ESAIM: M2AN 59 (2025) 749–787
https://doi.org/10.1051/m2an/2025006

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

A NEW ERROR ANALYSIS FOR PARABOLIC DIRICHLET BOUNDARY CONTROL PROBLEMS

Dongdong Liang[1,2], Wei Gong[3,*] and Xiaoping Xie[2]

Abstract. This paper investigates the finite element approximation of a parabolic Dirichlet boundary control problem, presenting a new *a priori* error estimate. We establish two main convergence results for both semi-discrete and fully discrete optimal control problems, under suitable assumptions. Specifically, we demonstrate convergence orders of $O(k^{\frac{1}{4}})$ and $O(k^{\frac{3}{4}-\varepsilon})$ ($\forall\varepsilon > 0$) for the temporal semi-discretization of control problems on polytopes and smooth domains, respectively. For control problems defined on polyhedra, we achieve a convergence rate of $O(k^{\frac{1}{4}}+h^{\frac{1}{2}})$ in the fully discrete setting. The contributions of this work are twofold. First, we provide an improved temporal convergence rate for parabolic Dirichlet boundary control problems on smooth domains, setting a foundation for further fully discrete error analysis. Second, we refine the existing fully discrete error estimate for boundary control problems on polyhedra by removing the artificial mesh size restriction $k = O(h^2)$. As an intermediate but essential result, we establish both the convergence order and stability of the finite element approximation for parabolic inhomogeneous boundary value problems. Importantly, these results hold under low regularity boundary conditions without imposing mesh size constraints.

Mathematics Subject Classification. 49J20, 65N15, 65N30.

Received January 2, 2024. Accepted January 20, 2025.

1. Introduction

Consider a bounded and convex polytopal, or a bounded smooth domain $\Omega \subseteq \mathbb{R}^N$ (where $N = 2, 3$) with boundary $\Gamma := \partial\Omega$. Let $T > 0$ be a constant, and denote $I := (0, T)$. This article focuses on the investigation of the following parabolic Dirichlet boundary optimal control problem:

$$\min_{u\in U_{ad},\ y\in L^2(I;L^2(\Omega))} \quad J(y,u) = \frac{1}{2}\|y - y_d\|^2_{L^2(I;L^2(\Omega))} + \frac{\alpha}{2}\|u\|^2_{L^2(I;L^2(\Gamma))}. \tag{1.1}$$

Keywords and phrases. Parabolic Dirichlet boundary control, smooth domains, convex polytopes, semi-discrete, fully discrete, finite element, error estimate.

[1] The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen 518057, P.R. China.

[2] School of Mathematics, Sichuan University, Chengdu 610064, P.R. China.

[3] NCMIS & LSEC, Institute of Computational Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, P.R. China.

*Corresponding author: wgong@lsec.cc.ac.cn

Here, $\alpha > 0$ represents the regularization parameter, $y_d \in L^2(I; L^2(\Omega))$ is a given target state, and U_{ad} denotes the admissible set of controls, which is of the following box type:

$$U_{ad} = \{u \in L^2(I; L^2(\Gamma)) : \ u_a \le u(t,x) \le u_b, \text{ a.e., } (t,x) \in \Sigma_T\},$$

with u_a, $u_b \in \mathbb{R} \cup \{\infty\}$ satisfying $u_a < u_b$. The state variable y and the control variable u in problem (1.1) are constrained by the following parabolic equation:

$$\begin{cases} \partial_t y - \Delta y = f & \text{in } \Omega_T := I \times \Omega, \\ y = u & \text{on } \Sigma_T := I \times \Gamma, \\ y(0) = y_0 & \text{in } \Omega, \end{cases} \tag{1.2}$$

where $f \in L^2(I; L^2(\Omega))$ and $y_0 \in L^2(\Omega)$ are given. The solution pair $(\bar{u}, \bar{y})$ with $\bar{u} \in U_{ad}$ minimizing the cost functional J is called the optimal pair of the optimal control problem (1.1).

Although Neumann controls are considered in real life applications most of the time, Dirichlet boundary control represents a crucial category of control problems with significant applications, particularly in fluid dynamics. Notably, it plays a pivotal role in practical scenarios such as active boundary control for fluid flows [19, 24, 26]. In this context, an example involves adjusting the velocity of fluid flowing over the surface of a cylinder by manipulating the angular velocity along its axes. The primary objective of such controls is often to mitigate vortex shedding or delay the onset of turbulence. It is noteworthy that controls with low regularity are permissible in these applications. Examples include controls involving the injection of fluids, the blowing, suction or mixing on a portion of the boundary, taking the form of Dirichlet boundary controls, which may exhibit discontinuities and adhere to pointwise constraints.

In recent decades, there has been extensive research on Dirichlet boundary control problems. Notable contributions include studies on elliptic Dirichlet boundary controls [2,9,10,15,18], boundary controls for Navier-Stokes equations [19,26], and parabolic Dirichlet boundary control problems [3,29,30,32,34]. One key aspect that distinguishes studies in Dirichlet boundary control problems is the choice of the control space. The $L^2(\Gamma)$-control space (referenced in [2,9,15,18]) is widely used due to its simplicity in implementation. However, it often results in solutions with lower regularity. On the other hand, the energy space method (as discussed in [40]) produces smoother solutions but involves a more intricate implementation. Regarding the parabolic Dirichlet boundary control problem (1.1), when formulated with the control space $L^2(I; L^2(\Gamma))$, the state equation needs to be interpreted in a very weak sense (as outlined in [5]). However, this choice often leads to less regular optimal controls [3,29,30,32,34], such as admitting only $H^{\frac{1}{2}}$-regularity in space and $H^{\frac{1}{4}}$-regularity in time on polytopes. To address this challenge, a proposed solution involves a Robin penalization approach, as introduced in [4].

When it comes to discretizations and error estimates for the Dirichlet boundary control problem, the finite element method has been widely employed, with numerous contributions focusing on elliptic control problems (see, for instance, [2,9,10,15,18,37]). However, limited attention has been given to parabolic Dirichlet boundary control problems, with only a handful of works available, including [21,23,29]. In [29], a fully discretized optimal control problem was tackled using the discontinuous Galerkin in time and continuous Galerkin in space (abbreviated as DG(r)-CG(s) with r and s respectively the polynomial orders of the temporal and spatial discretizations) method, and a semi-smooth Newton method was applied to solve the optimization problem. Another approach, presented in [23], utilized the DG(0)-CG(1) method in conjunction with variational discretization to derive the discrete optimal control problem in two dimensions. The study achieved convergence orders of $O(h^{\frac{1}{2}})$ in space and $O(k^{\frac{1}{4}})$ in time, under the condition $k = O(h^2)$, where k and h denote the time step and mesh size, respectively. In [21], an enhanced convergence order of $O(h^{1-\frac{1}{q}-r})$ was established for spatial semi-discretization. Here, $q > 0$ depends on the maximal interior angle of the corners, and $r > 0$ represents an arbitrarily small constant.

In this paper, we explore the temporal semi-discretization and time-space full discretization of the parabolic Dirichlet boundary control problem (1.1) and derive corresponding error estimates, which serves as the first main contribution of this work. For the temporal semi-discretization, we employ the DG(0) method for the state and variational discretization for the control. Under the conditions $f \in L^2(I; L^2(\Omega))$, $y_0 \in H^{2s-1}(\Omega)$, and

$y_d \in L^2(I; H^{2s'-\frac{1}{2}}(\Omega)) \cap H^{s'-\frac{1}{4}}(I; L^2(\Omega))$, with $s \in [\frac{1}{2}, \frac{3}{4})$ and $s' \in [s - \frac{1}{4}, s]$, in the case of a smooth domain, or $s = \frac{1}{2}$ and $s' = \frac{1}{4}$ for a polytope Ω, we establish the error estimate:

$$\|\bar{u} - \bar{u}_k\|_{L^2(I;L^2(\Gamma))} + \|\bar{y} - \bar{y}_k\|_{L^2(I;L^2(\Omega))} \leq C\Big(k^s + k^{s'}\Big). \tag{1.3}$$

Here, $(\bar{u}, \bar{y})$ and $(\bar{u}_k, \bar{y}_k) \in U_{ad} \times X_k$ represent the optimal pair of the optimal control problem (1.1) and the semi-discrete optimal control problem (4.7), respectively, under the non-increasing condition on time steps. For the full discretization, assuming Ω is a polytope, we employ the DG(0)-CG(1) method for the state and variational discretization for the control. This yields a convergence order of $O(h^{\frac{1}{2}})$ for the error between the semi-discrete and fully discrete solutions. Finally, by combining the two error estimates, we obtain the overall error estimate:

$$\|\bar{u} - \bar{u}_{kh}\|_{L^2(I;L^2(\Gamma))} + \|\bar{y} - \bar{y}_{kh}\|_{L^2(I;L^2(\Omega))} \leq C\Big(h^{\frac{1}{2}} + k^{\frac{1}{4}}\Big). \tag{1.4}$$

Here, $(\bar{u}, \bar{y})$ and $(\bar{u}_{kh}, \bar{y}_{kh})$ denote the continuous and fully discrete optimal pairs, respectively.

The key strategies employed to establish the error estimates (1.3) and (1.4) are rooted in their conversion into error estimates for finite element solutions to the state equation and discrete normal derivatives of the adjoint equation, as outlined in Proposition 6.6 and Theorem 6.9. These achievements rely on the stability estimate for finite element solutions to the parabolic equation with inhomogeneous Dirichlet data and the stability estimate for the discrete norm derivative of the adjoint equation. It is important to note that these stability estimates present heightened challenges when dealing with parabolic equations featuring inhomogeneous Dirichlet data and limited regularity, representing a departure from the classical stability results found in [38, 39, 43] and contributing to one of the main novelties of this paper. For the error estimate associated with the adjoint equation with homogeneous Dirichlet data, a notable innovation lies in the introduction of a new equivalent definition for discrete normal derivatives (see Def. 4.6), distinct from the standard Definition 4.7 introduced in [23]. This new definition facilitates the derivation of error estimates for the discrete normal derivatives between fully discrete and temporal semi-discrete adjoint equations without the imposition of the space-time mesh size condition $k = O(h^2)$. Furthermore, the error estimate for these discrete normal derivatives not only provides the stability of the fully discrete norm derivative but, when combined with the known stability of the semi-discrete normal derivative, also implies the stability of the fully discrete solution to the state equation concerning Dirichlet data with low regularity (refer to Prop. 6.8). This stability of finite element solutions serves as a crucial element in the error estimates of the state discretization, playing a pivotal role in proving (1.4).

To conduct the error analysis for finite element discretizations of the Dirichlet boundary control problem (1.1), a crucial and indispensable component is the error estimation for finite element approximations to parabolic equations with rough Dirichlet boundary conditions, as discussed in works such as [17,23,31]. In [31], convergence orders of $O(h^{\frac{1}{2}})$ (for $1 < p < \infty$) and $O(h^{\frac{1}{2}}\big|\ln h\big|)$ were achieved for spatial semi-discretization under the norm $\|\cdot\|_{L^p(I;L^2(\Omega))}$ for $p = 1$ and ∞, respectively. These results were obtained assuming Dirichlet data in $L^p(I; L^2(\Gamma))$. For fully discrete approximations to parabolic equations, convergence orders of $O(k^{\frac{1}{4}} + h^{\frac{1}{2}})$ and $O(k^{\frac{1}{2}} + h)$ were obtained in [17,23] with Dirichlet data in $L^2(I; L^2(\Gamma))$ and $L^2(I; H^{\frac{1}{2}}(\Gamma)) \cap H^{\frac{1}{4}}(I; L^2(\Gamma))$, respectively. However, the mesh size condition $k = O(h^2)$ in two dimensions, arising from the inverse estimate when the solution has low regularity, is restrictive. In this paper, we present improved convergence results without the need for the restrictive condition $k = O(h^2)$, which is of independent interest and gives the second main contribution of this work. Specifically, for given Dirichlet boundary condition $u \in H^{s-\frac{1}{4}}(I; L^2(\Gamma)) \cap L^2(I; H^{2s-\frac{1}{2}}(\Gamma))$ with $s \in [\frac{1}{2}, 1]$, letting $y \in H^s(I; L^2(\Omega)) \cap L^2(I; H^{2s}(\Omega))$ be the exact solution of equation (1.2), we establish the following error estimates:

$$\|y - y_k(u)\|_{L^2(I;L^2(\Omega))} \leq C\Big(k^s(\|y\|_{L^2(I;H^{2s}(\Omega))} + \|y\|_{H^s(I;L^2(\Omega))}) + k\|f\|_{L^2(I;L^2(\Omega))}\Big) \tag{1.5}$$

and

$$\|y - y_{kh}(u)\|_{L^2(I;L^2(\Omega))} \leq C\Big((h^{2s} + k^s)(\|y\|_{L^2(I;H^{2s}(\Omega))} + \|y\|_{H^s(I;L^2(\Omega))}) + k\|f\|_{L^2(I;L^2(\Omega))}\Big), \tag{1.6}$$

where $y_k(u)$ is the DG(0) semi-discretization of parabolic equations on smooth or convex polyhedral domains, and $y_{kh}(u)$ is the DG(0)-CG(1) discretization of parabolic equations on convex polytopes. Importantly, these analyses do not require the mesh size condition $k = O(h^2)$ and constitute the second novelty of this work compared to existing results. To derive these error estimates, we introduce a novel and more involved error analysis technique, especially suited for the challenges posed by inhomogeneous Dirichlet data and low regularity solutions. Specifically, we construct discrete error equations by homogenizing the Dirichlet data on the discrete level, as detailed in Section 5 for a comprehensive understanding.

The remainder of this paper is structured as follows. In Section 2 we provide essential preliminaries, including the definition of very weak solutions and the establishment of the well-posedness of parabolic equations. In Section 3 we derive the first-order optimality condition and explore the regularity of the optimal pair concerning the parabolic Dirichlet boundary control problem. Section 4 is dedicated to presenting the DG(0) semi-discretization and DG(0)-CG(1) full discretization methods for parabolic boundary control problems. Additionally, we introduce the corresponding discrete first-order optimality conditions. The focus of Section 5 is on the error analysis of both temporal discretization and spatial discretization for inhomogeneous parabolic boundary value problems. In the final section, we deduce error estimates for both the semi-discrete and fully discrete optimal control problems.

2. Preliminaries

Throughout this paper, we adopt standard notations for differential operators, function spaces, and norms, as commonly found in references such as [7, 36, 43]. We assume that $C > 0$ is a constant that remains independent of h, k, and the given data.

Recall that in the optimal control problem (1.1), we assume that

$$f \in L^2(I; L^2(\Omega)), \quad u \in L^2(I; L^2(\Gamma)), \quad y_0 \in L^2(\Omega), \tag{2.1}$$

the solution of the state equation (1.2) must be defined using the method of transposition, as outlined in [5, 36]. This type of solution is referred to as a very weak solution, and its definition is provided below.

Definition 2.1. For any given f, u and y_0 satisfying (2.1), a function $y \in L^2(I; L^2(\Omega))$ is called the very weak solution of (1.2), if y satisfies the identity

$$\int_{\Omega_T} y(-\partial_t z - \Delta z)\,\mathrm{d}x\,\mathrm{d}t = -\int_{\Sigma_T} u\partial_n z\,\mathrm{d}s\,\mathrm{d}t + \int_{\Omega_T} fz\,\mathrm{d}x\,\mathrm{d}t + \int_{\Omega} y_0 z(\cdot,0)\,\mathrm{d}x \tag{2.2}$$

for any $z \in L^2(I; H^2(\Omega) \cap H_0^1(\Omega)) \cap H^1(I; L^2(\Omega))$ with $z(\cdot, T) = 0$. Equivalently, for all $g \in L^2(I; L^2(\Omega))$ and z_g denoting the solution of

$$\begin{cases} -\partial_t z_g - \Delta z_g = g & \text{in } (0,T)\times\Omega, \\ z_g = 0 & \text{on } (0,T)\times\Gamma, \\ z_g(T) = 0 & \text{in } \Omega, \end{cases} \tag{2.3}$$

a function $y \in L^2(I; L^2(\Omega))$ is called the very weak solution of (1.2) if y satisfies the identity

$$\int_{\Omega_T} yg\,\mathrm{d}x\,\mathrm{d}t = -\int_{\Sigma_T} u\partial_n z_g\,\mathrm{d}s\,\mathrm{d}t + \int_{\Omega_T} fz_g\,\mathrm{d}x\,\mathrm{d}t + \int_{\Omega} y_0 z_g(\cdot,0)\,\mathrm{d}x \qquad \forall g \in L^2\big(I; L^2(\Omega)\big), \tag{2.4}$$

where $\partial_n z_g$ denotes the outward normal derivative of z_g.

Remark 2.2. By the well-posedness of parabolic equations (cf. [16], Chap. 7), we ascertain that $z_g \in L^2(I; H^2(\Omega) \cap H_0^1(\Omega)) \cap H^1(I; L^2(\Omega))$. Consequently, by the trace theorem (cf. [36], Chap. 4 and [9]), it follows that $\partial_n z_g \in L^2(I; H^{\frac{1}{2}}(\Gamma))$. Employing the embedding relation

$$L^2\big(I; H^2(\Omega) \cap H_0^1(\Omega)\big) \cap H^1\big(I; L^2(\Omega)\big) \hookrightarrow C\big(\bar{I}; H^1(\Omega)\big),$$

we have $z_g \in C(\bar{I}; L^2(\Omega))$. This establishes the well-defined nature of Definition 2.1.

The existence and uniqueness of a very weak solution of equation (1.2) is verified in the following lemma; see also [22, 23].

Lemma 2.3. *For any given y_0, u and f satisfying (2.1), there exists a unique very weak solution $y \in L^2(I; L^2(\Omega))$ of equation (1.2). Moreover, there holds*

$$\|y\|_{L^2(I;L^2(\Omega))} \le C\big(\|y_0\|_{L^2(\Omega)} + \|u\|_{L^2(I;L^2(\Gamma))} + \|f\|_{L^2(I;L^2(\Omega))}\big).$$

Proof. For any given $g \in L^2(I; L^2(\Omega))$ there exists a unique solution $z_g \in L^2(I; H^2(\Omega) \cap H^1_0(\Omega)) \cap H^1(I; L^2(\Omega))$ to the equation (2.3) such that the following estimate holds

$$\|z_g\|_{C(\bar{I};L^2(\Omega))} + \|z_g\|_{L^2(I;H^2(\Omega)\cap H^1_0(\Omega))} + \|z_g\|_{H^1(I;L^2(\Omega))} \le C\|g\|_{L^2(I;L^2(\Omega))}, \tag{2.5}$$

where $C > 0$ is a constant (cf. [16]).

Define the linear functional $\mathscr{L}$ on $L^2(I; L^2(\Omega))$ as follows

$$\mathscr{L}(g) := -\int_{\Sigma_T} u\frac{\partial z_g}{\partial n}\,\mathrm{d}s\,\mathrm{d}t + \int_{\Omega_T} f z_g\,\mathrm{d}x\,\mathrm{d}t + \int_\Omega y_0 z_g(\cdot, 0)\,\mathrm{d}x \qquad \forall g \in L^2(I; L^2(\Omega)).$$

By the trace theorem and (2.5) there holds

$$\left\|\frac{\partial z_g}{\partial n}\right\|_{L^2(I;L^2(\Gamma))} \le C\|g\|_{L^2(I;L^2(\Omega))}. \tag{2.6}$$

By using the estimates (2.5), (2.6) and the Schwarz inequality, we know that the functional $\mathscr{L}$ is bounded. Then by the Riesz representation theorem there exists a unique $y \in L^2(I; L^2(\Omega))$ such that

$$\int_{\Omega_T} yg\,\mathrm{d}x\,\mathrm{d}t = \mathscr{L}(g) \qquad \forall g \in L^2(I; L^2(\Omega)). \tag{2.7}$$

In other words, the function y satisfies the identity (2.4) in Definition 2.1. Therefore, y is the unique very weak solution of equation (1.2) while the estimate for y is a direct consequence of (2.5). This completes the proof. □

Furthermore, if the data for equation (1.2) exhibit higher regularity, we can anticipate an enhanced regularity of the solution, which is also influenced by the smoothness of Ω.

Lemma 2.4. *Suppose that Ω is a bounded smooth domain. For any given*

$$f \in L^2\big(I; L^2(\Omega)\big),\ u \in H^{s-\frac{1}{4}}\big(I; L^2(\Gamma)\big) \cap L^2\Big(I; H^{2s-\frac{1}{2}}(\Gamma)\Big) \text{ and } y_0 \in H^{2s-1}(\Omega), \qquad \frac{1}{4} \le s < \frac{3}{4},$$

the solution y of equation (1.2) belongs to $L^2(I; H^{2s}(\Omega)) \cap H^s(I; L^2(\Omega))$ and satisfies ([36], Page 78):

$$\|y\|_{L^2(I;H^{2s}(\Omega))} + \|y\|_{H^s(I;L^2(\Omega))} \le C\Big(\|f\|_{L^2(I;L^2(\Omega))} + \|y_0\|_{H^{2s-1}(\Omega)} + \|u\|_{H^{s-\frac{1}{4}}(I;L^2(\Gamma))} + \|u\|_{L^2(I;H^{2s-\frac{1}{2}}(\Gamma))}\Big).$$

Specifically, if $s = \frac{1}{2}$ there holds $y \in L^2(I; H^1(\Omega)) \cap H^1(I; H^{-1}(\Omega)) \hookrightarrow C(\bar{I}; L^2(\Omega))$ such that (cf. [36], Page 84 and [35], Page 19):

$$\begin{aligned}\|y\|_{L^2(I;H^1(\Omega))} &+ \|y\|_{C(\bar{I};L^2(\Omega))} + \|\partial_t y\|_{L^2(I;H^{-1}(\Omega))}\\ &\le C\Big(\|f\|_{L^2(I;L^2(\Omega))} + \|y_0\|_{L^2(\Omega)} + \|u\|_{H^{\frac{1}{4}}(I;L^2(\Gamma))} + \|u\|_{L^2(I;H^{\frac{1}{2}}(\Gamma))}\Big).\end{aligned}$$

The regularity of the solution to (2.3) can be further improved with sufficient smoothness in the data and the domain Ω. In the case where Ω is a convex polytope, the weak solution of equation (2.3) belongs to $H^1(I;L^2(\Omega))\cap L^2(I;H_0^1(\Omega)\cap H^2(\Omega))$ for any $g\in L^2(I;L^2(\Omega))$. Additionally, when Ω has a smooth boundary, the regularity of the solution to (2.3) is presented in the following lemma [16, 36].

Lemma 2.5. *Consider a smooth bounded domain Ω. For any given $0\le r<\frac{1}{2}$ and function $g\in L^2(I;H^{2r}(\Omega))\cap H^r(I;L^2(\Omega))$, the solution z of equation (2.3) belongs to $L^2(I;H^{2+2r}(\Omega)\cap H_0^1(\Omega))\cap H^{1+r}(I;L^2(\Omega))$, and the following estimate holds:*

$$\|z\|_{L^2(I;H^{2+2r}(\Omega))}+\|z\|_{H^{1+r}(I;L^2(\Omega))}\le C\big(\|g\|_{L^2(I;H^{2r}(\Omega))}+\|g\|_{H^r(I;L^2(\Omega))}\big).$$

If, in addition, $g\in L^2(I;H^1(\Omega))\cap H^{\frac{1}{2}}(I;L^2(\Omega))$ and

$$(T-t)^{-\frac{1}{2}}g\in L^2\big(I;L^2(\Omega)\big), \tag{2.8}$$

then $z\in L^2(I;H^3(\Omega)\cap H_0^1(\Omega))\cap H^{\frac{3}{2}}(I;L^2(\Omega))$ (cf. [36], Eq. (5.12)) and $\partial_t z\in H^{\frac{1}{2}}(I;L^2(\Omega))\cap L^2(I;H^1(\Omega))$ (cf. [36], Page 12). Moreover, there holds the estimate

$$\|z\|_{L^2(I;H^3(\Omega))}+\|z\|_{H^{\frac{3}{2}}(I;L^2(\Omega))}+\|\partial_t z\|_{L^2(I;H^1(\Omega))}\le C\Big(\|g\|_{L^2(I;H^1(\Omega))}+\|g\|_{H^{\frac{1}{2}}(I;L^2(\Omega))}\Big). \tag{2.9}$$

The aforementioned regularity results will be employed to investigate the regularity of solutions to the optimal control problem (1.1). In particular, Lemma 2.4 will be utilized to deduce the regularity of state variables, while Lemma 2.5 will be instrumental in enhancing the regularity of control variables.

3. The optimal control problem

For any $u\in L^2(I;L^2(\Gamma))$, let $y(u)$ denote the unique solution of the state equation (1.2), as defined in Definition 2.1. We can establish the following affine linear operator

$$S: L^2\big(I;L^2(\Gamma)\big)\to L^2\big(I;L^2(\Omega)\big),\qquad Su:=y(u), \tag{3.1}$$

which is bounded and Fréchet differentiable from $L^2(I;L^2(\Gamma))$ to $L^2(I;L^2(\Omega))$. In fact, the operator S can be decomposed as $Su=\hat{S}u+\hat{y}$ for any $u\in L^2(I;L^2(\Gamma))$, where $\hat{S}u$ solves the state equation with inhomogeneous Dirichlet data u and $f=0$, $y_0=0$, $\hat{y}$ solves the state equation with homogeneous Dirichlet data $u=0$ and inhomogeneous f, y_0. Obviously, the operator $\hat{S}$ is linear and bounded from $L^2(I;L^2(\Gamma))$ to $L^2(I;L^2(\Omega))$ and $\hat{y}$ is independent of u. Consequently, the optimal control problem (1.1) can be equivalently expressed as follows:

$$\min_{u\in U_{ad}}\hat{J}(u):=J(Su,u). \tag{3.2}$$

It is straightforward to verify that the above optimization problem admits a unique solution $\bar{u}$ (cf. [34], Chap. III or [27], Chap. 1). Denoting by $(\bar{u},\bar{y})$ the optimal pair, where $\bar{y}$ is the associated state.

For any $u\in L^2(I;L^2(\Gamma))$, the Fréchet derivative of the cost functional at u reads

$$\hat{J}'(u)v=\int_{\Sigma_T}\alpha uv\,\mathrm{d}s\,\mathrm{d}t+\int_{\Omega_T}(y-y_d)\tilde{y}(v)\,\mathrm{d}x\,\mathrm{d}t\qquad\forall v\in L^2\big(I;L^2(\Gamma)\big),$$

where $\tilde{y}(v)\in L^2(I;L^2(\Omega))$ is the very weak solution of the following equation:

$$\begin{cases}\partial_t\tilde{y}(v)-\Delta\tilde{y}(v)=0 & \text{in } (0,T)\times\Omega,\\ \tilde{y}(v)=v & \text{on } (0,T)\times\Gamma,\\ \tilde{y}(v)(\cdot,0)=0 & \text{in } \Omega.\end{cases}$$

The first order sufficient and necessary optimality condition is given in the following theorem.

Theorem 3.1. *The pair $(\bar u,\bar y)\in U_{ad}\times L^2(I;L^2(\Omega))$ is the optimal solution of the optimal control problem (1.1) if and only if $\bar y:=S\bar u=y(\bar u)$ and*

$$\hat J'(\bar u)(v-\bar u)\geq 0 \qquad \forall v\in U_{ad}. \tag{3.3}$$

Furthermore, (3.3) is equivalent to

$$\hat J'(\bar u)(v-\bar u)=\int_{\Sigma_T}(\alpha\bar u-\partial_n\bar z)(v-\bar u)\,\mathrm{d}s\,\mathrm{d}t\geq 0 \qquad \forall v\in U_{ad}, \tag{3.4}$$

where $\bar z\in L^2(I;H^2(\Omega)\cap H_0^1(\Omega))\cap H^1(I;L^2(\Omega))$ is the adjoint variable satisfying

$$\begin{cases}-\partial_t\bar z-\Delta\bar z=\bar y-y_d & \text{in } (0,T)\times\Omega,\\ \bar z=0 & \text{on } (0,T)\times\Gamma,\\ \bar z(T)=0 & \text{in } \Omega\,.\end{cases} \tag{3.5}$$

The optimality condition (3.4) can be expressed equivalently as:

$$\bar u=P_{U_{ad}}\left(\frac{1}{\alpha}\partial_n\bar z\right), \tag{3.6}$$

where $P_{U_{ad}}$ represents the orthogonal projection operator from $L^2(I;L^2(\Gamma))$ onto the admissible set U_{ad} that preserves H^s-regularity for $s\leq 1$ (cf. [44], Page 114). The above relationship between the optimal control and adjoint state allows for an enhancement in the regularity of solutions to the optimal control problem.

Theorem 3.2. *Assume that Ω is a bounded smooth domain and let $(\bar u,\bar y,\bar z)$ be the optimal solution of the optimal control problem (1.1). Then for any given $y_0\in H^{2s-1}(\Omega)$ $(\frac12\leq s<\frac34)$, $y_d\in L^2(I;H^{2s'-\frac12}(\Omega))\cap H^{s'-\frac14}(I;L^2(\Omega))$ $(s-\frac14\leq s'\leq s)$ and $f\in L^2(I;L^2(\Omega))$, there hold*

$$\begin{aligned}&\bar y\in L^2\big(I;H^{2s}(\Omega)\big)\cap H^s\big(I;L^2(\Omega)\big),\quad \bar u\in L^2\Big(I;H^{\min\{2s',1\}}(\Gamma)\Big)\cap H^{s'}\big(I;L^2(\Gamma)\big),\\ &\bar z\in L^2\Big(I;H^{2s'+\frac32}(\Omega)\cap H_0^1(\Omega)\Big)\cap H^{s'+\frac34}\big(I;L^2(\Omega)\big),\\ &\partial_t\bar z\in L^2\Big(I;H^{2(s'-\frac14)}(\Omega)\Big)\cap H^{s'-\frac14}\big(I;L^2(\Omega)\big).\end{aligned}$$

If, in addition, there holds the compatibility condition (2.8) with $g:=\bar y-y_d$ in the case $s'=\frac34$, then the adjoint state has the following improved regularity:

$$\bar z\in L^2\big(I;H^3(\Omega)\cap H_0^1(\Omega)\big)\cap H^{\frac32}\big(I;L^2(\Omega)\big),\quad \partial_t\bar z\in H^{\frac12}\big(I;L^2(\Omega)\big)\cap L^2\big(I;H^1(\Omega)\big).$$

Proof. For $\bar u\in L^2(I;L^2(\Gamma))$, $f\in L^2(I;L^2(\Omega))$ and $y_0\in L^2(\Omega)$, it follows from Lemma 2.3 that $\bar y\in L^2(I;L^2(\Omega))$. Choosing $r=0$ in Lemma 2.5 and combining with $y_d\in L^2(I;L^2(\Omega))$, we have $\bar z\in L^2(I;H^2(\Omega)\cap H_0^1(\Omega))\cap H^1(I;L^2(\Omega))$. Then $\bar u\in L^2(I;H^{\frac12}(\Gamma))\cap H^{\frac14}(I;L^2(\Gamma))$ by applying the identity (3.6) and the trace theorem (cf. [29]).

Next, choosing $s=\frac12$ in Lemma 2.4 and using the fact $y_0\in L^2(\Omega)$, $f\in L^2(I;L^2(\Omega))$, we obtain $\bar y\in L^2(I;H^1(\Omega))\cap H^{\frac12}(I;L^2(\Omega))$. That is, $\bar y\in L^2(I;H^{2s-1}(\Omega))\cap H^{s-\frac12}(I;L^2(\Omega))$ for $\frac12\leq s<\frac34$. Then, it follows $\bar z\in L^2(I;H^{2s+1}(\Omega)\cap H_0^1(\Omega))\cap H^{s+\frac12}(I;L^2(\Omega))$ from Lemma 2.5 and the fact $\bar y-y_d\in L^2(I;H^{2s-1}(\Omega))\cap H^{s-\frac12}(I;L^2(\Omega))$ for $y_d\in L^2(I;H^{2s'-\frac12}(\Omega))\cap H^{s'-\frac14}(I;L^2(\Omega))$ $\big(s-\frac14\leq s'\big)$, which implies that $\bar u\in L^2(I;H^{2s-\frac12}(\Gamma))\cap H^{s-\frac14}(I;L^2(\Gamma))$ by the identity (3.6) and the trace theorem (cf. [29]). Further, observing that $y_0\in H^{2s-1}(\Omega)$ and $f\in L^2(I;L^2(\Omega))$, then applying Lemma 2.4 to the state equation (1.2) implies that $\bar y\in L^2(I;H^{2s}(\Omega))\cap H^s(I;L^2(\Omega))$. Then, one derives $\bar z\in L^2(I;H^{2s'+\frac32}(\Omega)\cap H_0^1(\Omega))\cap H^{s'+\frac34}(I;L^2(\Omega))$ from Lemma 2.5 and the fact $\bar y-y_d\in L^2(I;H^{2s'-\frac12}(\Omega))\cap H^{s'-\frac14}(I;L^2(\Omega))$ for $y_d\in L^2(I;H^{2s'-\frac12}(\Omega))\cap H^{s'-\frac14}(I;L^2(\Omega))$ $(s'\leq s)$, which implies that $\bar u\in L^2(I;H^{\min(2s',1)}(\Gamma))\cap H^{s'}(I;L^2(\Gamma))$ by (3.6). At last, $\partial_t\bar z\in L^2(I;H^{2(s'-\frac14)}(\Omega))\cap H^{s'-\frac14}(I;L^2(\Omega))$ can be derived from [36], Page 12. This completes the proof. □

The regularity of solutions to the state and adjoint equations is strongly influenced by the smoothness of Ω. As evident from the above theorem, solutions to optimal control problems in smooth domains exhibit higher regularity when the data is smooth. However, this observation does not hold true for polytopal domains. The following theorem, extracted from [21,23], elucidates the regularity of the optimal control problem in polytopes.

Theorem 3.3. *Let Ω be a convex polytope and $(\bar{u},\bar{y},\bar{z})$ be the solution of the optimal control problem (1.1). Then for any given y_d, $f \in L^2(I;L^2(\Omega))$ and $y_0 \in L^2(\Omega)$ there hold*

$$\begin{aligned}&\bar{y} \in L^2\big(I;H^1(\Omega)\big)\cap H^{\frac{1}{2}}\big(I;L^2(\Omega)\big), \quad \bar{u} \in L^2\Big(I;H^{\frac{1}{2}}(\Gamma)\Big)\cap H^{\frac{1}{4}}\big(I;L^2(\Gamma)\big),\\ &\bar{z} \in L^2\big(I;H^2(\Omega)\cap H^1_0(\Omega)\big)\cap H^1\big(I;L^2(\Omega)\big).\end{aligned}$$

Remark 3.4. The outcome of Theorems 3.2 and 3.3 regarding the regularity of the optimal control problem (1.1) hold significant implications for the error estimation of discrete optimal control problems. The insight from Theorem 3.3 suggests that optimal convergence orders can be attained for the discrete control by employing the DG(0)-CG(1) discretization scheme. Moreover, it follows from Theorems 3.2 and 3.3 that the optimal control has the regularity $L^2(I;H^{\frac{1}{2}}(\Gamma))\cap H^{\frac{1}{4}}(I;L^2(\Gamma))$. Thus, the very weak solution to the state equation (1.2), defined in Definition 2.1, is equivalent to the standard weak solution outlined in Lemma 2.4.

4. Finite element discretization

4.1. Notations for finite element methods

To introduce the DG(0) semi-discretization in time and the DG(0)-CG(1) fully discretization in time-space of the state and adjoint equations, we begin with some notations.

Firstly, we divide the interval $\bar{I} := [0,T]$ into a family of subintervals $I_m := (t_{m-1},t_m]$ with step sizes $k_m := t_m - t_{m-1},, m = 1,\cdots,M$, where $0 = t_0 < t_1 < \cdots < t_M = T$. The maximal time step size is denoted by $k := \max_{1\le m\le M} k_m$.

Throughout this article, we assume the following two conditions for k_m and k:

(i) There exists a constant $C > 0$ independent of k and m such that

$$k/k_m \le C, \quad m = 1,\cdots,M. \tag{4.1}$$

(ii) The time step k_m is non-increasing for m, *i.e.*,

$$k_m/k_{m-1} \le 1, \quad m = 2,\cdots,M. \tag{4.2}$$

The first condition ensures that the partition for I is quasi-uniform, and the second one implies that the time step is non-increasing along the direction of T.

We define the temporal semi-discrete DG(0) space consisting of piecewise constants in time as follows:

$$\begin{aligned}X_k^0 &:= \big\{v_k \in L^2\big(I;H^1_0(\Omega)\big) : \ v_k|_{I_m} \in P_0\big(I_m;H^1_0(\Omega)\big),\ m = 1,\cdots,M\big\},\\ X_k &:= \big\{v_k \in L^2\big(I;H^1(\Omega)\big) : v_k|_{I_m} \in P_0\big(I_m;H^1(\Omega)\big),\ m = 1,\cdots,M\big\},\\ \tilde{X}_k &:= \big\{v_k \in L^2\big(I;L^2(\Omega)\big) : v_k|_{I_m} \in P_0\big(I_m;L^2(\Omega)\big),\ m = 1,\cdots,M\big\},\end{aligned}$$

where $P_0(I_m;H^1_0(\Omega))$, $P_0(I_m;H^1(\Omega))$ and $P_0(I_m;L^2(\Omega))$ denote function spaces of constants defined in I_m $(m = 1,\cdots,M)$ and valued in spaces $H^1_0(\Omega), H^1(\Omega)$ and $L^2(\Omega)$, respectively.

In addition, we introduce the following two temporal semi-discrete finite element spaces consisting of functions of piecewise constants in time:

$$\begin{aligned}X_k(\Gamma) &:= \Big\{v_k \in L^2\Big(I;H^{\frac{1}{2}}(\Gamma)\Big) : \exists\, \tilde{v}_k \in X_k, \ \ \text{s.t.}\, v_k|_{I_m} = \tilde{v}_k|_{I_m\times\Gamma}, m = 1,\cdots,M\Big\},\\ \tilde{X}_k(\Gamma) &:= \big\{v_k \in L^2\big(I;L^2(\Gamma)\big) : v_k|_{I_m} \in P_0\big(I_m;L^2(\Gamma)\big), m = 1,\cdots,M\big\}.\end{aligned}$$

When Ω is a polytope, we introduce a family of quasi-uniform and shape regular partitions $\mathcal{T}_h = \{\tau\}$ in the sense of Ciarlet [13], where τ is the N-simplex with diameter h_τ. We denote the mesh size of $\mathcal{T}_h$ by $h := \max_{\tau\in\mathcal{T}_h} h_\tau$. Define the P_1 finite element space

$$V_h := \{v_h \in C(\bar{\Omega}) : v_h|_\tau \in P_1(\tau),\ \forall \tau \in \mathcal{T}_h\}, \tag{4.3}$$

where $P_1(\tau)$ denotes the space of linear functions in τ. We set $V_h^0 := V_h \cap H_0^1(\Omega)$ and

$$V_h(\Gamma) := \{v_h|_\Gamma : v_h \in V_h\}.$$

Let $\pi_h : C(\bar{\Omega}) \to V_h$ be the Lagrange interpolation operator (cf. [7]) or the Clément interpolation operator from $L^1(\Omega)$ to V_h ([14]). Let $P_h : L^2(\Omega) \to V_h$ be the L^2-projection operator defined as follows: For any $y \in L^2(\Omega)$, $P_h y$ satisfies

$$(P_h y, v_h) = (y, v_h) \qquad \forall v_h \in V_h.$$

Assume that $R_h : H_0^1(\Omega) \to V_h^0$ is the Ritz projection operator, *i.e.*, for any $\varphi \in H_0^1(\Omega)$, $R_h\varphi \in V_h^0$ satisfies

$$(\nabla R_h\varphi, \nabla v_h) = (\nabla\varphi, \nabla v_h) \qquad \forall v_h \in V_h^0.$$

The following lemma gives error estimates for the Ritz projection (cf. [7]), which will be frequently used in this article.

Lemma 4.1. *Let R_h be the Ritz projection operator, then the following estimates hold:*

$$\|v - R_h v\|_{H^1(\Omega)} \le Ch\|\nabla^2 v\|_{L^2(\Omega)} \qquad \forall v \in H_0^1(\Omega) \cap H^2(\Omega),$$
$$\|v - R_h v\|_{L^2(\Omega)} \le Ch\|\nabla(v - R_h v)\|_{L^2(\Omega)} \qquad \forall v \in H_0^1(\Omega).$$

Remark 4.2. Although the operators R_h, P_h, and π_h are defined in the spaces of functions independent of time, it is possible to extend their definitions to the time-dependent case, which have to be understood pointwise in time. Then, Lemma 4.1 is also valid for the time-dependent case with corresponding estimates under time-space norms.

In order to define the DG(0)-CG(1) discrete scheme for parabolic equations, we introduce the following time-space finite element spaces:

$$X_{kh}^0 := \{v_{kh} \in X_k^0 : \ v_{kh}|_{I_m} \in P_0(I_m; V_h^0),\ m = 1, \cdots, M\},$$
$$X_{kh} := \{v_{kh} \in X_k : \ v_{kh}|_{I_m} \in P_0(I_m; V_h),\ m = 1, \cdots, M\},$$

where $P_0(I_m; V_h)$ denotes the function space of constants defined in I_m and valued in V_h, the definition of $P_0(I_m; V_h^0)$ is similar. In addition, we define

$$X_{kh}(\Gamma) := \{v_{kh} \in X_k(\Gamma) : \ v_{kh}|_{I_m} \in P_0(I_m; V_h(\Gamma)),\ m = 1, \cdots, M\},$$

where $P_0(I_m; V_h(\Gamma))$ is the function space of constants defined in I_m and valued in $V_h(\Gamma)$.

For the definition of the temporal DG(0) scheme for parabolic equations, we need the following notations: for any $v_k \in X_k$,

$$v_{k,m} = v_{k,m}^- := \lim_{t\to 0^+} v_k(t_m - t), \quad v_{k,m+1} = v_{k,m}^+ := \lim_{t\to 0^+} v_k(t_m + t), \quad [v_k]_m := v_{k,m}^+ - v_{k,m}^-, \quad m = 1, \cdots, M,$$

where $v_{k,m} := v_k|_{I_m}$ and $[v_k]_m$ denotes the jump of v_k at nodes t_m.

Given two piecewise-in-time H^1-functions $v,\, w \in L^2(I; H^1(\Omega))$, we define the bilinear form $B : (v, w) \to \mathbb{R}$ as follows:

$$B(v, w) := \sum_{m=1}^{M} (\partial_t v, w)_{I_m} + (\nabla v, \nabla w)_I + \sum_{m=2}^{M} \big([v]_{m-1}, w_{m-1}^+\big) + \big(v_0^+, w_0^+\big), \tag{4.4}$$

where $(\cdot,\cdot)_{I_m}$ denotes the inner product in space $L^2(I_m;L^2(\Omega))$ (cf. [38]). Applying integration by parts to the above bilinear form B, we obtain the following dual representation:

$$B(v,w) = -\sum_{m=1}^{M}(v,\partial_t w)_{I_m} + (\nabla v,\nabla w)_I - \sum_{m=1}^{M-1}\left(v_m^-,[w]_m\right) + \left(v_M^-,w_M^-\right). \tag{4.5}$$

Note that if $v,w\in X_k$, then the first term in B vanishes. Similarly, if v,w are continuous functions in time, then the terms of B containing jumps also vanish.

In the subsequent two subsections, we address the discretization of the optimal control problem (1.1). Here, the state variable is approximated using finite elements, while the control variable undergoes discretization *via* variational discretization, as outlined in [25]. In Section 4.2, we employ the DG(0) method for the temporal semi-discretization of the state variable, focusing on problems formulated in bounded polytopal/smooth domains. Section 4.3, on the other hand, employs the DG(0)-CG(1) method, utilizing piecewise constant functions in time and continuous piecewise linear polynomials in space, to discretize the state variable for problems posed on polytopal domains.

4.2. Discretization in time

To establish the temporal semi-discrete scheme for the parabolic equation (1.2), we introduce the boundary L^2-projection operator in time, denoted as $\tilde{P}_k : L^2(I;L^2(\Gamma))\to\tilde{X}_k(\Gamma)$. For any $w\in L^2(I;L^2(\Gamma))$, $\tilde{P}_k w\in\tilde{X}_k(\Gamma)$ is defined as follows:

$$\tilde{P}_k w|_{I_m} := \frac{1}{k_m}\int_{t_{m-1}}^{t_m} w(s)\,\mathrm{d}s, \qquad m=1,\cdots,M. \tag{4.6}$$

For simplicity, we set $\tilde{P}_k^m\omega := \tilde{P}_k\omega|_{I_m}$ for any $\omega\in L^2(I;L^2(\Gamma))$.

The semi-discrete parabolic Dirichlet boundary control problem is given by

$$\min_{u\in U_{ad},\ y_k(u)\in\tilde{X}_k} J_k(y_k(u),u) = \frac{1}{2}\|y_k(u)-y_d\|_I^2 + \frac{\alpha}{2}\|u\|^2_{L^2(I;L^2(\Gamma))}, \tag{4.7}$$

where $y_k(u)\in\tilde{X}_k$ is the semi-discrete state variable satisfying the scheme:

$$\tilde{B}(y_k(u),\varphi_k) = (f,\varphi_k)_I - (u,\partial_n\varphi_k)_{L^2(I;L^2(\Gamma))} + \left(y_0,\varphi_{k,0}^+\right) \tag{4.8}$$

for all $\varphi_k\in X_k^0\cap L^2(I;H^2(\Omega))$, and $\tilde{B}:\tilde{X}_k\times(X_k\cap L^2(I;H^2(\Omega)))\to\mathbb{R}$ is another bilinear form defined as

$$\tilde{B}(y_k,\varphi_k) := -(y_k,\Delta\varphi_k)_I + \sum_{m=2}^{M}\left([y_k]_{m-1},\varphi_{k,m-1}^+\right) + \left(y_{k,0}^+,\varphi_{k,0}^+\right).$$

Note that (4.8) is the very weak formulation of the temporal semi-discrete scheme for parabolic equations (2.4) with rough Dirichlet data $u\in L^2(I;L^2(\Gamma))$. If u has improved regularity in space, such as $u\in L^2(I;H^{\frac{1}{2}}(\Gamma))$, then (4.8) is equivalent to the following scheme: Find $y_k(u)\in X_k$ such that

$$B(y_k(u),\varphi_k) = (f,\varphi_k)_I + \left(y_0,\varphi_{k,0}^+\right) \qquad \forall\varphi_k\in X_k^0, \quad y_k(u)|_{I\times\Gamma} = \tilde{P}_k u. \tag{4.9}$$

Note that in the discrete problem (4.7) the control variable is not explicitly discretized, which is the so-called variational discretization proposed in [25] for optimal control problems.

Similar to Definition 2.1, one can demonstrate that the semi-discrete scheme (4.8) possesses a unique solution $y_k(u)\in\tilde{X}_k$ for any $u\in U_{ad}$. The control-to-discrete state mapping $S_k : L^2(I;L^2(\Gamma))\to\tilde{X}_k$ is defined as

$S_k u := y_k(u)$ for any $u \in L^2(I; L^2(\Gamma))$ and is Fréchet differentiable. Subsequently, we formulate the reduced optimization problem:

$$\min_{u\in U_{ad}} \hat{J}_k(u) := J(S_k u, u). \tag{4.10}$$

Similar to Proposition 3.2 in [38], combining Lemma 6.5 we can obtain that the semi-discrete optimal control problem (4.10) admits a unique solution $\bar{u}_k$ by using standard arguments (cf. [27], Page 53). Moreover, the reduced cost functional $\hat{J}_k$ is also Fréchet differentiable and

$$\hat{J}_k'(u)v = \int_{\Omega_T} (y_k(u) - y_d)\tilde{y}_k(v)\,\mathrm{d}x\,\mathrm{d}t + \alpha \int_{\Sigma_T} uv\,\mathrm{d}s\,\mathrm{d}t \qquad \forall v \in L^2\big(I; L^2(\Gamma)\big), \tag{4.11}$$

where $\tilde{y}_k(v) \in \tilde{X}_k$ is the solution of the following semi-discrete problem:

$$\tilde{B}(\tilde{y}_k(v), \varphi_k) = -(v, \partial_n \varphi_k)_{L^2(I;L^2(\Gamma))} \qquad \forall \varphi_k \in X_k^0 \cap L^2\big(I; H^2(\Omega)\big). \tag{4.12}$$

In the following theorem, we give the first order optimality condition for the semi-discrete optimal control problem (4.10).

Theorem 4.3. *The pair* $(\bar{u}_k, \bar{y}_k) \in L^2(I; H^{\frac{1}{2}}(\Gamma)) \cap X_k(\Gamma) \times X_k$ *is the optimal solution of the semi-discrete control problem* (4.7) *if and only if* $\bar{y}_k := S_k\bar{u}_k = y_k(\bar{u}_k)$ *and the following first order optimality condition holds:*

$$\hat{J}_k'(\bar{u}_k)(v - \bar{u}_k) \geq 0 \qquad \forall v \in U_{ad}. \tag{4.13}$$

Furthermore, there exists a semi-discrete adjoint state variable $\bar{z}_k \in X_k^0 \cap L^2(I; H^2(\Omega))$ *defined by*

$$B(\varphi_k, \bar{z}_k) = (\bar{y}_k - y_d, \varphi_k)_I \qquad \forall \varphi_k \in X_k^0, \tag{4.14}$$

such that (4.13) *can be equivalently written as*

$$\hat{J}_k'(\bar{u}_k)(v - \bar{u}_k) = \int_{\Sigma_T} (\alpha\bar{u}_k - \partial_n \bar{z}_k)(v - \bar{u}_k)\,\mathrm{d}s\,\mathrm{d}t \geq 0 \qquad \forall v \in U_{ad}, \tag{4.15}$$

where $\partial_n \bar{z}_k \in X_k(\Gamma)$ *is the outward normal derivative of* $\bar{z}_k$*, satisfies*

$$\int_{\Sigma_T} \partial_n \bar{z}_k \phi_k\,\mathrm{d}s\,\mathrm{d}t = -\int_{\Omega_T} (\bar{y}_k - y_d)p_k(\phi_k)\,\mathrm{d}x\,\mathrm{d}t \qquad \forall \phi_k \in \tilde{X}_k(\Gamma), \tag{4.16}$$

and $p_k(\phi_k) \in \tilde{X}_k$ *is the solution of the following semi-discrete scheme:*

$$\tilde{B}(p_k(\phi_k), \varphi_k) = -(\phi_k, \partial_n \varphi_k)_{L^2(I;L^2(\Gamma))} \qquad \forall \varphi_k \in X_k^0 \cap L^2\big(I; H^2(\Omega)\big). \tag{4.17}$$

Proof. Given that the optimization problem (4.7) is strictly convex, the first-order optimality condition (4.13) directly follows from the calculus of variations. Thus, we only need to validate (4.15). For any given $v \in U_{ad}$, by setting $\phi_k := \tilde{P}_k(v - \bar{u}_k)$ in (4.17), the superposition principle of linear equations implies that $p_k(\tilde{P}_k(v - \bar{u}_k)) = y_k(v) - \bar{y}_k$ satisfies (4.17). Subsequently, it follows from (4.16) that

$$\begin{aligned}
\hat{J}_k'(\bar{u}_k)(v - \bar{u}_k) &= \int_{\Omega_T} (\bar{y}_k - y_d)(y_k(v) - \bar{y}_k)\,\mathrm{d}x\,\mathrm{d}t + \alpha \int_{\Sigma_T} \bar{u}_k(v - \bar{u}_k)\,\mathrm{d}s\,\mathrm{d}t \\
&= \int_{\Omega_T} (\bar{y}_k - y_d)p_k\Big(\tilde{P}_k(v - \bar{u}_k)\Big)\,\mathrm{d}x\,\mathrm{d}t + \alpha \int_{\Sigma_T} \bar{u}_k(v - \bar{u}_k)\,\mathrm{d}s\,\mathrm{d}t \\
&= -\int_{\Sigma_T} \partial_n \bar{z}_k \tilde{P}_k(v - \bar{u}_k)\,\mathrm{d}s\,\mathrm{d}t + \alpha \int_{\Sigma_T} \bar{u}_k(v - \bar{u}_k)\,\mathrm{d}s\,\mathrm{d}t
\end{aligned}$$

$$
\begin{aligned}
&= -\int_{\Sigma_T} \partial_n \bar{z}_k (v - \bar{u}_k)\,\mathrm{d}s\,\mathrm{d}t + \alpha \int_{\Sigma_T} \bar{u}_k (v - \bar{u}_k)\,\mathrm{d}s\,\mathrm{d}t \\
&= \int_{\Sigma_T} (\alpha \bar{u}_k - \partial_n \bar{z}_k)(v - \bar{u}_k)\,\mathrm{d}s\,\mathrm{d}t \geq 0.
\end{aligned}
$$

The spatial regularity of the semi-discrete adjoint state $\bar{z}_k$ can be deduced from the standard elliptic regularity on each time interval. Consequently, we establish that $\bar{u}_k \in L^2(I; H^{\frac{1}{2}}(\Gamma))$ given that $\bar{u}_k$ is the projection of $\partial_n \bar{z}_k$ onto U_{ad}. Additionally, $\bar{y}_k$ satisfies the scheme (4.9), ensuring that $\bar{y}_k \in X_k$. This concludes the proof. □

Remark 4.4. In fact, one can easily check that the definition of the outward normal derivative in (4.16) is equivalent to the usual one (cf. [23]).

4.3. Discretization in space

When dealing with a polytope Ω, we can extend our analysis to include the time-space discretization of the optimal control problem (1.1) through the application of the DG(0)-CG(1) method. For this purpose, we introduce a family of L^2-projections.

Definition 4.5. Let $P_h : L^2(I; L^2(\Omega)) \to L^2(I; V_h)$ and $P_{kh} : L^2(I; L^2(\Omega)) \to X_{kh}$ be two orthogonal projection operators such that for any $w \in L^2(I; L^2(\Omega))$, $P_h w$ and $P_{kh} w$ satisfy respectively

$$
(w - P_h w, v_h)_{L^2(I;L^2(\Omega))} = 0 \qquad \forall v_h \in L^2(I; V_h), \quad (w - P_{kh} w, v_{kh})_{L^2(I;L^2(\Omega))} = 0 \qquad \forall v_{kh} \in X_{kh},
$$

where $L^2(I; V_h)$ denotes the space of functions defined in I and valued in V_h.

Similarly, we can define another two projections $\tilde{P}_h : L^2(I; L^2(\Gamma)) \to L^2(I; V_h(\Gamma))$ and $\tilde{P}_{kh} : L^2(I; L^2(\Gamma)) \to X_{kh}(\Gamma)$ with $L^2(I; V_h(\Gamma))$ and $X_{kh}(\Gamma)$ playing the role of $L^2(I; V_h)$ and X_{kh}, respectively.

The time-space discretization of the optimal control problem (1.1) is given by

$$
\min_{u \in U_{ad},\ y_{kh}(u) \in X_{kh}} J_{kh}(y_{kh}(u), u) = \frac{1}{2}\|y_{kh}(u) - y_d\|_I^2 + \frac{\alpha}{2}\|u\|^2_{L^2(I;L^2(\Gamma))}, \tag{4.18}
$$

where $y_{kh}(u) \in X_{kh}$ is the discrete state variable satisfying the following discrete state equation:

$$
B(y_{kh}(u), \varphi_{kh}) = (f, \varphi_{kh})_I + \left(y_0, \varphi^+_{kh,0}\right) \qquad \forall \varphi_{kh} \in X^0_{kh},\ y_{kh}(u)|_{I\times\Gamma} = \tilde{P}_{kh} u. \tag{4.19}
$$

Again, the control variable is not explicitly discretized in the above discrete control problem (4.18). However, the discrete adjoint state will yield an implicit discretization of the control.

For any given $u \in U_{ad}$, we can establish the unique solution $y_{kh}(u) \in X_{kh}$ to the discrete state equation (4.19). Subsequently, we define a discrete control-to-state linear operator $S_{kh} : L^2(I; L^2(\Gamma)) \to X_{kh}$ as $S_{kh} u := y_{kh}(u)$ for any $u \in L^2(I; L^2(\Gamma))$. This leads us to the following reduced optimization problem:

$$
\min_{u \in U_{ad}} \hat{J}_{kh}(u) := J(S_{kh} u, u). \tag{4.20}
$$

It is easy to check that the above discrete optimization problem has a unique solution, denoted by $\bar{u}_{kh}$. The first order Fréchet derivative of J_{kh} at $u \in L^2(I; L^2(\Gamma))$ can be calculated as

$$
\hat{J}'_{kh}(u)v = \int_{\Omega_T} (y_{kh}(u) - y_d)\tilde{y}_{kh}(v)\,\mathrm{d}x\,\mathrm{d}t + \alpha \int_{\Sigma_T} uv\,\mathrm{d}s\,\mathrm{d}t \qquad \forall v \in U_{ad},
$$

where $\tilde{y}_{kh}(v) \in X_{kh}$ is the solution of the following equation:

$$
B(\tilde{y}_{kh}(v), \varphi_{kh}) = 0 \qquad \forall \varphi_{kh} \in X^0_{kh}, \quad \tilde{y}_{kh}(v)|_{I\times\Gamma} = \tilde{P}_{kh} v.
$$

To simplify the Fréchet derivative of $\hat{J}_{kh}$, we first define the discrete outward normal derivative for the DG(0)-CG(1) finite element solution of the backward parabolic equation.

Definition 4.6. For any $g \in L^2(I; L^2(\Omega))$, let z be the solution of the backward parabolic equation (2.3). Let $z_{kh} \in X_{kh}^0$ be the DG(0)-CG(1) finite element solution of (2.3) defined by

$$B(\varphi_{kh}, z_{kh}) = (g, \varphi_{kh}) \qquad \forall \varphi_{kh} \in X_{kh}^0. \tag{4.21}$$

Then the time-space discrete normal derivative $\partial_n^h z_{kh} \in X_{kh}(\Gamma)$ of z_{kh} on Γ is defined by

$$\int_{\Sigma_T} \partial_n^h z_{kh} \phi_{kh} \,\mathrm{d}s\,\mathrm{d}t = -\int_{\Omega_T} g p_{kh}(\phi_{kh}) \,\mathrm{d}x\,\mathrm{d}t \qquad \forall \phi_{kh} \in X_{kh}(\Gamma), \tag{4.22}$$

where $p_{kh}(\phi_{kh}) \in X_{kh}$ satisfies the following equation for given $\phi_{kh} \in X_{kh}(\Gamma)$:

$$B(p_{kh}(\phi_{kh}), \varphi_{kh}) = 0 \qquad \forall \varphi_{kh} \in X_{kh}^0, \quad p_{kh}(\phi_{kh})|_{I\times\Gamma} = \phi_{kh}. \tag{4.23}$$

Definition 4.7. Let z be the solution of equation (2.3) and z_{kh} be the corresponding time-space fully discrete solution. The discrete function $\partial_n^h z_{kh} \in X_{kh}(\Gamma)$ is called the discrete normal derivative of z_{kh} on Γ, if $\partial_n^h z_{kh}$ satisfies

$$\int_{\Sigma_T} \partial_n^h z_{kh} \Phi_{kh} \,\mathrm{d}s\,\mathrm{d}t = B(\Phi_{kh}, z_{kh}) - \int_{\Omega_T} g\Phi_{kh} \,\mathrm{d}x\,\mathrm{d}t \qquad \forall\, \Phi_{kh} \in X_{kh}. \tag{4.24}$$

Remark 4.8. Similar to the principles outlined in Definition 2.1, the method of transposition is also employed in Definition 4.6. Specifically, (2.4) establishes the definition of the weak solution y using the test function g. In contrast, in (4.22), we provide the definition of $\partial_n^h z_{kh}$ with ϕ_{kh} serving as the test function. Meanwhile, Definition 4.7 adheres to the standard one found in [23]. It is important to note that Definitions 4.6 and 4.7 are equivalent, a fact that will be substantiated in the following proposition. The alternative Definition 4.6 facilitates the examination of error estimates between discrete and continuous normal derivatives of the adjoint equation using duality tricks, and it allows for the derivation of stability results for fully discrete solutions to parabolic equations with inhomogeneous Dirichlet data.

Proposition 4.9. *Definitions 4.6 and 4.7 are equivalent.*

Proof. Firstly, we prove that $\partial_n^h z_{kh}$ satisfying the equality (4.22) in Definition 4.6 implies (4.24) in Definition 4.7. For any $\Phi_{kh} \in X_{kh}$, we define $\Psi_{kh} \in X_{kh}^0$ by

$$B(\Psi_{kh}, \varphi_{kh}) = B(\Phi_{kh}, \varphi_{kh}) \qquad \forall \varphi_{kh} \in X_{kh}^0. \tag{4.25}$$

Let $\Theta_{kh} := \Phi_{kh} - \Psi_{kh}$, then Θ_{kh} is the solution of

$$B(\Theta_{kh}, \varphi_{kh}) = 0 \qquad \forall \varphi_{kh} \in X_{kh}^0, \quad \Theta_{kh}|_{I\times\Gamma} = \Phi_{kh}|_{I\times\Gamma}.$$

Setting $\phi_{kh} = \Phi_{kh}|_{I\times\Gamma}$ in the equality (4.22), there holds the following equality:

$$\int_{\Sigma_T} \partial_n^h z_{kh} \Phi_{kh} \,\mathrm{d}s\,\mathrm{d}t = -\int_{\Omega_T} g\Theta_{kh} \,\mathrm{d}x\,\mathrm{d}t = -\int_{\Omega_T} g\Phi_{kh} \,\mathrm{d}x\,\mathrm{d}t + \int_{\Omega_T} g\Psi_{kh} \,\mathrm{d}x\,\mathrm{d}t.$$

Note that z_{kh} is the solution of (4.21), it follows from (4.25) that

$$\int_{\Sigma_T} \partial_n^h z_{kh} \Phi_{kh} \,\mathrm{d}s\,\mathrm{d}t = B(\Phi_{kh}, z_{kh}) - \int_{\Omega_T} g\Phi_{kh} \,\mathrm{d}x\,\mathrm{d}t,$$

which implies that $\partial_n^h z_{kh}$ satisfies Definition 4.7. Reversely, $\partial_n^h z_{kh}$ given in Definition 4.7 satisfies the identity (4.22) in Definition 4.6. This completes the proof. ∎

Now we are ready to derive the discrete first order optimality system.

Theorem 4.10. *The pair* $(\bar{u}_{kh}, \bar{y}_{kh}) \in U_{ad} \times X_{kh}$ *is the optimal solution of the fully discrete optimal control problem* (4.20) *if and only if* $\bar{y}_{kh} := S_{kh}\bar{u}_{kh} = y_{kh}(\bar{u}_{kh})$ *and*

$$\hat{J}'_{kh}(\bar{u}_{kh})(v - \bar{u}_{kh}) \geq 0 \qquad \forall v \in U_{ad}. \tag{4.26}$$

Furthermore, we define the fully discrete adjoint state $\bar{z}_{kh} \in X^0_{kh}$ *as*

$$B(\varphi_{kh}, \bar{z}_{kh}) = (\bar{y}_{kh} - y_d, \varphi_{kh}) \qquad \forall \varphi_{kh} \in X^0_{kh}, \tag{4.27}$$

with which the optimality condition (4.26) *can be equivalently written as*

$$\hat{J}'_{kh}(\bar{u}_{kh})(v - \bar{u}_{kh}) = \int_{\Sigma_T} \big(\alpha \bar{u}_{kh} - \partial^h_n \bar{z}_{kh}\big)(v - \bar{u}_{kh})\,\mathrm{d}s\,\mathrm{d}t \geq 0 \qquad \forall v \in U_{ad}, \tag{4.28}$$

where $\partial^h_n \bar{z}_{kh} \in X_{kh}(\Gamma)$ *is the discrete normal derivative of* $\bar{z}_{kh} \in X^0_{kh}$ *defined in Definition 4.6.*

5. Error estimates for the parabolic equation

In order to conduct the error analysis for the discrete optimal control problem, it is essential to establish *a priori* error estimates for the finite element discretization of parabolic equations, assuming suitable regularity conditions on the solution. Since the optimal control has the regularity $\bar{u} \in L^2(I; H^{\frac{1}{2}}(\Gamma)) \cap H^{\frac{1}{4}}(I; L^2(\Gamma))$ (cf. Thms. 3.2 and 3.3), in the following we use the standard weak solution to the optimal state.

5.1. Analysis of the temporal discretization error

In this subsection, our primary focus is on the error estimation for temporal semi-discretizations of parabolic equations. Employing the Aubin–Nitsche technique [7], we transform the finite element error into the projection or interpolation error.

To this end, we introduce some projection and interpolation operators. Define the L^2-projection $P_k : L^2(I; L^2(\Omega)) \to \tilde{X}_k$ such that $P_k z$ satisfies

$$P_k z|_{I_m} := \frac{1}{k_m} \int_{t_{m-1}}^{t_m} z(s)\,\mathrm{d}s, \qquad m = 1, \cdots, M, \quad \forall z \in L^2(I; L^2(\Omega)). \tag{5.1}$$

For simplicity we write $P^m_k z := P_k z|_{I_m}$. For any $0 \leq s \leq 1$, there holds

$$\|z - P_k z\|_I \leq C k^s \|z\|_{H^s(I;L^2(\Omega))}, \qquad \forall z \in H^s(I; L^2(\Omega)). \tag{5.2}$$

In addition, we need the following two interpolation operators in time. Define the interpolation operators $\pi^r_k : C(\bar{I}; L^2(\Omega)) \to \tilde{X}_k$ and $\pi^l_k : C(\bar{I}; L^2(\Omega)) \to \tilde{X}_k$ such that for any $w \in C(\bar{I}; L^2(\Omega))$, $\pi^r_k w$ and $\pi^l_k w$ satisfy

$$\pi^r_k w|_{I_m} = w(t_m), \quad \pi^l_k w|_{I_m} = w(t_{m-1}), \qquad m = 1, \cdots, M. \tag{5.3}$$

Note that the above interpolations are defined by taking the end point values on each subinterval, and the interpolation error estimate is stated in the following lemma that is similar to (5.2).

Lemma 5.1. *For arbitrary* $s \in [\frac{1}{2}, 1]$, *if* $\omega \in H^s(I; L^2(\Omega)) \cap C(\bar{I}; L^2(\Omega))$, *then*

$$\big\|w - \pi^l_k w\big\|_I + \|w - \pi^r_k w\|_I \leq C k^s \|w\|_{H^s(I;L^2(\Omega))}. \tag{5.4}$$

Proof. The above interpolation error estimate for $s = \frac{1}{2}$ is a direct consequence of Theorem 5.20 in [28], see also [33, 41] for the case $s \in (\frac{1}{2}, 1]$. □

The scheme (4.8) constitutes a temporal DG(0) discretization of equation (1.2) (cf. [43]), yielding a unique solution denoted as $y_k(u) \in X_k$. It is noteworthy that while $f \in L^2(I;L^2(\Omega))$, $y_0 \in L^2(\Omega)$, and $u \in H^{\frac14}(I;L^2(\Gamma))\cap L^2(I;H^{\frac12}(\Gamma))$, the resulting y resides in $H^{\frac12}(I;L^2(\Omega))\cap L^2(I;H^1(\Omega))$. However, the DG(0) semi-discretization $y_k(u)\in X_k$ adopts a piecewise constant profile in time, falling short of belonging to $H^{\frac12}(I;L^2(\Omega))$, despite $\tilde P_k u \in H^{\frac14}(I;L^2(\Gamma))\cap L^2(I;H^{\frac12}(\Gamma))$. Indeed, it can be verified that $y_k(u)\in H^{\frac12-\varepsilon}(I;L^2(\Omega))$ for any $\varepsilon>0$ (cf. [42], Page 20). Therefore, the DG(0) semi-discretization (4.8) aligns with a nonconforming Galerkin method (*i.e.*, $X_k \nsubseteq H^{\frac12}(I;L^2(\Omega))$). Nevertheless, we still have the Galerkin orthogonality

$$B(y-y_k(u),\varphi_k)=0 \qquad \forall \varphi_k\in X_k^0.$$

Below, we present several lemmas pertaining to the stability and error estimation for the temporal semi-discretization of parabolic equations with general inhomogeneous Dirichlet boundary condition which may be of independent interest.

Lemma 5.2. *For any given $s\in[\frac12,1]$ and sufficiently smooth u, y_0 and f, assume that the solution to parabolic equations (1.2) satisfies $y\in H^s(I;L^2(\Omega))\cap L^2(I;H^{2s}(\Omega))$, then there holds*

$$\sum_{m=1}^M k_m^{1-(2s-1)}\|g\|^2_{L^2(I_m;L^2(\Omega))}\le C\|y\|^2_{H^s(I;L^2(\Omega))},$$

where g is a piecewise constant function defined by

$$g|_{I_m} := \frac{P_k^m y - y(t_m)}{k_m} - \frac{P_k^{m-1}y-y(t_{m-1})}{k_m}, \qquad m=1,\cdots,M, \quad P_k^0 y := y(t_0)=y_0.$$

Proof. Note that the solution $y\in H^s(I;L^2(\Omega))\cap L^2(I;H^{2s}(\Omega))$ ($s\in[\frac12,1]$) is continuous by recalling Lemma 2.4, then the function g is well-defined. Using the condition (4.1) we obtain

$$\sum_{m=1}^M k_m^{1-(2s-1)}\|g\|^2_{L^2(I_m;L^2(\Omega))}\le Ck^{1-(2s-1)}\sum_{m=1}^M \|g\|^2_{L^2(I_m;L^2(\Omega))}.$$

Therefore, we only need to estimate the right-hand side of the above inequality:

$$\begin{aligned}
\sum_{m=1}^M \|g\|^2_{L^2(I_m;L^2(\Omega))} &= \sum_{m=1}^M k_m^{-1}\big\|(P_k^m y-y(t_m)) - \big(P_k^{m-1}y-y(t_{m-1})\big)\big\|^2_{L^2(\Omega)}\\
&\le C\sum_{m=1}^M k_m^{-1}\|P_k^m y - y(t_m)\|^2_{L^2(\Omega)}\\
&= C\sum_{m=1}^M k_m^{-1}\int_\Omega\left(\frac{1}{k_m}\int_{t_{m-1}}^{t_m} y(s)\,\mathrm{d}s - y(t_m)\right)^2\mathrm{d}x\\
&= C\sum_{m=1}^M k_m^{-3}\int_\Omega\left(\int_{t_{m-1}}^{t_m}(y(s)-y(t_m))\,\mathrm{d}s\right)^2\mathrm{d}x\\
&\le Ck^{-2}\sum_{m=1}^M\int_\Omega\int_{t_{m-1}}^{t_m}(y(s)-y(t_m))^2\,\mathrm{d}s\,\mathrm{d}x\\
&= Ck^{-2}\|y-\pi_k^r y\|_I^2\\
&\le Ck^{2(s-1)}\|y\|^2_{H^s(I;L^2(\Omega))},
\end{aligned}$$

where we have used Lemma 5.1. This completes the proof. ∎

Lemma 5.3. *Let y and $y_k(u)$ be the solution of (1.2) and (4.8), respectively. For any given $s \in [\frac{1}{2}, 1]$ and sufficiently smooth u, y_0 and f, assume that $y \in H^s(I; L^2(\Omega)) \cap L^2(I; H^{2s}(\Omega))$, then we have*

$$\sum_{m=1}^{M} k_m^{1-(2s-1)} \big\|\nabla [P_k y - y_k(u)]_{m-1}\big\|_{L^2(\Omega)}^2 + \sum_{m=1}^{M} k_m^{-(2s-1)} \big\|[P_k y - y_k(u)]_{m-1}\big\|_{L^2(\Omega)}^2 \leq C\|y\|_{H^s(I;L^2(\Omega))}^2. \quad (5.5)$$

Proof. The fact $y \in H^s(I; L^2(\Omega)) \cap L^2(I; H^{2s}(\Omega))$ $(s \in [\frac{1}{2}, 1])$ also implies that $y \in H^1(I; H^{-1}(\Omega)) \cap L^2(I; H^1(\Omega))$ ([29], Page 84), so that the state y also satisfies the standard weak solution. Therefore, for each fixed $m = 1, \cdots, M$, from (1.2) it follows

$$(\nabla y, \nabla \varphi)_{I_m} + (\partial_t y, \varphi)_{I_m} = (f, \varphi)_{I_m} \qquad \forall \varphi \in P_0\big(I_m; H_0^1(\Omega)\big),$$

i.e.,

$$(\nabla P_k y, \nabla \varphi)_{I_m} + (y(t_m) - y(t_{m-1}), \varphi) = (f, \varphi)_{I_m} \qquad \forall \varphi \in P_0\big(I_m; H_0^1(\Omega)\big), \quad (5.6)$$

where P_k is defined in (5.1) and $y \in C(\bar{I}; L^2(\Omega))$ by Lemma 2.4 is used. Similarly, the temporal semi-discretization solution $y_k(u)$ of (4.8) on the subinterval I_m satisfies

$$(\nabla y_k(u), \nabla \varphi)_{I_m} + \big([y_k(u)]_{m-1}, \varphi\big) = (f, \varphi)_{I_m} \qquad \forall \varphi \in P_0\big(I_m; H_0^1(\Omega)\big), \quad (5.7)$$

for $m = 1, 2, \cdots, M$, where $[y_k(u)]_0 := y_{k,0}^+ - y_0$. Subtracting (5.7) from (5.6), we obtain

$$(\nabla (P_k y - y_k(u)), \nabla \varphi)_{I_m} + \big(y(t_m) - y(t_{m-1}) - [y_k(u)]_{m-1}, \varphi\big) = 0$$

for any $\varphi \in P_0(I_m; H_0^1(\Omega))$. Define $[P_k y]_0 := P_k^1 y - y_0$, $[P_k y - y_k(u)]_0 := P_k^1 y - y_{k,1}(u)$, then we can obtain:

$$(\nabla (P_k y - y_k(u)), \nabla \varphi)_{I_m} + \big([P_k y - y_k(u)]_{m-1}, \varphi\big) = ([P_k y]_{m-1}, \varphi) - (y(t_m) - y(t_{m-1})), \varphi = (g, \varphi)_{I_m} \quad (5.8)$$

for any $\varphi \in P_0(I_m; H_0^1(\Omega))$, where g is defined in Lemma 5.2.

By taking $\varphi|_{I_m} := [P_k y - y_k(u)]_{m-1}$ in equality (5.8), there has

$$\big(\nabla (P_k y - y_k(u)), \nabla [P_k y - y_k(u)]_{m-1}\big)_{I_m} + \big\|[P_k y - y_k(u)]_{m-1}\big\|_{L^2(\Omega)}^2 = \big(g, [P_k y - y_k(u)]_{m-1}\big)_{I_m},$$

i.e.,

$$\begin{aligned} &\frac{1}{2}\Big(\big\|\nabla [P_k y - y_k(u)]_{m-1}\big\|_{I_m}^2 + \|\nabla (P_k y - y_k(u))_m\|_{I_m}^2 - \big\|\nabla (P_k y - y_k(u))_{m-1}\big\|_{I_m}^2\Big) \\ &\qquad + \big\|[P_k y - y_k(u)]_{m-1}\big\|_{L^2(\Omega)}^2 = \big(g, [P_k y - y_k(u)]_{m-1}\big)_{I_m} \end{aligned} \quad (5.9)$$

where $(P_k y - y_k(u))_m := P_k^m y - y_{k,m}(u)$ $(m \neq 0)$ and $(P_k y - y_k(u))_0 := 0$. It remains to estimate the right-hand side of the above equality.

We use the first identity of (5.9) and the Cauchy–Schwarz inequality to obtain

$$\begin{aligned} &\frac{k_m}{2}\Big(\big\|\nabla [P_k y - y_k(u)]_{m-1}\big\|_{L^2(\Omega)}^2 + \|\nabla (P_k y - y_k(u))_m\|_{L^2(\Omega)}^2 - \big\|\nabla (P_k y - y_k(u))_{m-1}\big\|_{L^2(\Omega)}^2\Big) \\ &\qquad + \big\|[P_k y - y_k(u)]_{m-1}\big\|_{L^2(\Omega)}^2 \leq \frac{k_m}{2}\|g\|_{I_m}^2 + \frac{\big\|[P_k y - y_k(u)]_{m-1}\big\|_{I_m}^2}{2k_m}. \end{aligned}$$

Multiplying by $k_m^{-(2s-1)}$ on both sides of the above estimate and absorbing the second term of the right-hand side to the left, we obtain

$$k_m^{1-(2s-1)}\Big(\big\|\nabla [P_k y - y_k(u)]_{m-1}\big\|_{L^2(\Omega)}^2 + \|\nabla (P_k y - y_k(u))_m\|_{L^2(\Omega)}^2$$

$$- \left\|\nabla(P_k y - y_k(u))_{m-1}\right\|^2_{L^2(\Omega)}\Big) + k_m^{-(2s-1)}\left\|[P_k y - y_k(u)]_{m-1}\right\|^2_{L^2(\Omega)} \le k_m^{1-(2s-1)}\|g\|^2_{I_m}.$$

By using the step size condition (4.2), the above estimate can be rewritten as

$$\begin{aligned} &k_m^{1-(2s-1)}\left\|\nabla[P_k y - y_k(u)]_{m-1}\right\|^2_{L^2(\Omega)} + k_m^{1-(2s-1)}\left\|\nabla(P_k y - y_k(u))_m\right\|^2_{L^2(\Omega)} \\ &\quad - k_{m-1}^{1-(2s-1)}\left\|\nabla(P_k y - y_k(u))_{m-1}\right\|^2_{L^2(\Omega)} + k_m^{-(2s-1)}\left\|[P_k y - y_k(u)]_{m-1}\right\|^2_{L^2(\Omega)} \\ &\quad \le k_m^{1-(2s-1)}\|g\|^2_{I_m}, \end{aligned}$$

where $k_0 := 1$. Summing up the above estimate for m and using Lemma 5.2, we can obtain the result. This completes the proof of the lemma. ☐

Based on the previous lemma, we obtain the following result.

Lemma 5.4. *Assume that y and $y_k(u)$ are the solutions of equations (1.2) and (4.8) respectively. For given $s \in [\frac{1}{2}, 1]$ and sufficiently smooth u, y_0 and f, assume that $y \in H^s(I; L^2(\Omega)) \cap L^2(I; H^{2s}(\Omega))$, then there holds*

$$\sum_{m=1}^{M} k_m^{-(2s-1)}\left\|[y_k(u)]_{m-1}\right\|^2_{L^2(\Omega)} \le C\|y\|^2_{H^s(I;L^2(\Omega))} \tag{5.10}$$

Proof. Note that

$$\sum_{m=1}^{M} k_m^{-(2s-1)}\left\|[y_k(u)]_{m-1}\right\|^2_{L^2(\Omega)} \le 2\sum_{m=1}^{M} k_m^{-(2s-1)}\Big(\left\|[P_k y - y_k(u)]_{m-1}\right\|^2_{L^2(\Omega)} + \left\|[P_k y]_{m-1}\right\|^2_{L^2(\Omega)}\Big).$$

By Lemma 5.3, it suffices to estimate the second term on the right-hand side of the above estimate.

For $s \in [\frac{1}{2}, 1]$, we apply the Cauchy–Schwarz inequality and the interpolation operators in (5.3) to obtain

$$\begin{aligned} &\sum_{m=1}^{M} k_m^{-(2s-1)}\left\|[P_k y]_{m-1}\right\|^2_{L^2(\Omega)} \\ &\quad = \sum_{m=2}^{M} k_m^{-(2s-1)}\left\|\frac{1}{k_m}\int_{t_{m-1}}^{t_m} y(s)\,\mathrm{d}s - \frac{1}{k_{m-1}}\int_{t_{m-2}}^{t_{m-1}} y(s)\,\mathrm{d}s\right\|^2_{L^2(\Omega)} + k_1^{-2s+1}\left\|P_k^1 y - y_0\right\|^2_{L^2(\Omega)} \\ &\quad = \sum_{m=2}^{M} k_m^{-(2s-1)}\int_\Omega\left(\frac{1}{k_m}\int_{t_{m-1}}^{t_m} y(s)\,\mathrm{d}s - \frac{1}{k_{m-1}}\int_{t_{m-2}}^{t_{m-1}} y(s)\,\mathrm{d}s\right)^2 + k_1^{-2s+1}\int_\Omega\left(P_k^1 y - y_0\right)^2 \\ &\quad = \sum_{m=2}^{M} k_m^{-(2s-1)}\int_\Omega\left(\frac{1}{k_m}\int_{t_{m-1}}^{t_m} (y(s) - y(t_{m-1}))\,\mathrm{d}s + \frac{1}{k_{m-1}}\int_{t_{m-2}}^{t_{m-1}} (y(t_{m-1}) - y(s))\,\mathrm{d}s\right)^2 \\ &\quad\quad + k_1^{-2s+1}\int_\Omega\left(\frac{1}{k_1}\int_{t_0}^{t_1}(y(s) - y_0)\,\mathrm{d}s\right)^2\mathrm{d}x \\ &\quad \le 2\sum_{m=2}^{M} k_m^{-(2s-1)}\Bigg(\int_\Omega\left(\frac{1}{k_m}\int_{t_{m-1}}^{t_m}(y(s) - y(t_{m-1}))\,\mathrm{d}s\right)^2\mathrm{d}x \\ &\quad\quad + \int_\Omega\left(\frac{1}{k_{m-1}}\int_{t_{m-2}}^{t_{m-1}}(y(t_{m-1}) - y(s))\,\mathrm{d}s\right)^2\mathrm{d}x\Bigg) + k_1^{-2s+1}\int_\Omega\left(\frac{1}{k_1}\int_{t_0}^{t_1}(y(s) - y_0)\,\mathrm{d}s\right)^2\mathrm{d}x \end{aligned}$$

$$
\begin{aligned}
&\leq Ck^{-2s}\left(\sum_{m=1}^{M}\int_{\Omega}\int_{t_{m-1}}^{t_m}(y(s)-y(t_{m-1}))^2\,\mathrm{d}s\,\mathrm{d}x+\sum_{m=1}^{M}\int_{\Omega}\int_{t_{m-1}}^{t_m}(y(s)-y(t_m))^2\,\mathrm{d}s\,\mathrm{d}x\right)\\
&= Ck^{-2s}\left(\left\|y-\pi_k^l y\right\|_I^2+\|y-\pi_k^r y\|_I^2\right)\\
&\leq C\|y\|^2_{H^s(I;L^2(\Omega))},
\end{aligned}
$$

where we have used Lemma 5.1. This completes the proof. □

The following theorem employs the previously established lemmas and applies the Aubin–Nitsche trick to estimate the error in the DG(0) semi-discretization under the norm $\|\cdot\|_{L^2(I,L^2(\Omega))}$.

Theorem 5.5. *Let y and $y_k(u)$ be the solution of equations (1.2) and (4.8), respectively. For any given $s\in[\frac{1}{2},1]$ and sufficiently smooth u, y_0 and f, assume that $y\in H^s(I;L^2(\Omega))\cap L^2(I;H^{2s}(\Omega))$, then there holds*

$$
\|y-y_k(u)\|_{L^2(I;L^2(\Omega))}\leq C\left(k^s\left(\|u\|_{H^{s-\frac{1}{4}}(I;L^2(\Gamma))}+\|y\|_{H^s(I;L^2(\Omega))}\right)+k\|f\|_{L^2(I;L^2(\Omega))}\right). \tag{5.11}
$$

Proof. Set $e_k:=y-y_k(u)$ and let $z\in L^2(I;H^2(\Omega)\cap H_0^1(\Omega))\cap H^1(I;L^2(\Omega))$ be the solution of (2.3) with $g:=e_k$. We use the Aubin–Nitsche trick to estimate $\|e_k\|_{L^2(I,L^2(\Omega))}$. Note that

$$
\begin{aligned}
\|e_k\|^2_{L^2(I;L^2(\Omega))}&=(e_k,-\partial_t z-\Delta z)_I\\
&=\int_{\Omega_T}(-\partial_t z-\Delta z)y\,\mathrm{d}x\,\mathrm{d}t+\int_{\Omega_T}y_k(u)(\partial_t z+\Delta z)\,\mathrm{d}x\,\mathrm{d}t\\
&=\int_{\Sigma_T}(y_k(u)-u)\partial_n z\,\mathrm{d}s\,\mathrm{d}t+\int_{\Omega}y_0 z(0)\,\mathrm{d}x+\int_{\Omega_T}(fz+y_k(u)\partial_t z-\nabla y_k(u)\cdot\nabla z)\,\mathrm{d}x\,\mathrm{d}t\\
&=\int_{\Sigma_T}\left(\tilde{P}_k u-u\right)\partial_n z\,\mathrm{d}s\,\mathrm{d}t+\int_{\Omega}y_0 z(0)\,\mathrm{d}x+\int_{\Omega_T}(fz+y_k(u)\partial_t z-\nabla y_k(u)\cdot\nabla z)\,\mathrm{d}x\,\mathrm{d}t,
\end{aligned}
$$

where we have used Definition 2.1 and the equation (2.3). The first term in the above estimate can be bounded by projection errors, so it remains to consider other terms. Note that

$$
\begin{aligned}
&\int_{\Omega_T}(y_k(u)\partial_t z-\nabla y_k(u)\cdot\nabla z)\,\mathrm{d}x\,\mathrm{d}t=\sum_{n=1}^{M}\int_{t_{n-1}}^{t_n}(y_k(u),\partial_t z)\,\mathrm{d}t-(\nabla y_k(u),\nabla z)_I\\
&\quad=\sum_{n=1}^{M}(y_{k,n}(u),z(t_n)-z(t_{n-1}))-(\nabla y_k(u),\nabla z)_I\\
&\quad=\sum_{n=2}^{M}(y_{k,n-1}(u),z(t_{n-1}))-\sum_{n=2}^{M}(y_{k,n}(u),z(t_{n-1}))-\left(y_{k,0}^+(u),z(0)\right)-(\nabla y_k(u),\nabla z)_I\\
&\quad=-\sum_{n=2}^{M}([y_k(u)]_{n-1},z(t_{n-1}))-\left(y_{k,0}^+(u),z(0)\right)-(\nabla y_k(u),\nabla z)_I\\
&\quad=-\sum_{n=2}^{M}\left([y_k(u)]_{n-1},(P_k z)_{n-1}^+\right)-(\nabla y_k(u),\nabla P_k z)_I-\left(y_{k,0}^+(u),(P_k z)_0^+\right)\\
&\qquad+\sum_{n=2}^{M}\left([y_k(u)]_{n-1},(P_k z)_{n-1}^+-z(t_{n-1})\right)+\left(y_{k,0}^+(u),(P_k z)_0^+-z(0)\right)\\
&\quad=-B(y_k(u),P_k z)+\sum_{n=1}^{M}\left([y_k(u)]_{n-1},(P_k z)_{n-1}^+-z(t_{n-1})\right)+\left(y_0,(P_k z)_0^+-z(0)\right),
\end{aligned}
$$

where we have used (4.4), (4.8) and the definition of P_k. Combining the above estimates, we have

$$\|e_k\|^2_{L^2(I;L^2(\Omega))} = \int_{\Sigma_T} \Big(\tilde{P}_k u - u\Big)\partial_n z \,\mathrm{d}s\,\mathrm{d}t + \int_{\Omega_T} f(z - P_k z)\,\mathrm{d}x\,\mathrm{d}t + \sum_{n=1}^{M}\Big([y_k(u)]_{n-1}, (P_k z)^+_{n-1} - z(t_{n-1})\Big)$$
$$=: I_1 + I_2 + I_3. \tag{5.12}$$

Now we estimate the terms I_1, I_2 and I_3 in the above equality. Using the fact $u \in H^{s-\frac{1}{4}}(I;L^2(\Gamma))$ and the Cauchy–Schwarz inequality, we are led to

$$|I_1| = \left|\int_{\Sigma_T}\Big(\tilde{P}_k u - u\Big)\partial_n z\,\mathrm{d}s\,\mathrm{d}t\right| \le \left\|\tilde{P}_k u - u\right\|_{L^2(I;L^2(\Gamma))}\|\partial_n z - P_k\partial_n z\|_{L^2(I;L^2(\Gamma))} \le Ck^s\|u\|_{H^{s-\frac{1}{4}}(I;L^2(\Gamma))}\|e_k\|_I,$$

where the following two estimates have been used:

$$\left\|\tilde{P}_k u - u\right\|_{L^2(I;L^2(\Gamma))} \le Ck^{s-\frac{1}{4}}\|u\|_{H^{s-\frac{1}{4}}(I;L^2(\Gamma))}, \quad \|\partial_n z - P_k\partial_n z\|_{L^2(I;L^2(\Gamma))} \le Ck^{\frac{1}{4}}\|e_k\|_I.$$

The first estimate is the usual projection error estimate (cf. [7]), while the second can be derived as follows. Since $z \in H^1(I;L^2(\Omega)) \cap L^2(I;H^2(\Omega))$, we have $\partial_n z \in H^{\frac{1}{4}}(I;L^2(\Gamma))$ and

$$\|\partial_n z\|_{H^{\frac{1}{4}}(I;L^2(\Gamma))} \le C\big(\|z\|_{L^2(I;H^2(\Omega))} + \|z\|_{H^1(I;L^2(\Omega))}\big) \le C\|e_k\|_I,$$

where we have used the trace theorem [36] and the *a priori* estimate for parabolic equation [16]. Then we obtain

$$\|\partial_n z - P_k\partial_n z\|_{L^2(I;L^2(\Gamma))} \le Ck^{\frac{1}{4}}\|\partial_n z\|_{H^{\frac{1}{4}}(I;L^2(\Gamma))} \le Ck^{\frac{1}{4}}\|e_k\|_I.$$

Similarly, I_2 can be bounded as

$$|I_2| \le \|f\|_{L^2(I;L^2(\Omega))}\|z - P_k z\|_I \le Ck\|f\|_{L^2(I;L^2(\Omega))}\|z\|_{H^1(I;L^2(\Omega))} \le Ck\|f\|_{L^2(I;L^2(\Omega))}\|e_k\|_I.$$

The term I_3 can be estimated by

$$\begin{aligned}
|I_3| &= \left|\sum_{n=1}^{M}\Big([y_k(u)]_{n-1}, (P_k z)^+_{n-1} - z(t_{n-1})\Big)\right| \\
&\le \left(\sum_{n=1}^{M} k_n^{-(2s-1)}\big\|[y_k(u)]_{n-1}\big\|^2_{L^2(\Omega)}\right)^{\frac{1}{2}}\left(\sum_{n=1}^{M} k_n^{2s-1}\|P_k^n z - z(t_{n-1})\|^2_{L^2(\Omega)}\right)^{\frac{1}{2}} \\
&\le C\left(\sum_{n=1}^{M} k_n^{-(2s-1)}\big\|[y_k(u)]_{n-1}\big\|^2_{L^2(\Omega)}\right)^{\frac{1}{2}}\left(\sum_{n=1}^{M} k_n^{2s}\|\partial_t z\|^2_{L^2(I_n;L^2(\Omega))}\right)^{\frac{1}{2}} \\
&\le Ck^s\left(\sum_{n=1}^{M} k_n^{-(2s-1)}\big\|[y_k(u)]_{n-1}\big\|^2_{L^2(\Omega)}\right)^{\frac{1}{2}}\|\partial_t z\|_I \\
&\le Ck^s\left(\sum_{n=1}^{M} k_n^{-(2s-1)}\big\|[y_k(u)]_{n-1}\big\|^2_{L^2(\Omega)}\right)^{\frac{1}{2}}\|e_k\|_I,
\end{aligned}$$

where we have used the Cauchy–Schwarz inequality and the following estimate:

$$\|P_k^n z - z(t_{n-1})\|^2_{L^2(\Omega)} \le Ck_n\|\partial_t z\|_{L^2(I_n;L^2(\Omega))}, \qquad n = 1, \cdots, M.$$

Applying Lemma 5.4 in the above inequality, we finish the estimate for I_3. Inserting the estimates for I_1, I_2, I_3 into (5.12), we finally obtain the desired result of this theorem. □

Remark 5.6. By using the trace theorem in [36], we can rewrite the error estimate (5.11) as

$$\|y - y_k(u)\|_{L^2(I;L^2(\Omega))} \le C\big(k^s\big(\|y\|_{L^2(I;H^{2s}(\Omega))} + \|y\|_{H^s(I;L^2(\Omega))}\big) + k\|f\|_{L^2(I;L^2(\Omega))}\big), \tag{5.13}$$

where $y \in H^s(I; L^2(\Omega)) \cap L^2(I; H^{2s}(\Omega))$ for $s \in [\frac{1}{2}, 1]$. The last term in above estimate is a higher order term, which will vanish when $f \in \tilde{X}_k$.

When Ω is a convex polytope, we will consider the error estimate for the fully discrete solution of the optimal control problem (1.1). Therefore, as the first step to study finite element error estimates of fully discretizations to parabolic equations, we have to provide the following stability of temporal semi-discretization $y_k(u)$.

Proposition 5.7. *Let y and $y_k(u) \in X_k$ be the solution of equations (1.2) and (4.8), respectively. For any given $s \in [\frac{1}{2}, 1]$ and sufficiently smooth u, y_0 and f, assume that $y \in H^s(I; L^2(\Omega)) \cap L^2(I; H^{2s}(\Omega))$, then $y_k(u) \in L^2(I; H^{2s}(\Omega))$, and there holds*

$$\|y_k(u)\|_{L^2(I;H^{2s}(\Omega))} \le C\big(\|y\|_{L^2(I;H^{2s}(\Omega))} + \|y\|_{H^s(I;L^2(\Omega))}\big).$$

Proof. For each fixed $m = 1, \cdots, M$, we can rewrite the semi-discrete scheme (4.8) into the following time stepping scheme:

$$\begin{cases} (\nabla y_{k,m}(u), \nabla\varphi) + \Big(\frac{1}{k_m} y_{k,m}(u), \varphi\Big) = \Big(k_m \bar{f}_m + \frac{y_{k,m-1}(u)}{k_m}, \varphi\Big) & \forall\, \varphi \in H_0^1(\Omega),\\ y_{k,0}(u) = y_0 & \text{in } \Omega,\\ y_{k,m}(u)|_\Gamma = \tilde{P}_k^m u & \text{on } \Gamma, \end{cases}$$

where $\bar{f}_m = \frac{1}{k_m}\int_{t_{m-1}}^{t_m} f(s)\,\mathrm{d}s$. Since $\tilde{P}_k^m u \in H^{2s-\frac{1}{2}}(\Gamma)$ and $k_m\bar{f}_m + \frac{y_{k,m-1}(u)}{k_m} \in L^2(\Omega)$, we obtain $y_{k,m}(u) \in H^{2s}(\Omega),\ m = 1, \cdots, M$ by using the regularity of elliptic equations (cf. [16]).

Note that

$$\begin{aligned} \|y_k(u)\|_{L^2(I;H^{2s}(\Omega))} &\le \|P_k y\|_{L^2(I;H^{2s}(\Omega))} + \|P_k y - y_k(u)\|_{L^2(I;H^{2s}(\Omega))}\\ &\le C\big(\|y\|_{L^2(I;H^{2s}(\Omega))} + \|P_k y - y_k(u)\|_{L^2(I;H^{2s}(\Omega))}\big). \end{aligned} \tag{5.14}$$

Hence it suffices to estimate the second term of the above inequality. The proof can be divided into $u = 0$, $y_0 = 0$, $f \neq 0$ and $u \neq 0$, $y_0 \neq 0$, $f = 0$. We only provide the proof for the latter case. The remainder of proof is divided into three steps.

Step 1. First, we consider the case $s = \frac{1}{2}$. Choosing $\varphi|_{I_m} := (P_k y - y_k(u))|_{I_m}$ as the test function in the identity (5.8) of Lemma 5.3, we have

$$k_m\|\nabla(P_k y - y_k(u))_m\|^2_{L^2(\Omega)} + \big([P_k y - y_k(u)]_{m-1}, (P_k y - y_k(u))_m\big) = (g, P_k y - y_k(u))_{I_m}.$$

Using the Cauchy–Schwarz inequality we obtain

$$\begin{aligned} k_m\|\nabla(P_k y - y_k(u))_m\|^2_{L^2(\Omega)} + \frac{1}{2}\Big(\big\|[P_k y - y_k(u)]_{m-1}\big\|^2_{L^2(\Omega)} + \|(P_k y - y_k(u))_m\|^2_{L^2(\Omega)}\\ - \big\|(P_k y - y_k(u))_{m-1}\big\|^2_{L^2(\Omega)}\Big) \le \|g\|_{I_m}\|P_k y - y_k(u)\|_{I_m}. \end{aligned}$$

Summing up the above inequality for $m = 1, \cdots, M$ and applying the Cauchy–Schwarz inequality, we derive from Theorem 5.5 and Lemma 5.2 for the case $s = \frac{1}{2}$ that

$$\sum_{m=1}^M k_m\|\nabla(P_k y - y_k(u))_m\|^2_{L^2(\Omega)} + \sum_{m=1}^M \big\|[P_k y - y_k(u)]_{m-1}\big\|^2_{L^2(\Omega)}$$

$$
\begin{aligned}
&\le 2\left(\sum_{m=1}^{M} k_m \|g\|_{I_m}^2\right)^{\frac{1}{2}} \left(\sum_{m=1}^{M} k_m^{-1} \|P_k y - y_k(u)\|_{I_m}^2\right)^{\frac{1}{2}} \\
&\le C k^{-\frac{1}{2}} \|y\|_{H^{\frac{1}{2}}(I;L^2(\Omega))} (\|P_k y - y\|_I + \|y - y_k(u)\|_I) \\
&\le C \|y\|_{H^{\frac{1}{2}}(I;L^2(\Omega))} \Big(\|y\|_{H^{\frac{1}{2}}(I;L^2(\Omega))} + \|y\|_{L^2(I;H^1(\Omega))} \Big) \\
&\le C \Big(\|y\|_{H^{\frac{1}{2}}(I;L^2(\Omega))}^2 + \|y\|_{L^2(I;H^1(\Omega))}^2 \Big).
\end{aligned}
$$

Step 2. Next, we consider the case $s = 1$. Note that $(P_k y - y_k(u))|_{I_m} \in H^2(\Omega) \cap H_0^1(\Omega)$. Therefore, using the identity (5.8) of Lemma 5.3 one obtains the following identity:

$$
(-\Delta(P_k y - y_k(u)), \varphi)_{I_m} + \big([P_k y - y_k(u)]_{m-1}, \varphi\big) = (g, \varphi)_{I_m} \qquad \forall \varphi \in P_0\big(I_m; H_0^1(\Omega)\big). \tag{5.15}
$$

Since there are no spatial derivatives for the test function, the scheme (5.15) holds not only for all $\varphi \in P_0(I_m; H_0^1(\Omega))$, but also, by the density of $H_0^1(\Omega)$ in $L^2(\Omega)$, for all $\varphi \in P_0(I_m; L^2(\Omega))$.
Choosing $\varphi = -\Delta(P_k y - y_k(u))|_{I_m}$ as test functions in (5.15), we have

$$
\|\Delta(P_k y - y_k(u))\|_{I_m}^2 + \big(\nabla[P_k y - y_k(u)]_{m-1}, \nabla(P_k y - y_k(u))\big) = (g, -\Delta(P_k y - y_k(u)))_{I_m}.
$$

By using the Cauchy–Schwarz inequality, we obtain

$$
\begin{aligned}
\|\Delta(P_k y - y_k(u))\|_{I_m}^2 &+ \frac{1}{2}\Big(\|\nabla[P_k y - y_k(u)]_{m-1}\|_{L^2(\Omega)}^2 + \|\nabla(P_k y - y_k(u))_m\|_{L^2(\Omega)}^2 \\
&- \|\nabla(P_k y - y_k(u))_{m-1}\|_{L^2(\Omega)}^2 \Big) \\
&= (g, -\Delta(P_k y - y_k(u)))_{I_m} \le \frac{1}{2}\Big(\|g\|_{I_m}^2 + \|\Delta(P_k y - y_k(u))\|_{I_m}^2 \Big),
\end{aligned}
$$

that is,

$$
\begin{aligned}
\|\Delta(P_k y - y_k(u))\|_{I_m}^2 &+ \big(\|\nabla[P_k y - y_k(u)]_{m-1}\|_{L^2(\Omega)}^2 + \|\nabla(P_k y - y_k(u))_m\|_{L^2(\Omega)}^2 \\
&- \|\nabla(P_k y - y_k(u))_{m-1}\|_{L^2(\Omega)}^2 \big) \le \|g\|_{I_m}^2.
\end{aligned}
$$

Summing up the above inequality for $m = 1, 2, \cdots, M$, using $(P_k y - y_k(u))_0 = 0$ and Lemma 5.2 for the case $s = 1$, we obtain

$$
\|\Delta(P_k y - y_k(u))\|_I^2 + \sum_{m=1}^{M} \|\nabla[P_k y - y_k(u)]_{m-1}\|_{L^2(\Omega)}^2 \le C \|y\|_{H^1(I;L^2(\Omega))}^2.
$$

Utilizing

$$
\|P_k y - y_k(u)\|_{L^2(I;H^2(\Omega))}^2 \le C \|\Delta(P_k y - y_k(u))\|_I^2 \le C \|y\|_{H^1(I;L^2(\Omega))}^2
$$

and (5.14) finishes the proof of the case $s = 1$.

Step 3. Now, the case $s \in (\frac{1}{2}, 1)$ follows from interpolations. To do this, let $\mathcal{T}$ be a linear operator on the stability of $y_k(u)$ in terms of y. The above two cases imply that $\|\mathcal{T} y\|_{H^{\frac{1}{2}}(I;L^2(\Omega)) \cap L^2(I;H^1(\Omega)) \to L^2(I;H^1(\Omega))}$ and $\|\mathcal{T} y\|_{H^1(I;L^2(\Omega)) \cap L^2(I;H^2(\Omega)) \to L^2(I;H^2(\Omega))}$ are both bounded. Therefore, $\mathcal{T}$ is also bounded from $H^s(I; L^2(\Omega)) \cap L^2(I; H^{2s}(\Omega))$ to $L^2(I; H^{2s}(\Omega))$ by the interpolation estimate in [36], where we need the following two interpolation spaces:

$$
\begin{aligned}
\big[L^2\big(I; H^2(\Omega)\big), L^2\big(I; H^1(\Omega)\big)\big]_{2(1-s)} &= L^2\big(I; H^{2s}(\Omega)\big), \\
\Big[L^2\big(I; H^2(\Omega)\big) \cap H^1\big(I; L^2(\Omega)\big), L^2\big(I; H^1(\Omega)\big) \cap H^{\frac{1}{2}}\big(I; L^2(\Omega)\big)\Big]_{2(1-s)} &= L^2\big(I; H^{2s}(\Omega)\big) \cap H^s\big(I; L^2(\Omega)\big).
\end{aligned}
$$

This finishes the proof. ∎

Remark 5.8. For any given $s \in [\frac{1}{2}, 1]$ and sufficiently smooth u, y_0 and f, assume that $y(u) \in H^s(I; L^2(\Omega)) \cap L^2(I; H^{2s}(\Omega))$ holds, then Theorem 5.5 ensures the convergence order $O(k^s)$ for the temporal semi-discretization of parabolic equations. However, for Dirichlet boundary control problems studied in current paper, according to Theorem 3.2 the state and the control have only limited regularity, *i.e.*, $\bar{y} \in H^s(I; L^2(\Omega)) \cap L^2(I; H^{2s}(\Omega))$ and $\bar{u} \in H^{s'}(I; L^2(\Omega)) \cap L^2(I; H^{\min\{2s',1\}}(\Gamma))$ with $s \in [\frac{1}{2}, \frac{3}{4})$ and $s - \frac{1}{4} \le s' \le s$. Therefore, we only use the result for the case $s \in [\frac{1}{2}, \frac{3}{4})$ of Theorem 5.5 to derive the error estimate for optimal control problems.

5.2. Analysis of the spatial discretization error

When Ω is a convex polytope, we consider the spatial discretization error estimate for the solution of (4.19) under the norm $\|\cdot\|_{L^2(I,L^2(\Omega))}$. For this purpose, we use the Aubin–Nitsche technique to convert the finite element estimate into projection errors.

We remark that the projection operators P_k and $\tilde{P}_k$ defined respectively in (5.1) and (4.6) can be extended such that $P_k : L^2(I; H^1(\Omega)) \to X_k$ and $\tilde{P}_k : L^2(I; H^{\frac{1}{2}}(\Gamma)) \to X_k(\Gamma)$ are well-defined. The following lemma establishes the relationships between these projections, which will be used in the error estimate of the spatial discretization.

Lemma 5.9. *The projections P_k, $\tilde{P}_k$, P_h, $\tilde{P}_h$, P_{kh} and $\tilde{P}_{kh}$ in Definition 4.5 satisfy*

$$P_{kh} = P_k P_h, \quad \tilde{P}_{kh} = \tilde{P}_k \tilde{P}_h, \quad \gamma_0 P_k = \tilde{P}_k \gamma_0,$$

where $\gamma_0 : L^2(I; H^1(\Omega)) \to L^2(I; H^{\frac{1}{2}}(\Gamma))$ is the trace operator. Particularly, when $\omega \in L^2(I; H^1(\Omega))$ and $\tilde{\omega} \in L^2(I; H^{\frac{1}{2}}(\Gamma))$, there hold

$$P_{kh}\omega = P_h P_k \omega, \quad \tilde{P}_{kh}\tilde{\omega} = \tilde{P}_h \tilde{P}_k \tilde{\omega}.$$

Proof. We first prove $P_{kh} = P_k P_h$. For any given $\omega \in L^2(I; L^2(\Omega))$ and $v_{kh} \in X_{kh}$, there holds

$$\begin{aligned}\int_\Omega \int_0^T (\omega - P_k P_h \omega) v_{kh} \, \mathrm{d}t \, \mathrm{d}x &= \int_\Omega \int_0^T (P_k \omega - P_k P_h \omega) v_{kh} \, \mathrm{d}t \, \mathrm{d}x \\ &= \sum_{m=1}^M \int_\Omega \int_{I_m} P_k(\omega - P_h \omega) v_{kh} \, \mathrm{d}t \, \mathrm{d}x \\ &= \sum_{m=1}^M \int_\Omega \int_{I_m} (\omega - P_h \omega) \, \mathrm{d}t v_{kh} \, \mathrm{d}x \\ &= (\omega - P_h \omega, v_{kh})_{L^2(I;L^2(\Omega))},\end{aligned}$$

where we have used the orthogonality of P_k. Note that $X_{kh} \subseteq L^2(I; V_h)$, if follows from the definition of P_h that the above identity is zero. By the definition of P_{kh} and the uniqueness of $P_{kh}\omega$, we obtain $P_{kh}\omega = P_k P_h \omega$ for any $\omega \in L^2(I; L^2(\omega))$, *i.e.*, $P_{kh} = P_k P_h$. Similarly, we can show that $\tilde{P}_{kh} = \tilde{P}_h \tilde{P}_k$.

Now, we claim $P_{kh}\omega = P_h P_k \omega$ for any $\omega \in L^2(I; H^1(\Omega))$. For any $v_{kh} \in X_{kh}$, there holds

$$\begin{aligned}\int_\Omega \int_0^T (\omega - P_h P_k \omega) v_{kh} \, \mathrm{d}t \, \mathrm{d}x &= \int_\Omega \int_0^T [(\omega - P_h \omega) + P_h(\omega - P_k \omega)] v_{kh} \, \mathrm{d}t \, \mathrm{d}x \\ &= \sum_{m=1}^M \int_\Omega \int_{I_m} P_h(\omega - P_k \omega) v_{kh} \, \mathrm{d}t \, \mathrm{d}x \\ &= \sum_{m=1}^M \int_{I_m} \int_\Omega (\omega - P_k \omega) \, \mathrm{d}x v_{kh} \, \mathrm{d}t \\ &= (\omega - P_k \omega, v_{kh})_{L^2(I;L^2(\Omega))},\end{aligned}$$

where the orthogonality of P_h is used. By the orthogonality of P_k and the inclusion $X_{kh} \subseteq X_k$, the above equality is zero. Using the definition of P_{kh} and the uniqueness of $P_{kh}\omega$, we obtain $P_{kh}\omega = P_h P_k \omega$ for any $\omega \in L^2(I; H^1(\Omega))$. Similar to the above process, there holds $\tilde{P}_{kh}\tilde{\omega} = \tilde{P}_h \tilde{P}_k \tilde{\omega}$ for any $\omega \in L^2(I; H^{\frac{1}{2}}(\Gamma))$. The relation $\gamma_0 P_k = \tilde{P}_k \gamma_0$ is obvious. This completes the proof. □

For any given $\omega \in L^2(I; H^1(\Omega))$, let $P_h\omega$ be the projection of ω defined in Definition 4.5. We split $P_h\omega$ into

$$P_h\omega = P_h^I\omega + P_h^B\omega, \quad P_h^I\omega \in L^2\big(I; V_h^0\big), \quad P_h^B\omega \in L^2\big(I; V_h^\partial\big),$$

where $L^2(I; V_h^0)$ and $L^2(I; V_h)$ are function spaces defined in I, and valued in V_h^0 and V_h respectively, while V_h^0 and V_h^∂ are subspaces of V_h spanned by interior node basis functions and boundary node basis functions, respectively.

Then we define a modified projection

$$\hat{P}_h : L^2\big(I; H^1(\Omega)\big) \to L^2(I; V_h), \quad \omega \mapsto \hat{P}_h\omega$$

such that $\hat{P}_h\omega := P_h^I\omega + P_h^\partial\omega$, where $P_h^\partial\omega$ satisfies $P_h^\partial\omega \in L^2(I; V_h^\partial)$ and $P_h^\partial\omega|_{I\times\Gamma} := \tilde{P}_h\gamma_0\omega$, *i.e.*,

$$\big(\omega|_{I\times\Gamma} - P_h^\partial\omega|_{I\times\Gamma}, v_h\big)_{L^2(I;L^2(\Gamma))} = 0 \qquad \forall v_h \in L^2(I; V_h(\Gamma)).$$

Note that the above projection is obtained by replacing the boundary components of the usual L^2-projection with the boundary L^2-projection. Therefore, the orthogonality is only valid on the boundary. In the following proposition, we study the projection error estimation of the temporal DG(0) discretization for parabolic equations.

Proposition 5.10. *Let $y_k(u) \in X_k$ be the solution of (4.8). For given $s \in [\frac{1}{2}, 1]$, if $y_k(u) \in L^2(I; H^{2s}(\Omega))$, then there holds*

$$\Big\|y_k(u) - \hat{P}_h y_k(u)\Big\|_I + h\Big\|\nabla\Big(y_k(u) - \hat{P}_h y_k(u)\Big)\Big\|_I \le Ch^{2s}\|y_k(u)\|_{L^2(I;H^{2s}(\Omega))}. \tag{5.16}$$

Proof. By the error estimate of the L^2-projection P_h, we have

$$\begin{aligned}
&\Big\|y_k(u) - \hat{P}_h y_k(u)\Big\|_I + h\Big\|\nabla(y_k(u) - \hat{P}_h y_k(u))\Big\|_I \\
&\quad \le (\|y_k(u) - P_h y_k(u)\|_I + h\|\nabla(y_k(u) - P_h y_k(u))\|_I) + \Big\|P_h y_k(u) - \hat{P}_h y_k(u)\Big\|_I + h\Big\|\nabla\Big(P_h y_k(u) - \hat{P}_h y_k(u)\Big)\Big\|_I \\
&\quad \le Ch^{2s}\|y_k(u)\|_{L^2(I;H^{2s}(\Omega))} + \Big\|P_h y_k(u) - \hat{P}_h y_k(u)\Big\|_I + h\Big\|\nabla\Big(P_h y_k(u) - \hat{P}_h y_k(u)\Big)\Big\|_I.
\end{aligned}$$

Then we estimate the latter two terms of the above inequality as follows:

$$\begin{aligned}
&\Big\|P_h y_k(u) - \hat{P}_h y_k(u)\Big\|_I^2 + h^2\Big\|\nabla\Big(P_h y_k(u) - \hat{P}_h y_k(u)\Big)\Big\|_I^2 \\
&\quad = \sum_{m=1}^M k_m \Big\|P_h^m y_k(u) - \hat{P}_h^m y_k(u)\Big\|_{L^2(\Omega)}^2 + h^2 \sum_{m=1}^M k_m \Big\|\nabla\Big(P_h^m y_k(u) - \hat{P}_h^m y_k(u)\Big)\Big\|_{L^2(\Omega)}^2 \\
&\quad = \sum_{m=1}^M k_m \sum_{j=1}^{M_h}\Big(P_h^m y_k(u)(x_j) - \hat{P}_h^m y_k(u)(x_j)\Big)^2 \|\varphi_j\|_{L^2(\Omega)}^2 \\
&\qquad + h^2 \sum_{m=1}^M k_m \sum_{j=1}^{M_h}\Big(P_h^m y_k(u)(x_j) - \hat{P}_h^m y_k(u)(x_j)\Big)^2 \|\nabla\varphi_j\|_{L^2(\Omega)}^2 \\
&\quad \le Ch^N \sum_{m=1}^M k_m \sum_{j=1}^{M_h}\Big(P_h^m y_k(u)(x_j) - \hat{P}_h^m y_k(u)(x_j)\Big)^2
\end{aligned}$$

$$\leq Ch\sum_{m=1}^{M}k_m\Big\|P_h^m y_k(u)-\hat{P}_h^m y_k(u)\Big\|_{L^2(\Gamma)}^2$$
$$= Ch\Big\|P_h y_k(u)-\hat{P}_h y_k(u)\Big\|_{L^2(I;L^2(\Gamma))}^2,$$

where $\hat{P}_h^m y_k(u) := \hat{P}_h y_k(u)|_{I_m}$, $m=1,\cdots,M$, and $\{\varphi_j\}_{j=1}^{M_h}$ is a family of basis functions corresponding to the boundary nodes $\{x_j\}_{j=1}^{M_h}$. Subsequently,

$$\begin{aligned}
&Ch^{\frac{1}{2}}\Big\|P_h y_k(u)-\hat{P}_h y_k(u)\Big\|_{L^2(I;L^2(\Gamma))}\\
&\leq Ch^{\frac{1}{2}}\|P_h y_k(u)-y_k(u)\|_{L^2(I;L^2(\Gamma))}+Ch^{\frac{1}{2}}\Big\|y_k(u)-\hat{P}_h y_k(u)\Big\|_{L^2(I;L^2(\Gamma))}\\
&\leq C\bigg(\|y_k(u)-P_h y_k(u)\|_{L^2(I;L^2(\Omega))}+h\|y_k(u)-P_h y_k(u)\|_{L^2(I;H^1(\Omega))}+h^{\frac{1}{2}}\Big\|\tilde{P}_k u-\tilde{P}_h\tilde{P}_k u\Big\|_{L^2(I;L^2(\Gamma))}\bigg)\\
&\leq Ch^{2s}\|y_k(u)\|_{L^2(I;H^{2s}(\Omega))}+Ch^{\frac{1}{2}}\Big\|\Big(I-\tilde{P}_h\Big)\tilde{P}_k u)\Big\|_{L^2(I;L^2(\Gamma))}\\
&\leq Ch^{2s}\bigg(\|y_k(u)\|_{L^2(I;H^{2s}(\Omega))}+\Big\|\tilde{P}_k u\Big\|_{L^2(I;H^{2s-\frac{1}{2}}(\Gamma))}\bigg)\\
&\leq Ch^{2s}\|y_k(u)\|_{L^2(I;H^{2s}(\Omega))},
\end{aligned}$$

where we have used Lemma 5.9, the trace theorem, the stability and approximation properties of projection operators. This completes the proof. □

The fully discrete scheme (4.19) for parabolic equations is called the DG(0)-CG(1) scheme, which can be viewed as the spatial discretization of (4.8). Since $X_{kh}^0\subseteq X_k^0$, there holds the following orthogonality relation:

$$B(y_k(u)-y_{kh}(u),\varphi_{kh})=B(y-y_{kh}(u),\varphi_{kh})=0\qquad\forall\varphi_{kh}\in X_{kh}^0.$$

Theorem 5.11. *Let $y_k(u)\in X_k$ and $y_{kh}(u)\in X_{kh}$ be the solution of (4.8) and (4.19), respectively. For given $s\in[\frac{1}{2},1]$, if $y_k(u)\in L^2(I;H^{2s}(\Omega))$, then we have*

$$\|y_k(u)-y_{kh}(u)\|_{L^2(I;L^2(\Omega))}\leq Ch^{2s}\|y_k(u)\|_{L^2(I;H^{2s}(\Omega))}.$$

Proof. Let $e_h := y_k(u)-y_{kh}(u)=\eta_h+\zeta_h$, $\eta_h := y_k(u)-\hat{P}_h y_k(u)$, $\zeta_h := \hat{P}_h y_k(u)-y_{kh}(u)$. Then

$$\|e_h\|_{L^2(I;L^2(\Omega))}^2=(e_h,\eta_h)_{L^2(I;L^2(\Omega))}+(e_h,\zeta_h)_{L^2(I;L^2(\Omega))}.\tag{5.17}$$

By using the Cauchy–Schwarz inequality, it suffices to estimate the second term. To this end, we use the Aubin–Nitsche technique and let z be the solution of (2.3) with $g := e_h$.

Let $z_k\in X_k^0$ and $z_{kh}\in X_{kh}^0$ be the temporal semi-discretization and fully discretization of the equation (2.3). By Lemma 5.9, there holds $\zeta_h\in X_{kh}^0$. Note that $X_{kh}^0\subseteq X_k^0$, then

$$\begin{aligned}
-(e_h,\zeta_h)_{L^2(I;L^2(\Omega))}&=-B(\zeta_h,z_{kh})=B(\eta_h,z_{kh})\\
&=(\nabla\eta_h,\nabla z_{kh})_I-\sum_{m=1}^{M-1}\Big(\eta_{h,m}^-,[z_{kh}]_m\Big)+\Big(\eta_{h,M}^-,z_{kh,M}^-\Big)\\
&=(\nabla\eta_h,\nabla z_k)_I-(\nabla\eta_h,\nabla(z_k-z_{kh}))_I-\sum_{m=1}^{M}\Big(\eta_{h,m}^-,[z_{kh}]_m\Big)
\end{aligned}$$

$$
\begin{aligned}
&= (\eta_h, \partial_n z_k)_{L^2(I;L^2(\Gamma))} - (\eta_h, \Delta z_k)_I - (\nabla \eta_h, \nabla(z_k - z_{kh}))_I - \sum_{m=1}^{M} \left(\eta_{h,m}^-, [z_{kh}]_m\right) \\
&= \left(u - \tilde{P}_h u, \partial_n z_k - \tilde{P}_h \partial_n z_k\right)_{L^2(I;L^2(\Gamma))} - (\eta_h, \Delta z_k)_I \\
&\quad - (\nabla \eta_h, \nabla(z_k - z_{kh}))_I - \sum_{m=1}^{M} \left(\eta_{h,m}^-, [z_{kh}]_m\right),
\end{aligned}
$$

where $[z_{kh}]_M = -z_{kh,M}^-$, and we have used $(\eta_h, \partial_n z_k)_{L^2(I;L^2(\Gamma))} = (u - \tilde{P}_h u, \partial_n z_k - \tilde{P}_h \partial_n z_k)_{L^2(I;L^2(\Gamma))}$. In fact, it follows from Lemma 5.9 and the orthogonality of the projection $\tilde{P}_h$ that

$$
\begin{aligned}
(\eta_h, \partial_n z_k)_{L^2(I;L^2(\Gamma))} &= \left(y_k(u) - \tilde{P}_h y_k(u), \partial_n z_k\right)_{L^2(I;L^2(\Gamma))} \\
&= \left(\tilde{P}_k u - \tilde{P}_h \tilde{P}_k u, \partial_n z_k\right)_{L^2(I;L^2(\Gamma))} \\
&= \left(\tilde{P}_k \left(u - \tilde{P}_h u\right), \partial_n z_k\right)_{L^2(I;L^2(\Gamma))} \\
&= \left(u - \tilde{P}_h u, \partial_n z_k - \tilde{P}_h \partial_n z_k\right)_{L^2(I;L^2(\Gamma))}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\left|(e_h, \zeta_h)_{L^2(I;L^2(\Omega))}\right| &\leq \|\eta_h\|_I \|\Delta z_k\|_I + \left\|u - \tilde{P}_h u\right\|_{L^2(I;L^2(\Gamma))} \left\|\left(I - \tilde{P}_h\right)\partial_n z_k\right\|_{L^2(I;L^2(\Gamma))} \\
&\quad + \|\nabla \eta_h\|_I \|\nabla(z_k - z_{kh})\|_I + \left(\sum_{m=1}^{M} k_m \|\eta_{h,m}\|_{L^2(\Omega)}^2\right)^{\frac{1}{2}} \left(\sum_{m=1}^{M} k_m^{-1} \|[z_{kh}]_m\|_{L^2(\Omega)}^2\right)^{\frac{1}{2}} \\
&\leq C\left(\|\eta_h\|_I + h^{\frac{1}{2}} \left\|u - \tilde{P}_h u\right\|_{L^2(I;L^2(\Gamma))} + h\|\nabla \eta_h\|_I\right) \|e_h\|_I.
\end{aligned}
$$

Here we have used the following three estimates:

$$
\begin{aligned}
\|\nabla(z_k - z_{kh})\|_I &\leq Ch\|e_h\|_I, \\
\left\|\left(I - \tilde{P}_h\right)\partial_n z_k\right\|_{L^2(I;L^2(\Gamma))} &\leq Ch^{\frac{1}{2}}\|e_h\|_I, \\
\|\Delta z_k\|_I + \|z_k\|_I + \left(\sum_{m=1}^{M} k_m^{-1} \|[z_{kh}]_m\|_{L^2(\Omega)}^2\right)^{\frac{1}{2}} &\leq C\|e_h\|_I,
\end{aligned}
$$

where the third one can be found in Corollaries 4.2 and 4.7 in [38]. Therefore, we only verify the first two estimates. The first estimate can be derived by using the inverse estimate and the projection error estimate:

$$
\begin{aligned}
\|\nabla(z_k - z_{kh})\|_I &\leq \|\nabla(z_k - P_h z_k)\|_I + Ch^{-1}\|P_h z_k - z_{kh}\|_I \\
&\leq Ch\left\|\nabla^2 z_k\right\|_I + Ch^{-1}(\|P_h z_k - z_k\|_I + \|z_k - z_{kh}\|_I) \\
&\leq Ch\|z_k\|_{L^2(I;H^2(\Omega))} \\
&\leq Ch(\|z_k\|_I + \|\Delta z_k\|_I) \\
&\leq Ch\|e_h\|_I,
\end{aligned}
$$

where the error estimate and the stability of the temporal semi-discretization (cf. [38], Thm. 5.1 and Cor. 4.2) have been applied. Again, the stability of the temporal semi-discretization (cf. [38], Cor. 4.2) implies the second estimate

$$\left\|\left(I-\tilde{P}_h\right)\partial_n z_k\right\|_{L^2(I;L^2(\Gamma))} \leq Ch^{\frac{1}{2}}\|\partial_n z_k\|_{L^2(I;H^{\frac{1}{2}}(\Gamma))} \leq Ch^{\frac{1}{2}}\|z_k\|_{L^2(I;H^2(\Omega))} \leq Ch^{\frac{1}{2}}\|e_h\|_I.$$

Finally, combining (5.17), Proposition 5.10, the Cauchy–Schwarz and Young's inequalities, and the projection error estimate and trace theorem completes the proof. □

Remark 5.12. Combining Theorem 5.5, Proposition 5.7 and Theorem 5.11, we obtain *a priori* error estimate $O(h^{2s}+k^s)$ for the fully discretization of parabolic equations with inhomogeneous Dirichlet data posed on convex polytopes for given $s\in[\frac{1}{2},1]$, if the solution admits the regularity $L^2(I;H^{2s}(\Omega))\cap H^s(I;L^2(\Omega))$. These estimates improve the result in [17] and remove the mesh size condition $k=O(h^2)$.

6. Error estimates for the optimal control problem

In this section, we present the primary findings of this article, centering on the error estimates between the solutions of the continuous optimal control problem and their discrete counterparts. The first subsection delivers the optimal order of convergence for the temporal semi-discrete optimal control problem (4.7), considering both smooth and polytopal domains for Ω. The subsequent subsection is devoted to the error estimation between the continuous problem (1.1) and the fully discrete optimal control problem (4.18), particularly when dealing with a polytopal domain.

6.1. Estimates for the semi-discrete adjoint state

Lemma 6.1. *Let $\bar{z}\in L^2(I;H^2(\Omega)\cap H_0^1(\Omega))\cap H^1(I;L^2(\Omega))$ be the adjoint state of the optimal control problem (1.1) and $\hat{z}_k$ be its semi-discrete finite element solution, i.e., $\hat{z}_k\in X_k^0$ satisfies*

$$B(\varphi_k,\hat{z}_k)=(\bar{y}-y_d,\varphi_k)_I \qquad \forall\varphi_k\in X_k^0. \tag{6.1}$$

Let $\pi_k^r\bar{z}\in X_k$ be the pointwise in time interpolation of $\bar{z}$ defined by (5.3). Then we have

$$\|\nabla(\bar{z}-\hat{z}_k)\|_I \leq C\|\nabla(\bar{z}-\pi_k^r\bar{z})\|_I. \tag{6.2}$$

Proof. Observing that

$$\|\nabla(\bar{z}-\hat{z}_k)\|_I \leq \|\nabla(\bar{z}-\pi_k^r\bar{z})\|_I+\|\nabla(\pi_k^r\bar{z}-\hat{z}_k)\|_I, \tag{6.3}$$

we only need to estimate the second term on the right-hand side of the above inequality.

Adding the expression (4.4) and the dual expression (4.5) of the bilinear form B for $v=w=\pi_k^r\bar{z}-\hat{z}_k\in X_k$, we obtain

$$\begin{aligned}\|\nabla(\pi_k^r\bar{z}-\hat{z}_k)\|_I^2 &\leq B(\pi_k^r\bar{z}-\hat{z}_k,\pi_k^r\bar{z}-\hat{z}_k)\\ &= B(\pi_k^r\bar{z}-\hat{z}_k,\pi_k^r\bar{z}-\bar{z})\\ &= (\nabla(\pi_k^r\bar{z}-\bar{z}),\nabla(\pi_k^r\bar{z}-\hat{z}_k))_I-\sum_{m=1}^{M}\Big((\pi_k^r\bar{z}-\bar{z})_m^-,[\pi_k^r\bar{z}-\hat{z}_k]_m\Big)\\ &\leq (\nabla(\pi_k^r\bar{z}-\bar{z}),\nabla(\pi_k^r\bar{z}-\hat{z}_k))_I\\ &\leq \|\nabla(\bar{z}-\pi_k^r\bar{z})\|_I\|\nabla(\hat{z}_k-\pi_k^r\bar{z})\|_I,\end{aligned}$$

where we have used $\bar{z}(t_m^-)=\pi_k^r\bar{z}|_{I_m}=\bar{z}(t_m)$. Then, combining the above estimate with (6.3) gives the desired estimate (6.2). This completes the proof. □

Lemma 6.2. *Assuming that Ω is a bounded smooth domain and the compatibility condition (2.8) holds for the adjoint equation (3.5) with $g := \bar{y} - y_d$, where $\bar{y}$ is the optimal state and $y_d \in L^2(I; H^1(\Omega)) \cap H^{\frac{1}{2}}(I; L^2(\Omega))$. Let $\hat{z}_k \in X_k$ be the finite element approximation of $\bar{z}$ defined by (6.1), then we have $\hat{z}_k \in L^2(I; H^3(\Omega) \cap H_0^1(\Omega))$, and the following estimate holds:*

$$\sum_{m=1}^{M} k_m^{-1} \|\nabla[\hat{z}_k]_m\|_{L^2(\Omega)}^2 \leq C\|\bar{z}\|_{H^1(I;H^1(\Omega))}^2, \tag{6.4}$$

where $[\hat{z}_k]_M := -\hat{z}_{k,M-1} = -\hat{z}_k|_{I_M}$.

Proof. For each fixed $m = 1, \cdots, M$, we denote by $\hat{z}_{k,m-1} := \hat{z}_k|_{I_m}$ and rewrite the equation (6.1) as the following time-stepping scheme: For any $\varphi_k \in P_0(I_m, H_0^1(\Omega))$ there holds

$$\begin{cases} -\left(\frac{\hat{z}_{k,m} - \hat{z}_{k,m-1}}{k_m}, \varphi_k\right)_{I_m} + (\nabla \hat{z}_{k,m-1}, \nabla \varphi_k)_{I_m} = (P_k(\bar{y} - y_d), \varphi_k)_{I_m}, \\ \hat{z}_{k,M} = 0. \end{cases} \tag{6.5}$$

Since $\frac{\hat{z}_{k,m}}{k_m} + P_k(\bar{y} - y_d) \in H^1(\Omega)$, there holds $\hat{z}_{k,m-1} \in H^3(\Omega) \cap H_0^1(\Omega)$ by the regularity result of elliptic equations (cf. [20], Thm. 1.8, Chap. I).

On the other hand, the adjoint equation (3.5) satisfies the following weak form on I_m: for any $\varphi_k \in P_0(I_m, H_0^1(\Omega))$ there holds

$$\begin{cases} -\left(\frac{[\pi_k^l \bar{z}]_m}{k_m}, \varphi_k\right)_{I_m} + (\nabla \bar{z}, \nabla \varphi_k)_{I_m} = (P_k(\bar{y} - y_d), \varphi_k)_{I_m}, \\ \bar{z}(t_M) = 0, \end{cases} \tag{6.6}$$

where $[\pi_k^l \bar{z}]_m = \bar{z}(t_m) - \bar{z}(t_{m-1})$. Subtracting (6.6) from (6.5), we obtain

$$-\left(\frac{\left[\pi_k^l \bar{z} - \hat{z}_k\right]_m}{k_m}, \varphi_k\right)_{I_m} + (\nabla(\bar{z} - \hat{z}_{k,m-1}), \nabla \varphi_k)_{I_m} = 0 \qquad \forall \varphi_k \in P_0\big(I_m, H_0^1(\Omega)\big),$$

i.e.,

$$-\left(\frac{\left[\pi_k^l \bar{z} - \hat{z}_k\right]_m}{k_m}, \varphi_k\right)_{I_m} + \big(\nabla\big(\pi_k^l \bar{z} - \hat{z}_{k,m-1}\big), \nabla \varphi_k\big)_{I_m} = \big(\nabla\big(\pi_k^l \bar{z} - \bar{z}\big), \nabla \varphi_k\big)_{I_m}.$$

Taking $\varphi_k = [\pi_k^l \bar{z} - \hat{z}_k]_m$ in the above equality, we obtain

$$\big(\nabla\big(\pi_k^l \bar{z} - \hat{z}_{k,m-1}\big), \nabla\big[\pi_k^l \bar{z} - \hat{z}_k\big]_m\big)_{I_m} - \frac{\big\|\big[\pi_k^l \bar{z} - \hat{z}_k\big]_m\big\|_{I_m}^2}{k_m} = \big(\nabla\big(\pi_k^l \bar{z} - \bar{z}\big), \nabla\big[\pi_k^l \bar{z} - \hat{z}_k\big]_m\big)_{I_m}. \tag{6.7}$$

Note that

$$\begin{aligned} \big(\nabla\big(\pi_k^l \bar{z} - \hat{z}_{k,m-1}\big), \nabla\big[\pi_k^l \bar{z} - \hat{z}_k\big]_m\big)_{I_m} &= -\frac{1}{2}\big\|\nabla\big[\pi_k^l \bar{z} - \hat{z}_k\big]_m\big\|_{I_m}^2 \\ &\quad + \frac{1}{2}\left(\Big\|\nabla\big(\pi_k^l \bar{z} - \hat{z}_k\big)_m^+\Big\|_{I_m}^2 - \Big\|\nabla\big(\pi_k^l \bar{z} - \hat{z}_k\big)_m^-\Big\|_{I_m}^2\right), \end{aligned}$$

where $(\pi_k^l \bar{z} - \hat{z}_k)_M^+ := 0$. Therefore, the equality (6.7) is equivalent to

$$-\frac{\big\|\big[\pi_k^l \bar{z} - \hat{z}_k\big]_m\big\|_{I_m}^2}{k_m} - \frac{1}{2}\left(\big\|\nabla\big[\pi_k^l \bar{z} - \hat{z}_k\big]_m\big\|_{I_m}^2 - \Big\|\nabla\big(\pi_k^l \bar{z} - \hat{z}_k\big)_m^+\Big\|_{I_m}^2 + \Big\|\nabla\big(\pi_k^l \bar{z} - \hat{z}_k\big)_m^-\Big\|_{I_m}^2\right)$$

$$= \big(\nabla\big(\pi_k^l\bar z-\bar z\big),\nabla\big[\pi_k^l\bar z-\hat z_k\big]_m\big)_{I_m}. \tag{6.8}$$

Thus, we obtain the following estimate:

$$\frac{\big\|\big[\pi_k^l\bar z-\hat z_k\big]_m\big\|_{I_m}^2}{k_m}+\frac12\big\|\nabla\big[\pi_k^l\bar z-\hat z_k\big]_m\big\|_{I_m}^2-\big\|\nabla\big(\big(\pi_k^l\bar z\big)(t_m)-\hat z_{k,m}\big)\big\|_{I_m}^2 + \big\|\nabla\big(\big(\pi_k^l\bar z\big)(t_{m-1})-\hat z_{k,m-1}\big)\big\|_{I_m}^2\le C\big\|\nabla\big(\pi_k^l\bar z-\bar z\big)\big\|_{I_m}^2. \tag{6.9}$$

Using the condition (4.1) we are led to

$$\begin{aligned}\frac{\big\|\big[\pi_k^l\bar z-\hat z_k\big]_m\big\|_{I_m}^2}{k_m}+\frac12\big\|\nabla\big[\pi_k^l\bar z-\hat z_k\big]_m\big\|_{I_m}^2 &\le C\big\|\nabla\big(\pi_k^l\bar z-\bar z\big)\big\|_{I_m}^2+\big\|\nabla\big(\big(\pi_k^l\bar z\big)(t_m)-\hat z_{k,m}\big)\big\|_{I_m}^2\\ &\le C\big\|\nabla\big(\pi_k^l\bar z-\bar z\big)\big\|_{I_m}^2+\frac{k_m}{k_{m+1}}\big\|\nabla\big(\pi_k^l\bar z-\hat z_k\big)\big\|_{I_{m+1}}^2\\ &\le C\Big(\big\|\nabla\big(\pi_k^l\bar z-\bar z\big)\big\|_{I_m}^2+\big\|\nabla\big(\pi_k^l\bar z-\bar z\big)\big\|_{I_{m+1}}^2+\|\nabla(\bar z-\hat z_k)\|_{I_{m+1}}^2\Big).\end{aligned}$$

Summing up for $m=1,2,\cdots,M$ on both sides of the above estimate, we obtain

$$\begin{aligned}\sum_{m=1}^M\big\|\big[\pi_k^l\bar z-\hat z_k\big]_m\big\|_{L^2(\Omega)}^2+k_m\big\|\nabla\big[\pi_k^l\bar z-\hat z_k\big]_m\big\|_{L^2(\Omega)}^2 &\le C\Big(\big\|\nabla\big(\pi_k^l\bar z-\bar z\big)\big\|_I^2+\|\nabla(\bar z-\hat z_k)\|_I^2\Big)\\ &\le Ck^2\|\bar z\|_{H^1(I;H^1(\Omega))}^2,\end{aligned} \tag{6.10}$$

where Lemma 6.1 has been used. Besides, the term $\sum_{m=1}^M\|\nabla[\pi_k^l\bar z]_m\|_{L^2(\Omega)}^2$ can be estimated as

$$\begin{aligned}\sum_{m=1}^M k_m^{-1}\big\|\nabla\big[\pi_k^l\bar z\big]_m\big\|_{L^2(\Omega)}^2 &= \sum_{m=1}^M k_m^{-1}\|\nabla\bar z(t_m)-\nabla\bar z(t_{m-1})\|_{L^2(\Omega)}^2\\ &=\sum_{m=1}^M k_m^{-1}\bigg\|\nabla\int_{t_{m-1}}^{t_m}\partial_t\bar z\,\mathrm dt\bigg\|_{L^2(\Omega)}^2\\ &\le\sum_{m=1}^M\int_{t_{m-1}}^{t_m}\|\nabla\partial_t\bar z\|_{L^2(\Omega)}^2\,\mathrm dt\\ &\le C\|\bar z\|_{H^1(I;H^1(\Omega))}.\end{aligned} \tag{6.11}$$

Finally, using (6.10), (6.11) and (4.1) we get the desired result (6.4). □

Lemma 6.3. *Let Ω be a bounded smooth domain, and $\hat z_k\in X_k$ be the solution of (6.1). For a given $s\in[\frac14,\frac34]$, if $\bar y-y_d\in L^2(I;H^{2s-\frac12}(\Omega))\cap H^{s-\frac14}(I;L^2(\Omega))$ and the compatibility condition (2.8) holds for the adjoint equation (3.5) with $g:=\bar y-y_d$, where $\bar y$ is the optimal state, then $\bar z\in L^2(I;H^{2s+\frac32}(\Omega)\cap H_0^1(\Omega))\cap H^{s+\frac34}(I;L^2(\Omega))$ with $\partial_t\bar z\in L^2(I;H^{2(s-\frac14)}(\Omega))\cap H^{s-\frac14}(I;L^2(\Omega))$, and the following estimate holds:*

$$\begin{aligned}\|\bar z-\hat z_k\|_{L^2(I;H^{2(s-\frac14)}(\Omega))}&\le Ck\Big(\|\bar y-y_d\|_{L^2(I;H^{2s-\frac12}(\Omega))}+\|\bar y-y_d\|_{H^{s-\frac14}(I;L^2(\Omega))}\Big),\\ \|\hat z_k\|_{L^2(I;H^{2s+\frac32}(\Omega))}&\le C\Big(\|\bar y-y_d\|_{L^2(I;H^{2s-\frac12}(\Omega))}+\|\bar y-y_d\|_{H^{s-\frac14}(I;L^2(\Omega))}\Big).\end{aligned}$$

Proof. First, the regularity result $\bar z\in L^2(I;H^{2s+\frac32}(\Omega)\cap H_0^1(\Omega))\cap H^{s+\frac34}(I;L^2(\Omega))$ with $\partial_t\bar z\in L^2(I;H^{2(s-\frac14)}(\Omega))\cap H^{s-\frac14}(I;L^2(\Omega))$ is a direct consequence of Lemma 2.5.

Second, we estimate the finite element error between $\bar z$ and $\hat z_k$. The proof is divided into three steps.

Step 1. We consider the case $s = \frac{1}{4}$, *i.e.*,

$$\|\bar{z} - \hat{z}_k\|_{L^2(I;L^2(\Omega))} \le Ck\|\bar{z}\|_{H^1(I;L^2(\Omega))}, \tag{6.12}$$

which is a direct consequence of Theorem 5.1 in [38].

Step 2. We consider the case $s = \frac{3}{4}$. By Poincaré's inequality and Lemma 6.1, we obtain

$$\|\bar{z} - \hat{z}_k\|_{L^2(I;H^1(\Omega))} \le C\|\nabla(\bar{z} - \pi_k^r \bar{z})\|_I \le Ck\|\bar{z}\|_{H^1(I;H^1(\Omega))}. \tag{6.13}$$

Step 3. The case $\frac{1}{4} < s < \frac{3}{4}$ follows from interpolations. To do this, denote by $\mathcal{G}$ a linear operator from $\bar{z}$ to the error $\bar{z} - \hat{z}_k$. The above two estimates imply that $\|\mathcal{G}\|_{H^1(I;L^2(\Omega))\to L^2(I;L^2(\Omega))} \le Ck$ and $\|\mathcal{G}\|_{H^1(I;H^1(\Omega))\to L^2(I;H^1(\Omega))} \le Ck$, so that the desired error estimate follows from Proposition 14.1.5 and Theorem 14.2.3 in [7].

Next, we prove the stability estimate. Similar to the above analysis, the estimate is proved by interpolation estimates and the main procedure includes three steps. Setting $\hat{z}_{k,m-1} := \hat{z}_k|_{I_m} \in H^2(\Omega) \cap H_0^1(\Omega)$, the equation (6.1) can be rewritten as the time-stepping scheme:

$$\begin{cases} -\frac{\hat{z}_{k,m} - \hat{z}_{k,m-1}}{k_m} - \Delta \hat{z}_{k,m-1} = P_k(\bar{y} - y_d) & \text{in} \quad \Omega, \ m = 1, \cdots, M, \\ \hat{z}_{k,m} = 0 & \text{on} \quad \Gamma, \ m = 1, \cdots, M, \\ \hat{z}_{k,M} = 0 & \text{in} \quad \Omega, \end{cases}$$

which is a family of elliptic equations on the interval I_m.

We first consider the case $s = \frac{3}{4}$. By the regularity of elliptic equations, there holds $\hat{z}_{k,m-1} \in H^3(\Omega) \cap H_0^1(\Omega)$. Moreover, we can obtain that

$$\|\hat{z}_{k,m-1}\|^2_{H^3(\Omega)} \le C\Big(\|P_k(\bar{y} - y_d)\|^2_{H^1(\Omega)} + k_m^{-2}\|[\hat{z}_k]_m\|^2_{H^1(\Omega)}\Big) \le C\Big(\|P_k(\bar{y} - y_d)\|^2_{H^1(\Omega)} + k_m^{-2}\|\nabla[\hat{z}_k]_m\|^2_{L^2(\Omega)}\Big),$$

where $[\hat{z}_k]_M := -\hat{z}_{k,M-1}$. Integration in I_m on both sides of the above inequality and summing up for m, we can obtain

$$\begin{aligned} \|\hat{z}_k\|^2_{L^2(I;H^3(\Omega))} &\le C\left(\|\bar{y} - y_d\|^2_{L^2(I;H^1(\Omega))} + \sum_{m=1}^{M} k_m^{-1}\|\nabla[\hat{z}_k]_m\|^2_{L^2(\Omega)}\right) \\ &\le C\Big(\|\bar{y} - y_d\|^2_{L^2(I;H^1(\Omega))} + \|\bar{z}\|^2_{H^1(I;H^1(\Omega))}\Big) \\ &\le C\Big(\|\bar{y} - y_d\|^2_{L^2(I;H^1(\Omega))} + \|\bar{y} - y_d\|^2_{H^{\frac{1}{2}}(I;L^2(\Omega))}\Big), \end{aligned}$$

where we have used Lemmas 6.2 and 2.5.

The cases $s = \frac{1}{4}$ can be shown by a similar argument. In fact, we can derive the following stability estimate (cf. [38], Cor. 4.2):

$$\|\hat{z}_k\|^2_{L^2(I;H^2(\Omega))} \le C\|\bar{y} - y_d\|^2_{L^2(I;L^2(\Omega))}. \tag{6.14}$$

The case $\frac{1}{4} < s < \frac{3}{4}$ now follows from interpolations, where the following interpolation spaces will be used in this process:

$$\begin{aligned} \big[L^2\big(I;H^3(\Omega)\big), L^2\big(I;H^2(\Omega)\big)\big]_{\frac{3}{2}-2s} &= L^2\Big(I;H^{2s+\frac{3}{2}}(\Omega)\Big), \\ \Big[L^2\big(I;H^1(\Omega)\big) \cap H^{\frac{1}{2}}\big(I;L^2(\Omega)\big), L^2\big(I;L^2(\Omega)\big)\Big]_{\frac{3}{2}-2s} &= L^2\Big(I;H^{2s-\frac{1}{2}}(\Omega)\Big) \cap H^{s-\frac{1}{4}}\big(I;L^2(\Omega)\big), \end{aligned}$$

which are classical and can be found in many literatures, *e.g.*, [36]. This completes the proof. □

Remark 6.4. Utilizing the regularity results from Theorems 3.2 and 3.3, we observe that Lemma 6.3 holds for $s \in [\frac{1}{4}, \frac{3}{4}]$ when Ω is a smooth domain, and there holds the compatibility condition (2.8) with $g := \bar{y} - y_d$ for $s' = \frac{3}{4}$, and for $s = \frac{1}{4}$ when Ω is a polytope. Building upon this lemma, we are able to establish optimal orders of convergence for the solution of temporal semi-discrete control problems in the next subsection.

6.2. Error estimates for the temporal semi-discretization

In order to estimate the error for the temporal semi-discrete solution of (4.7), we establish the stability of the semi-discrete solution of the parabolic inhomogeneous boundary value problem with respect to the Dirichlet data in the following lemma.

Lemma 6.5. *For any given $u_k \in X_k(\Gamma)$, let $y_k \in X_k$ be the solution of the following equation:*

$$B(y_k, \varphi_k) = 0 \qquad \forall \varphi_k \in X_k^0, \quad y_k|_{I\times\Gamma} = u_k. \tag{6.15}$$

Then there holds

$$\|y_k\|_I \le C\|u_k\|_{L^2(I;L^2(\Gamma))}.$$

Proof. Let $z_k \in X_k^0$ satisfy the equation

$$B(\varphi_k, z_k) = (y_k, \varphi_k)_I \qquad \forall \varphi_k \in X_k^0.$$

For each fixed $m = 1, \cdots, M$, setting $z_{k,m-1} := z_k|_{I_m} \in H^2(\Omega) \cap H_0^1(\Omega)$, then the above equation can be rewritten as the following time-stepping scheme:

$$\begin{cases} -\frac{z_{k,m}-z_{k,m-1}}{k_m} - \Delta z_{k,m-1} = y_{k,m} & \text{in} \quad \Omega, \\ z_{k,M} = 0 & \text{in} \quad \Omega, \\ z_{k,m} = 0 & \text{on} \quad \Gamma. \end{cases}$$

Multiplying $y_{k,m}$ on both sides of the above equation, integration by parts in $I_m \times \Omega$ yields

$$-([z_k]_m, y_{k,m}) + (\nabla z_k, \nabla y_k)_{L^2(I_m;L^2(\Omega))} - (\partial_n z_k, u_k)_{L^2(I_m;L^2(\Gamma))} = \|y_k\|^2_{L^2(I_m;L^2(\Omega))}.$$

Summing up the above identity for m and observing $z_{k,M} = z(T) = 0$, we have

$$\left(z_{k,M}^-, y_{k,M}\right) - \sum_{m=1}^{M-1}([z_k]_m, y_{k,m}) + (\nabla z_k, \nabla y_k)_{L^2(I;L^2(\Omega))} - (\partial_n z_k, u_k)_{L^2(I;L^2(\Gamma))} = \|y_k\|_I^2.$$

That is,

$$B(y_k, z_k) - (\partial_n z_k, u_k)_{L^2(I;L^2(\Gamma))} = \|y_k\|_I^2.$$

By using the equation (6.15), we derive

$$\begin{aligned} \|y_k\|_I^2 &= -(\partial_n z_k, u_k)_{L^2(I;L^2(\Gamma))} \\ &\le \|\partial_n z_k\|_{L^2(I;L^2(\Gamma))}\|u_k\|_{L^2(I;L^2(\Gamma))} \\ &\le C\|z_k\|_{L^2(I;H^2(\Omega))}\|u_k\|_{L^2(I;L^2(\Gamma))} \\ &\le C(\|z_k\|_I + \|\Delta z_k\|_I)\|u_k\|_{L^2(I;L^2(\Gamma))} \\ &\le C\|y_k\|_I\|u_k\|_{L^2(I;L^2(\Gamma))}, \end{aligned}$$

where we have used the stability of the temporal semi-discrete solution of parabolic equations (cf. [38], Cor. 4.2). This completes the proof. □

Now we are ready to state the result on the error estimation for the solution of the temporal semi-discrete optimal control problem (4.7).

Proposition 6.6. *Assume that Ω is a bounded, smooth domain or convex polytope. Let $(\bar{u}, \bar{y})$ and $(\bar{u}_k, \bar{y}_k) \in U_{ad} \times X_k$ be the optimal solution of the control problem (1.1) and the semi-discrete problem (4.7), respectively. Assume that $\bar{z} \in L^2(I; H^2(\Omega) \cap H_0^1(\Omega))$ is the adjoint state, $y_k(\bar{u})$ is the solution of (4.8) with u replaced by $\bar{u}$, and $\hat{z}_k \in X_k^0$ satisfies (6.1). Then*

$$\|\bar{u}_k - \bar{u}\|_{L^2(I;L^2(\Gamma))} + \|\bar{y}_k - \bar{y}\|_{L^2(I;L^2(\Omega))} \leq C\Big(\|\partial_n(\bar{z} - \hat{z}_k)\|_{L^2(I;L^2(\Gamma))} + \|\bar{y} - y_k(\bar{u})\|_{L^2(I;L^2(\Omega))}\Big). \quad (6.16)$$

Proof. Since $\hat{J}_k$ is a quadratic functional, $\hat{J}_k''(u)$ is independent of u for all $u \in U_{ad}$. Note that $\hat{J}_k''(u)(v, v) \geq \alpha\|v\|^2_{L^2(I;L^2(\Gamma))}$, $\forall v \in L^2(I; L^2(\Gamma))$. Define the auxiliary variable $\tilde{z}_k \in X_k^0$ such that

$$B(\varphi_k, \tilde{z}_k) = (y_k(\bar{u}) - y_d, \varphi_k)_I \qquad \forall \varphi_k \in X_k^0.$$

Then

$$\begin{aligned}
\alpha\|\bar{u} - \bar{u}_k\|^2_{L^2(I;L^2(\Gamma))} &\leq \hat{J}_k''(\bar{u}_k)(\bar{u} - \bar{u}_k, \bar{u} - \bar{u}_k) \\
&= \hat{J}_k'(\bar{u})(\bar{u} - \bar{u}_k) - \hat{J}_k'(\bar{u}_k)(\bar{u} - \bar{u}_k) \\
&\leq \hat{J}_k'(\bar{u})(\bar{u} - \bar{u}_k) - \hat{J}'(\bar{u})(\bar{u} - \bar{u}_k) \\
&\leq \int_{\Sigma_T} (\partial_n \bar{z} - \partial_n \tilde{z}_k)(\bar{u} - \bar{u}_k) \\
&\leq \|\partial_n \bar{z} - \partial_n \tilde{z}_k\|_{L^2(I;L^2(\Gamma))}\|\bar{u} - \bar{u}_k\|_{L^2(I;L^2(\Gamma))},
\end{aligned} \quad (6.17)$$

where we have used $-\hat{J}_k'(\bar{u}_k)(\bar{u} - \bar{u}_k) \leq 0 \leq -\hat{J}'(\bar{u})(\bar{u} - \bar{u}_k)$ which can be easily checked by taking $v = \bar{u}$ in (4.15) and $v = \bar{u}_k$ in (3.4). From the inequality (6.17), we can obtain

$$\alpha\|\bar{u} - \bar{u}_k\|_{L^2(I;L^2(\Gamma))} \leq \|\partial_n \bar{z} - \partial_n \tilde{z}_k\|_{L^2(I;L^2(\Gamma))} \leq \|\partial_n \bar{z} - \partial_n \hat{z}_k\|_{L^2(I;L^2(\Gamma))} + \|\partial_n \hat{z}_k - \partial_n \tilde{z}_k\|_{L^2(I;L^2(\Gamma))}, \quad (6.18)$$

where $\hat{z}_k \in X_k^0$ satisfies (6.1). It follows from the stability estimate of the temporal semi-discrete solution of adjoint equations (cf. [38], Cor. 4.2) that

$$\begin{aligned}
\|\partial_n \hat{z}_k - \partial_n \tilde{z}_k\|_{L^2(I;L^2(\Gamma))} &\leq C\|\hat{z}_k - \tilde{z}_k\|_{L^2(I;H^2(\Omega))} \\
&\leq C\Big(\|\hat{z}_k - \tilde{z}_k\|_{L^2(I;L^2(\Omega))} + \|\Delta(\hat{z}_k - \tilde{z}_k)\|_{L^2(I;L^2(\Omega))}\Big) \\
&\leq C\|\bar{y} - y_k(\bar{u})\|_I.
\end{aligned} \quad (6.19)$$

Now, we consider the error estimate for the semi-discrete optimal state:

$$\begin{aligned}
\|\bar{y} - \bar{y}_k\|_I &\leq \|\bar{y} - y_k(\bar{u})\|_I + \|y_k(\bar{u}) - \bar{y}_k\|_I \\
&\leq \|\bar{y} - y_k(\bar{u})\|_I + C\Big\|\tilde{P}_k\bar{u} - \tilde{P}_k\bar{u}_k\Big\|_{L^2(I;L^2(\Gamma))} \\
&\leq \|\bar{y} - y_k(\bar{u})\|_I + C\|\bar{u} - \bar{u}_k\|_{L^2(I;L^2(\Gamma))},
\end{aligned} \quad (6.20)$$

where we have used Lemma 6.5 and the stability of L^2-projection $\tilde{P}_k$. Combining the estimates (6.18)–(6.20), we finally obtain the estimate (6.16). This completes the proof. □

From Proposition 6.6, it remains to estimate the right-hand terms of (6.16), which depends heavily on the smoothness of Ω.

Theorem 6.7. *Assume $f \in L^2(I; L^2(\Omega))$, $y_0 \in H^{2s-1}(\Omega)$ and $y_d \in L^2(I; H^{2s'-\frac{1}{2}}(\Omega)) \cap H^{s'-\frac{1}{4}}(I; L^2(\Omega))$, where $s \in [\frac{1}{2}, \frac{3}{4})$, $s' \in [s - \frac{1}{4}, s]$ in case that Ω is smooth with the compatibility condition (2.8) for the adjoint equation*

(3.5) *for* $g := \bar{y} - y_d$, *and* $s = \frac{1}{2}$, $s' = \frac{1}{4}$ *in case that* Ω *is a convex polytope. Let* $(\bar{u}, \bar{y})$ *and* $(\bar{u}_k, \bar{y}_k) \in U_{ad} \times X_k$ *be the optimal solution of the optimal control problem* (1.1) *and the semi-discrete optimal control problem* (4.7)*, respectively. Then*

$$\|\bar{u}_k - \bar{u}\|_{L^2(I;L^2(\Gamma))} + \|\bar{y}_k - \bar{y}\|_{L^2(I;L^2(\Omega))} \leq C\Big(k^s + k^{s'}\Big). \tag{6.21}$$

Proof. By Proposition 6.6, we only need to consider the two terms on the right-hand side of (6.16). In viewing of the regularity result of the optimal pair $(\bar{u}, \bar{y})$ in Theorems 3.2 and 3.3, we first use Theorem 5.5 to deduce the following estimate:

$$\|\bar{y} - y_k(\bar{u})\|_{L^2(I;L^2(\Omega))} \leq Ck^s, \tag{6.22}$$

where $s = \frac{1}{2}$ for a convex polytopal domain Ω and $\frac{1}{2} \leq s < \frac{3}{4}$ for a smooth domain.

Based on the regularity of the adjoint state $\bar{z}$ in Theorems 3.2 and 3.3, we estimate the first term on the right-hand side of (6.16) by taking $s' = \frac{1}{4}$ for a polytopal Ω, and $s - \frac{1}{4} \leq s' \leq s$ for a smooth domain. By the trace theorem, we have

$$\begin{aligned}
\|\partial_n \bar{z} - \partial_n \hat{z}_k\|_{L^2(I;L^2(\Gamma))} &\leq C\|\bar{z} - \hat{z}_k\|^{\frac{1}{2}}_{L^2(I;H^2(\Omega))}\|\bar{z} - \hat{z}_k\|^{\frac{1}{2}}_{L^2(I;H^1(\Omega))} \\
&\leq C\|\bar{z} - \hat{z}_k\|^{1-s'}_{L^2(I;H^{2s'+\frac{3}{2}}(\Omega))}\|\bar{z} - \hat{z}_k\|^{s'}_{L^2(I;H^{2(s'-\frac{1}{4})}(\Omega))} \\
&\leq Ck^{s'},
\end{aligned} \tag{6.23}$$

where we have used Lemma 6.3 and the following interpolation result (see, *e.g.*, [36]):

$$\Big[L^2\Big(I; H^{2s'+\frac{3}{2}}(\Omega)\Big), L^2\Big(I; H^{2(s'-\frac{1}{4})}(\Omega)\Big)\Big]_{s'-\frac{1}{4}} = L^2\big(I; H^2(\Omega)\big),$$
$$\Big[L^2\Big(I; H^{2s'+\frac{3}{2}}(\Omega)\Big), L^2\Big(I; H^{2(s'-\frac{1}{4})}(\Omega)\Big)\Big]_{s'+\frac{1}{4}} = L^2\big(I; H^1(\Omega)\big).$$

Combining the estimates (6.22) and (6.23), we finish the proof. □

6.3. Error estimates for the spatial discretization

In this subsection, we aim to estimate the error between the fully discrete solution of the optimal control problem (4.18) and the temporal semi-discrete solution of the optimal control problem (4.7). To achieve this, we first establish the stability of the fully discrete solution to the parabolic equation concerning inhomogeneous Dirichlet data.

Proposition 6.8. *For any given* $g \in L^2(I; L^2(\Omega))$*, let* $z \in L^2(I; H^2(\Omega) \cap H^1_0(\Omega)) \cap H^1(I; L^2(\Omega))$ *be the solution of equation* (2.3) *and* z_k *be the corresponding temporal semi-discretization. By the identity* (4.16)*, the normal derivative* $\partial_n z_k$ *satisfies*

$$\int_{\Sigma_T} \partial_n z_k \phi_k \,\mathrm{d}s\,\mathrm{d}t = -\int_{\Omega_T} g p_k(\phi_k)\,\mathrm{d}x\,\mathrm{d}t \qquad \forall \phi_k \in X_k(\Gamma), \tag{6.24}$$

where $p_k(\phi_k) \in X_k$ *is the solution of* (4.17)*. Denote by* z_{kh} *the fully discrete solution of* (2.3) *and by* $\partial^h_n z_{kh}$ *the fully discrete normal derivative of* z_{kh}*. Let* $y_{kh} \in X_{kh}$ *be the solution of*

$$B(y_{kh}, \varphi_{kh}) = 0 \qquad \forall \varphi_{kh} \in X^0_{kh}, \quad y_{kh}|_{I\times\Gamma} = u_{kh} \in X_{kh}(\Gamma).$$

Then there holds

$$\begin{aligned}
\big\|\partial^h_n z_{kh}\big\|_{L^2(I;L^2(\Gamma))} &\leq C\|g\|_I, \\
\|y_{kh}\|_I &\leq C\|u_{kh}\|_{L^2(I;L^2(\Gamma))}, \\
\big\|\partial_n z_k - \partial^h_n z_{kh}\big\|_{L^2(I;L^2(\Gamma))} &\leq Ch^{\frac{1}{2}}\|g\|_I.
\end{aligned} \tag{6.25}$$

Proof. We first prove the third estimate of (6.25). For any given $\omega_{kh} \in X_{kh}(\Gamma)$, setting $\phi_k = \omega_{kh}$ in (6.24) and $\phi_{kh} = \omega_{kh}$ in (4.22), then there holds

$$\begin{aligned} \int_{\Sigma_T} \partial_n z_k \omega_{kh} \,\mathrm{d}s\,\mathrm{d}t &= -\int_{\Omega_T} g p_k(\omega_{kh}) \,\mathrm{d}x\,\mathrm{d}t \\ &= -\int_{\Omega_T} g p_{kh}(\omega_{kh}) \,\mathrm{d}x\,\mathrm{d}t + \int_{\Omega_T} g(p_{kh}(\omega_{kh}) - p_k(\omega_{kh})) \,\mathrm{d}x\,\mathrm{d}t \\ &= \int_{\Sigma_T} \partial_n^h z_{kh} \omega_{kh} \,\mathrm{d}s\,\mathrm{d}t + \int_{\Omega_T} g(p_{kh}(\omega_{kh}) - p_k(\omega_{kh})) \,\mathrm{d}x\,\mathrm{d}t. \end{aligned} \tag{6.26}$$

Define the Ritz projection $R_h : X_{kh} \to X_{kh}^0$ such that $R_h^n p_{kh,n} := R_h p_{kh}|_{I_n} \in V_h^0$ satisfies

$$(\nabla R_h^n p_{kh,n}, \nabla v_h) = (\nabla p_{kh,n}, \nabla v_h) \qquad \forall v_h \in V_h^0, \qquad n = 1, \cdots, M. \tag{6.27}$$

Using the fact $p_{kh}(\omega_{kh}) - p_k(\omega_{kh}) \in X_k^0$ and $B(p_k(\omega_{kh}), z_k) = B(p_{kh}(\omega_{kh}), z_{kh}) = B(R_h p_{kh}(\omega_{kh}), z_k - z_{kh}) = 0$, we have

$$\begin{aligned} &\left| \int_{\Sigma_T} (\partial_n z_k - \partial_n^h z_{kh}) \omega_{kh} \,\mathrm{d}s\,\mathrm{d}t \right| = \left| \int_{\Omega_T} g(p_{kh}(\omega_{kh}) - p_k(\omega_{kh})) \,\mathrm{d}x\,\mathrm{d}t \right| \\ &\quad = |B(p_{kh}(\omega_{kh}) - p_k(\omega_{kh}), z_k)| \\ &\quad = |B(p_{kh}(\omega_{kh}) - R_h p_{kh}(\omega_{kh}), z_k - z_{kh})| \\ &\quad = \left| (\nabla(p_{kh}(\omega_{kh}) - R_h p_{kh}(\omega_{kh})), \nabla(z_k - z_{kh}))_I - \sum_{m=1}^{M} \left((p_{kh}(\omega_{kh}) - R_h p_{kh}(\omega_{kh}))_m^-, [z_k - z_{kh}]_m \right) \right| \\ &\quad \le \|\nabla(p_{kh}(\omega_{kh}) - R_h p_{kh}(\omega_{kh}))\|_I \|\nabla(z_k - z_{kh})\|_I \\ &\qquad + \left(\sum_{m=1}^{M} k_m \|p_{kh,m}(\omega_{kh}) - R_h^m p_{kh,m}(\omega_{kh})\|_{L^2(\Omega)}^2 \right)^{\frac{1}{2}} \left(\sum_{m=1}^{M} \frac{\|[z_k - z_{kh}]_m\|_{L^2(\Omega)}^2}{k_m} \right)^{\frac{1}{2}} \\ &\quad = \|\nabla(p_{kh}(\omega_{kh}) - R_h p_{kh}(\omega_{kh}))\|_I \|\nabla(z_k - z_{kh})\|_I \\ &\qquad + \|p_{kh}(\omega_{kh}) - R_h p_{kh}(\omega_{kh})\|_I \left(\sum_{m=1}^{M} \frac{\|[z_k - z_{kh}]_m\|_{L^2(\Omega)}^2}{k_m} \right)^{\frac{1}{2}} \\ &\quad \le C h^{\frac{1}{2}} \|g\|_I \|\omega_{kh}\|_{L^2(I;L^2(\Gamma))}, \end{aligned} \tag{6.28}$$

where we have used the following estimates:

$$\begin{aligned} \left(\sum_{m=1}^{M} \frac{\|[z_k - z_{kh}]_m\|_{L^2(\Omega)}^2}{k_m} \right)^{\frac{1}{2}} &\le C\|g\|_I, \\ \|\nabla(z_k - z_{kh})\|_I &\le Ch\|g\|_I, \\ \|p_{kh}(\omega_{kh}) - R_h p_{kh}(\omega_{kh})\|_I &\le C h^{\frac{1}{2}} \|\omega_{kh}\|_{L^2(I;L^2(\Gamma))}, \\ \|\nabla(p_{kh}(\omega_{kh}) - R_h p_{kh}(\omega_{kh}))\|_I &\le C h^{-\frac{1}{2}} \|\omega_{kh}\|_{L^2(I;L^2(\Gamma))}, \end{aligned}$$

in which the first two inequalities can be found in [38]. The last two estimates can be derived similar to the proof of Lemma 3 in [23]. That is,

$$\begin{aligned} \|p_{kh}(\omega_{kh}) - R_h p_{kh}(\omega_{kh})\|_I &= \|(I - R_h) p_{kh}(\omega_{kh})\|_I \\ &= \|(I - R_h)((I - Q_h) p_{kh}(\omega_{kh}) + Q_h p_{kh}(\omega_{kh}))\|_I \\ &= \|(I - R_h)(I - Q_h) p_{kh}(\omega_{kh})\|_I \\ &\le Ch \|\nabla(I - Q_h) p_{kh}(\omega_{kh})\|_I \\ &\le C \|(I - Q_h) p_{kh}(\omega_{kh})\|_I, \end{aligned}$$

where we have used the error estimate and stability of R_h as well as the inverse estimate, while $Q_h : X_{kh} \to X_{kh}^0$ is the L^2-projection. By setting $y_2^n = (I - Q_h)p_{kh}(\omega_{kh})$ in the proof of Lemma 3 from [23], we have the following estimate

$$\|(I - Q_h)p_{kh}(\omega_{kh})\|_I \leq Ch^{\frac{1}{2}}\|\omega_{kh}\|_{L^2(I;L^2(\Gamma))},$$

then we can obtain the desired estimate. The last one can be obtained by using the inverse estimate combining with the penultimate one.

Particularly, setting $\omega_{kh} = \tilde{P}_{kh}\partial_n z_k - \partial_n^h z_{kh}$ in (6.28), we obtain

$$\left\|\tilde{P}_{kh}\partial_n z_k - \partial_n^h z_{kh}\right\|^2_{L^2(I;L^2(\Gamma))} = \left|\int_{\Sigma_T}(\partial_n z_k - \partial_n^h z_{kh})\omega_{kh}\,\mathrm{d}s\,\mathrm{d}t\right| \leq Ch^{\frac{1}{2}}\left\|\tilde{P}_{kh}\partial_n z_k - \partial_n^h z_{kh}\right\|_{L^2(I;L^2(\Gamma))}\|g\|_I.$$

Using the triangle inequality yields

$$\begin{aligned}\left\|\partial_n z_k - \partial_n^h z_{kh}\right\|_{L^2(I;L^2(\Gamma))} &\leq \left\|\partial_n z_k - \tilde{P}_h\partial_n z_k\right\|_{L^2(I;L^2(\Gamma))} + \left\|\tilde{P}_h\partial_n z_k - \partial_n^h z_{kh}\right\|_{L^2(I;L^2(\Gamma))}\\ &\leq Ch^{\frac{1}{2}}\|\partial_n z_k\|_{L^2(I;H^{\frac{1}{2}}(\Gamma))} + Ch^{\frac{1}{2}}\|g\|_I\\ &\leq Ch^{\frac{1}{2}}\|g\|_I.\end{aligned}$$

This verifies the third estimate in (6.25).

Next, we prove the first estimate in (6.25). It follows from the triangle inequality that

$$\left\|\partial_n^h z_{kh}\right\|_{L^2(I;L^2(\Gamma))} \leq \|\partial_n z_k\|_{L^2(I;L^2(\Gamma))} + \left\|\partial_n z_k - \partial_n^h z_{kh}\right\|_{L^2(I;L^2(\Gamma))} \leq C\Big(\|g\|_I + h^{\frac{1}{2}}\|g\|_I\Big) \leq C\|g\|_I.$$

Finally, we verify the second estimate in (6.25). Taking $g = y_{kh}$ on the right-hand side of (2.3) and denoting the fully discrete solution of (2.3) by z_{kh}, then the solution $p_{kh}(\phi_{kh})$ of (4.23) satisfies $p_{kh}(\phi_{kh}) = y_{kh}$ by taking $\phi_{kh} = u_{kh}$ in (4.22). Therefore, there holds

$$\int_{\Omega_T} y_{kh}^2\,\mathrm{d}x\,\mathrm{d}t = -\int_{\Sigma_T}\partial_n^h z_{kh}u_{kh}\,\mathrm{d}s\,\mathrm{d}t \leq \left\|\partial_n^h z_{kh}\right\|_{L^2(I;L^2(\Gamma))}\|u_{kh}\|_{L^2(I;L^2(\Gamma))} \leq C\|y_{kh}\|_I\|u_{kh}\|_{L^2(I;L^2(\Gamma))},$$

where we used the first and third estimates in (6.25). This completes the proof. □

Theorem 6.9. *Let $(\bar{u}_k, \bar{y}_k) \in U_{ad} \times X_k$ and $(\bar{u}_{kh}, \bar{y}_{kh}) \in U_{ad} \times X_{kh}$ be the optimal pair of the temporal semi-discrete control problem* (4.7) *and the fully discrete optimal control problem* (4.18)*, respectively. Then*

$$\|\bar{u}_k - \bar{u}_{kh}\|_{L^2(I;L^2(\Gamma))} + \|\bar{y}_k - \bar{y}_{kh}\|_{L^2(I;L^2(\Omega))} \leq Ch^{\frac{1}{2}}.$$

Proof. Since $\hat{J}_{kh}(u)$ is a quadratic functional with respect to u, $\hat{J}''_{kh}(u)$ is a constant operator and satisfies $\hat{J}''_{kh}(u) \geq \alpha$. Taking $v = \bar{u}_k$ and $v = \bar{u}_{kh}$ in (4.28) and (4.15), respectively, there holds

$$-\hat{J}'_{kh}(\bar{u}_{kh})(\bar{u}_k - \bar{u}_{kh}) \leq 0 \leq -\hat{J}'_k(\bar{u}_k)(\bar{u}_k - \bar{u}_{kh}).$$

Hence

$$\begin{aligned}\alpha\|\bar{u}_k - \bar{u}_{kh}\|^2_{L^2(I;L^2(\Gamma))} &\leq \hat{J}''_{kh}(\bar{u}_{kh})(\bar{u}_k - \bar{u}_{kh}, \bar{u}_k - \bar{u}_{kh})\\ &= \hat{J}'_{kh}(\bar{u}_k)(\bar{u}_k - \bar{u}_{kh}) - \hat{J}'_{kh}(\bar{u}_{kh})(\bar{u}_k - \bar{u}_{kh})\\ &\leq \hat{J}'_{kh}(\bar{u}_k)(\bar{u}_k - \bar{u}_{kh}) - \hat{J}'_k(\bar{u}_k)(\bar{u}_k - \bar{u}_{kh})\\ &\leq \int_{\Sigma_T}\big(\partial_n\bar{z}_k - \partial_n^h z_{kh}\big)(\bar{u}_k - \bar{u}_{kh})\,\mathrm{d}s\,\mathrm{d}t\\ &\leq \left\|\partial_n\bar{z}_k - \partial_n^h z_{kh}\right\|_{L^2(I;L^2(\Gamma))}\|\bar{u}_k - \bar{u}_{kh}\|_{L^2(I;L^2(\Gamma))},\end{aligned} \tag{6.29}$$

where $z_{kh} \in X_{kh}^0$ satisfies

$$B(\varphi_{kh}, z_{kh}) = (\varphi_{kh}, y_{kh}(\bar{u}_k) - y_d)_I \qquad \forall \varphi_{kh} \in X_{kh}^0.$$

Let $\hat{z}_{kh} \in X_{kh}^0$ be the solution of the equation

$$B(\varphi_{kh}, \hat{z}_{kh}) = (\varphi_{kh}, \bar{y}_k - y_d)_I \qquad \forall \varphi_{kh} \in X_{kh}^0.$$

By using the estimate (6.29) and the triangle inequality, there holds

$$\begin{aligned}
\alpha\|\bar{u}_k - \bar{u}_{kh}\|_{L^2(I;L^2(\Gamma))} &\le \left\|\partial_n \bar{z}_k - \partial_n^h z_{kh}\right\|_{L^2(I;L^2(\Gamma))} \\
&\le \left\|\partial_n \bar{z}_k - \partial_n^h \hat{z}_{kh}\right\|_{L^2(I;L^2(\Gamma))} + \left\|\partial_n^h \hat{z}_{kh} - \partial_n^h z_{kh}\right\|_{L^2(I;L^2(\Gamma))} \\
&\le Ch^{\frac{1}{2}}\|\bar{y}_k - y_d\|_I + C\|\bar{y}_k - y_{kh}(\bar{u}_k)\|_I \\
&\le Ch^{\frac{1}{2}}\|\bar{y}_k - y_d\|_I + Ch\Big(\|\bar{y}_k\|_{L^2(I;H^1(\Omega))} + \|\bar{u}\|_{L^2(I;H^{\frac{1}{2}}(\Gamma))}\Big) \\
&\le Ch^{\frac{1}{2}},
\end{aligned}$$

where we have used Lemma 2.5, Theorem 5.11, Propositions 6.8 and 5.7.

Finally, the error between the semi-discrete and the fully discrete states reads

$$\begin{aligned}
\|\bar{y}_k - \bar{y}_{kh}\|_I &\le \|\bar{y}_k - y_{kh}(\bar{u}_k)\|_I + \|y_{kh}(\bar{u}_k) - \bar{y}_{kh}\|_I \\
&\le Ch + C\left\|\tilde{P}_{kh}(\bar{u}_k - \bar{u}_{kh})\right\|_I \\
&\le Ch + C\|\bar{u}_k - \bar{u}_{kh}\|_I \\
&\le Ch^{\frac{1}{2}},
\end{aligned}$$

where we used Proposition 6.8. This completes the proof. ☐

Combining Theorems 6.7 and 6.9, we can obtain the error estimate for the optimal pairs between the continuous and fully discrete control problems.

Theorem 6.10. *Let $(\bar{u}, \bar{y}) \in U_{ad} \times L^2(I; H^1(\Omega)) \cap H^{\frac{1}{2}}(I; L^2(\Omega))$ and $(\bar{u}_{kh}, \bar{y}_{kh}) \in U_{ad} \times X_{kh}$ be the solution of the continuous control problem (1.1) and the fully discrete control problem (6.9), respectively. Then we have*

$$\|\bar{u} - \bar{u}_{kh}\|_{L^2(I;L^2(\Gamma))} + \|\bar{y} - \bar{y}_{kh}\|_{L^2(I;L^2(\Omega))} \le C\Big(h^{\frac{1}{2}} + k^{\frac{1}{4}}\Big). \tag{6.30}$$

Remark 6.11. There is an alternative error estimate for $(\bar{u}_{kh}, \bar{y}_{kh})$ presented in [23], specifically in the two-dimensional case. In that work, the condition $k = O(h^2)$ is imposed to ensure a convergence order similar to (6.30). More precisely, the estimate for $(\bar{u}_{kh}, \bar{y}_{kh})$ in [23] is given as follows:

$$\|\bar{u} - \bar{u}_{kh}\|_{L^2(I;L^2(\Gamma))} + \|\bar{y} - \bar{y}_{kh}\|_{L^2(I;L^2(\Omega))} \le C\Big(h^{\frac{1}{2}} + k^{\frac{1}{4}} + h^{\frac{3}{2}}k^{-\frac{1}{2}} + h^{-\frac{1}{2}}k^{\frac{1}{2}} + h^{\frac{5}{2}}k^{-1}\Big).$$

Here, it is important to note that the time step k and mesh size h are coupled in the previous analysis. Consequently, the condition $k = O(h^2)$ is deemed necessary to achieve the optimal convergence order. However, we emphasize that this condition is unnecessary both in theory and numerical computations. In this paper, we eliminate this mesh size condition by developing a new *a priori* error analysis, resulting in an improved error estimate compared to [23].

TABLE 1. Convergence orders of the control and state with respect to spatial discretizations for Example 7.1 with fixed time step size $k = \frac{1}{512}$.

Dof	$\|\bar{y} - \bar{y}_{kh}\|_{L^2(I,L^2(\Omega))}$	Rate	$\|\bar{u} - \bar{u}_{kh}\|_{L^2(I,L^2(\Gamma))}$	Rate
16	1.85708e−2	–	6.96347e−2	–
51	5.81066e−3	1.67626	4.65893e−2	0.57981
181	1.59735e−3	1.86303	2.51189e−2	0.89123
681	4.36846e−4	1.87048	1.22724e−2	1.03335
2641	1.24153e−4	1.81501	5.38633e−3	1.18805
10 401	3.97606e−5	1.64270	1.82002e−3	1.56535

Remark 6.12. Since the regularity of the optimal control $\bar{u}$ is determined by that of the adjoint state $\bar{z}$, so the convergence order $h^{\frac{1}{2}} + k^{\frac{1}{4}}$ is sharp with respect to the regularity $\partial_n \bar{z} \in L^2(I; H^{\frac{1}{2}}(\Gamma)) \cap H^{\frac{1}{4}}(I; L^2(\Gamma))$, cf. Proposition 6.6. However, this regularity for $\partial_n \bar{z}$ is not sharp for convex polytopal domains because we can expect $L^q(I; L^p(\Omega))$ regularity for parabolic equations with homogeneous Dirichlet data. Therefore, we should work with this Banach space setting other than the Hilbert space setting in current paper. However, to the best of our knowledge, it is an open question to obtain optimal *a priori* error estimate for the fully discrete approximation to parabolic Dirichlet boundary control problems in convex polytopal domains, we refer to [2,9] for the elliptic case and [21] for a spatial semi-discrete approximation.

Remark 6.13. For the full discretization of parabolic Dirichlet boundary control problems in smooth domains, traditional polygonal approximation introduces geometric errors, complicating the error estimate. The error analysis is deferred to future work, with reference to [15] for the elliptic case.

7. NUMERICAL EXPERIMENTS

In this section we carry out some numerical experiments to support our theoretical findings. We consider an example with pointwise control constraints where the exact solutions are unknown, and take the numerical solutions with time step size $k = \frac{1}{512}$ and degrees of freedom $Dof = 41281$ as reference solutions to compute the convergence order. We set $\alpha = 1$ and $T = 1$.

Example 7.1. We consider the parabolic Dirichlet boundary control problem posed on the polygonal domain

$$\Omega = (0,1) \times (0,1) \cup \Big\{(x_1, x_2) \in \mathbb{R}^2,\ -x_1 \le x_2 \le 1,\ -1 \le x_1 \le 1\Big\}$$

with the maximum interior angle $\omega = \frac{3\pi}{4}$, and the control constraint

$$U_{ad} := \{u \in L^2((0,1), L^2(\Gamma)) :\ -1 \le u(t,x) \le 1,\ a.e.\ (t,x) \in (0,1) \times \Gamma\}.$$

The data is chosen as $f = 1$ and

$$y_d(t,x) = \begin{cases} t^2(x_1(1-x_1) + x_2(1-x_2)) & 0 \le t < 0.5, \\ -t^2(x_1(1-x_1) + x_2(1-x_2)) & 0.5 \le t \le 1. \end{cases}$$

The convergence orders of the control and state variables with respect to spatial and temporal discretizations are listed in Tables 1 and 2, respectively. We can observe first order convergence rates for both the spatial and temporal discretizations of the control variable that are higher than our theoretical results. This indicates that the *a priori* error estimates obtained in this paper are not optimal. To obtain optimal *a priori* error estimates for

TABLE 2. Convergence orders of the control and state with respect to temporal discretizations for Example 7.1 with fixed spatial degrees of freedom 41 281.

N	$\|\bar{y}-\bar{y}_{kh}\|_{L^2(I,L^2(\Omega))}$	Rate	$\|\bar{u}-\bar{u}_{kh}\|_{L^2(I,L^2(\Gamma))}$	Rate
4	1.01315e−2	–	1.59023e−2	–
8	7.38646e−3	0.45590	1.28935e−2	0.30259
16	4.74827e−3	0.63748	8.33539e−3	0.62932
32	2.75298e−3	0.78641	4.73513e−3	0.81585
64	1.47759e−3	0.89774	2.45471e−3	0.94785
128	7.30976e−4	1.01536	1.16730e−3	1.07238
256	3.09579e−4	1.23952	4.73034e−4	1.30316

Dirichlet boundary control problems posed on polygonal domains we need to derive $L^q(I;L^p(\Omega))$ regularity for parabolic equations, we refer to Remark 6.12 for an explanation and to [2,9] for the elliptic case. On the other hand, the convergence rates for the state variable are higher than that of the control variable, which is very common in numerical analysis for PDE-constrained optimal control problems. Optimal *a priori* error estimate for the state variable is also a delicate and difficult task, we refer to [37] for an attempt for elliptic Dirichlet boundary control problems without control constraints.

Acknowledgments

The authors would like to thank two anonymous referees for their helpful suggestions which lead to a great improvement of the paper.

Funding

W. Gong was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDB 41000000) and the National Natural Science Foundation of China (Grant No. 12071468). X. Xie was supported in part by the National Natural Science Foundation of China (Grant No. 12171340).

Data availability statement

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

References

[1] T. Apel, M. Mateos, J. Pfefferer and A. Rösch, On the regularity of the solutions of Dirichlet optimal control problems in polygonal domains. *SIAM J. Control Optim.* **53** (2015) 3620–3641.

[2] T. Apel, M. Mateos, J. Pfefferer and A. Rösch, Error estimates for Dirichlet control problems in polygonal domains: quasi-uniform meshes. *Math. Control Related Fields* **8** (2018) 217–245.

[3] N. Arada and J.P. Raymond, Dirichlet boundary control of semilinear parabolic equations. Part 1: problems with no state constraints. *Appl. Math. Optim.* **45** (2002) 125–143.

[4] F.B. Belgacem, C. Bernardi and H.E. Fekih, Dirichlet boundary control for a parabolic equation with a final observation I: a space-time mixed formulation and penalization. *Asymptot. Anal.* **71** (2011) 101–121.

[5] M. Berggren, Approximation of very weak solutions to boundary value problems. *SIAM J. Numer. Anal.* **42** (2004) 860–877.

[6] K. Bhandari, J. Lemoine and A. Münch, Global boundary null-controllability of one-dimensional semilinear heat equations. *Discrete Contin. Dyn. Syst. Ser. S* **17** (2024) 2251–2297.

[7] S.C. Brenner and L.R. Scott, The Mathematical Theory of Finite Element Methods, 3rd edition. Springer, New York (2008).

[8] E. Casas and K. Chrysafinos, A discontinuous Galerkin time-stepping scheme for the velocity tracking problem. *SIAM J. Numer. Anal.* **50** (2012) 2281–2306.

[9] E. Casas and J.P. Raymond, Error estimates for the numerical approximation of Dirichlet boundary control for semilinear elliptic equations. *SIAM J. Control Optim.* **45** (2006) 1586–1611.

[10] E. Casas and J. Sokolowski, Approximation of boundary control problems on curved domains. *SIAM J. Control Optim.* **48** (2010) 3746–3780.

[11] E. Casas, M. Mateos and J.P. Raymond, Penalization of Dirichlet optimal control problems. *ESAIM Control Optim. Calc. Var.* **15** (2009) 782–809.

[12] K. Chrysafinos, Convergence of discontinuous Galerkin approximations of an optimal control problem associated to semilinear parabolic PDEs. *ESAIM: Math. Model. Numer. Anal.* **44** (2010) 189–206.

[13] P.G. Ciarlet, The Finite Element Methods for Elliptic Problems, 1st edition. Vol. 4. North-Holland (1978).

[14] P. Clément, Approximation by finite element functions using local regularization. *RAIRO Anal. Numer.* **9** (1975) 77–84.

[15] K. Deckelnick, A. Günther and M. Hinze, Finite element approximation of Dirichlet boundary control for elliptic PDEs on two and three-dimensional curved domains. *SIAM J. Control Optim.* **48** (2009) 2798–2819.

[16] L.C. Evans, Partial Differential Equations, 2nd edition. Vol. 19. AMS (2002).

[17] D.A. French and J.T. King, Analysis of a robust finite element approximation for a parabolic equation with rough boundary data. *Math. Comput.* **60** (1993) 79–104.

[18] D.A. French and J.T. King, Approximation of an elliptic control problem by the finite element method. *Numer. Funct. Anal. Optim.* **12** (1991) 299–314.

[19] A.V. Fursikov, M.D. Gunzburger and L.S. Hou, Boundary value problems and optimal boundary control for the Navier–Stokes systems: the two-dimensional case. *SIAM J. Control Optim.* **36** (1998) 852–894.

[20] V. Girault and P.A. Raviart, Finite Element Methods for Navier–Stokes Equations: Theory and Algorithms. Springer (1986).

[21] W. Gong and B.Y. Li, Improved error estimates for semidiscrete finite element solutions of parabolic Dirichlet boundary control problems. *IMA J. Numer. Anal.* **40** (2020) 2898–2939.

[22] W. Gong, M. Hinze and Z.J. Zhou, A priori error analysis for finite element approximation of parabolic optimal control problems with pointwise control. *SIAM J. Control Optim.* **52** (2014) 97–119.

[23] W. Gong, M. Hinze and Z.J. Zhou, Finite element method and a priori error estimates for Dirichlet boundary control problems governed by parabolic PDEs. *J. Sci. Comput.* **66** (2016) 941–967.

[24] M.D. Gunzburger, L.S. Hou and T.P. Svobodny, Analysis and finite element approximation of optimal control problems for the stationary Navier–Stokes equations with Dirichlet controls. *RAIRO Modél. Math. Anal. Numér.* **25** (1991) 711–748.

[25] M. Hinze, A variational discretization concept in control constrained optimization: the linear-quadratic case. *Comput. Optim. Appl.* **30** (2005) 45–61.

[26] M. Hinze and K. Kunisch, Second order methods for boundary control of the instationary Navier–Stokes system. *Z. Angew. Math. Mech.* **84** (2004) 171–187.

[27] M. Hinze, R. Pinnau, M. Ulbrich and S. Ulbrich, Optimization with PDE Constraints. Vol. 23. Springer-Netherlands (2009).

[28] A. Kufner, L. Persson and N. Samko, Weighted Inequalities of Hardy Type, 2nd edition. World Scientific Publishing Company (2017).

[29] K. Kunisch and B. Vexler, Constrained Dirichlet boundary control in L^2 for a class of evolution equations. *SIAM J. Control Optim.* **46** (2007) 1726–1753.

[30] I. Lasiecka, Unified theory for abstract parabolic boundary problems – a semigroup approach. *Appl. Math. Optim.* **6** (1980) 287–333.

[31] I. Lasiecka, Galerkin approximations of abstract parabolic boundary value problems with rough boundary data – L^p theory. *Math. Comput.* **47** (1986) 55–75.

[32] I. Lasiecka and K. Malanowski, On discrete-time Ritz-Galerkin approximation of control constrained optimal control problems for parabolic systems. *Control Cybern.* **7** (1978) 21–36.

[33] B.J. Li, H. Luo and X.P. Xie, Error estimation of a discontinuous Galerkin method for time fractional subdiffusion problems with nonsmooth data. *Fract. Calc. Appl. Anal.* **25** (2022) 747–782.

[34] J.L. Lions, Optimal Control of Systems Governed by Partial Differential Equations, 1st edition. Vol. 170. Springer, Berlin (1971).

[35] J.L. Lions and E. Magenes, Nonhomogeneous Boundary Value Problem and Applications. Vol. 1. Springer, New York (1972).

[36] J.L. Lions and E. Magenes, Nonhomogeneous Boundary Value Problem and Applications. Vol. 2. Springer, New York (1972).

[37] S. May, R. Rannacher and B. Vexler, Error analysis fo a finite element approximation of elliptic Dirichlet boundary control problems. *SIAM J. Control Optim.* **51** (2013) 2585–2611.

[38] D. Meidner and B. Vexler, A priori error estimates for space-time finite element discretization of parabolic optimal control problems. I: Problems without control constraints. *SIAM J. Control Optim.* **47** (2008) 1150–1177.

[39] D. Meidner and B. Vexler, A priori error estimates for space-time finite element discretization of parabolic optimal control problems. II. Problems with control constraints. *SIAM J. Control Optim.* **47** (2008) 1301–1329.

[40] G. Of, T.X. Phan and O. Steinbach, An energy space finite element approach for elliptic Dirichlet boundary control problems. *Numer. Math.* **129** (2015) 723–748.

[41] D. Schötzau, *Hp-DGFEM for Parabolic Evolution Problems.* Ph.D thesis, Swiss Federal Institute of Technology, Zürich (1999).

[42] J. Sokolowski and J.P. Zolesio, Introduction to Shape Optimization: Shape Sensitivity Analysis, 1st edition. Springer-Verlag, Berlin (1992).

[43] V. Thomée, Galerkin Finite Element Methods for Parabolic Problems, 2nd edition. Springer-Verlag, Berlin (1997).

[44] F. Tröltzsch, Optimal Control of Partial Differential Equations: Theory, Methods and Applications. Vol. 12. AMS, Providence, Rhode Island (2010).

[45] N. von Daniels, M. Hinze and M. Vierling, Crank–Nicolson time stepping and variational discretization of control-constrained parabolic optimal control problems. *SIAM J. Control Optim.* **53** (2015) 1182–1198.