(3.18), we obtain

$$\begin{aligned}
\|R^{\boldsymbol{p}}\theta\|_{L^\infty(I_n;L^2(\Omega))}^2 &\le C\left(\frac{\tau_n}{2}\right)^{2k_n-1}\frac{\Gamma(q_n+2-k_n)}{q_n\Gamma(q_n+k_n)}\|u\|_{H^{k_n}(I_n;H^1(\Omega))}^2 \\
&\le C\left(\frac{\tau_n}{2}\right)^{2k_n}\frac{\Gamma(q_n+2-k_n)}{q_n\Gamma(q_n+k_n)}\|u\|_{W^{k_n,\infty}(I_n;H^1(\Omega))}^2,
\end{aligned}$$

which leads to (3.14).

Finally, applying Stirling's formula to (3.14), we derive (3.15). This completes the proof. □

As the final step, we estimate the term $\psi = \Pi^{\boldsymbol{q}}R^{\boldsymbol{p}}u - u_{h\tau}$. The following lemma demonstrates that ψ can be bounded by $\partial_t\eta$ and θ.

Lemma 3.8. *Assume that $u \in H^1(I;H_0^1(\Omega)\cap H^2(\Omega))$. Then, the following estimates hold*

$$\|\psi\|_{L^2(I;L^2(\Omega))} \le C\|\partial_t\eta\|_{L^2(I;L^2(\Omega))} + C\|\theta\|_{L^2(I;H^2(\Omega))}, \tag{3.19}$$

$$\|\psi\|_{L^\infty(I;L^2(\Omega))} \le C\|\partial_t\eta\|_{L^2(I;L^2(\Omega))} + C\|\theta\|_{L^2(I;H^2(\Omega))}. \tag{3.20}$$

Proof. First, replacing $u_{h\tau}$ in (2.3) with the exact solution u of the problem (1.1) yields

$$\int_I \big\{(\partial_t u, v_{h\tau}) + a(u, v_{h\tau})\big\}\,\mathrm{d}t = \int_I (f, v_{h\tau})\,\mathrm{d}t \tag{3.21}$$

for any $v_{h\tau} \in S^{\boldsymbol{p},\boldsymbol{q}-1,0}(\mathcal{M}_h, \mathcal{J}_\tau)$. Next, subtracting (2.3) from (3.21), we obtain

$$\int_I \big\{(\partial_t e, v_{h\tau}) + a(e, v_{h\tau})\big\}\,\mathrm{d}t = 0, \quad \forall v_{h\tau} \in S^{\boldsymbol{p},\boldsymbol{q}-1,0}(\mathcal{M}_h, \mathcal{J}_\tau). \tag{3.22}$$

Using the definitions of $R^{\boldsymbol{p}}$ and $\Pi^{\boldsymbol{q}}$ from (3.2) and (3.5), respectively, we derive

$$\int_I a(\eta, v_{h\tau})\,\mathrm{d}t = \int_I a\big(u - R^{\boldsymbol{p}}u, v_{h\tau}\big)\,\mathrm{d}t = 0 \tag{3.23}$$

and

$$\int_I \big(\partial_t R^{\boldsymbol{p}}\theta, v_{h\tau}\big)\,\mathrm{d}t = \sum_{n=1}^{N_t} \int_{I_n} \big(\partial_t\big(R^{\boldsymbol{p}}u - \Pi_{I_n}^{q_n} R^{\boldsymbol{p}}u\big), v_{h\tau}\big)\,\mathrm{d}t = 0. \tag{3.24}$$

Then, substituting $e = \eta + R^{\boldsymbol{p}}\theta + \psi$ into (3.22), and using (3.23) and (3.24), we obtain

$$\begin{aligned} \int_I \big\{(\partial_t \psi, v_{h\tau}) + a(\psi, v_{h\tau})\big\}\,\mathrm{d}t &= -\int_I (\partial_t \eta + \partial_t R^{\boldsymbol{p}}\theta, v_{h\tau})\,\mathrm{d}t - \int_I a(\eta + R^{\boldsymbol{p}}\theta, v_{h\tau})\,\mathrm{d}t \\ &= -\int_I (\partial_t \eta, v_{h\tau})\,\mathrm{d}t - \int_I a(R^{\boldsymbol{p}}\theta, v_{h\tau})\,\mathrm{d}t \end{aligned} \tag{3.25}$$

for any $v_{h\tau} \in S^{\boldsymbol{p},\boldsymbol{q}-1,0}(\mathcal{M}_h, \mathcal{J}_\tau)$.

Using the definition of $R^{\boldsymbol{p}}$ from (3.2) and applying integration by parts, we have

$$\int_I a\big(R^{\boldsymbol{p}}\theta, v_{h\tau}\big)\,\mathrm{d}t = \int_I a(\theta, v_{h\tau})\,\mathrm{d}t = \int_I (\mathcal{A}\theta, v_{h\tau})\,\mathrm{d}t.$$

Substituting this into (3.25), we get

$$\int_I \big\{(\partial_t \psi, v_{h\tau}) + a(\psi, v_{h\tau})\big\}\,\mathrm{d}t = \int_I (-\partial_t \eta - \mathcal{A}\theta, v_{h\tau})\,\mathrm{d}t. \tag{3.26}$$

Moreover, using (2.3), (3.7), and (3.5), we obtain

$$\psi(\boldsymbol{x}, 0) = \big(\Pi^{\boldsymbol{q}} R^{\boldsymbol{p}}u - u_{h\tau}\big)(\boldsymbol{x}, 0) = \big(\Pi^{\boldsymbol{q}} R^{\boldsymbol{p}}u - R^{\boldsymbol{p}}u\big)(\boldsymbol{x}, 0) = 0. \tag{3.27}$$

Let $g := -\partial_t \eta - \mathcal{A}\theta$. Combing (3.26) and (3.27) yields

$$\begin{cases} \displaystyle\int_I \big\{(\partial_t \psi, v_{h\tau}) + a(\psi, v_{h\tau})\big\}\,\mathrm{d}t = \int_I (g, v_{h\tau})\,\mathrm{d}t, \\ \psi(\boldsymbol{x}, 0) = 0 \end{cases} \tag{3.28}$$

for any $v_{h\tau} \in S^{\boldsymbol{p},\boldsymbol{q}-1,0}(\mathcal{M}_h, \mathcal{J}_\tau)$. By comparing (3.28) and (2.3), it is clear that ψ satisfies a variational equation almost identical to that of $u_{h\tau}$, except for the right-hand side and the initial condition. Therefore, ψ also satisfies a stability estimate similar to (2.13), namely,

$$\|\psi\|_{L^2(I;L^2(\Omega))} + \|\partial_t \psi\|_{L^2(I;L^2(\Omega))} \le C\|g\|_{L^2(I;L^2(\Omega))} + C\|\psi(\boldsymbol{x}, 0)\|_{H^1(\Omega)}.$$

Thus, we obtain

$$\begin{aligned} \|\psi\|_{L^2(I;L^2(\Omega))} + \|\partial_t\psi\|_{L^2(I;L^2(\Omega))} &\le C\|\partial_t\eta\|_{L^2(I;L^2(\Omega))} + C\|\mathcal{A}\theta\|_{L^2(I;L^2(\Omega))} \\ &\le C\|\partial_t\eta\|_{L^2(I;L^2(\Omega))} + C\|\theta\|_{L^2(I;H^2(\Omega))}, \end{aligned} \tag{3.29}$$

which implies (3.19).

According to the Sobolev embedding theorem

$$\|v\|_{L^\infty(a,b)} \le C\|v\|_{H^1(a,b)}, \quad \forall v \in H^1(a,b), \tag{3.30}$$

and from (3.29), we deduce that

$$\|\psi\|_{L^\infty(I;L^2(\Omega))} \le C\|\psi\|_{H^1(I;L^2(\Omega))} \le C\|\partial_t\eta\|_{L^2(I;L^2(\Omega))} + C\|\theta\|_{L^2(I;H^2(\Omega)},$$

which leads to (3.20). □

By combining the estimates from Lemmas 3.8, we can immediately derive the following abstract error bounds for the *hp*-version fully discrete Galerkin approximation.

Theorem 3.9. *Assume that $u \in H^1(I; H_0^1(\Omega) \cap H^2(\Omega))$ is the exact solution of* (1.1)*, and $u_{h\tau}$ is the hp-version fully discrete Galerkin approximation defined in* (2.3)*. Then, we have the following error estimates*

$$\|u - u_{h\tau}\|_{L^2(I;L^2(\Omega))} \le C\|\eta\|_{H^1(I;L^2(\Omega))} + C\|\theta\|_{L^2(I;H^2(\Omega))}, \tag{3.31}$$

$$\|u - u_{h\tau}\|_{L^\infty(I;L^2(\Omega))} \le C\|\eta\|_{H^1(I;L^2(\Omega))} + \|R^{\boldsymbol{p}}\theta\|_{L^\infty(I;L^2(\Omega))} + C\|\theta\|_{L^2(I;H^2(\Omega))}, \tag{3.32}$$

where η, θ, and $R^{\boldsymbol{p}}\theta$ are defined in (3.10)*.*

Proof. Using (3.10), the triangle inequality, (3.19), and (3.4), we have

$$\begin{aligned} \|u - u_{h\tau}\|_{L^2(I;L^2(\Omega))} &\le \|\eta\|_{L^2(I;L^2(\Omega))} + \|R^{\boldsymbol{p}}\theta\|_{L^2(I;L^2(\Omega))} + \|\psi\|_{L^2(I;L^2(\Omega))} \\ &\le \|\eta\|_{L^2(I;L^2(\Omega))} + \|R^{\boldsymbol{p}}\theta\|_{L^2(I;H^1(\Omega))} + C\|\partial_t\eta\|_{L^2(I;L^2(\Omega))} + C\|\theta\|_{L^2(I;H^2(\Omega))} \\ &\le C\|\eta\|_{H^1(I;L^2(\Omega))} + \|R^{\boldsymbol{p}}\theta\|_{L^2(I;H^1(\Omega))} + C\|\theta\|_{L^2(I;H^2(\Omega))} \\ &\le C\|\eta\|_{H^1(I;L^2(\Omega))} + C\|\theta\|_{L^2(I;H^1(\Omega))} + C\|\theta\|_{L^2(I;H^2(\Omega))} \\ &\le C\|\eta\|_{H^1(I;L^2(\Omega))} + C\|\theta\|_{L^2(I;H^2(\Omega))}, \end{aligned}$$

which implies (3.31).

Next, applying the Sobolev inequality (3.30), we have

$$\|\eta\|_{L^\infty(I;L^2(\Omega))} \le C\|\eta\|_{H^1(I;L^2(\Omega))}.$$

Combining this with (3.10), the triangle inequality, and (3.20), we obtain

$$\begin{aligned} \|u - u_{h\tau}\|_{L^\infty(I;L^2(\Omega))} &\le \|\eta\|_{L^\infty(I;L^2(\Omega))} + \|R^{\boldsymbol{p}}\theta\|_{L^\infty(I;L^2(\Omega))} + \|\psi\|_{L^\infty(I;L^2(\Omega))} \\ &\le \|\eta\|_{L^\infty(I;L^2(\Omega))} + \|R^{\boldsymbol{p}}\theta\|_{L^\infty(I;L^2(\Omega))} + C\|\partial_t\eta\|_{L^2(I;L^2(\Omega))} + C\|\theta\|_{L^2(I;H^2(\Omega))} \\ &\le C\|\eta\|_{H^1(I;L^2(\Omega))} + \|R^{\boldsymbol{p}}\theta\|_{L^\infty(I;L^2(\Omega))} + C\|\theta\|_{L^2(I;H^2(\Omega))}, \end{aligned}$$

which yields (3.32). The proof is complete. □

3.3. Convergence rates for smooth solutions

We now combine the abstract error estimates from Theorem 3.9 with the estimates for η, θ, and $R^{\boldsymbol{p}}\theta$ to derive the following $L^2(I;L^2(\Omega))$- and $L^\infty(I;L^2(\Omega))$-error estimates for the hp-version fully discrete Galerkin approximation.

Theorem 3.10. *Let u be the exact solution of the problem (1.1), and $u_{h\tau}$ be the hp-version fully discrete Galerkin approximation defined in (2.3). Let the tuple $(\mathcal{M}_h,\boldsymbol{p})$ be a quasi-uniform hp-discretization of the spatial domain Ω, and $(\mathcal{J}_\tau,\boldsymbol{q})$ be an arbitrary hp-discretization of the time interval I. Assume that $u\in H^1(I;H_0^1(\Omega)\cap H^r(\Omega))$ with $r\geq 2$ satisfies $u|_{\Omega\times I_n}\in H^{k_{0,n}}(I_n;H^2(\Omega))$ for $k_{0,n}\geq 1$. Then, for $1\leq k_n\leq \min\{q_n+1,k_{0,n}\}$, we have the following error estimate*

$$\begin{aligned}\|u-u_{h\tau}\|^2_{L^2(I;L^2(\Omega))}\leq{}& C\frac{h^{2\min\{p+1,r\}}}{p^{2r}}\|u\|^2_{H^1(I;H^r(\Omega))}\\&+C\sum_{n=1}^{N_t}\Big(\frac{\tau_n}{2}\Big)^{2k_n}\frac{\Gamma(q_n+2-k_n)}{q_n(q_n+1)\Gamma(q_n+k_n)}\|u\|^2_{H^{k_n}(I_n;H^2(\Omega))}.\end{aligned}\tag{3.33}$$

Furthermore, if u satisfies $u|_{\Omega\times I_n}\in H^{k_{0,n}}(I_n;H^2(\Omega))\cap W^{k_{0,n},\infty}(I_n;H^1(\Omega))$ for $k_{0,n}\geq 1$, then for $1\leq k_n\leq \min\{q_n+1,k_{0,n}\}$, we have the following error estimate

$$\begin{aligned}\|u-u_{h\tau}\|^2_{L^\infty(I;L^2(\Omega))}\leq{}& C\frac{h^{2\min\{p+1,r\}}}{p^{2r}}\|u\|^2_{H^1(I;H^r(\Omega))}\\&+C\sum_{n=1}^{N_t}\Big(\frac{\tau_n}{2}\Big)^{2k_n}\frac{\Gamma(q_n+2-k_n)}{q_n(q_n+1)\Gamma(q_n+k_n)}\|u\|^2_{H^{k_n}(I_n;H^2(\Omega))}\\&+C\max_{1\leq n\leq N_t}\Big\{\Big(\frac{\tau_n}{2}\Big)^{2k_n}\frac{\Gamma(q_n+2-k_n)}{q_n\Gamma(q_n+k_n)}\|u\|^2_{W^{k_n,\infty}(I_n;H^1(\Omega))}\Big\}.\end{aligned}\tag{3.34}$$

Proof. By combining (3.31), (3.13), and (3.11), we can derive the estimate (3.33). Similarly, the estimate (3.34) follows from (3.32), (3.13), (3.11), and (3.14). □

Based on Theorem 3.10, by employing quasi-uniform mesh partitions and quasi-uniform approximation degrees in both space and time, and by applying Stirling's formula, we derive the following typical hp-version error estimates for both space and time.

Corollary 3.11. *Let $(\mathcal{M}_h,\boldsymbol{p})$ and $(\mathcal{J}_\tau,\boldsymbol{q})$ denote quasi-uniform hp-discretizations of the spatial domain Ω and the time interval I, respectively. Assume that $u\in H^1(I;H_0^1(\Omega)\cap H^r(\Omega))\cap H^k(I;H^2(\Omega))$ with $k\geq 1$ and $r\geq 2$. Then, we have the following error estimate*

$$\|u-u_{h\tau}\|_{L^2(I;L^2(\Omega))}\leq C\frac{h^{\min\{p+1,r\}}}{p^r}\|u\|_{H^1(I;H^r(\Omega))}+C\frac{\tau^{\min\{q+1,k\}}}{q^k}\|u\|_{H^k(I;H^2(\Omega))}.\tag{3.35}$$

Moreover, if $u\in H^1(I;H_0^1(\Omega)\cap H^r(\Omega))\cap W^{k,\infty}(I;H^2(\Omega))$ with $k\geq 1$ and $r\geq 2$, we have the following error estimate

$$\|u-u_{h\tau}\|_{L^\infty(I;L^2(\Omega))}\leq C\frac{h^{\min\{p+1,r\}}}{p^r}\|u\|_{H^1(I;H^r(\Omega))}+C\frac{\tau^{\min\{q+1,k\}}}{q^{k-1/2}}\|u\|_{W^{k,\infty}(I;H^2(\Omega))}.\tag{3.36}$$

Remark 3.12. The error estimates (3.35) and (3.36) demonstrate that the hp-version fully discrete Galerkin scheme converges as the mesh sizes h (in space) and τ (in time) are refined, or as the approximation degrees p (in space) and q (in time) are increased.

Remark 3.13. The error estimates (3.35) and (3.36) also imply that the p-version fully discrete Galerkin method (*i.e.*, using fixed and relatively coarse space and time mesh partitions) can achieve spectral convergence if the exact solution u is sufficiently smooth. In other words, the smoother the solution, the higher the convergence order with respect to the approximation degrees p and q. Specifically, if u is analytic in $\Omega \times [0,T]$, exponential convergence can be achieved in both the space and time directions.

3.4. Exponential convergence rates for non-smooth solutions

Next, we consider the *hp*-version fully discrete Galerkin method for the problem (1.1) with solutions exhibiting startup singularities at $t=0$, but analytic behavior for $t>0$.

Specifically, we assume that the source term f and initial data u_0 are chosen such that the exact solution u satisfies the following regularity estimate

$$\left\|\partial_t^j u\right\|_{H^2(\Omega)} \leq C d^j \Gamma(j+1) t^{\mu-j}, \quad t \in (0,T], \quad j \in \mathbb{N}_0, \quad \mu \geq 1, \tag{3.37}$$

where C and d are positive constants dependent on u, but independent of t, j, and μ. The exponent μ characterizes the strength of the singularity at $t=0$. For a more detailed discussion of the regularity estimate (3.37), please refer to [46].

For solutions satisfying the regularity assumption (3.37), the optimal convergence rates in Theorem 3.10 cannot be achieved with a quasi-uniform time mesh over $(0,T)$, owing to the initial singularity at $t=0$.

Below, we demonstrate that exponential convergence in the temporal direction can be achieved using a specialized *hp*-version discretization with geometric time partitions and linearly increasing approximation degrees. To this end, we introduce the following definitions (cf. [12, 45, 46, 55]).

Definition 3.14 (Basic geometric mesh)**.** We define $\hat{\mathcal{J}}_{M,\sigma}$ as the basic geometric partition of the interval $\hat{I}=(0,1)$ with a grading parameter $\sigma \in (0,1)$. The partition is determined by the nodal points

$$t_0 = 0, \quad t_m = \sigma^{M-m+1}, \quad 1 \leq m \leq M+1.$$

Definition 3.15 (General geometric mesh)**.** A geometric partition $\mathcal{J}_{M,\sigma}$ of the interval $(0,T)$, with grading factor $\sigma \in (0,1)$ and M levels of refinement, is constructed by first partitioning $(0,T)$ quasi-uniformly into intervals $\{J_l\}_{l=1}^{L}$. The first interval, $J_1 = (0,t_1)$, is then subdivided into $M+1$ subintervals $\{I_m\}_{m=1}^{M+1}$ by applying a linear mapping of the basic geometric mesh $\hat{\mathcal{J}}_{M,\sigma}$ onto J_1.

Definition 3.16 (Linearly increasing approximation degrees)**.** The approximation degree vector $\boldsymbol{q}$ on the geometric mesh $\mathcal{J}_{M,\sigma}$ is said to be linear with slope $\nu > 0$ if it satisfies the following conditions: (i) On the geometrically refined elements $\{I_m\}_{m=1}^{M+1}$, the approximation degrees are given by $q_m = \lceil \nu m \rceil$ for $1 \leq m \leq M+1$. (ii) On the coarse elements $\{J_l\}_{l=2}^{L}$, the approximation degrees remain constant, defined as $q_{M+l} = \lceil \nu(M+1) \rceil$.

We now present the main results on exponential convergence rates in the temporal direction for solutions exhibiting a start-up singularity at $t=0$.

Theorem 3.17. *Let u be the exact solution of problem (1.1), and let $u_{h\tau}$ denote the hp-version fully discrete Galerkin approximation defined in (2.3). Assume that $(\mathcal{M}_h, \boldsymbol{p})$ is a quasi-uniform hp-discretization of Ω, and the geometric mesh $\mathcal{J}_{M,\sigma}$ combined with a linearly increasing approximation degree $\boldsymbol{q}$ is used for time discretization. Suppose further that $u \in H^1(I; H_0^1(\Omega) \cap H^r(\Omega))$ with $r \geq 2$ and satisfies the regularity assumption (3.37). Then, there exists a slope $\nu_0 > 0$ such that for all linearly increasing approximation degree vectors $\boldsymbol{q}$ with slope $\nu \geq \nu_0$, the following error estimates hold:*

$$\|u - u_{h\tau}\|_{L^2(I;L^2(\Omega))} \leq C \frac{h^{\min\{p+1,r\}}}{p^r} + C e^{-b\sqrt{D_t}}, \tag{3.38}$$

TABLE 1. Example 1: temporal convergence of the h-version CPG time-stepping method.

q	τ	$\|e\|_{L^2(I,L^2(\Omega))}$	Order	$\|e\|_{L^\infty(I,L^2(\Omega))}$	Order
	1/512	8.25e−06	2.00	9.00e−05	1.95
1	1/1024	2.06e−06	2.00	2.29e−05	1.98
	1/2048	5.15e−07	2.00	5.76e−06	1.99
	1/256	2.10e−07	3.00	1.79e−06	2.96
2	1/512	2.62e−08	3.00	2.27e−07	2.98
	1/1024	3.28e−09	3.00	2.85e−08	2.99
	1/128	1.49e−08	3.99	1.36e−07	3.89
3	1/256	9.34e−10	4.00	8.86e−09	3.94
	1/512	5.84e−11	4.00	5.64e−10	3.97

$$\|u - u_{h\tau}\|_{L^\infty(I;L^2(\Omega))} \le C\frac{h^{\min\{p+1,r\}}}{p^r} + Ce^{-b\sqrt{D_t}}, \tag{3.39}$$

where D_t represents the degrees of freedom of the hp-CPG discretization in the temporal direction, and C and b are positive constants independent of h, p, and D_t.

Proof. Theorem 3.9 shows that the error of the hp-version fully discrete Galerkin approximation is controlled by η (Ritz–Galerkin projection error) and θ (temporal projection error). Using the regularity assumption (3.37), the approximation properties of θ (cf. Lem. 3.5) and $R^p\theta$ (cf. Lem. 3.7), and the similar proof techniques from [12, 45, 46, 55], the exponential convergence in time for (3.38) and (3.39) is established. The hp-version convergence in space follows from the proof of Corollary 3.11. □

4. NUMERICAL EXPERIMENTS

In this section, we present numerical results to validate the theoretical findings. The spatial domain Ω is uniformly partitioned into $N_{\boldsymbol{x}}$ rectangular elements with mesh size denoted by h, while the time interval $I = (0, T)$ is divided into either uniform meshes (for smooth solutions) or geometric meshes (for singular solutions), with the number of time steps denoted by N_t and the uniform step-size denoted by τ. The error function is defined as $e := u - u_{h\tau}$, where $u_{h\tau}$ represents the hp-version fully discrete Galerkin approximation as specified in (2.3). Additionally, the term "order" refers to the experimentally observed convergence rates.

4.1. Example 1: heat equation with smooth solution

Consider the two-dimensional heat equation

$$\begin{cases} \partial_t u - \triangle u = f(\boldsymbol{x}, t) & \text{in } \Omega \times I, \\ \quad u(\boldsymbol{x}, t) = 0 & \text{on } \partial\Omega \times I, \\ \quad u(\boldsymbol{x}, 0) = u_0(\boldsymbol{x}) & \text{in } \Omega, \end{cases}$$

with $\Omega = (0, 1) \times (0, 1)$ and $I = (0, 1)$. The right-hand side f and the initial condition u_0 are chosen such that

$$u = e^{-2\pi^2 t} \sin(\pi x) \sin(\pi y).$$

We first investigate the convergence orders in time for the h- and p-version CPG time-stepping methods, respectively. To this end, we employ the p-version CG method with only one element (*i.e.*, $h = 1$) and the polynomial degree of $p = 24$ for spatial discretization, ensuring that the spatial error is negligible. From Table 1,

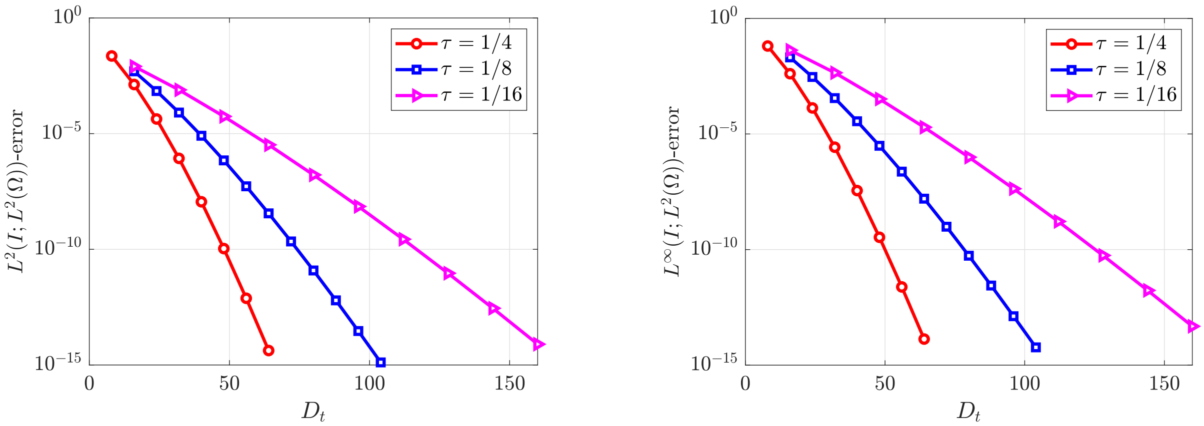

FIGURE 1. Example 1: temporal convergence of the p-version CPG time-stepping method.

TABLE 2. Example 1: spatial convergence of the h-version CG method.

p	h	$\|e\|_{L^2(I,L^2(\Omega))}$	Order	$\|e\|_{L^\infty(I,L^2(\Omega))}$	Order
	1/16	4.35e−04	1.99	1.90e−03	2.00
1	1/32	1.09e−04	2.00	4.75e−04	2.00
	1/64	2.72e−05	2.00	1.19e−04	2.00
	1/8	3.91e−05	3.00	2.45e−04	2.98
2	1/16	4.90e−06	3.00	3.07e−05	2.99
	1/32	6.12e−07	3.00	3.85e−06	3.00
	1/4	1.42e−05	3.98	8.81e−05	3.95
3	1/8	8.88e−07	4.00	5.56e−06	3.99
	1/16	5.55e−08	4.00	3.49e−07	4.00

we observe that the convergence orders of the h-version CPG method with $q = 1, 2, 3$ achieve the expected order $q + 1$ for both the $L^2(L^2)$- and $L^\infty(L^2)$-errors, aligning well with the theoretical results in Corollary 3.11.

In Figure 1, we display the $L^2(L^2)$- and $L^\infty(L^2)$-errors of the p-version CPG time-stepping method with a fixed temporal mesh partition (with uniform step-size $\tau = 1/4$, $1/8$, and $1/16$, respectively). The errors are plotted on a semi-log scale, where the horizontal axis represents the degrees of freedom in the time direction D_t (corresponding to the temporal approximation degree q). From Figure 1, it is evident that the p-version CPG time-stepping method achieves exponential convergence, consistent with the discussion in Remark 3.13.

We next consider the spatial convergence of the h- and p-version CG methods, respectively. To this end, we employ the p-version CPG method with $\tau = 1$ and a fixed polynomial degree of $q = 24$ for time discretization, ensuring that the temporal discretization error is sufficiently small. From Table 2, we observe that the h-version CG method (with uniform approximation degrees $p = 1, 2, 3$) achieves the expected order of $p+1$ in the spatial direction, consistent with the prediction of Corollary 3.11.

In Figure 2, we display the convergence rates of the p-version CG method (with 1, 4, and 16 elements, respectively) in the spatial direction. Evidently, it achieves exponential convergence, aligning well with Remark 3.13. Furthermore, we find that to achieve the same level of accuracy, the degrees of freedom required by the p-version method are significantly fewer than those required by the h-version method. Therefore, for smooth solutions, the p-version method offers greater advantages compared to the h-version method.

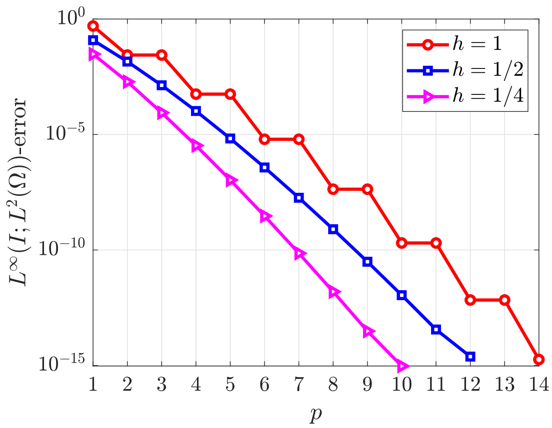

FIGURE 2. Example 1: spatial convergence of the p-version CG method.

4.2. Example 2: variable coefficient problem with smooth solution

Consider the two-dimensional parabolic equation

$$\begin{cases} \partial_t u - \nabla \cdot (a \nabla u) + bu = f(\boldsymbol{x}, t) & \text{in } \Omega \times I, \\ u(\boldsymbol{x}, t) = 0 & \text{on } \partial\Omega \times I, \\ u(\boldsymbol{x}, 0) = u_0(\boldsymbol{x}) & \text{in } \Omega, \end{cases}$$

with $\Omega = (0,1) \times (0,1)$, $I = (0,1)$, and the coefficients

$$a = \begin{bmatrix} x^2 + y^2 + 1 & xy \\ xy & x^2 + y^2 + 1 \end{bmatrix} \quad \text{and} \quad b = x + y.$$

The right-hand side f and the initial condition u_0 are chosen such that

$$u = e^{t+x+y} x(x-1)y(y-1).$$

We first examine the temporal convergence of the h- and p-version CPG time-stepping methods. To minimize spatial errors, the p-version CG method with $h = 1$ and the polynomial degree of $p = 24$ is used for spatial discretization. From Table 3, it is evident that the h-version CPG method with $q = 1, 2, 3$ achieves the expected convergence order of $q + 1$ for both the $L^2(L^2)$- and $L^\infty(L^2)$-errors, in agreement with Corollary 3.11.

Figure 3 illustrates the $L^2(L^2)$- and $L^\infty(L^2)$-errors for the p-version CPG time-stepping method using fixed temporal partitions (with $\tau = 1$, $1/2$, and $1/4$, respectively). The errors are plotted on a semi-log scale with the horizontal axis denoting the temporal degrees of freedom D_t (corresponding to the approximation degree q). The results clearly demonstrate exponential convergence of the p-version method, consistent with Remark 3.13.

We next analyze the spatial convergence of the h- and p-version CG methods. To suppress temporal errors, the p-version CPG method with $\tau = 1$ and $q = 24$ is employed for time discretization. From Table 4, we observe that the h-version CG method (with polynomial degrees $p = 1, 2, 3$) achieves the expected order of $p + 1$, verifying the theoretical results in Corollary 3.11.

In Figure 4, we illustrate the convergence rates of the p-version CG method (with 1, 4, and 16 elements) in the spatial direction. Once again, the method exhibits exponential convergence, consistent with Remark 3.13.

TABLE 3. Example 2: temporal convergence of the h-version CPG time-stepping method.

q	τ	$\|e\|_{L^2(I,L^2(\Omega))}$	Order	$\|e\|_{L^\infty(I,L^2(\Omega))}$	Order
	1/64	1.60e−06	2.00	5.24e−06	2.00
1	1/128	4.01e−07	2.00	1.31e−06	2.00
	1/256	1.00e−07	2.00	3.27e−07	2.00
	1/32	3.10e−08	3.05	6.43e−08	3.02
2	1/64	3.83e−09	3.02	8.02e−09	3.00
	1/128	4.77e−10	3.00	1.01e−09	3.00
	1/16	8.57e−10	3.93	1.95e−09	4.02
3	1/32	5.46e−11	3.97	1.27e−10	3.94
	1/64	3.43e−12	3.99	8.10e−12	3.97

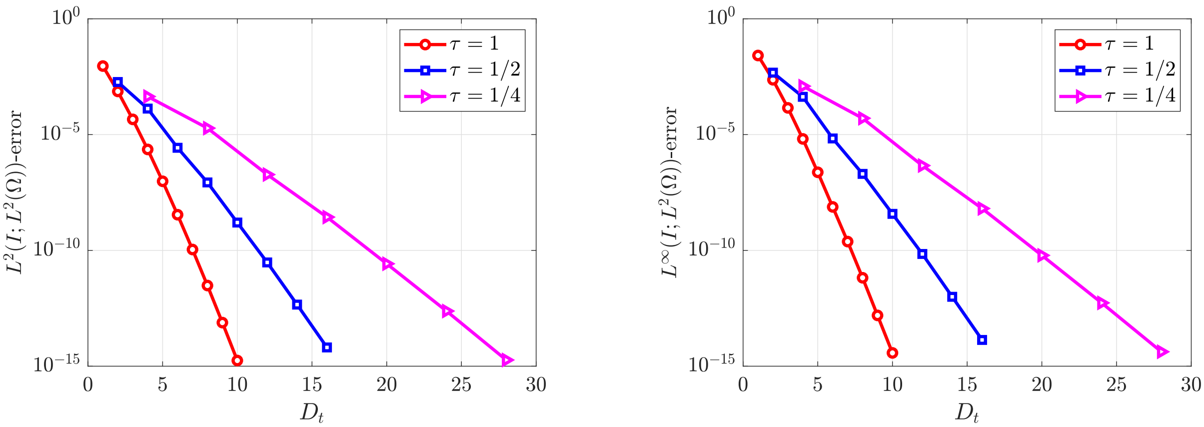

FIGURE 3. Example 2: temporal convergence of the p-version CPG time-stepping method.

TABLE 4. Example 2: spatial convergence of the h-version CG method.

p	h	$\|e\|_{L^2(I,L^2(\Omega))}$	Order	$\|e\|_{L^\infty(I,L^2(\Omega))}$	Order
	1/16	1.06e−03	2.00	1.61e−03	2.00
1	1/32	2.65e−04	2.00	4.02e−04	2.00
	1/64	6.61e−05	2.00	1.01e−04	2.00
	1/8	1.10e−04	2.98	1.67e−04	2.98
2	1/16	1.38e−05	3.00	2.10e−05	3.00
	1/32	1.72e−06	3.00	2.62e−06	3.00
	1/4	2.39e−05	3.94	3.63e−05	3.94
3	1/8	1.51e−06	3.98	2.30e−06	3.98
	1/16	9.46e−08	4.00	1.44e−07	4.00

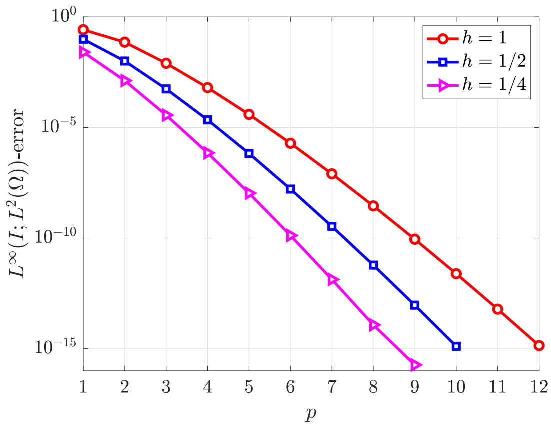

FIGURE 4. Example 2: spatial convergence of the p-version CG method.

4.3. Example 3: variable coefficient problem with non-smooth solution

Consider the two-dimensional parabolic equation

$$\begin{cases} \partial_t u - \nabla\cdot(a\nabla u) + bu = f(\boldsymbol{x},t) & \text{in } \Omega\times I, \\ u(\boldsymbol{x},t)=0 & \text{on } \partial\Omega\times I, \\ u(\boldsymbol{x},0)=u_0(\boldsymbol{x}) & \text{in } \Omega, \end{cases}$$

with $\Omega=(-1,1)\times(-1,1)$, $I=(0,1)$, and the coefficients

$$a=\begin{bmatrix} 3+e^{-x^2} & \sin x\cos y \\ \sin x\cos y & 3+e^{-y^2}\end{bmatrix} \quad\text{and}\quad b=x^2+y^2+1.$$

The right-hand side f and the initial condition u_0 are chosen such that

$$u=t^\mu e^{-(x+y)}(x^2-1)(y^2-1)$$

with $\mu\ge 1$. It can be readily seen that u satisfies the regularity condition (3.37).

Since this test solution is sufficiently smooth in the spatial direction, we focus solely on the performance of the CPG method in the temporal direction for this example. To achieve this, we use the p-version CG method with $p=24$ and $h=2$ for spatial discretization, ensuring that the spatial error is negligible.

Consider the case $\mu=1.5$, where the solution u belongs to $H^{2-\varepsilon}(I;H^{r+1}(\Omega))$ for any $\varepsilon>0$ and satisfies $u\in W^{1.5,\infty}(I;H^{r+1}(\Omega))$, with r being an arbitrary positive integer.

We first examine the performance of the h- and p-version CPG methods on uniform temporal meshes, respectively. As shown in Table 5, the temporal convergence rates for the $L^2(L^2)$- and $L^\infty(L^2)$-errors are 2 and 1.5, respectively, aligning well with the predictions of Corollary 3.11. This observation highlights that for non-smooth solutions, increasing the polynomial degree q does not lead to the optimal convergence order of $q+1$.

In Table 6, we present the convergence rates of the p-version CPG method. It is evident that the $L^2(L^2)$- and $L^\infty(L^2)$-errors in the temporal direction achieve orders of 3.5 and 3, respectively. These convergence rates not only exceed those of the h-version but also surpass the expectations outlined in Corollary 3.11. However, this result is not surprising, as it aligns with the theoretical findings in [8,48,55]. Specifically, for solutions with endpoint singularities, the convergence rate of the p-version method is known to be twice that of the h-version

TABLE 5. Example 3: temporal convergence of the h-version CPG time-stepping method with uniform mesh for $\mu = 1.5$.

q	τ	$\|e\|_{L^2(I,L^2(\Omega))}$	Order	$\|e\|_{L^\infty(I,L^2(\Omega))}$	Order
1	1/256	2.93e−06	1.88	4.94e−05	1.46
	1/512	7.88e−07	1.89	1.77e−05	1.48
	1/1024	2.10e−07	1.90	6.31e−06	1.49
2	1/128	8.50e−07	2.00	1.46e−05	1.47
	1/256	2.12e−07	2.00	5.23e−06	1.48
	1/512	5.30e−08	2.00	1.86e−06	1.49
3	1/64	9.10e−07	2.01	1.19e−05	1.47
	1/128	2.27e−07	2.01	4.24e−06	1.48
	1/256	5.65e−08	2.00	1.51e−06	1.49

TABLE 6. Example 3: temporal convergence of the p-version CPG time-stepping method with uniform mesh for $\mu = 1.5$.

τ	q	$\|e\|_{L^2(I,L^2(\Omega))}$	Order	$\|e\|_{L^\infty(I,L^2(\Omega))}$	Order
1	20	7.57e−06	3.48	1.99e−05	2.95
	22	5.43e−06	3.48	1.50e−05	2.96
	24	4.01e−06	3.49	1.16e−05	2.96
	26	3.03e−06	3.50	9.13e−06	2.97
1/2	20	1.81e−06	3.46	7.10e−06	2.97
	22	1.30e−06	3.47	5.35e−06	2.97
	24	9.59e−07	3.47	4.13e−06	2.97
	26	7.26e−07	3.48	3.26e−06	2.98
1/4	20	4.38e−07	3.44	2.53e−06	2.98
	22	3.15e−07	3.45	1.91e−06	2.98
	24	2.33e−07	3.45	1.47e−06	2.99
	26	1.77e−07	3.46	1.16e−06	2.99

method. For ease of comparison, Figure 5 illustrates the temporal L^∞-errors for both the h- and p-version CPG methods, plotted against the degrees of freedom in time on a log–log scale.

To accurately capture the singular behavior of the exact solution near $t = 0$, we employ the hp-version CPG method in the temporal direction. This method employs basic geometric time partitions $\hat{\mathcal{J}}_{M,\sigma}$ of $I = (0, 1)$, combined with linearly increasing approximation degrees, characterized by a slope ν, as outlined in Section 3.4. Figure 6 presents the $L^\infty(L^2)$-errors plotted against the square root of the degrees of freedom in time for various values of σ and ν on a semi-log scale. The results clearly demonstrate exponential convergence rates for all tested σ and ν, validating the theoretical findings in Theorem 3.17. While Figure 6 suggests that the optimal discrete parameters are approximately $\sigma = 0.3$ and $\nu = 1$, determining the optimal selection of σ and ν remains an open question and warrants further investigation.

Finally, in Figure 7, we evaluate the hp-version CPG method on basic geometric time partitions $\hat{\mathcal{J}}_{M,\sigma}$ with linearly increasing approximation degrees for various values of μ. The grading factor is set to $\sigma = 0.3$, and the slope parameter to $\nu = 1$. The results demonstrate that exponential convergence rates are achieved for all tested values of μ, which align closely with the theoretical predictions.

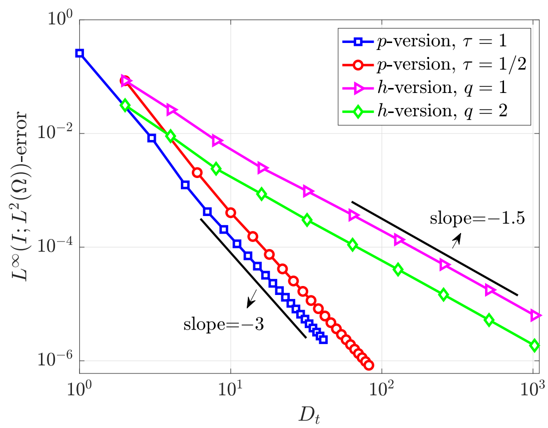

FIGURE 5. Example 3: temporal convergence of the h- and p-version CPG time-stepping methods with uniform meshes for $\mu = 1.5$.

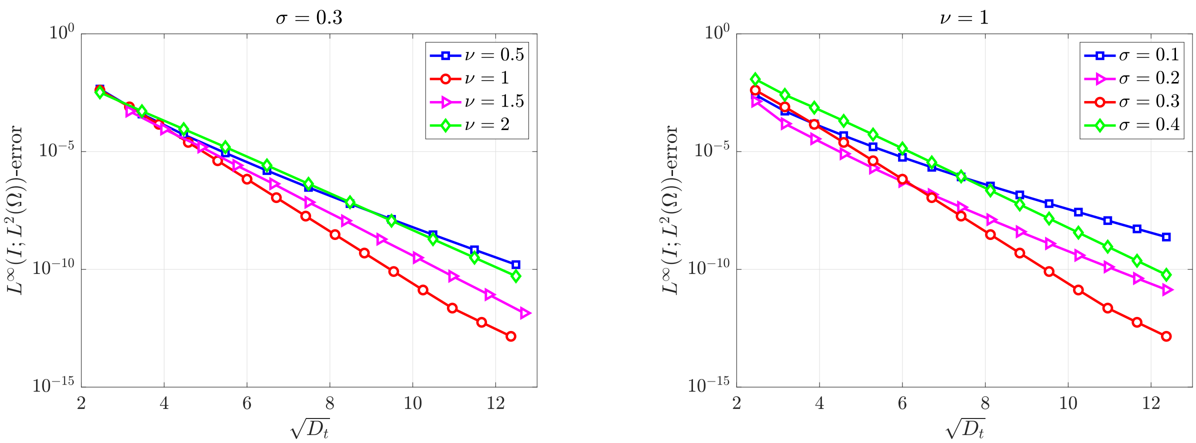

FIGURE 6. Example 3: temporal convergence of the hp-version CPG time-stepping method with fixed σ (*left*) or ν (*right*) for $\mu = 1.5$.

5. CONCLUDING REMARKS

In this work, we developed a fully discrete hp-version Galerkin method for second-order parabolic equations, combining an hp-version CPG time-stepping scheme in the temporal direction with an hp-version CG method for spatial discretization. We derived *a priori* error estimates in the $L^2(L^2)$- and $L^\infty(L^2)$-norms, with explicit dependence on the mesh sizes and polynomial degrees. For solutions with initial singularities, we demonstrated exponential convergence rates using geometric time partitions with linearly increasing approximation degrees. Future research will focus on extending the hp-version CPG time-stepping method to nonlinear parabolic equations and fourth-order parabolic equations.

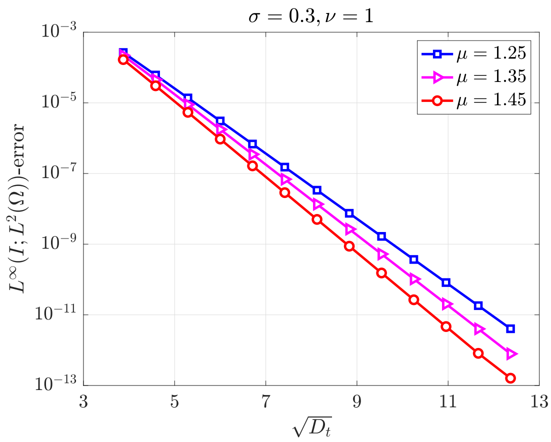

FIGURE 7. Example 3: temporal convergence of the *hp*-version CPG time-stepping method with fixed $\sigma = 0.3$ and $\nu = 1$ for different μ.

Funding

The work of L. Yi is supported by the National Natural Science Foundation of China (Grant Nos. 12171322, 12571391, and 12271366).

Data availability statement

The research data associated with this article are included in the article.

References

[1] N. Ahmed and G. Matthies, Higher order continuous Galerkin–Petrov time stepping schemes for transient convection-diffusion-reaction equations. *ESAIM: Math. Model. Numer. Anal.* **49** (2015) 1429–1450.

[2] N. Ahmed, G. Matthies, L. Tobiska and H. Xie, Discontinuous Galerkin time stepping with local projection stabilization for transient convection-diffusion-reaction problems. *Comput. Methods Appl. Mech. Eng.* **200** (2011) 1747–1756.

[3] G. Akrivis and C. Makridakis, Galerkin time-stepping methods for nonlinear parabolic equations. *ESAIM: Math. Model. Numer. Anal.* **38** (2004) 261–289.

[4] S.M. Allen and J.W. Cahn, A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening. *Acta Metall.* **27** (1979) 1085–1095.

[5] P.F. Antonietti, I. Mazzieri, N. Dal Santo and A. Quarteroni, A high-order discontinuous Galerkin approximation to ordinary differential equations with applications to elastodynamics. *IMA J. Numer. Anal.* **38** (2018) 1709–1734.

[6] A.K. Aziz and P. Monk, Continuous finite elements in space and time for the heat equation. *Math. Comput.* **52** (1989) 255–274.

[7] I. Babuška and T. Janik, The *h-p* version of the finite element method for parabolic equations: Part II. The *h-p* version in time. *Numer. Methods Part. Differ. Equ.* **6** (1990) 343–369.

[8] I. Babuška and M. Suri, The *p* and *h-p* versions of the finite element method, basic principles and properties. *SIAM Rev.* **36** (1994) 578–632.

[9] M. Bause, F.A. Radu and U. Köcher, Error analysis for discretizations of parabolic problems using continuous finite elements in time and mixed finite elements in space. *Numer. Math.* **137** (2017) 773–818.

[10] J. Becker, A second order backward difference method with variable steps for a parabolic problem. *BIT* **38** (1998) 644–662.

[11] D. Braess, Finite Elements: Theory, Fast Solvers and Applications in Solid Mechanics, 3rd edition. Cambridge University Press, Cambridge (2007).

[12] H. Brunner and D. Schötzau, *hp*-discontinuous Galerkin time-stepping for Volterra integrodifferential equations. *SIAM J. Numer. Anal.* **44** (2006) 224–245.

[13] G. Buffoni, A. Griffa and E. Zambianchi, Modelling of dispersion processes in a tide-forced flow. *Hydrobiologia* **393** (1999) 19–24.

[14] Z.M. Chen and Y. Liu, Efficient and parallel solution of high-order continuous time Galerkin for dissipative and wave propagation problems. *SIAM J. Sci. Comput.* **46** (2024) A2073–A2100.

[15] M. Delfour, W. Hager and F. Trochu, Discontinuous Galerkin methods for ordinary differential equations. *Math. Comput.* **36** (1981) 455–473.

[16] J. Douglas and T. Dupont, Galerkin methods for parabolic equations. *SIAM J. Numer. Anal.* **7** (1970) 575–626.

[17] Y.H. Du ang Z.M. Guo, The Stefan problem for the Fisher-KPP equation. *J. Differ. Equ.* **253** (2012) 996–1035.

[18] K. Eriksson, C. Johnson and V. Thomée, Time discretization of parabolic problems by the discontinuous Galerkin method. *RAIRO Modél. Math. Anal. Numér.* **19** (1985) 611–643.

[19] A. Ern and J.-L. Guermond, Finite Elements III: First-Order and Time-Dependent PDEs. Springer, Cham (2021).

[20] A. Ern and M. Vohralík, A posteriori error estimation based on potential and flux reconstruction for the heat equation. *SIAM J. Numer. Anal.* **48** (2010) 198–223.

[21] L. Evans, Partial Differential Equations. American Mathematical Society, Providence, RI (2010).

[22] F. Fakhar-Izadi and M. Dehghan, A spectral element method using the modal basis and its application in solving second-order nonlinear partial differential equations. *Math. Methods Appl. Sci.* **38** (2015) 478–504.

[23] E.H. Georgoulis, O. Lakkis and J.M. Virtanen, A posteriori error control for discontinuous Galerkin methods for parabolic problems. *SIAM J. Numer. Anal.* **49** (2011) 427–458.

[24] E.H. Georgoulis, O. Lakkis and T.P. Wihler, A posteriori error bounds for fully-discrete *hp*-discontinuous Galerkin time-stepping methods for parabolic problems. *Numer. Math.* **148** (2021) 363–368.

[25] B.L. Hulme, One-step piecewise polynomial Galerkin methods for initial value problems. *Math. Comput.* **26** (1972) 415–426.

[26] B.L. Hulme, Discrete Galerkin and related one-step methods for ordinary differential equations. *Math. Comput.* **26** (1972) 881–891.

[27] S. Hussain, F. Schieweck and S. Turek, Higher order Galerkin time discretizations and fast multigrid solvers for the heat equation. *J. Numer. Math.* **19** (2011) 41–61.

[28] S. Hussain, F. Schieweck and S. Turek, A note on accurate and efficient higher order Galerkin time stepping schemes for the nonstationary Stokes equations. *Open Numer. Methods J.* **4** (2012) 35–45.

[29] A. Ilyin, A. Laptev, M. Loss and S. Zelik, One-dimensional interpolation inequalities, Carlson–Landau inequalities, and magnetic Schrödinger operators. *Int. Math. Res. Notes* **2016** (2016) 1190–1222.

[30] J. Isenberg and C. Gutfinger, Heat transfer to a draining film. *Int. J. Heat Transfer* **16** (1972) 505–512.

[31] P. Jamet, Galerkin-type approximations which are discontinuous in time for parabolic equations in a variable domain. *SIAM J. Numer. Anal.* **15** (1978) 91–928.

[32] B. Janssen and T.P. Wihler, Computational comparison of continuous and discontinuous Galerkin time-stepping methods for nonlinear initial value problems. *Lect. Notes Comput. Sci. Eng.* **106** (2015) 103–114.

[33] L.S. Kimpton, J.P. Whiteley, S.L. Waters and J.M. Oliver, Approaches to myosin modelling in a two-phase flow model for cell motility. *Phys. D* **318/319** (2016) 34–49.

[34] O. Lakkis and C. Makridakis, Elliptic reconstruction and a posteriori error estimates for fully discrete linear parabolic problems. *Math. Comput.* **75** (2006) 1627–1658.

[35] H. Li, C.Y. Du and Z.H. Zhao, Continuous space-time finite element method for a reaction-diffusion equation. *Math. Numer. Sin.* **39** (2017) 167–178.

[36] C. Makridakis and I. Babuška, On the stability of the discontinuous Galerkin method for the heat equation. *SIAM J. Numer. Anal.* **34** (1997) 389–401.

[37] H. Matsuzawa, A free boundary problem for the fisher-KPP equation with a given moving boundary. *Commun. Pure Appl. Anal.* **17** (2018) 1821–1852.

[38] S. Metcalfe and T.P. Wihler, Conditional a posteriori error bounds for high order discontinuous Galerkin time stepping approximations of semilinear heat models with blow-up. *SIAM J. Sci. Comput.* **44** (2022) A1337–A1357.

[39] A. Miranville and R. Quintanilla, A generalization of the Allen–Cahn equation. *IMA J. Appl. Math.* **80** (2015) 410–430.

[40] S. Nicaise and N. Soualem, A posteriori error estimates for a nonconforming finite element discretization of the heat equation. *ESAIM: Math. Model. Numer. Anal.* **39** (2005) 319–348.

[41] E.M. Rønquist and A.T. Patera, A Legendre spectral element method for the Stefan problem. *Int. J. Numer. Methods Eng.* **24** (1987) 2273–2299.

[42] N. Saito, Variational analysis of the discontinuous Galerkin time-stepping method for parabolic equations. *IMA J. Numer. Anal.* **41** (2021) 1267–1292.

[43] J.R. Salmon, J.A. Liggett and R.H. Gallager, Dispersion analysis in homogeneous lakes. *Int. J. Numer. Methods Eng.* **15** (1980) 1627–1642.

[44] F. Schieweck, A-stable discontinuous Galerkin–Petrov time discretization of higher order. *J. Numer. Math.* **18** (2010) 25–57.

[45] D. Schötzau and C. Schwab, An *hp* a priori error analysis of the DG time-stepping method for initial value problems. *Calcolo* **37** (2000) 207–232.

[46] D. Schötzau and C. Schwab, Time discretization of parabolic problems by the *hp*-version of the discontinuous Galerkin finite element method. *SIAM J. Numer. Anal.* **38** (2000) 837–875.

[47] D. Schötzau and T.P. Wihler, A posteriori error estimation for *hp*-version time-stepping methods for parabolic partial differential equations. *Numer. Math.* **115** (2010) 475–509.

[48] C. Schwab, *p*-and *hp*-Finite Element Methods. Oxford University Press, New York (1998).

[49] V. Thomée, Galerkin Finite Element Methods for Parabolic Problems. Vol. 25 of *Springer Series in Computational Mathematics*, 2nd edition. Springer-Verlag, Berlin (2006).

[50] L.N. Wang, M.Z. Zhang, H.J. Tian and L.J. Yi, *hp*-version C^1-continuous Petrov–Galerkin method for nonlinear second-order initial value problems with application to wave equations. *IMA J. Numer. Anal.* **45** (2025) 1455–1500.

[51] Y.C. Wei and L.J. Yi, An *hp*-version of the C^0-continuous Galerkin time-stepping method for nonlinear second-order initial value problems. *Adv. Comput. Math.* **46** (2020) 56.

[52] T. Werder, K. Gerdes, D. Schötzau and C. Schwab, *hp*-discontinuous Galerkin time stepping for parabolic problems. *Comput. Methods Appl. Mech. Eng.* **190** (2001) 6685–6708.

[53] T.P. Wihler, An a priori error analysis of the *hp*-version of the continuous Galerkin FEM for nonlinear initial value problems. *J. Sci. Comput.* **25** (2005) 523–549.

[54] L.J. Yi, An L^∞-error estimate for the *h-p* version continuous Petrov–Galerkin method for nonlinear initial value problems. *East Asian J. Appl. Math.* **5** (2015) 301–311.

[55] L.J. Yi and B.Q. Guo, An *h-p* version of the continuous Petrov–Galerkin finite element method for Volterra integro-differential equations with smooth and nonsmooth kernels. *SIAM J. Numer. Anal.* **53** (2015) 2677–2704.

[56] J. Zhu and J. X. Qiu, Local DG method using WENO type limiters for convection-diffusion problems. *J. Comput. Phys.* **230** (2011) 4353–4375.

[57] Z. Zlatev, R. Berkowicz and L.P. Prahm, Implementation of a variable stepsize variable formula in the time-integration part of a code for treatment of long-range transport of air pollutants. *J. Comput. Phys.* **55** (1984) 278–301.