FINITE ELEMENT APPROXIMATIONS OF PARABOLIC OPTIMAL CONTROL PROBLEMS WITH CONTROLS ACTING ON A LOWER DIMENSIONAL MANIFOLD

WEI GONG ◇ AND NINGNING YAN ‡

Abstract: This paper is devoted to the study of finite element approximations to parabolic optimal control problems with controls acting on a lower dimensional manifold. The manifold can be a point, a curve or a surface which may be independent of time or evolve in the time horizon, and is assumed to be strictly contained in the space domain. At first, we obtain the first order optimality conditions for the control problems and the corresponding regularity results. Then, for the control problems we consider the fully discrete finite element approximations based on the dG(0) scheme for time discretization and piecewise linear finite elements for space discretization, and variational discretization to the control variable. A priori error estimates are finally obtained for the fully discretised control problems and supported by numerical examples.

Keywords: finite element method, parabolic equation, optimal control problem, moving manifold, fully discrete error estimates.

Subject Classification: 49J20, 49K20, 65N15, 65N30.

1. Introduction

The aim of this paper is to analyze the finite element approximations of parabolic optimal control problems with controls acting on a lower dimensional manifold. Let $\Omega_T = \Omega \times I$, $\Gamma_T = \partial\Omega \times I$ with time interval $I = [0, T]$, Ω is an open bounded domain in $\mathbb{R}^n$ ($n = 2$ or 3) with boundary $\Gamma = \partial\Omega$. We consider the following parabolic optimal control problems

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

subject to

$$\begin{cases} \partial_t y - \Delta y = u(x,t)\delta_{\gamma(t)}(x) & \text{in } \Omega_T, \\ y = 0 & \text{on } \Gamma_T, \\ y(\cdot, 0) = y_0 & \text{in } \Omega, \end{cases} \tag{1.2}$$

where $\partial_t y = \frac{\partial y}{\partial t}$, $y_0 \in L^2(\Omega)$ and $T > 0$ are fixed. U is the control space which will be specified later. $\alpha > 0$ is a regularization parameter and $y_d \in L^2(I; L^2(\Omega))$ is the desired state. The admissible control set U_{ad} is of the following type:

$$U_{ad} := \big\{u \in U : \quad a \leqslant u(x,t) \leqslant b \quad \text{a.e. on } \gamma(t), \text{ a.a. } t \in I\big\}, \tag{1.3}$$

where $a < b$ are constants or constant vectors depending on the dimension of the manifold $\gamma(t)$.

Here we assume that $\gamma(t)$ is a lower dimensional continuous manifold which is strictly contained in Ω for all $t \in [0, T]$. $\delta_{\gamma(t)}$ denotes the Dirac measure on $\gamma(t)$. We stress that $\gamma(t)$ can be a point, a curve if $n \geqslant 2$ or even a surface if $n = 3$, it can be independent of time t or evolves in the time horizon. To ensure the well-posedness of the above optimal control problem and for the

---

Date: February 11, 2016.

◇ NCMIS, LSEC, Institute of Computational Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China. Email: wgong@lsec.cc.ac.cn.

‡ NCMIS, LSEC, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China. Email: ynn@amss.ac.cn.

convenience of error estimates and numerical computation, we assume the following hypotheses throughout the paper for the lower dimensional manifold $\gamma(t)$ according to [10]:

($\mathcal{A}_1$) $\gamma(t)$ is a Lipschitz-continuous k-dimensional manifold in Ω with $0 \leqslant k \leqslant n-1$ for all $t \in [0,T]$;
($\mathcal{A}_2$) The distance between $\gamma(t)$ and $\partial\Omega$ is positive for all $t \in [0,T]$;
($\mathcal{A}_3$) The set $\{\gamma(t)\}_{0\leqslant t\leqslant T}$ is a time-continuous family of manifolds in the sense of Castro and Zuazua (see Definition 3.1 in [10]). Furthermore, we assume that $\gamma(t)$ is of class $\mathcal{C}^0$ with respect to the time variable;
($\mathcal{A}_4$) The k-dimensional Hausdorff measure of $\gamma(t) \subset \Omega$ in $\mathbb{R}^n$ is finite for all $t \in [0,T]$, this means the measure of $\gamma(t)$ is uniformly bounded in $t \in [0,T]$ in view of Definition 3.1 in [10];
($\mathcal{A}_5$) When $k=0$, $\gamma(t)$ will reduce to a single point or a finite number of points for each $t \in [0,T]$; when $k=1$, $\gamma(t)$ is either a $\mathcal{C}^2$-curve s.t. $\gamma(t) \subset \partial D$ for some n-dimensional $\mathcal{C}^2$-domain $D \subset\subset \Omega$ or Lipschitz s.t. $\gamma(t) \subset \partial D$ for some 2-dimensional Lipschitz domain $D \subset\subset \Omega$ in the case that $n=2$ for each $t \in [0,T]$; when $k=2$, $\gamma(t)$ is either a $\mathcal{C}^2$-surface s.t. $\gamma(t) \subset \partial D$ for some 3-dimensional $\mathcal{C}^2$-domain $D \subset\subset \Omega$ or Lipschitz s.t. $\gamma(t) \subset \partial D$ for some 3-dimensional Lipschitz domain $D \subset\subset \Omega$ for each $t \in [0,T]$.

The motivation to consider optimal control problems with controls acting on a lower dimensional manifold comes from the fact that, the support of the controls needs to be very small compared to the total size of the domain Ω if we are restricted by the cost of controls. So it would be a good choice to consider the control to be located in such lower dimensional manifolds. There seems to be only few contributions in the literature focusing on this subject, and most of results are concentrated on the case that controls act on the whole domain Ω (see [40, 41]), or at least on a subdomain $\omega \subset \Omega$ ([35, 36]), or on the boundary of the domain Ω (see, e.g., [9, 20]). Only few papers can be found to our knowledge, among them we should mention the work of Castro and Zuazua who considered the approximate controllability of the heat equation with controls acting on an oscillating lower dimensional manifold in [10], and the work [27] of Khapalov who considered the controllability of the semilinear parabolic equation. Nguyen and Raymond also considered the control localized on thin structures for semilinear parabolic and convection-diffusion equations in [42] and [43].

In the past decades, numerical methods for optimal control problems attracted a lot of attentions from the fields of both control theory and numerical analysis. Among these numerical methods finite element plays an increasing role in the numerical analysis of optimal control problems governed by partial differential equations. The earliest work can be traced back to Falk in [16], since then a lot of achievements have been made in the aspect of a priori and a posteriori error estimates (see, e.g., [29, 35, 36, 37, 38, 39, 40, 41, 44, 48]). We refer to monographs [37] and [25] for recent developments.

For optimal control problems with controls acting on the boundary of the domain, Dirichlet and Neumann boundary control problems have been studied extensively, see for example [9], [22] and [25] for elliptic case and [20] for parabolic case and the references therein. In [21] the authors studied the finite element approximations to elliptic control problems with controls acting on a lower dimensional manifold, here we would like to generalise the results to parabolic case where the manifold may evolve in the time horizon. As a special case, Gong, Hinze and Zhou studied in [19] the finite element approximations to pointwise control of parabolic equations with control acting on finitely many spatial points which are independent of the time, the error estimates presented there are subsequently improved by Leykekhman and Vexler in [30] for two dimensional case. For optimal controls with compact support and sparsity we should mention the work of Kunisch and coauthors, who studied in [7], [8] and [28] the elliptic and parabolic optimal control problems in measure space.

In this paper we intend to consider the finite element approximations to parabolic optimal control problems with controls acting on a lower dimensional manifold. The manifold can be independent of time or evolve in the time horizon. The results of this paper cover the pointwise control of parabolic equations as a special case, and thus form a general framework. The controls for traditional control problems studied in the literature act on a subdomain of Ω, thus the error estimates involve only global errors. For the control problems acting on lower dimensional manifold, one needs to rely on some local error estimates to derive improved error estimates compared to traditional techniques (see [30]). As indicated in [21], when the dimension of manifold $\gamma(t)$ is one order lower than that of Ω, we are able to derive optimal a priori error estimates up to a logarithmic factor. Otherwise, we derive a priori error estimates for the control with reasonable reduced order. In this paper, we derive a priori error estimates for the optimal control problems in both two and three dimensions, but only optimal error estimates are obtained in two dimension following the idea of [30]. The main results of this paper are summarised as follows. In the case $n = 2$ and $k = 0$ or 1, the optimal error estimate

$$\sqrt{\alpha}\|u - u_{h\tau}\|_U \leqslant C|\log h|^{\frac{7}{2}}(h^2 + \tau) \tag{1.4}$$

holds for all $0 < h < h_0$ with some $0 < h_0 < 1$. Moreover, we have the suboptimal error estimates

$$\sqrt{\alpha}\|u - u_{h\tau}\|_U + \|y - y_{h\tau}\|_{L^2(I;L^2(\Omega))} \leqslant Ch^{\frac{1}{2}}, \quad \text{if } k = 0,\ n = 3; \tag{1.5}$$

$$\sqrt{\alpha}\|u - u_{h\tau}\|_U + \|y - y_{h\tau}\|_{L^2(I;L^2(\Omega))} \leqslant Ch^{\frac{5}{2}-\frac{3}{\sigma}},\ \sigma \in (\frac{3}{2}, 2) \quad \text{if } k = 1,\ n = 3; \tag{1.6}$$

$$\sqrt{\alpha}\|u - u_{h\tau}\|_U + \|y - y_{h\tau}\|_{L^2(I;L^2(\Omega))} \leqslant Ch^{\frac{3}{2}}, \quad \text{if } k = 2,\ n = 3 \tag{1.7}$$

under the coupling $\tau = O(h^2)$. We also present some two and three dimensional numerical experiments to support our theoretical results.

The structure of this paper is as follows. At first, in Section 2 we analyse the well-posedness of the state equation in order to obtain the first order optimality conditions for the control problems and the corresponding regularity results. Then, we consider in Section 3 the fully discrete finite element approximations to the state equation based on dG(0) scheme for time discretization and piecewise linear finite elements for space discretization. A priori error estimates for the state approximation is also derived. For the fully discretised control problems we use variational discretization to the control variable and finally derive a priori error estimates in Section 4. The last section is devoted to numerical experiments.

2. Theoretical analysis for the optimal control problems

2.1. **Notations.** Assume that $\Omega \subset \mathbb{R}^n$, $n = 2$ or 3 is a convex polygonal or polyhedral domain, or domain with a $\mathcal{C}^{1,1}$ boundary. We denote by $W^{m,p}(\Omega)$ the usual Sobolev space of order $m \geqslant 0$, $1 \leqslant p < \infty$ with norm $\|\cdot\|_{m,p,\Omega}$ and seminorm $|\cdot|_{m,p,\Omega}$, and the standard modification for $p = \infty$. For $p = 2$ we denote $W^{m,p}(\Omega)$ by $H^m(\Omega)$ and $\|\cdot\|_{m,\Omega} = \|\cdot\|_{m,2,\Omega}$, which is a Hilbert space. Note that $H^0(\Omega) = L^2(\Omega)$ and $H_0^1(\Omega) = \{v \in H^1(\Omega) :\ v = 0 \text{ on } \partial\Omega\}$.

For $p \in [1, \infty)$, the interval $I \subset \mathbb{R}$ and the Banach space A with norm $\|\cdot\|_A$, we denote by $L^p(I; A)$ the set of measurable functions $y : I \to A$ such that $\int_I \|y\|_A^p dt \leqslant \infty$. The norm on $L^p(I; A)$ is defined by

$$\|y(t)\|_{L^p(I;A)} = \begin{cases} \Big(\int_I \|y\|_A^p dt\Big)^{\frac{1}{p}} & 1 \leqslant p < \infty, \\ \operatorname*{ess\,sup}_{t\in I} \|y(t)\|_A & p = \infty. \end{cases}$$

We denote the L^2-inner products on $L^2(\Omega)$ and $L^2(\Omega_T)$ by

$$(v, w) = \int_\Omega vw dx \quad \forall\ v, w \in L^2(\Omega)$$

and

$$(v, w)_I = \int_I \int_\Omega vw dx dt \quad \forall\ v, w \in L^2(I; L^2(\Omega)),$$

respectively. In addition, c and C denote generic positive constants.

We next assume that the control $u(x,t)$ (or u if $k=0$) in (1.2) belongs to $L^2(I;L^2(\gamma))$ (resp. $L^2(I;\mathbb{R}^m)$), i.e.

$$\int_I\int_{\gamma(t)}|u(x,t)|^2dxdt<\infty \qquad (\text{resp.}\quad \int_I\|u(t)\|_{\mathbb{R}^m}^2dt<\infty),$$

where $\|\cdot\|_{\mathbb{R}^m}$ denotes the Euclidean norm in $\mathbb{R}^m$. From now on, we denote the control space $U:=L^2(I;\mathbb{R}^m)$ if $k=0$ and $U:=L^2(I;L^2(\gamma))$ if $k\geqslant 1$, with $(\cdot,\cdot)_U$ the inner-product between U and its dual space.

2.2. **Analysis of the state equation.** For $f\in L^2(I;L^2(\Omega))$, we assume that ψ is the solution of following backward in time parabolic problems:

$$\begin{cases} -\partial_t\psi-\Delta\psi=f & \text{in } \Omega_T,\\ \psi=0 & \text{on } \Gamma_T,\\ \psi(T)=0 & \text{in } \Omega. \end{cases} \tag{2.1}$$

Then the following standard stability estimates can be found in, e.g. [15, Ch.7, Theorem 5] and [34, Ch.4, Sec.6].

Lemma 2.1. *Let ψ denote the solution of problem (2.1). Then there holds $\psi\in L^2(I;H^2(\Omega)\cap H_0^1(\Omega))\cap H^1(I;L^2(\Omega))\hookrightarrow\mathcal{C}(I;H_0^1(\Omega))$ and*

$$\|\psi\|_{L^2(I;H^2(\Omega))}+\|\partial_t\psi\|_{L^2(I;L^2(\Omega))}\leqslant C\|f\|_{L^2(I;L^2(\Omega))} \tag{2.2}$$

and

$$\|\psi(\cdot,0)\|_{1,\Omega}\leqslant C\|f\|_{L^2(I;L^2(\Omega))}. \tag{2.3}$$

Now we will argue the existence and uniqueness of solution to the state equation (1.2). To begin with, we introduce the notations given in [10], i.e.,

$$H_0=\begin{cases} H^{-1}(\Omega), & \text{if } n-k=1,\\ L^2(\Omega), & \text{if } n-k>1; \end{cases}\qquad H_2=\begin{cases} H_0^1(\Omega), & \text{if } n-k=1,\\ H^2(\Omega)\cap H_0^1(\Omega), & \text{if } n-k>1 \end{cases}$$

and H_i', with $i=0,2$, stands for the dual space of H_i. We can verify that the righthand side of (1.2) satisfies

$$u(x,t)\delta_{\gamma(t)}\in L^2(I;H_2'). \tag{2.4}$$

In fact, we have for $\psi\in L^2(I;H_2)$ that

$$\langle u(x,t)\delta_{\gamma(t)},\psi\rangle_I=\begin{cases} \displaystyle\int_I\int_{\gamma(t)}u(x,t)\psi(x,t)dxdt, & \text{if } k\geqslant 1,\\ \displaystyle\sum_{j=1}^m\int_I u_j(t)\psi(\gamma_j(t),t)dt, & \text{if } k=0, \end{cases}$$

where $\langle\cdot,\cdot\rangle_I$ denote the duality pairing between $L^2(I;H_2)$ and its dual space. Note that

$$|\int_I\int_{\gamma(t)}u(x,t)\psi(x,t)dxdt|\leqslant\|u\|_{L^2(I;L^2(\gamma(t)))}\|\psi\|_{L^2(I;L^2(\gamma(t)))},\quad \text{if } k\geqslant 1,$$
$$|\sum_{j=1}^m\int_I u_j(t)\psi(\gamma_j(t),t)dt|\leqslant\|u\|_{L^2(I;\mathbb{R}^m)}\|\psi\|_{L^2(I;L^\infty(\Omega))},\quad \text{if } k=0.$$

We can conclude that

$$\|\psi\|_{L^2(I;L^2(\gamma(t)))}\leqslant\begin{cases} C(\gamma)\|\psi\|_{L^2(I;L^\infty(\Omega))}\leqslant C(\gamma)\|\psi\|_{L^2(I;H^2\cap H_0^1(\Omega))}, \text{ if } n=3;\ k=1,\\ C(\gamma)\|\psi\|_{L^2(I;H_0^1(\Omega))}, \text{ if } n=2,3;\ n-k=1 \end{cases}$$

and

$$\|\psi\|_{L^2(I;L^\infty(\Omega))}\leqslant C(\gamma)\|\psi\|_{L^2(I;H^2\cap H_0^1(\Omega))},\quad \text{if } n=2,3,$$

where we used the well-known embedding theorem

$$H^2(\Omega)\cap H_0^1(\Omega)\hookrightarrow \mathcal{C}(\bar\Omega),\ \text{ if } n=2,3$$

and the trace theorem

$$\|\psi|_{\gamma(t)}(\cdot,t)\|_{L^2(\gamma(t))}\leqslant C(\gamma(t))\|\psi(\cdot,t)\|_{H_0^1(\Omega)}\ \ \text{for } n=2,3 \text{ and } n-k=1.$$

This implies that

$$\langle u(x,t)\delta_{\gamma(t)},\psi\rangle_I\leqslant\begin{cases} C\|u\|_{L^2(I;L^2(\gamma(t)))}\|\psi\|_{L^2(I;H_2)}, & \text{if } k\geqslant 1,\\ C\|u\|_{L^2(I;\mathbb{R}^m)}\|\psi\|_{L^2(I;H_2)}, & \text{if } k=0,\end{cases}$$

this proves (2.4).

Then we are in the position to define the weak solution of problem (1.2), this is done by using transposition technique (see [33, Ch.2, Sec.5.2] and [34, Ch.4, Sec.9]). We define the solution to equation (1.2) with $u\in U_{ad}$ as follows: *Say that $y\in L^2(I;H_0')$ is a very weak solution to equation (1.2) if*

$$\text{(2.5)}\quad \langle u\delta_{\gamma(t)},\psi\rangle_{L^2(I;H_2'),L^2(I;H_2)}+(y_0,\psi(\cdot,0))=\langle y,f\rangle_{L^2(I;H_0'),L^2(I;H_0)}\quad\forall\ f\in L^2(I;H_0),$$

where $\psi\in L^2(I;H_2)$ is the unique solution of problem (2.1) with right hand side $f\in L^2(I;H_0)$. From (2.4) and the standard argument involving the Lax-Milgram theorem we can prove that equation (1.2) has a unique solution $y\in L^2(I;H_0')$ in the sense of (2.5).

Now we collect the regularity of solution to equation (1.2) under different cases in the following theorem.

Theorem 2.2. *Assume that $y_0\in H_0^1(\Omega)$ and $y\in L^2(I;L^2(\Omega))$ is the very weak solution of the state equation (1.2) defined by (2.5). Then it holds that*

$$y\in L^2(I;W_0^{1,s}(\Omega))\cap H^1(I;W^{-1,s}(\Omega))\qquad s\in(1,\frac{n}{n-1})\ \ \textit{when } k=0,\ n=2,3;$$

$$y\in L^2(I;W_0^{1,\sigma}(\Omega))\cap H^1(I;W^{-1,\sigma}(\Omega))\qquad \sigma\in(1,2)\ \ \textit{when } k=1,\ n=3;$$

$$y\in L^2(I;H^{\frac{3-\epsilon}{2}}(\Omega)\cap H_0^1(\Omega))\cap H^1(I;H^{-\frac{1+\epsilon}{2}}(\Omega))\qquad \textit{for all } \epsilon>0,\quad \textit{when } k\geqslant 1,\ n-k=1.$$

Proof. We show the regularity from several cases.

Case 1: $k=0$ and $n=2,3$

For each $s\in(1,\frac{n}{n-1})$, we let $s'>n$ be its conjugate number, i.e., $\frac{1}{s}+\frac{1}{s'}=1$. By the embedding theorem we have $W_0^{1,s'}(\Omega)\hookrightarrow\mathcal{C}(\bar\Omega)$.

By this, we can use the same argument showing (2.4) to verify that $u(x,t)\delta_{\gamma(t)}$ can be identified as an element of $L^2(I;W^{-1,s}(\Omega))$. Since $\Delta: W_0^{1,s}(\Omega)\to W^{-1,s}(\Omega)$ is an isomorphism for $s\in(s_0,\frac{n}{n-1})$ for some $s_0\in[1,\frac{n}{n-1})$; see [26] for details on the choice of s_0. By using the result on maximal parabolic regularity (see [14] and [30]) equation (1.2) admits a unique solution $y\in L^2(I;W_0^{1,s}(\Omega))$ and $\partial_t y\in L^2(I;W^{-1,s}(\Omega))$ for all $s\in(s_0,\frac{n}{n-1})$ in the sense that

$$\text{(2.6)}\qquad \langle\partial_t y,v\rangle_I+(\nabla y,\nabla v)_I=\sum_{j=1}^m\int_I u_j(t)v(\gamma_j(t),t)dt\qquad\forall\ v\in L^2(I;W_0^{1,s'}(\Omega)),$$

where $\langle\cdot,\cdot\rangle_I$ denotes the duality pairing between $L^2(I;W_0^{-1,s}(\Omega))$ and $L^2(I;W_0^{1,s'}(\Omega))$. By slightly abusing the notation, we do not distinguish $\langle\cdot,\cdot\rangle_I$ in different duality pairings in this paper if no confusion is involved.

Case 2: $k=1$ and $n=3$

By using the similar argument as in Case 1, we can prove that in this case $y\in L^2(I;W_0^{1,s}(\Omega))\cap H^1(I;W^{-1,s}(\Omega))$ for all $s\in(1,\frac{3}{2})$. However, here we will derive higher regularity.

By the assumption $(\mathcal{A}_5)$, $\gamma(t)\subset\partial D$ for some 3-dimensional $\mathcal{C}^2$-domain $D\subset\subset\Omega$ for each $t\in[0,T]$. Then by Proposition 2.3 in [43] (see also Theorem 7.42 and Remark 7.45 in [1]), the trace mapping T_γ is continuous from $W^{r,p}(D)$ into $L^q(\gamma(t))$ for each $t\in[0,T]$, when $0\leqslant r\leqslant 2$,

$0 < 3-rp < k$ and $p \leqslant q < \frac{kp}{3-rp}$. Hence, it can be extended into a continuous mapping from $W^{r,p}(\Omega)$ into $L^q(\gamma(t))$. Thus, for each $\frac{3}{2} < \sigma < 2$, T_γ is continuous from $W_0^{1,\sigma'}(\Omega)$ to $L^q(\gamma(t))$, when $2 < \sigma' \leqslant q < \frac{\sigma'}{3-\sigma'} = \frac{\sigma}{2\sigma-3}$, where $2 < \sigma' < 3$ is the conjugate number of σ. Hence

$$
\begin{aligned}
|\int_I \int_{\gamma(t)} u(x,t)v(x,t)dxdt| &\leqslant C\int_I \|u\|_{L^{q'}(\gamma(t))}\|v\|_{L^q(\gamma(t))}dt \\
&\leqslant C\int_I \|u\|_{L^{q'}(\gamma(t))}\|v\|_{W_0^{1,\sigma'}(\Omega)}dt \\
&\leqslant C\|u\|_{L^2(I;L^2(\gamma(t)))}\|v\|_{L^2(I;W_0^{1,\sigma'}(\Omega))} \qquad (2.7)
\end{aligned}
$$

for each $v \in L^2(I;W_0^{1,\sigma'}(\Omega))$, which leads to that $u(x,t)\delta_{\gamma(t)} \in L^2(I;W^{-1,\sigma}(\Omega))$ for all $\frac{3}{2} < \sigma < 2$. From this, we can use the same way used in Case 1 to get that $y \in L^2(I;W_0^{1,\sigma}(\Omega)) \cap H^1(I;W^{-1,\sigma}(\Omega))$ for all $\frac{3}{2} < \sigma < 2$. This, combining with $y \in L^2(I;W_0^{1,s}(\Omega)) \cap H^1(I;W^{-1,s}(\Omega))$ for all $s \in (1,\frac{3}{2})$, implies $y \in L^2(I;W_0^{1,\sigma}(\Omega)) \cap H^1(I;W^{-1,\sigma}(\Omega))$ for all $1 < \sigma < 2$.

Case 3: $k \geqslant 1$ and $n-k=1$

Clearly, $u(x,t)\delta_{\gamma(t)}$ can be identified as an element of $L^2(I;H^{-1}(\Omega))$, thus we conclude that $y \in L^2(I;H_0^1(\Omega))$ and $\partial_t y \in L^2(I;H^{-1}(\Omega))$ in the sense that

$$
(\partial_t y, v)_I + (\nabla y, \nabla v)_I = \int_I \int_{\gamma(t)} u(x,t)v(x,t)dxdt \quad \forall\, v \in L^2(I;H_0^1(\Omega)). \qquad (2.8)
$$

We aim to show that $y \in L^2(I;H^{\frac{3-\epsilon}{2}}(\Omega) \cap H_0^1(\Omega)) \cap H^1(I;H^{-\frac{1+\epsilon}{2}}(\Omega))$ for each $\epsilon > 0$. By the assumption $(\mathcal{A}_5)$, for each $t \in [0,T]$, $\gamma(t) \subset \hat{\gamma}(t) \triangleq \partial D(t)$ for some n-dimensional Lipschitz or $\mathcal{C}^2$ domain $D(t) \subset\subset \Omega$. Arbitrarily fix an $\epsilon > 0$. By the trace theorem (see [23, Thm.1.5.1.2] and [34, Vol.I, Section 9.2]), one can observe that

$$
\begin{aligned}
\int_I \int_{\gamma(t)} u(x,t)v(x,t)dxdt &\leqslant C\int_I \|u\|_{L^2(\gamma(t))}\|v\|_{L^2(\gamma(t))}dt \\
&\leqslant C(\epsilon)\int_I \|u\|_{L^2(\gamma(t))}\|v\|_{H^{\frac{1+\epsilon}{2}}(D(t))}dt \\
&\leqslant C(\epsilon)\|u\|_{L^2(I;L^2(\gamma(t)))}\|v\|_{L^2(I;H^{\frac{1+\epsilon}{2}}(\Omega))} \qquad (2.9)
\end{aligned}
$$

for all $v \in L^2(I;H^{\frac{1+\epsilon}{2}}(\Omega))$, which implies that $u(x,t)\delta_{\gamma(t)} \in L^2(I;H^{-\frac{1+\epsilon}{2}}(\Omega))$ defines a bounded linear functional on $L^2(I;H^{\frac{1+\epsilon}{2}}(\Omega))$. By the result on maximal parabolic regularity theory (see [14]) one can deduce that $y \in L^2(I;H^{\frac{3-\epsilon}{2}}(\Omega) \cap H_0^1(\Omega)) \cap H^1(I;H^{-\frac{1+\epsilon}{2}}(\Omega))$ for each $\epsilon > 0$. $\square$

2.3. **Analysis of the optimal control problems.** We denote the control-to-state mapping of the state equation by $y := Su$. Then problem (1.1) can be reduced to the optimization problem:

$$
\min_{u \in U_{ad}} \quad J(u) \quad = \frac{1}{2}\|Su - y_d\|^2_{L^2(I;L^2(\Omega))} + \frac{\alpha}{2}\|u\|^2_U. \qquad (2.10)
$$

Since U_{ad} is bounded, convex and S is affine linear, by the standard argument ([33, Ch.2, Sec.1.2]) we can prove that the problem (2.10) admits a unique solution $u \in U_{ad}$. Moreover, there exists an adjoint state $z \in L^2(I;H_0^1(\Omega))$ such that the following first order optimality condition holds:

$$
\begin{cases}
-\partial_t z - \Delta z = y - y_d & \text{in } \Omega_T, \\
z = 0 & \text{on } \Gamma_T, \\
z(T) = 0 & \text{in } \Omega
\end{cases} \qquad (2.11)
$$

and

$$
J'(u)(v-u) \geqslant 0 \quad \forall\, v \in U_{ad},
$$

or more precisely,

$$
(\alpha u + z|_\gamma, v-u)_U \geqslant 0 \quad \forall\, v \in U_{ad}. \qquad (2.12)
$$

Here $z|_\gamma$ denotes the restriction of z on the manifold γ. *We say that (y, u, z) is the solution to the problem (2.10) when (y, u) is the optimal pair and z is the corresponding adjoint state.*

The precise regularity of the solutions to the problem (2.10) are collected in the following theorem.

Theorem 2.3. *Let $(y, u, z) \in L^2(I; L^2(\Omega)) \times U_{ad} \times L^2(I; H_0^1(\Omega))$ be the solution to the problem (2.10). Assume that $y_0 \in H_0^1(\Omega)$. Then the following regularity results hold:*

$$y \in L^2(I; W_0^{1,s}(\Omega)) \cap H^1(I; W^{-1,s}(\Omega)), \quad s \in (1, \frac{n}{n-1}),$$
$$z \in L^2(I; H^2(\Omega) \cap H_0^1(\Omega)) \cap H^1(I; L^2(\Omega)), \quad u \in L^2(I; R^m), \quad \textit{if } k = 0, n = 2, 3;$$
$$y \in L^2(I; W_0^{1,\sigma}(\Omega)) \cap H^1(I; W^{-1,\sigma}(\Omega)), \quad \sigma \in (1, 2),$$
$$z \in L^2(I; H^2(\Omega) \cap H_0^1(\Omega)) \cap H^1(I; L^2(\Omega)), \quad u \in L^2(I; H^1(\gamma(t))), \quad \textit{if } k = 1, n = 3;$$
$$y \in L^2(I; H^{\frac{3-\epsilon}{2}}(\Omega) \cap H_0^1(\Omega)) \cap H^1(I; H^{-\frac{1+\epsilon}{2}}(\Omega)), \quad \textit{for any } \epsilon > 0,$$
$$z \in L^2(I; H^2(\Omega) \cap H_0^1(\Omega)) \cap H^1(I; L^2(\Omega)), \quad u \in L^2(I; H^1(\gamma(t))), \quad \textit{if } k \geqslant 1, n - k = 1.$$

Proof. The desired regularity for y has been proved in Theorem 2.2. Since Ω is convex and $y - y_d \in L^2(I; L^2(\Omega))$ we have $z \in L^2(I; H^2(\Omega) \cap H_0^1(\Omega)) \cap H^1(I; L^2(\Omega))$ for all the cases. By the Sobolev imbedding $H^2(\Omega) \cap H_0^1(\Omega) \hookrightarrow \mathcal{C}(\bar{\Omega})$, the restriction of z on the manifold $\gamma(t)$ is well defined. From this and (2.12), the control u can be represented via the adjoint state z by

$$u(t) = P_{U_{ad}}\Big(- \frac{1}{\alpha} z(\gamma_i(t))(t)\Big)_{i=1}^m, \text{ when } k = 0; \tag{2.13}$$

$$u(x, t) = P_{U_{ad}}(-\frac{1}{\alpha} z(x, t)|_\gamma), \text{ when } k \geqslant 1, \tag{2.14}$$

where $P_{U_{ad}}$ is the orthogonal projection onto U_{ad}. Thus $u \in L^2(I; R^m)$ if $k = 0$. By the trace theorem one derive that $z|_{\gamma(t)} \in L^2(I; H^1(\gamma(t)))$, this yields that $u \in L^2(I; H^1(\gamma(t)))$ if $k \neq 0$. $\square$

Note, that when $k \geqslant 1$ and $n - k = 1$ we have $y \in L^2(I; H^{\frac{3-\epsilon}{2}}(\Omega) \cap H_0^1(\Omega))$. Thus, we can conclude from the embedding theorem that $y \in L^2(I; L^\infty(\Omega))$ when $n = 2$, and $y \in L^2(I; L^q(\Omega))$ for any $q \in [1, \infty)$ as $\epsilon \to 0$ when $n = 3$.

In the case when $k = 0$ and $n = 2$ or 3, the solution y of problem (1.2) belongs to $L^2(I; W_0^{1,s}(\Omega))$ for all $s \in [1, \frac{n}{n-1})$, thus from the embedding theorem we have $y \in L^2(I; L^q(\Omega))$ for $q = \frac{ns}{n-s}$, which means $q < +\infty$ when $n = 2$ and $q < 3$ when $n = 3$.

In the case when $k = 1$ and $n = 3$, we have $y \in L^2(I; W_0^{1,\sigma}(\Omega))$ for any $\sigma \in [1, 2)$. Then it follows by the well-known embedding theorem that $y \in L^2(I; L^q(\Omega))$ for $q = \frac{3\sigma}{3-\sigma}$.

If we assume in addition that $n = 2$, $y_d \in L^2(I; L^\infty(\Omega))$ (which is obviously not restrictive), for any subdomain $\Omega_0 \subset\subset \Omega$ we can conclude from Lemma 2.2 in [30] that $z \in L^2(I; W^{2,q}(\Omega_0)) \cap H^1(I; L^q(\Omega_0))$ and

$$\|z\|_{L^2(I;W^{2,q}(\Omega_0))} + \|\partial_t z\|_{L^2(I;L^q(\Omega_0))} \leqslant Cq(\|y\|_{L^2(I;L^q(\Omega))} + \|y_d\|_{L^2(I;L^\infty(\Omega))}) \tag{2.15}$$

holds for any $2 \leqslant q < \infty$.

3. Error estimates for fully discrete finite element approximations of the state equation

3.1. **Finite element spaces.** Let us now consider the finite element approximations to the state equation (1.2). To this aim, we consider a family of triangulation $\mathcal{T}^h$ of $\overline{\Omega}$, such that $\overline{\Omega} = \bigcup_{e \in \mathcal{T}^h} \overline{e}$. We suppose that $\overline{\Omega}$ is the union of the elements of $\mathcal{T}^h$ so that element edges lying on the boundary may be curved if Ω has curved boundary. This triangulation is supposed to be shape regular in the usual sense (see [12]). For each element $e \in \mathcal{T}^h$ we associate two parameters $\rho(e)$ and $\sigma(e)$, where $\rho(e)$ denotes the diameter of the element e and $\sigma(e)$ is the supremum of the diameters of all circles contained in e. Define the size of the mesh by $h = \max_{e\in\mathcal{T}^h} \rho(e)$. We suppose that the following regularity assumptions are satisfied: There exists a positive constant C such that

$$\frac{\rho(e)}{\sigma(e)} \leqslant C, \quad \frac{h}{\rho(e)} \leqslant C \tag{3.1}$$

hold for all $e \in \mathcal{T}^h$ and all $h > 0$.

Here we consider only n-simplex elements, as they are among the most widely used ones. Associated with $\mathcal{T}^h$ is a finite dimensional subspace V^h of $\mathcal{C}(\overline{\Omega})$, such that $\chi|_e$ is linear for $\forall\ \chi \in V^h$ and $e \in \mathcal{T}^h$. We also set $V_0^h = V^h \cap H_0^1(\Omega)$.

Note that the regular assumption (3.1) guarantees the following inverse properties for $v_h \in V^h$ (see [4, Sec.4.5] and [12, Sec.3.2]):

$$\|v_h\|_{s,\Omega} \leqslant Ch^{l-s}\|v_h\|_{l,\Omega} \quad 0 \leqslant l \leqslant s \leqslant 1 \tag{3.2}$$

and

$$\|v_h\|_{0,\infty,\Omega} \leqslant Ch^{-\frac{n}{q}}\|v_h\|_{0,q,\Omega}, \ 1 \leqslant q < \infty, \tag{3.3}$$

$$\|v_h\|_{0,\infty,\Omega} \leqslant C\rho(n,h)\|v_h\|_{1,\Omega}, \tag{3.4}$$

where

$$\rho(n,h) = \begin{cases} \sqrt{|\log h|}, & n = 2; \\ h^{-\frac{1}{2}}, & n = 3. \end{cases} \tag{3.5}$$

Let $\pi_h : \mathcal{C}(\overline{\Omega}) \to V^h$ denote the standard Lagrange interpolation operator, then interpolation error estimate implies that for $y \in W^{2,q}(\Omega)$, $\frac{n}{2} < q < \infty$ there holds (see, e.g., [4, Sec.4.4] and [12, Sec.3.1])

$$\|y - \pi_h y\|_{0,q,\Omega} + h\|y - \pi_h y\|_{1,q,\Omega} \leqslant Ch^2\|y\|_{2,q,\Omega} \tag{3.6}$$

and

$$\|y - \pi_h y\|_{0,\infty,\Omega} \leqslant Ch^{2-\frac{n}{q}}\|y\|_{2,q,\Omega}. \tag{3.7}$$

Let $\mathcal{P}_h$ be the $L^2(\Omega)$-projection operator defined from $L^2(\Omega)$ to V^h:

$$(\mathcal{P}_h y, v_h) = (y, v_h) \quad \forall\ v_h \in V^h \tag{3.8}$$

and $\mathcal{R}_h : H_0^1(\Omega) \to V_0^h$ denote the Ritz projection operator defined as

$$(\nabla\mathcal{R}_h y, \nabla v_h) = (\nabla y, \nabla v_h) \quad \forall\ v_h \in V_0^h. \tag{3.9}$$

Then we have the following error estimates (see, e.g., [4, Sec.5.4, 5.8]):

$$\|y - \mathcal{P}_h y\|_{-1,\Omega} + h\|y - \mathcal{P}_h y\|_{0,\Omega} \leqslant Ch^2\|y\|_{1,\Omega}, \tag{3.10}$$

$$\|y - \mathcal{R}_h y\|_{0,\Omega} + h\|y - \mathcal{R}_h y\|_{1,\Omega} \leqslant Ch^2\|y\|_{2,\Omega}. \tag{3.11}$$

Moreover, we have (see [3], [12, p. 168] and the references cited therein)

$$\|y - \mathcal{R}_h y\|_{0,\infty,\Omega} \leqslant Ch^{2-\frac{n}{2}}\|y\|_{2,\Omega}. \tag{3.12}$$

3.2. **Fully discrete finite element approximations to parabolic equations.** The semi-discrete finite element approximation to the state equation (1.2) is to find $y_h \in L^2(I; V_0^h)$ such that

$$(\partial_t y_h, v_h)_I + (\nabla y_h, \nabla v_h)_I = \langle u(x,t)\delta_{\gamma(t)}, v_h\rangle_I \quad \forall\ v_h \in V_0^h \tag{3.13}$$

with $y_h(0) = \mathcal{P}_h y_0$ the L^2-projection of y_0.

We next consider the fully discrete approximations to the state equation (1.2) by using the piecewise constant discontinuous Galerkin method (dG(0) for short, see [41] and [47]). We consider a partitioning of the time interval $I = [0, T]$ as

$$\overline{I} = I_1 \cup I_2 \cup \cdots \cup I_N$$

with subintervals $I_i = [t_{i-1}, t_i]$ of size τ_i and time points

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

Let $\tau = \max_{1 \leqslant i \leqslant N} \tau_i$. We assume that the time partitioning is quasi-uniform, i.e., there exists positive constant c such that $\tau \leqslant c\tau_i$ holds for each $i = 1, 2, \cdots, N$. For simplicity we consider the same finite element space on each time step.

We set

$$V_{h,\tau} := \Big\{ v : \bar{\Omega} \times I \to \mathbb{R},\ v(\cdot, t)|_{\bar{\Omega}} \in V_0^h,\ v(x, \cdot)|_{I_i} \in \mathbb{P}_0 \text{ for } i = 1, \cdots, N \Big\},$$

i.e., $v \in V_{h,\tau}$ is a piecewise constant polynomial with respect to time. To introduce the fully discretization we define the bilinear form

$$A(y, w) := \sum_{i=1}^{N} \langle y_t, w \rangle_{I_i} + (\nabla y, \nabla w)_I + \sum_{i=2}^{N} (y^i - y^{i-1}, w^i) + (y_+^0, w_+^0),$$

where $\langle \cdot, \cdot \rangle_{I_i}$ denotes the duality pairing between $L^2(I_i; W_0^{-1,s}(\Omega))$ and $L^2(I_i; W_0^{1,s'}(\Omega))$, and $w^i := w_-^i = \lim_{r \to 0^+} w(t_i - r)$, $w_+^i = \lim_{r \to 0^+} w(t_i + r)$. Then for $y, w \in V_{h,\tau}$ we have

$$A(y, w) = \sum_{i=1}^{N} \tau_i (\nabla y^i, \nabla w^i) + \sum_{i=2}^{N} (y^i - y^{i-1}, w^i) + (y_+^0, w_+^0).$$

We define $\pi_\tau : L^2(I) \to \mathbb{P}_0(I)$ as the L^2-projection operator on the variable t such that $\pi_\tau v|_{I_i}$ is pieceise constant on I_i for each $v \in L^2(I)$. Then there holds $\pi_\tau v|_{I_i} = \frac{1}{\tau_i} \int_{I_i} v dt$. Combining with the spatial interpolation operator π_h we can define the space-time projection operator $\pi_{h\tau} : L^2(I; \mathcal{C}(\overline{\Omega})) \to V_{h,\tau}$ such that $\pi_{h\tau} = \pi_h \pi_\tau = \pi_\tau \pi_h$.

At first, we consider the fully discrete finite element approximation of the backward parabolic equation (2.1). The discrete approximation can be stated as: Find $\psi_{h\tau} \in V_{h,\tau}$ such that

$$A(w_{h\tau}, \psi_{h\tau}) = \sum_{i=1}^{N} \int_{I_i} (f, w_{h\tau}), \quad \forall\ w_{h\tau} \in V_{h,\tau}. \tag{3.14}$$

Since $\psi_{h\tau}$ is the standard fully discrete finite element approximation of ψ, the following a priori error estimates can be found in the literature.

Lemma 3.1. *Assume that ψ and $\psi_{h\tau}$ are the solutions of problems (2.1) and (3.14), respectively. Then there holds*

$$\|\psi - \psi_{h\tau}\|_{L^2(I; L^2(\Omega))} \leqslant C(h^2 + \tau) \|f\|_{L^2(I; L^2(\Omega))}. \tag{3.15}$$

Moreover, the following global pointwise in space error estimate

$$\begin{aligned} \sup_{x \in \bar{\Omega}} \int_I |(\psi - \psi_{h\tau})(x, t)|^2 dt \quad &\leqslant \quad C |\log h|^2 \inf_{v_{h\tau} \in V_{h,\tau}} \big(\|\psi - v_{h\tau}\|_{L^2(I; L^\infty(\Omega))}^2 \\ &\qquad + h^{-\frac{4}{q}} \|\pi_\tau \psi - v_{h\tau}\|_{L^2(I; L^q(\Omega))}^2 \big) \end{aligned} \tag{3.16}$$

holds for $n = 2$ and $\forall 1 \leqslant q \leqslant \infty$.

Proof. The proof of (3.15) is quite standard and can be found in many papers, e.g., [17], the proof of (3.16) in 2D is given in Theorem 3.1 of [30]. □

As a consequence, we can derive the corresponding error estimate on the manifold $\gamma(t)$ from Lemma 3.1 and the trace theorem.

Lemma 3.2. *Assume that $\gamma(t)$ is a $k-$dimensional manifold with $k \geqslant 1$ and $n - k = 1$. Let $\psi \in L^2(I; H^2(\Omega) \cap H_0^1(\Omega))$ be the solution of problem (2.1), and $\psi_{h\tau} \in V_{h,\tau}$ be the solution of problem (3.14)). Then we have*

$$\|\psi - \psi_{h\tau}\|_{L^2(I; L^2(\gamma(t)))} \leqslant C(h^{\frac{3}{2}} + h\tau^{\frac{1}{4}} + h^{\frac{1}{2}}\tau^{\frac{1}{2}} + \tau^{\frac{3}{4}}) \|f\|_{L^2(I; L^2(\Omega))}. \tag{3.17}$$

Proof. It follows from [11, eq. (4.11), p.2851] that

$$\|\psi-\psi_{h\tau}\|_{L^2(I;H^1(\Omega))}\leqslant C(h+\tau^{\frac{1}{2}})\|f\|_{L^2(I;L^2(\Omega))}. \tag{3.18}$$

This combining with (3.15) and the trace theorem ([4, Theorem 1.6.6]) gives

$$\begin{aligned}\|\psi-\psi_{h\tau}\|_{L^2(I;L^2(\gamma(t)))} &\leqslant C\|\psi-\psi_{h\tau}\|^{\frac{1}{2}}_{L^2(I;L^2(\Omega))}\|\psi-\psi_{h\tau}\|^{\frac{1}{2}}_{L^2(I;H^1(\Omega))}\\ &\leqslant C(h^{\frac{3}{2}}+h\tau^{\frac{1}{4}}+h^{\frac{1}{2}}\tau^{\frac{1}{2}}+\tau^{\frac{3}{4}})\|f\|_{L^2(I;L^2(\Omega))},\end{aligned}$$

which gives the result. □

Since the manifold $\gamma(t)$ is strictly contained in Ω for all $t\in[0,T]$, there exists a subdomain $\Omega_0\subset\subset\Omega$ such that $\gamma(t)\subset\Omega_0$ for each $t\in[0,T]$. Moreover, there exists another subdomain Ω_1 such that $\Omega_0\subset\subset\Omega_1\subset\subset\Omega$. For the following purpose we need to derive a local maximum norm error estimate on subdomain Ω_0.

Lemma 3.3. *Let $n=2$. Assume that ψ and $\psi_{h\tau}$ are the solutions of problems (2.1) and (3.14), respectively. Then there holds*

$$\begin{aligned}\sup_{x\in\bar{\Omega}_0}\int_I|(\psi-\psi_{h\tau})(x,t)|^2dt &\leqslant C|\log h|^3\inf_{v_{h\tau}\in V_{h,\tau}}\big(\|\psi-v_{h\tau}\|^2_{L^2(I;L^\infty(\Omega_1))}\\ &\quad +h^{-\frac{4}{q}}\|\pi_\tau\psi-v_{h\tau}\|^2_{L^2(I;L^q(\Omega_1))}\big)+C|\log h|\|\psi-\psi_{h\tau}\|^2_{L^2(I;L^2(\Omega))}\end{aligned} \tag{3.19}$$

for all $0<h<h_0$ with some fixed $0<h_0<1$.

Proof. The proof follows the approaches of [45] and [30]. Let $x_0\in\bar{\Omega}_0$ and $D_0\subset\subset\Omega_1$ be a sphere with center at x_0 and diameter l, the diameter l is chosen as large as possible such that the sphere D_1 with center at x_0 and diameter $2l$ satisfies $D_1\subset\Omega_1$. In the following we may not trace the dependence of constant C on l. Let ω be a smooth cut-off function which is 1 in D_0 and compactly supported in Ω_1 such that $\|\omega\|_{W^{j,\infty}(\Omega_1)}\leqslant Cl^{-j}$, $0\leqslant j\leqslant 2$. Such functions can be constructed by the standard mollifier and are extensively used for local estimates (see [30, 45] and the references cited therein). We set $\tilde{\psi}=\omega\psi$ and let $\tilde{\psi}_{h\tau}\in V_{h,\tau}$ satisfy

$$A(v_{h\tau},\tilde{\psi}-\tilde{\psi}_{h\tau})=0\quad\forall v_{h\tau}\in V_{h,\tau},$$

which means that $\tilde{\psi}_{h\tau}$ is the fully discrete finite element approximation of $\tilde{\psi}$. Then we have

$$(\psi-\psi_{h\tau})(x_0,t)=(\tilde{\psi}-\psi_{h\tau})(x_0,t)=(\tilde{\psi}-\tilde{\psi}_{h\tau})(x_0,t)+(\tilde{\psi}_{h\tau}-\psi_{h\tau})(x_0,t).$$

From the global maximum norm error estimate (3.16) we have

$$\begin{aligned}\int_I|(\tilde{\psi}-\tilde{\psi}_{h\tau})(x_0,t)|^2dt &\leqslant C|\log h|^2\Big(\|\tilde{\psi}\|^2_{L^2(I;L^\infty(\Omega_1))}+h^{-\frac{4}{q}}\|\pi_\tau\tilde{\psi}\|^2_{L^2(I;L^q(\Omega_1))}\Big)\\ &\leqslant C|\log h|^2\Big(\|\psi\|^2_{L^2(I;L^\infty(\Omega_1))}+h^{-\frac{4}{q}}\|\pi_\tau\psi\|^2_{L^2(I;L^q(\Omega_1))}\Big).\end{aligned} \tag{3.20}$$

Note that $\tilde{\psi}_{h\tau}-\psi_{h\tau}$ satisfies the interior equation

$$A(v_{h\tau},\tilde{\psi}_{h\tau}-\psi_{h\tau})=0\quad\forall v_{h\tau}\in V_{h,\tau}(D_0), \tag{3.21}$$

where $V_{h,\tau}(D_0)$ is the subspace of $V_{h,\tau}$ which contains functions vanishing outside of D_0. It has been proved in [30, P. 2813] that

$$\int_I|(\tilde{\psi}_{h\tau}-\psi_{h\tau})(x_0,t)|^2dt\leqslant C|\log h|l^{-2}\|\tilde{\psi}_{h\tau}-\psi_{h\tau}\|^2_{L^2(I;L^2(D_0))} \tag{3.22}$$

provided that $h<cl$ for 2D case. Applying the triangle inequality we can derive

$$\|\tilde{\psi}_{h\tau}-\psi_{h\tau}\|_{L^2(I;L^2(D_0))}\leqslant\|\tilde{\psi}-\tilde{\psi}_{h\tau}\|_{L^2(I;L^2(D_0))}+\|\psi-\psi_{h\tau}\|_{L^2(I;L^2(D_0))}. \tag{3.23}$$

Moreover, using $|D_0|\leqslant Cl^2$ one has

$$\|\tilde{\psi}-\tilde{\psi}_{h\tau}\|_{L^2(I;L^2(D_0))}\leqslant Cl\|\tilde{\psi}-\tilde{\psi}_{h\tau}\|_{L^2(I;L^\infty(D_0))}. \tag{3.24}$$

Combining above estimates and using (3.16) again we arrive at

$$\begin{aligned}\int_I |(\tilde{\psi}_{h\tau}-\psi_{h\tau})(x_0,t)|^2dt &\leqslant C|\log h|^3(\|\psi\|^2_{L^2(I;L^\infty(\Omega_1))}+h^{-\frac{4}{q}}\|\pi_\tau\psi\|^2_{L^2(I;L^q(\Omega_1))})\\ &\quad +Cl^{-2}|\log h|\|\psi-\psi_{h\tau}\|^2_{L^2(I;L^2(\Omega))}.\end{aligned} \tag{3.25}$$

Combining (3.20) and (3.25) we have

$$\begin{aligned}\int_I |(\psi-\psi_{h\tau})(x_0,t)|^2dt &\leqslant C|\log h|^3(\|\psi\|^2_{L^2(I;L^\infty(\Omega_1))}+h^{-\frac{4}{q}}\|\pi_\tau\psi\|^2_{L^2(I;L^q(\Omega_1))})\\ &\quad +C|\log h|\|\psi-\psi_{h\tau}\|^2_{L^2(I;L^2(\Omega))}\end{aligned} \tag{3.26}$$

holds for all $0<h<h_0$ with some fixed $0<h_0<1$ depending on l. Let $x_0\in\bar{\Omega}_0$ be such that

$$\sup_{x\in\bar{\Omega}_0}\int_I |(\psi-\psi_{h\tau})(x,t)|^2dt=\int_I |(\psi-\psi_{h\tau})(x_0,t)|^2dt.$$

By replacing ψ and $\psi_{h\tau}$ in (3.26) by $\psi-v_{h\tau}$ and $\psi_{h\tau}-v_{h\tau}$ for any $v_{h\tau}\in V_{h,\tau}$, we complete the proof. □

3.3. **Fully discrete finite element approximations to the state equation.** The fully discrete finite element approximation based on dG(0) scheme for time discretisation and continuous, piecewise linear finite element method for spatial discretisation to the state equation (1.2) reads: Find $y_{h\tau}\in V_{h,\tau}$ such that

$$A(y_{h\tau},w_{h\tau})=\langle u\delta_{\gamma(t)},w_{h\tau}\rangle_I+(y_0,w^0_{h\tau,+}),\quad \forall\ w_{h\tau}\in V_{h,\tau}. \tag{3.27}$$

Note that on each time interval I_i, the solution $y^i_{h\tau}\in V^h_0$ satisfies

$$\begin{cases}(\dfrac{y^i_{h\tau}-y^{i-1}_{h\tau}}{\tau_i},w_h)+(\nabla y^i_{h\tau},\nabla w_h)=\langle u\delta_{\gamma(t)},w_h\rangle_{I_i},\forall\ w_h\in V^h_0, i=1,\cdots,N,\\ y^0_{h\tau}(x)=y^h_0(x)=\mathcal{P}_hy_0\quad x\in\Omega.\end{cases} \tag{3.28}$$

Here

$$\langle u\delta_{\gamma(t)},v_h\rangle_{I_i}:=\begin{cases}\dfrac{1}{\tau_i}\displaystyle\int_{I_i}\int_{\gamma(t)}u(x,t)v_h(x)dxdt\quad\forall\ v_h\in V^h\quad\text{if } k\geqslant 1,\\ \dfrac{1}{\tau_i}\displaystyle\sum_{j=1}^m\int_{I_i}u_j(t)v_h(\gamma_j(t))dt\quad\forall\ v_h\in V^h\quad\text{if } k=0.\end{cases}$$

If $\gamma(t)$ is a $k-$dimensional moving manifold with $n-k>1$, then we need to estimate the error between continuous and discrete problem carefully. In [18] the author derived a priori error estimate for parabolic equation with measure data in space where the measure data is independent of the time, which covers the cases in this paper of $n-k>1$ and $\gamma(t)$ being independent of time. Here we extend the results of [18] to problems with time-dependent space measures and derive a priori error estimates with variable time steps and separated spatial and time discretisation errors although the proof is very similar. We note that the error estimates for this kind of problems are of independent interest for, e.g., the identification of moving pointwise source of parabolic equations and other related topics (see [2] and [32]). We mention that the error estimates for the elliptic equation with measure data were derived in [5].

Now we are in a position to estimate the error between the solutions of problem (1.2) and (3.28) in the case $n-k>1$. For compactness we will include the proof of this theorem in the Appendix.

Theorem 3.4. *Assume that $\gamma(t)$ is a $k-$dimensional manifold with $n-k>1$. Assume that $u\in L^2(I;L^2(\gamma(t)))$ when $k\geqslant 1$ and $u\in L^2(0,T;\mathbb{R}^m)$ when $k=0$, $y_0\in L^2(\Omega)$. Let $y\in L^2(I;L^2(\Omega))$ be the solution of problem (1.2), and $y_{h\tau}\in V_{h,\tau}$ be the solution of problem (3.27)). Then it holds that*

$$\|y-y_{h\tau}\|_{L^2(I;L^2(\Omega))}\leqslant C(h^{2-\frac{n}{2}}+h^{-\frac{n}{2}}\tau)\Big(\|u\|_U+\|y_0\|_{0,\Omega}\Big) \tag{3.29}$$

when $k=0$ and $n=2$ or 3. Furthermore, when $n=3$ and $k=1$, it holds that

$$\|y-y_{h\tau}\|_{L^2(I;L^2(\Omega))} \leqslant C(h^{\frac{5}{2}-\frac{3}{\sigma}}+h^{\frac{1}{2}-\frac{3}{\sigma}}\tau)\Big(\|u\|_U+\|y_0\|_{0,\Omega}\Big) \tag{3.30}$$

for each $\sigma \in (\frac{3}{2},2)$.

As indicated in [30], one can obtain improved error estimates for the optimal control in two dimension by exploiting higher order convergence of the state under norm $L^2(I;L^1(\Omega))$ instead of the $L^2(I;L^2(\Omega))$-norm used in [19]. For the purpose of deriving optimal a priori error estimate we need to obtain a sharper error estimate for the state approximation under a weaker norm $L^2(I;L^1(\Omega))$ in two dimension, which goes back to the idea of [30].

Theorem 3.5. *Assume that $\gamma(t)$ consists of m points ($k=0$) which are strictly contained in Ω with $n=2$. Let the assumptions in Theorem 3.4 be valid. Let $y\in L^2(I;L^2(\Omega))$ be the solution of problem (1.2), and $y_{h\tau}$ be the solution of problem (3.27)). Then it holds that*

$$\|y-y_{h\tau}\|_{L^2(I;L^1(\Omega))} \leqslant C|\log h|^{\frac{5}{2}}(h^2+\tau)\|u\|_{L^2(I;\mathbb{R}^m)} \tag{3.31}$$

for all $0<h<h_0$ with some $0<h_0<1$.

Proof. To begin with we use duality argument following [30]. Let ψ be the solution of the backward parabolic equation (2.1) with right hand side $f(x,t)=\mathrm{sign}(e(x,t))\|e(\cdot,t)\|_{L^1(\Omega)}$, where $e(x,t)=y-y_{h\tau}$. The corresponding fully discrete finite element approximation $\psi_{h\tau}\in V_{h,\tau}$ of ψ is defined by (3.14). Thus, we have the Galerkin orthogonality

$$A(w_{h\tau},\psi-\psi_{h\tau})=0, \quad \forall\ w_{h\tau}\in V_{h,\tau}. \tag{3.32}$$

Using the Galerkin orthogonal properties of $y-y_{h\tau}$ and $\psi-\psi_{h\tau}$ we can derive

$$\begin{aligned}
\int_I \|e(\cdot,t)\|^2_{L^1(\Omega)}dt &= \int_I(\mathrm{sign}(e(x,t))\|e(\cdot,t)\|_{L^1(\Omega)},e)dt=(f,e)\\
&= A(y-y_{h\tau},\psi)=A(y-y_{h\tau},\psi-\psi_{h\tau})\\
&= A(y,\psi-\psi_{h\tau})=\langle u(x,t)\delta_{\gamma(t)},\psi-\psi_{h\tau}\rangle_I\\
&= \sum_{j=1}^m\int_I u_j(\gamma_j(t),t)(\psi-\psi_{h\tau})(\gamma_j(t),t)dt\\
&\leqslant C\|u\|_{L^2(I;\mathbb{R}^m))}\big(\sup_{x\in\bar\Omega_0}\int_I|(\psi-\psi_{h\tau})(x,t)|^2dt\big)^{\frac{1}{2}}. \qquad (3.33)
\end{aligned}$$

From the local maximum norm error estimate (3.19) in Lemma 3.3 we are able to obtain

$$\begin{aligned}
&\big(\sup_{x\in\bar\Omega_0}\int_I|(\psi-\psi_{h\tau})(x,t)|^2dt\big)^{\frac{1}{2}}\leqslant C|\log h|^{\frac{3}{2}}\inf_{v_{h\tau}\in V_{h,\tau}}(\|\psi-v_{h\tau}\|_{L^2(I;L^\infty(\Omega_1))}\\
&+h^{-\frac{2}{q}}\|\pi_\tau\psi-v_{h\tau}\|_{L^2(I;L^q(\Omega_1))})+C|\log h|^{\frac{1}{2}}\|\psi-\psi_{h\tau}\|_{L^2(I;L^2(\Omega))} \qquad (3.34)
\end{aligned}$$

for all $0<h<h_0$ with some $0<h_0<1$. Note that $f\in L^2(I;L^\infty(\Omega))$, then we can conclude from Lemma 2.2 in [30] that for any subdomain $\Omega_1\subset\subset\Omega$, $\psi\in L^2(I;W^{2,q}(\Omega_1))\cap H^1(I;L^q(\Omega_1))$, $1\leqslant q<\infty$ and satisfies

$$\|\psi\|_{L^2(I;W^{2,q}(\Omega_1))}+\|\psi_t\|_{L^2(I;L^q(\Omega_1))}\leqslant Cq\|f\|_{L^2(I;L^\infty(\Omega))}. \tag{3.35}$$

Taking $v_{h\tau}=\pi_{h\tau}\psi$ in (3.34) we now estimate the right hand side terms one by one. From the triangle inequality, the interpolation error estimate, the inverse inequality and proceed as in the proof of (A.12) we have

$$\begin{aligned}
\|\psi-\pi_h\pi_\tau\psi\|_{L^2(I;L^\infty(\Omega_1))} &\leqslant \|\psi-\pi_h\psi\|_{L^2(I;L^\infty(\Omega_1))}+\|\pi_h\psi-\pi_h\pi_\tau\psi\|_{L^2(I;L^\infty(\Omega_1))}\\
&\leqslant Ch^{2-\frac{2}{q}}\|\psi\|_{L^2(I;W^{2,q}(\Omega_1))}+Ch^{-\frac{2}{q}}\|\pi_h\psi-\pi_h\pi_\tau\psi\|_{L^2(I;L^q(\Omega_1))}\\
&\leqslant Ch^{2-\frac{2}{q}}\|\psi\|_{L^2(I;W^{2,q}(\Omega_1))}+Ch^{-\frac{2}{q}}\tau\|\partial_t\psi\|_{L^2(I;L^q(\Omega_1))}. \qquad (3.36)
\end{aligned}$$

For the second term we have

$$
\begin{aligned}
& \|\pi_\tau\psi - \pi_h\pi_\tau\psi\|_{L^2(I;L^q(\Omega_1))} \\
\leqslant\ & \|\pi_\tau\psi - \psi\|_{L^2(I;L^q(\Omega_1))} + \|\psi - \pi_h\psi\|_{L^2(I;L^q(\Omega_1))} + \|\pi_h\psi - \pi_h\pi_\tau\psi\|_{L^2(I;L^q(\Omega_1))} \\
\leqslant\ & C\tau\|\partial_t\psi\|_{L^2(I;L^q(\Omega_1))} + Ch^2\|\psi\|_{L^2(I;W^{2,q}(\Omega_1))}. \qquad (3.37)
\end{aligned}
$$

For the third term we conclude from (3.15) that

$$\|\psi - \psi_{h\tau}\|_{L^2(I;L^2(\Omega))} \leqslant C(h^2+\tau)\|f\|_{L^2(I;L^2(\Omega))}. \qquad (3.38)$$

Collecting (3.34)-(3.38) we obtain

$$
\begin{aligned}
\big(\sup_{x\in\bar{\Omega}_0}\int_I |(\psi-\psi_{h\tau})(x,t)|^2dt\big)^{\frac{1}{2}} & \leqslant C|\log h|^{\frac{3}{2}}h^{-\frac{2}{q}}(h^2+\tau)\big(\|\psi\|_{L^2(I;W^{2,q}(\Omega_1))} + \|\partial_t\psi\|_{L^2(I;L^q(\Omega_1))}\big) \\
& \leqslant C|\log h|^{\frac{3}{2}}qh^{-\frac{2}{q}}(h^2+\tau)\|f\|_{L^2(I;L^\infty(\Omega))} \\
& \leqslant C|\log h|^{\frac{3}{2}}qh^{-\frac{2}{q}}(h^2+\tau)\|e\|_{L^2(I;L^1(\Omega))}. \qquad (3.39)
\end{aligned}
$$

Setting $q = |\log h|$ we complete the proof. □

Note that if $\gamma(t)$ is a $k-$dimensional moving manifold with $k \geqslant 1$ and $n-k=1$, then we can derive from Theorem 2.3 that $y \in L^2(I;H^{\frac{3-\epsilon}{2}}(\Omega)\cap H_0^1(\Omega))\cap H^1(I;H^{-\frac{1+\epsilon}{2}}(\Omega))$. Thus, we can derive the following a priori error estimate.

Theorem 3.6. *Assume that $\gamma(t)$ is a $k-$dimensional manifold with $n-k=1$. Let $y_0 \in H_0^1(\Omega)$ and $y \in L^2(I;H^{\frac{3-\epsilon}{2}}(\Omega)\cap H_0^1(\Omega))\cap H^1(I;H^{-\frac{1+\epsilon}{2}}(\Omega))$ be the solution of problem (1.2) for any $\epsilon > 0$, and $y_{h\tau} \in V_{h,\tau}$ be the solution of problem (3.27). Then it holds that*

$$\|y - y_{h\tau}\|_{L^2(I;L^2(\Omega))} \leqslant C(h^{\frac{3}{2}} + h^{\frac{1}{2}}\tau^{\frac{1}{2}} + h^{-\frac{1}{2}}\tau)\Big(\|u\|_{L^2(I;L^2(\gamma(t)))} + \|y_0\|_{1,\Omega}\Big). \qquad (3.40)$$

Proof. To prove this theorem we use the similar approach as in the proof of Theorem 3.4. At first we derive the stability estimate for numerical scheme (3.28) in case that $n-k=1$.

Let $y_{h\tau}^i \in V_0^h,\ i = 1,2,\cdots,N$ be the solutions of fully discrete scheme (3.28). Then there exists a constant C independent of h,τ and the data u such that

$$\sum_{i=1}^N \|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{0,\Omega}^2 + \tau_N\|y_{h\tau}^N\|_{1,\Omega}^2 \leqslant C\tau\|y_0\|_{1,\Omega}^2 + C\tau h^{-1}\|u(x,t)\|_{L^2(I;L^2(\gamma(t)))}^2. \qquad (3.41)$$

The proof is a slight modification of the proof in [18], see also (A.1)-(A.2). In fact, let $w_h = \tau_i(y_{h\tau}^i - y_{h\tau}^{i-1})$ in (3.28) we get

$$(y_{h\tau}^i - y_{h\tau}^{i-1}, y_{h\tau}^i - y_{h\tau}^{i-1}) + \tau_i(\nabla y_{h\tau}^i, \nabla(y_{h\tau}^i - y_{h\tau}^{i-1})) = \tau_i\langle u\delta_{\gamma(t)}, y_{h\tau}^i - y_{h\tau}^{i-1}\rangle_{I_i}.$$

Since $y_{h\tau}^i - y_{h\tau}^{i-1}$ is piecewise constant on each time interval I_i w.r.t time, thus it follows from the trace theorem ([4, Theorem 1.6.6]) and the inverse estimate that

$$
\begin{aligned}
& \|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{0,\Omega}^2 + \tau_i\|y_{h\tau}^i\|_{1,\Omega}^2 \\
\leqslant\ & \tau_i(\nabla y_{h\tau}^i, \nabla y_{h\tau}^{i-1}) + \int_{I_i}\int_{\gamma(t)} u(x,t)(y_{h\tau}^i - y_{h\tau}^{i-1})dxdt \\
\leqslant\ & \frac{1}{2}\tau_i\|y_{h\tau}^i\|_{1,\Omega}^2 + \frac{1}{2}\tau_i\|y_{h\tau}^{i-1}\|_{1,\Omega}^2 + \int_{I_i}\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{L^2(\gamma(t))}\|u(x,t)\|_{L^2(\gamma(t))}dt \\
\leqslant\ & \frac{1}{2}\tau_i\|y_{h\tau}^i\|_{1,\Omega}^2 + \frac{1}{2}\tau_i\|y_{h\tau}^{i-1}\|_{1,\Omega}^2 + C\int_{I_i}\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{L^2(\Omega)}^{\frac{1}{2}}\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{H_0^1(\Omega)}^{\frac{1}{2}}\|u(x,t)\|_{L^2(\gamma(t))}dt \\
\leqslant\ & \frac{1}{2}\tau_i\|y_{h\tau}^i\|_{1,\Omega}^2 + \frac{1}{2}\tau_i\|y_{h\tau}^{i-1}\|_{1,\Omega}^2 + C\tau_i h^{-1}\|u(x,t)\|_{L^2(I_i;L^2(\gamma(t)))}^2 + \frac{1}{2}\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{0,\Omega}^2.
\end{aligned}
$$

Summing the above equations over i from 1 to N we get

$$\begin{aligned}\sum_{i=1}^{N}\|y_{h\tau}^{i}-y_{h\tau}^{i-1}\|_{0,\Omega}^{2}+\tau\|y_{h\tau}^{N}\|_{1,\Omega}^{2} &\leqslant \tau\|y_{h\tau}^{0}\|_{1,\Omega}^{2}+C\tau h^{-1}\sum_{i=1}^{N}\|u(x,t)\|_{L^{2}(I_i;L^{2}(\gamma(t)))}^{2}\\ &\leqslant C\tau\|y_0\|_{1,\Omega}^{2}+C\tau h^{-1}\|u(x,t)\|_{L^{2}(I;L^{2}(\gamma(t)))}^{2}.\end{aligned}$$

This gives the results.

Let ψ be the solution of problem (2.1) with $f\in L^2(I;L^2(\Omega))$. Note that $\psi=0$ on $\partial\Omega$, $\psi^N=\psi(T)=0$, it follows from (2.8) and (A.8) that

$$\begin{aligned}\int_{\Omega_T}(y-y_{h\tau})f\,dxdt &= -\sum_{i=1}^{N}\int_{I_i}\tau_i^{-1}(y_{h\tau}^{i}-y_{h\tau}^{i-1},\psi^{i-1}-\pi_\tau\mathcal{R}_h\psi)dt\\ &\quad+\Big(\langle u\delta_{\gamma(t)},\psi\rangle_I-\sum_{i=1}^{N}\int_{I_i}\langle u\delta_{\gamma(t)},\pi_\tau\mathcal{R}_h\psi\rangle_{I_i}\Big)+(y_0-y_{h\tau}^{0},\psi(\cdot,0))\\ &=:\ \widetilde{G}_1+\widetilde{G}_2+\widetilde{G}_3.\end{aligned}\tag{3.42}$$

From (3.10) it is obvious that

$$|\widetilde{G}_3|=|(y_0-y_{h\tau}^{0},\psi(\cdot,0))|\leqslant Ch^2\|y_0\|_{1,\Omega}\|\psi(\cdot,0)\|_{1,\Omega}.\tag{3.43}$$

Now it remains to estimate $\widetilde{G}_1$ and $\widetilde{G}_2$. We can deduce from the trace theorem ([4, Theorem 1.6.6]) that

$$\begin{aligned}|\widetilde{G}_2| &= \Big|\langle u\delta_{\gamma(t)},\psi\rangle_I-\sum_{i=1}^{N}\int_{I_i}\langle u\delta_{\gamma(t)},\pi_\tau\mathcal{R}_h\psi\rangle_{I_i}\Big|\\ &= \Big|\sum_{i=1}^{N}\int_{I_i}\int_{\gamma(t)}u(x,t)(\psi-\pi_\tau\mathcal{R}_h\psi)(x,t)dxdt\Big|\\ &\leqslant C\|u\|_U\|\psi-\pi_\tau\mathcal{R}_h\psi\|_{L^2(I;L^2(\gamma(t)))}\\ &\leqslant C\|u\|_U\|\psi-\pi_\tau\mathcal{R}_h\psi\|_{L^2(I;L^2(\Omega))}^{\frac12}\|\psi-\pi_\tau\mathcal{R}_h\psi\|_{L^2(I;H_0^1(\Omega))}^{\frac12}.\end{aligned}\tag{3.44}$$

Standard error estimates (3.11) yield

$$\|\psi-\mathcal{R}_h\psi\|_{L^2(I;L^2(\Omega))}+h\|\psi-\mathcal{R}_h\psi\|_{L^2(I;H_0^1(\Omega))}\leqslant Ch^2\|\psi\|_{L^2(I;H^2(\Omega))}.\tag{3.45}$$

Moreover, we can conclude from (A.12) that (see [8])

$$\begin{aligned}\|\mathcal{R}_h\psi-\pi_\tau\mathcal{R}_h\psi\|_{L^2(I;L^2(\Omega))}^2 &\leqslant C\tau^{-1}\sum_{i=1}^{N}\int_{I_i}\int_{I_i}\|\mathcal{R}_h\psi(t)-\mathcal{R}_h\psi(s)\|_{L^2(\Omega)}^2dsdt\\ &\leqslant Ch^4\|\psi\|_{L^2(I;H^2(\Omega))}^2+C\tau^2\|\partial_t\psi\|_{L^2(I;L^2(\Omega))}^2.\end{aligned}\tag{3.46}$$

Similar to (A.12), by using the inverse estimate we can prove

$$\|\mathcal{R}_h\psi-\pi_\tau\mathcal{R}_h\psi\|_{L^2(I;H_0^1(\Omega))}^2\leqslant Ch^2\|\psi\|_{L^2(I;H^2(\Omega))}^2+Ch^{-2}\tau^2\|\partial_t\psi\|_{L^2(I;L^2(\Omega))}^2.\tag{3.47}$$

The triangle inequality together with (3.44)-(3.47) implies

$$|\widetilde{G}_2|\leqslant C(h^{\frac32}+h^{\frac12}\tau^{\frac12}+h^{-\frac12}\tau)\|u\|_U\|f\|_{L^2(I;L^2(\Omega))}.\tag{3.48}$$

The Cauchy-Schwarz inequality, the stability (3.41) and the standard error estimate yield (see [18])

$$\begin{aligned}|\widetilde{G}_1| &\leqslant \Big(\sum_{i=1}^{N}\|y_{h\tau}^{i}-y_{h\tau}^{i-1}\|_{0,\Omega}^2\Big)^{\frac12}\Big(\sum_{i=1}^{N}\|\psi^{i-1}-\pi_\tau\mathcal{R}_h\psi\|_{0,\Omega}^2\Big)^{\frac12}\\ &\leqslant C(h^2\tau^{-\frac12}+\tau^{\frac12})\|\psi\|_{L^2(I;H^2(\Omega))\cap H^1(I;L^2(\Omega))}h^{-\frac12}\tau^{\frac12}(\|y_0\|_{1,\Omega}+\|u\|_U)\\ &\leqslant C(h^{\frac32}+h^{-\frac12}\tau)\|f\|_{L^2(I;L^2(\Omega))}(\|y_0\|_{1,\Omega}+\|u\|_U).\end{aligned}\tag{3.49}$$

Then from Lemma 2.1, (3.42), (3.43), (3.48) and (3.49) we have

$$\begin{aligned}\|y-y_{h\tau}\|_{L^2(I;L^2(\Omega))} &= \sup_{f\in L^2(I;L^2(\Omega)),f\neq 0}\frac{(f,y-y_{h\tau})_{\Omega_T}}{\|f\|_{L^2(I;L^2(\Omega))}}\\ &\leqslant C(h^{\frac{3}{2}}+h^{\frac{1}{2}}\tau^{\frac{1}{2}}+h^{-\frac{1}{2}}\tau)(\|y_0\|_{1,\Omega}+\|u\|_U),\end{aligned}$$

which completes the proof of (3.40).

□

4. Error estimates for the optimal control problems

In this section we consider the fully discrete finite element approximations to the optimal control problems. As presented in the above section, for the discretisation of the state equation we use piecewise linear continuous finite elements for spatial discretisation and dG(0) scheme for time discretisation. For the discretisation of the control variable we adopt the variational discretisation approach proposed by Hinze in [24]. Due to the fact that $\gamma(t)$ may evolve in the time horizon, we do not require that $\mathcal{T}^h$ restricted to $\gamma(t)$ gives a triangulation of $\gamma(t)$. We consider the "discrete-then-optimize" approach to discretize the optimal control problems, which means the discretised optimality conditions are derived from the finite dimensional optimisation problems.

In addition, we restrict our study to the case that $\gamma(t)$ at each t is one of the following objects: a combination of some points; an interval in case $n=3$; a polygonal line in case $n=2$; and a polyhedral plane. The studies on the cases where $\gamma(t)$ on some t is a curve or a surface will be very similar to the above case with some technical modifications. Therefore, we omit the latter case and refer the reader to paper [21] for the details of curved cases.

The discrete optimal control problems now reads

$$\min_{u_{h\tau}\in U_{ad}} J(y_{h\tau},u_{h\tau}) = \tfrac{1}{2}\|y_{h\tau}-y_d\|^2_{L^2(I;L^2(\Omega))}+\tfrac{\alpha}{2}\|u_{h\tau}\|^2_U \tag{4.1}$$

subject to

$$A(y_{h\tau},w_{h\tau})=\langle u_{h\tau}\delta_{\gamma(t)},w_{h\tau}\rangle_I+(y_0,w^0_{h\tau,+}),\quad \forall\, w_{h\tau}\in V_{h,\tau}. \tag{4.2}$$

We can prove by standard arguments ([33, Ch.2, Sec.1.2]) that the above optimisation problems admit a unique solution $u_{h\tau}\in U_{ad}$. Moreover, $(y_{h\tau},u_{h\tau})\in V_{h,\tau}\times U_{ad}$ is the solution of problem (4.1)-(4.2) if and only if there exists an adjoint state $z_{h\tau}\in V_{h,\tau}$ such that the triplet $(y_{h\tau},u_{h\tau},z_{h\tau})$ satisfies the following first order optimality conditions:

$$\left\{\begin{aligned} &A(y_{h\tau},w_{h\tau})=\langle u_{h\tau}\delta_{\gamma(t)},w_{h\tau}\rangle_I+(y_0,w^0_{h\tau,+}),\ \forall\, w_{h\tau}\in V_{h,\tau},\\ &A(w_{h\tau},z_{h\tau})=\sum_{i=1}^N\int_{I_i}(y_{h\tau}-y_d,w_{h\tau}),\ \forall\, w_{h\tau}\in V_{h,\tau},\\ &(\alpha u_{h\tau}+z_{h\tau}|_{\gamma(t)},v-u_{h\tau})_U\geqslant 0\ \ \forall\, v\in U_{ad}.\end{aligned}\right. \tag{4.3}$$

We say that $(y_{h\tau},u_{h\tau},z_{h\tau})$ is the solution to the problem (4.1)-(4.2) if $(y_{h\tau},u_{h\tau})$ is the optimal pair to this problem and $z_{h\tau}$ is the adjoint state corresponding to this pair. Furthermore, analogue to (2.13)-(2.14) the control $u_{h\tau}$ can be represented by the discrete adjoint state $z_{h\tau}$ on each time interval I_i as

$$u_{h\tau}|_{I_i}=P_{U_{ad}}\Big(-\frac{1}{\alpha}z^i_{h\tau}(\gamma_j(t|_{I_i}))\Big)_{j=1}^m,\quad \text{when } k=0;$$

$$u_{h\tau}|_{I_i}=P_{U_{ad}}\Big(-\frac{1}{\alpha}z^i_{h\tau}|_{\gamma(t|_{I_i})}\Big),\quad \text{when } k\geqslant 1.$$

Now we are in the position to derive a priori error estimates for the fully discrete finite element approximation to the optimal control problems. At first, we consider the case $k\geqslant 1$ and $n-k=1$.

Theorem 4.1. *Assume that $\gamma(t)$ is a k-dimensional manifold strictly contained in Ω with $n-k=1$ and $U=L^2(I;L^2(\gamma(t)))$. Let $(y,u,z)\in L^2(I;H^1_0(\Omega))\times U_{ad}\times L^2(I;H^1_0(\Omega))$ and $(y_{h\tau},u_{h\tau},z_{h\tau})\in$*

$V_{h,\tau} \times U_{ad} \times V_{h,\tau}$ *be the solutions of the continuous and discrete optimal control problems (2.10)-(2.12) and (4.3), respectively. Assume that* $\tau = O(h^2)$. *Then it holds that*

$$\sqrt{\alpha}\|u - u_{h\tau}\|_{L^2(I;L^2(\gamma(t)))} + \|y - y_{h\tau}\|_{L^2(I;L^2(\Omega))} \leqslant Ch^{\frac{3}{2}}. \tag{4.4}$$

Moreover, if $y_d \in L^2(I; L^\infty(\Omega))$, $k = 1$ *and* $n = 2$, *the optimal error estimate*

$$\sqrt{\alpha}\|u - u_{h\tau}\|_{L^2(I;L^2(\gamma(t)))} \leqslant C|\log h|^{\frac{7}{2}}(h^2 + \tau) \tag{4.5}$$

holds for all $0 < h < h_0$ *with some* $0 < h_0 < 1$.

Proof. It follows from the continuous and discrete optimality conditions that

$$\int_I (\alpha u + z|_\gamma)(v - u)dt \geqslant 0 \quad \forall\, v \in U_{ad} \tag{4.6}$$

and

$$\int_I (\alpha u_{h\tau} + z_{h\tau}|_\gamma)(v - u_{h\tau})dt \geqslant 0 \quad \forall\, v \in U_{ad}. \tag{4.7}$$

Choosing $v = u_{h\tau}$ in (4.6) and $v = u$ in (4.7), and adding the two inequalities yields

$$\begin{aligned}
&\alpha\|u - u_{h\tau}\|^2_{L^2(I;L^2(\gamma(t)))} \\
\leqslant\ & \int_I\int_{\gamma(t)} (z_{h\tau} - z)(u - u_{h\tau})dxdt \\
\leqslant\ & \int_I\int_{\gamma(t)} (z_{h\tau} - z_{h\tau}(y))(u - u_{h\tau})dxdt + \int_I\int_{\gamma(t)} (z_{h\tau}(y) - z)(u - u_{h\tau})dxdt,
\end{aligned} \tag{4.8}$$

where $z_{h\tau}(y) \in V_{h,\tau}$ solves the following auxiliary problem

$$A(w_{h\tau}, z_{h\tau}(y)) = \sum_{i=1}^N \int_{I_i} (y - y_d, w_{h\tau}), \qquad \forall\, w_{h\tau} \in V_{h,\tau}. \tag{4.9}$$

Note that

$$\begin{aligned}
\int_I\int_{\gamma(t)} (z_{h\tau} - z_{h\tau}(y))(u - u_{h\tau})dxdt &= \int_I (z_{h\tau} - z_{h\tau}(y), (u - u_{h\tau})\delta_{\gamma(t)})dt \\
&= A(y_{h\tau}(u) - y_{h\tau}, z_{h\tau} - z_{h\tau}(y)) \\
&= \int_I (y_{h\tau}(u) - y_{h\tau}, y_{h\tau} - y) \\
&= -\|y - y_{h\tau}\|^2_{L^2(I;L^2(\Omega))} + \int_I (y_{h\tau}(u) - y, y_{h\tau} - y),
\end{aligned} \tag{4.10}$$

where $y_{h\tau}(u) \in V_{h,\tau}$ is the solution of the following problem:

$$A(y_{h\tau}(u), w_{h\tau}) = \langle u(x,t)\delta_{\gamma(t)}, w_{h\tau}\rangle_I + (y_0, w^0_{h\tau,+}), \qquad \forall\, w_{h\tau} \in V_{h,\tau}. \tag{4.11}$$

Making use of the Young's inequality in (4.8) and (4.10) we obtain

$$\begin{aligned}
&\alpha\|u - u_{h\tau}\|^2_{L^2(I;L^2(\gamma(t)))} + \|y - y_{h\tau}\|^2_{L^2(I;L^2(\Omega))} \\
\leqslant\ & C(\|y - y_{h\tau}(u)\|^2_{L^2(I;L^2(\Omega))} + \|z - z_{h\tau}(y)\|^2_{L^2(I;L^2(\gamma(t)))}).
\end{aligned} \tag{4.12}$$

From Theorem 3.6 we have

$$\|y - y_{h\tau}(u)\|_{L^2(I;L^2(\Omega))} \leqslant C(h^{\frac{3}{2}} + h^{\frac{1}{2}}\tau^{\frac{1}{2}} + h^{-\frac{1}{2}}\tau). \tag{4.13}$$

If we estimate $\|z - z_{h\tau}(y)\|_{L^2(I;L^2(\gamma(t)))}$ by (3.17) we can obtain (4.4) under the coupling $\tau = O(h^2)$.

Now we prove (4.5) in the case $k = 1$ and $n = 2$. To begin with, we introduce the following auxiliary problems: Find $z_{h\tau}(u) \in V_{h,\tau}$ such that

$$A(w_{h\tau}, z_{h\tau}(u)) = \sum_{i=1}^N \int_{I_i} (y_{h\tau}(u) - y_d, w_{h\tau}), \qquad \forall\, w_{h\tau} \in V_{h,\tau}. \tag{4.14}$$

From (4.3), (4.11) and (4.14) we can conclude that

$$\begin{aligned}\int_I\int_{\gamma(t)}(z_{h\tau}-z_{h\tau}(u))(u-u_{h\tau})dsdt &= ((z_{h\tau}-z_{h\tau}(u))|_\gamma, u-u_{h\tau})_U\\ &= \langle (u-u_{h\tau})\delta_\gamma, z_{h\tau}-z_{h\tau}(u)\rangle_I\\ &= A(y_{h\tau}(u)-y_{h\tau}, z_{h\tau}-z_{h\tau}(u))\\ &= (y_{h\tau}(u)-y_{h\tau}, y_{h\tau}-y_{h\tau}(u))_I\\ &= -\|y_{h\tau}(u)-y_{h\tau}\|^2_{L^2(I;L^2(\Omega))}. \end{aligned}\tag{4.15}$$

It follows from (4.8) that

$$\begin{aligned}&\alpha\|u-u_{h\tau}\|^2_{L^2(I;L^2(\gamma(t)))}\\ \leqslant\;& \int_I\int_{\gamma(t)}(z_{h\tau}-z)(u-u_{h\tau})dxdt\\ \leqslant\;& \int_I\int_{\gamma(t)}(z_{h\tau}-z_{h\tau}(u))(u-u_{h\tau})dxdt+\int_I\int_{\gamma(t)}(z_{h\tau}(u)-z_{h\tau}(y))(u-u_{h\tau})dxdt\\ &+\int_I\int_{\gamma(t)}(z_{h\tau}(y)-z)(u-u_{h\tau})dxdt,\end{aligned}\tag{4.16}$$

Then by (4.16), (4.15) and the Young's inequality, we see that

$$\begin{aligned}&\alpha\|u-u_{h\tau}\|^2_{L^2(I;L^2(\gamma(t)))}+\|y_{h\tau}(u)-y_{h\tau}\|^2_{L^2(I;L^2(\Omega))}\\ \leqslant\;& \int_I\int_{\gamma(t)}(z_{h\tau}-z_{h\tau}(u))(u-u_{h\tau})dxdt+\int_I\int_{\gamma(t)}(z_{h\tau}(u)-z_{h\tau}(y))(u-u_{h\tau})dxdt\\ \leqslant\;& \epsilon\alpha\|u-u_{h\tau}\|^2_{L^2(I;L^2(\gamma(t)))}+C(\epsilon,\alpha)\|z_{h\tau}(u)-z_{h\tau}(y)\|^2_{L^2(I;L^2(\gamma(t)))}\\ &+C(\epsilon,\alpha)\|z_{h\tau}(y)-z\|^2_{L^2(I;L^2(\gamma(t)))}.\end{aligned}\tag{4.17}$$

Moreover, from the embedding theorem we have $y\in L^2(I;L^\infty(\Omega))$ when $n=2$. If we assume in addition that $y_d\in L^2(I;L^\infty(\Omega))$, then for any subdomain $\Omega_0\subset\subset\Omega$ we can conclude from Lemma 2.2 in [30] that $z\in L^2(I;W^{2,q}(\Omega_0))\cap H^1(I;L^q(\Omega_0))$ for any $2\leqslant q<\infty$. Then from Lemma 3.3 and proceeding as in the estimate of (3.34) we can derive

$$\begin{aligned}\|z-z_{h\tau}(y)\|^2_{L^2(I;L^2(\gamma(t)))} &\leqslant C\sup_{x\in\bar\Omega_0}\int_I|(z-z_{h\tau}(y))(x,t)|^2dt\\ &\leqslant C|\log h|^5(h^4+\tau^2).\end{aligned}\tag{4.18}$$

Now we estimate $\|z_{h\tau}(u)-z_{h\tau}(y)\|_{L^2(I;L^2(\gamma(t)))}$. Let ϕ be the solution of the following problem

$$\begin{cases}\partial_t\phi-\Delta\phi=(z_{h\tau}(u)-z_{h\tau}(y))\delta_{\gamma(t)}(x) & \text{in } \Omega_T,\\ \phi=0 & \text{on } \Gamma_T,\\ \phi(\cdot,0)=0 & \text{in } \Omega.\end{cases}\tag{4.19}$$

Then it follows from Theorem 2.2 that $\phi\in L^2(I;H^{\frac{3-\epsilon}{2}}(\Omega)\cap H^1_0(\Omega))\cap H^1(I;H^{-\frac{1+\epsilon}{2}}(\Omega))$ for any $\epsilon>0$ and there holds

$$\|\phi\|_{L^2(I;H^{\frac{3-\epsilon}{2}}(\Omega)\cap H^1_0(\Omega))}+\|\partial_t\phi\|_{L^2(I;H^{-\frac{1+\epsilon}{2}}(\Omega))}\leqslant C\|z_{h\tau}(u)-z_{h\tau}(y)\|_{L^2(I;L^2(\gamma(t)))}.\tag{4.20}$$

Note that $z_{h\tau}(u)-z_{h\tau}(y)\in V_{h,\tau}$ satisfies the following equation

$$A(w_{h\tau},z_{h\tau}(u)-z_{h\tau}(y))=\sum_{i=1}^N\int_{I_i}(y_{h\tau}(u)-y,w_{h\tau})dt,\quad\forall\; w_{h\tau}\in V_{h,\tau}.\tag{4.21}$$

Let $\phi_{h\tau} \in V_{h,\tau}$ be the fully discrete finite element approximation of ϕ. Then we can derive from the orthogonality and (4.21) that

$$\begin{aligned}
\|z_{h\tau}(u) - z_{h\tau}(y)\|^2_{L^2(I;L^2(\gamma(t)))} &= \langle (z_{h\tau}(u) - z_{h\tau}(y))\delta_\gamma, z_{h\tau}(u) - z_{h\tau}(y)\rangle_I \\
&= A(\phi, z_{h\tau}(u) - z_{h\tau}(y)) \\
&= A(\phi_{h\tau}, z_{h\tau}(u) - z_{h\tau}(y)) \\
&= (y_{h\tau}(u) - y, \phi_{h\tau})_I \\
&= (y_{h\tau}(u) - y, \phi_{h\tau} - \phi)_I + (y_{h\tau}(u) - y, \phi)_I. \qquad (4.22)
\end{aligned}$$

It follows from Theorem 3.6 that

$$\begin{aligned}
(y_{h\tau}(u) - y, \phi_{h\tau} - \phi)_I &\leqslant C\|y_{h\tau}(u) - y\|_{L^2(I;L^2(\Omega))}\|\phi_{h\tau} - \phi\|_{L^2(I;L^2(\Omega))} \\
&\leqslant C(h^3 + h\tau + h^{-1}\tau^2)\|u\|_U\|z_{h\tau}(u) - z_{h\tau}(y)\|_{L^2(I;L^2(\gamma(t)))}. \qquad (4.23)
\end{aligned}$$

Now it remains to estimate $(y_{h\tau}(u) - y, \phi)_I$. Let $\psi \in L^2(I; H^2(\Omega) \cap H^1_0(\Omega)) \cap H^1(I; L^2(\Omega))$ be the solution of problem (2.1) with right hand side ϕ. Note that from the embedding theorem we have $\phi \in L^2(I; L^\infty(\Omega))$. For any subdomain $\Omega_0 \subset\subset \Omega$ we can conclude from Lemma 2.2 in [30] that $\psi \in L^2(I; W^{2,q}(\Omega_0)) \cap H^1(I; L^q(\Omega_0))$ for any $2 \leqslant q < \infty$ with the estimate

$$\begin{aligned}
\|\psi\|_{L^2(I;W^{2,q}(\Omega_0))} + \|\partial_t\psi\|_{L^2(I;L^q(\Omega_0))} &\leqslant Cq\|\phi\|_{L^2(I;L^q(\Omega))} \\
&\leqslant Cq^2\|z_{h\tau}(u) - z_{h\tau}(y)\|_{L^2(I;L^2(\gamma(t)))}. \qquad (4.24)
\end{aligned}$$

Let $\psi_{h\tau} \in V_{h,\tau}$ be the finite element approximation of ψ. Then from orthogonality of A we have

$$\begin{aligned}
(y_{h\tau}(u) - y, \phi)_I &= A(y_{h\tau}(u) - y, \psi) = A(y_{h\tau}(u) - y, \psi - \psi_{h\tau}) \\
&= A(y, \psi_{h\tau} - \psi) = \langle u\delta_\gamma, \psi_{h\tau} - \psi\rangle_I \\
&\leqslant C\|u\|_U\big(\sup_{x\in\bar{\Omega}_0}\int_I |(\psi_{h\tau} - \psi)(x,t)|^2 dt\big)^{\frac{1}{2}}. \qquad (4.25)
\end{aligned}$$

Similar to the proof of Lemma 3.3 and Theorem 3.5, we are led to

$$(y_{h\tau}(u) - y, \phi)_I \leqslant C|\log h|^{\frac{7}{2}}(h^2 + \tau)\|u\|_U\|z_{h\tau}(u) - z_{h\tau}(y)\|_{L^2(I;L^2(\gamma(t)))} \qquad (4.26)$$

by setting $q = |\log h|$. Combining (4.23) and (4.26) one can deduce that

$$\|z_{h\tau}(u) - z_{h\tau}(y)\|_{L^2(I;L^2(\gamma(t)))} \leqslant C|\log h|^{\frac{7}{2}}(h^2 + \tau)\|u\|_U. \qquad (4.27)$$

Collecting (4.17), (4.18) and (4.27) we can prove

$$\alpha\|u - u_{h\tau}\|^2_{L^2(I;L^2(\gamma(t)))} \leqslant C|\log h|^7(h^4 + \tau^2), \qquad (4.28)$$

which completes the proof. □

Then, we consider the case $n - k > 1$. To derive an optimal error estimate in two dimension we follow the idea of [30].

Theorem 4.2. *Assume that $\gamma(t)$ is a k-dimensional manifold strictly contained in Ω with $n-k > 1$. Let $(y, u, z) \in L^2(I; W^{1,s}_0(\Omega)) \times U_{ad} \times L^2(I; H^1_0(\Omega))$ and $(y_{h\tau}, u_{h\tau}, z_{h\tau}) \times V_{h,\tau} \times U_{ad} \times V_{h,\tau}$ be the solutions of the continuous and discrete optimal control problems (2.10)-(2.12) and (4.3), respectively. Assume that $\tau = O(h^2)$. Then the following statements stand:*
(i) When $k = 0$, the suboptimal error estimate

$$\sqrt{\alpha}\|u - u_{h\tau}\|_U + \|y - y_{h\tau}\|_{L^2(I;L^2(\Omega))} \leqslant Ch^{2-\frac{n}{2}} \qquad (4.29)$$

holds. If it is further assumed that $y_d \in L^2(I; L^\infty(\Omega))$ and $n = 2$, the optimal error estimate

$$\sqrt{\alpha}\|u - u_{h\tau}\|_U \leqslant C|\log h|^{\frac{7}{2}}(h^2 + \tau) \qquad (4.30)$$

holds for all $0 < h < h_0$ with some $0 < h_0 < 1$.
(ii) When $k = 1$ and $n = 3$, the suboptimal error estimate

$$\sqrt{\alpha}\|u - u_{h\tau}\|_U + \|y - y_{h\tau}\|_{L^2(I;L^2(\Omega))} \leqslant Ch^{\frac{5}{2}-\frac{3}{\sigma}}, \quad \sigma \in (3/2, 2) \qquad (4.31)$$

holds.

Proof. The estimates of (4.29) and (4.31) follow the similar idea of [19] by using $L^2(I;L^2(\Omega))$-error of the state y and the $L^2(I;L^\infty(\Omega))$-error of the adjoint state z. Now we prove (4.30). Similar to the proof of (4.8) we get

$$\begin{aligned} & \alpha\|u-u_{h\tau}\|_U^2 \leqslant (u-u_{h\tau},(z_{h\tau}-z)|_{\gamma(t)})_U \\ = \;& (u-u_{h\tau},(z_{h\tau}-z_{h\tau}(u))|_{\gamma(t)})_U + (u-u_{h\tau},(z_{h\tau}(u)-z_{h\tau}(y))|_{\gamma(t)})_U \\ & +(u-u_{h\tau},(z_{h\tau}(y)-z)|_{\gamma(t)})_U. \end{aligned} \tag{4.32}$$

Then from (4.15) and the Young's inequality we arrive at

$$\begin{aligned} & \alpha\|u-u_{h\tau}\|_U^2 + \|y_{h\tau}(u)-y_{h\tau}\|^2_{L^2(I;L^2(\Omega))} \\ \leqslant\; & (u-u_{h\tau},(z_{h\tau}(u)-z_{h\tau}(y))|_{\gamma(t)})_U + (u-u_{h\tau},(z_{h\tau}(y)-z)|_{\gamma(t)})_U \\ \leqslant\; & \epsilon\alpha\|u-u_{h\tau}\|_U^2 + C(\epsilon,\alpha)(\|z_{h\tau}(u)-z_{h\tau}(y)\|^2_{L^2(I;L^\infty(\Omega))} + \sup_{x\in\bar\Omega_0}\int_I |(z-z_{h\tau}(y))(x,t)|^2dt). \end{aligned} \tag{4.33}$$

Using the coercivity property of $A(\cdot,\cdot)$ and the inverse estimate we are led to

$$\begin{aligned} & c\|z_{h\tau}(u)-z_{h\tau}(y)\|^2_{L^2(I;H^1(\Omega))} \\ \leqslant\; & A(z_{h\tau}(u)-z_{h\tau}(y), z_{h\tau}(u)-z_{h\tau}(y)) \\ =\; & \sum_{i=1}^N\int_{I_i}(y_{h\tau}(u)-y, z_{h\tau}(u)-z_{h\tau}(y))dt \\ \leqslant\; & \|y_{h\tau}(u)-y\|_{L^2(I;L^1(\Omega))}\|z_{h\tau}(u)-z_{h\tau}(y)\|_{L^2(I;L^\infty(\Omega))} \\ \leqslant\; & C\rho(2,h)\|y_{h\tau}(u)-y\|_{L^2(I;L^1(\Omega))}\|z_{h\tau}(u)-z_{h\tau}(y)\|_{L^2(I;H^1(\Omega))}, \end{aligned}$$

which yields

$$c\|z_{h\tau}(u)-z_{h\tau}(y)\|_{L^2(I;H^1(\Omega))} \leqslant C\rho(2,h)\|y_{h\tau}(u)-y\|_{L^2(I;L^1(\Omega))}. \tag{4.34}$$

From the inverse estimate again we can obtain

$$\begin{aligned} \|z_{h\tau}(u)-z_{h\tau}(y)\|_{L^2(I;L^\infty(\Omega))} & \leqslant C\rho(2,h)\|z_{h\tau}(u)-z_{h\tau}(y)\|_{L^2(I;H^1(\Omega))} \\ & \leqslant C\rho^2(2,h)\|y_{h\tau}(u)-y\|_{L^2(I;L^1(\Omega))}. \end{aligned} \tag{4.35}$$

From (4.33) and (4.35) we have

$$\begin{aligned} & \alpha\|u-u_{h\tau}\|^2_{L^2(I;L^2(\gamma(t)))} + \|y_{h\tau}(u)-y_{h\tau}\|^2_{L^2(I;L^2(\Omega))} \\ \leqslant\; & C(\epsilon,\alpha)(\|z_{h\tau}(u)-z_{h\tau}(y)\|^2_{L^2(I;L^\infty(\Omega))} + \sup_{x\in\bar\Omega_0}\int_I |(z-z_{h\tau}(y))(x,t)|^2dt) \\ \leqslant\; & C(\epsilon,\alpha)\rho^4(2,h)\|y_{h\tau}(u)-y\|^2_{L^2(I;L^1(\Omega))} + \sup_{x\in\bar\Omega_0}\int_I |(z-z_{h\tau}(y))(x,t)|^2dt. \end{aligned} \tag{4.36}$$

As indicated in Section 2 we have $z\in L^2(I;W^{2,q}(\Omega_0))\cap H^1(I;L^q(\Omega_0))$ for any $q\leqslant \frac{n}{n-s}$. Then applying the local maximum norm error estimate (3.19) and proceeding as in the estimate of (3.34) we obtain

$$\sup_{x\in\bar\Omega_0}\int_I |(z-z_{h\tau}(y))(x,t)|^2dt \leqslant C|\log h|^3q^4h^{-\frac{4}{q}}(h^4+\tau^2). \tag{4.37}$$

Specifically, we can set $q=|\log h|$ in (4.37). Then we are able to derive

$$\sup_{x\in\bar\Omega_0}\int_I |(z-z_{h\tau}(y))(x,t)|^2dt \leqslant C|\log h|^7(h^4+\tau^2). \tag{4.38}$$

Combining (4.36)-(4.37) and Theorem 3.5 and considering (3.5) we arrive at

$$
\begin{aligned}
&\alpha\|u-u_{h\tau}\|^2_{L^2(I;L^2(\gamma(t)))}+\|y_{h\tau}(u)-y_{h\tau}\|^2_{L^2(I;L^2(\Omega))}\\
\leqslant\ & C\rho^4(2,h)|\log h|^5(h^4+\tau^2)+C|\log h|^7(h^4+\tau^2)\\
\leqslant\ & C|\log h|^7(h^4+\tau^2). \qquad (4.39)
\end{aligned}
$$

This completes the proof. □

Remark 4.3. *A number of remarks for the error estimates derived in Theorems 4.1 and 4.2 are in order. Firstly, compared to* [19] *we are able to derive an optimal error estimate for the optimal control in two dimension, but the estimate for the state in (4.4), (4.29) and (4.31) are sharp due to the limited regularity of the state. Secondly, in present paper we only derived optimal error estimates for the optimal control in two dimension following the idea of* [30]*, although in* [21] *we studied the finite element approximations of elliptic optimal control problems with controls acting on a lower dimensional manifold and derived optimal error estimates for the optimal control in both two- and three dimensions. This is caused by the lack of a maximum norm error estimates for parabolic equation in three dimension (see Lemma 3.3) as proved in* [30] *for two dimensional case. Thirdly, we expect the optimal orders of convergence $O(h+\tau^{\frac{1}{2}})$ when $k=0$ and $n=3$, and $O(h^2+\tau)$ when $k=1$ and $n=3$ hold for three dimensional problems but it still needs a rigorous proof. Due to this reason the improved error estimates for the optimal control in three dimension may be postponed to a future work.*

5. Numerical Examples

In this section we will carry out some numerical experiments to support our theoretical findings. For the computation the software package AFEPack ([31]) has been used. To validate the estimates developed in the previous section, we may show the convergence order by separating the discretization errors in space and time.

Pointwise control problems for parabolic equations can be viewed as a special case of controls acting on a lower dimensional manifold. In [19] the authors presented some numerical examples for pointwise control problems where the spatial point is independent of time. In the following we will consider the case that the lower dimensional manifold may move around in the space domain as time evolves. In the numerical experiments we may illustrate the convergence orders with respect to the spatial and time discretizations separately by setting h and τ small enough respectively, although some a priori error estimates are derived with coupling $\tau=O(h^2)$. The numerical tests indicate that such a coupling of τ and h seems not to be needed. We expect that an according analysis is possible with adapting the techniques of [40] and [41] to the present setting.

Example 5.1. *In the first example we consider the case $n=2$ and $k=0$ with γ a moving point contained in Ω. Let $\Omega_T=B(0,1)\times[0,1]$. We set $\alpha=1$, $\gamma(t)=(\gamma_1(t),\gamma_2(t))=(\frac{1}{2}\cos(2\pi t),\frac{1}{2}\sin(2\pi t))$ and $U_{ad}:=\{u\in L^2[0,T]:\ 0.5\leqslant u(t)\leqslant 1\ a.e.\ t\in[0,T]\}$. We take the exact solutions as*

$$
\begin{aligned}
y(x,t)=-\frac{1}{2\pi}\log|x-\gamma(t)|\cdot(e-e^t),\quad u(x,t)=P_{U_{ad}}(e-e^t),\\
z(x,t)=-\cos(\frac{\pi}{2}|x-\gamma(t)|^2)\sin(2\pi|x|^2)(e-e^t),
\end{aligned}
$$

with corresponding f and y_d.

We show in Table 1 the convergence oder for the L^2-norms of the control, state and adjoint state. The time step is set to be $O(h^2)$. We can observe second order convergence for the optimal control which confirms our theoretical results. However, only first order convergence can be observed for the L^2-norm of the state, which is caused by the low regularity of the state equation due to the Dirac measure. In Figure 1 we present the computed state $y_{h\tau}$ at different times on fine mesh with 66049 Dofs and 256 time steps for Example 5.1. We can see that the Dirac measure can be exactly captured by the profiles of the state.

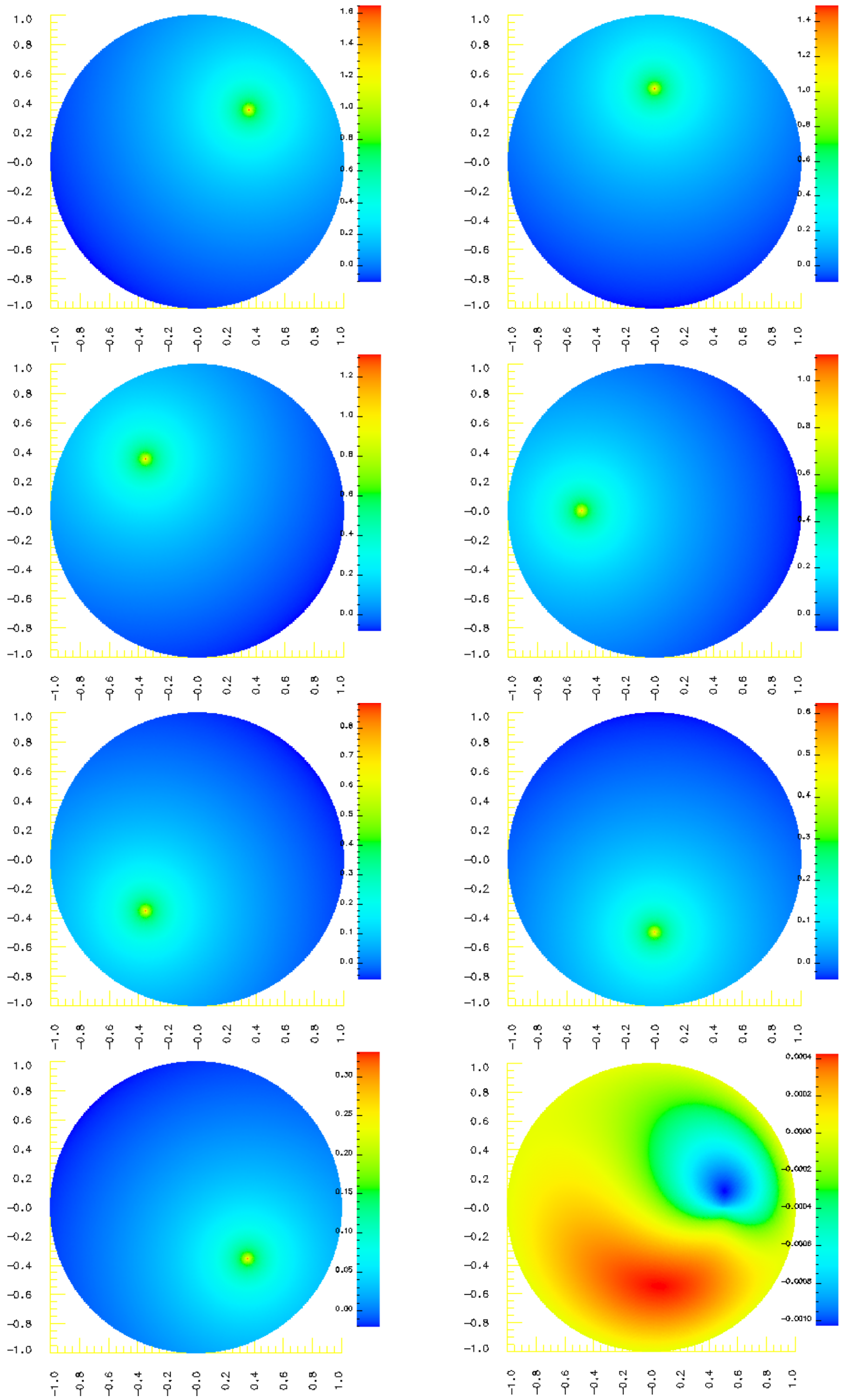

FIGURE 1. The computed state $y_{h\tau}$ on fine mesh with 66049 Dofs and 256 time steps for Example 5.1. The subplots show the solutions of the state at different time $t = \frac{1}{8}, \frac{1}{4}, \frac{3}{8}, \frac{1}{2}, \frac{5}{8}, \frac{3}{4}, \frac{7}{8}, 1$ (from top left to bottom right).

TABLE 1. Errors of control u, state y and adjoint state z for Example 5.1 with respect to space and time.

Dof	N	$\|u-u_{h\tau}\|_{L^2(0,T)}$	Rate	$\|y-y_{h\tau}\|_{L^2(\Omega_T)}$	Rate	$\|z-z_{h\tau}\|_{L^2(\Omega_T)}$	Rate
25	8	0.266876039146	\	0.048518775420	\	0.963378762924	\
81	32	0.038745367936	2.7841	0.027784961679	0.8042	0.643763751918	0.5816
289	128	0.008367663349	2.2111	0.011153618365	1.3168	0.165069654498	1.9635
1089	512	0.001862279212	2.1678	0.005219041154	1.0957	0.043065698090	1.9385
4225	2048	0.000437728257	2.0890	0.002572033248	1.0209	0.010929899309	1.9783
16641	8192	0.000106024846	2.0456	0.001281382684	1.0052	0.002743510676	1.9942

TABLE 2. Errors of control u, state y and adjoint state z for Example 5.2 with respect to time with fixed space triangulation.

Dof	N	$\|u-u_{h\tau}\|_{L^2(I;L^2(\gamma))}$	Rate	$\|y-y_{h\tau}\|_{L^2(\Omega_T)}$	Rate	$\|z-z_{h\tau}\|_{L^2(\Omega_T)}$	Rate
263169	4	0.002909173531	\	0.008003724330	\	0.013997680754	\
263169	8	0.001711455367	0.7654	0.005023153257	0.6721	0.008462220210	0.7261
263169	16	0.000927892146	0.8832	0.002737692566	0.8756	0.004745487754	0.8345
263169	32	0.000469988319	0.9813	0.001379795139	0.9885	0.002470170020	0.9419
263169	64	0.000219877153	1.0959	0.000632363584	1.1256	0.001180761839	1.0649
263169	128	0.000087233130	1.3337	0.000240513292	1.3946	0.000479449034	1.3003

TABLE 3. Errors of control u, state y and adjoint state z for Example 5.2 with respect to space with fixed time step.

Dof	N	$\|u-u_{h\tau}\|_{L^2(I;L^2(\gamma))}$	Rate	$\|y-y_{h\tau}\|_{L^2(\Omega_T)}$	Rate	$\|z-z_{h\tau}\|_{L^2(\Omega_T)}$	Rate
81	256	1.797384265e-3	\	1.1884261247e-2	\	1.0371803861e-2	\
289	256	4.450005900e-4	2.0140	3.3545615070e-3	1.8249	2.8060229020e-3	1.8861
1089	256	1.108811560e-4	2.0048	9.7769836600e-4	1.7787	7.1713594300e-4	1.9682
4225	256	2.749660800e-5	2.0117	2.9367647900e-4	1.7352	1.7872158100e-4	2.0045
16641	256	6.707969000e-6	2.0353	9.0345879000e-5	1.7007	4.2876307000e-5	2.0595
66049	256	1.527385000e-6	2.1348	2.7487056000e-5	1.7167	8.8247380000e-6	2.2806

Example 5.2. *In the second example we consider the case* $n = 2$ *and* $k = 1$ *with* γ *a moving interval strictly contained in* Ω. *Let* $\Omega = (-1, 1)^2$, $T = 0.5$, $\gamma = \{2t - 0.5\} \times [-0.25, 0.25]$. *We set* $y_d = \pi \sin(\pi t)\sin(0.5\pi x_1)\sin(0.5\pi x_2)$, $a = -1$, $b = 1$ *and* $\alpha = 0.01$.

In the second example we have no explicit solutions for the optimal control, state and adjoint state. We compute the solutions on fine mesh with 263169 Dofs and 256 time steps as referee solutions to calculate the convergence order. At first we consider the behavior of the errors for a sequence of discretizations with different mesh sizes and fixed 256 time steps. Then we show the behavior of the errors for different time steps but a fixed spatial triangulation with 263169 Dofs. From Table 2 and 3 we can observe second order convergence of the optimal control for the spatial discretisation and first order convergence of both the optimal control and state for the temporal discretisation, which are the optimal convergence rates for the linear finite element spatial discretizations and dG(0) time discretisation scheme. However, the convergence order of the state y for spatial discretisation is nearly only 1.7, which is reasonable because of the limited regularity due to the linear measure. In Figure 2 we present the computed state $y_{h\tau}$ at different times on fine mesh with 263169 Dofs and 64 time steps for Example 5.2 and we can observe the solution changes as the line measure evolves.

Example 5.3. *In the third example we consider the case* $n = 3$ *and* $k = 0$ *with* γ *a fixed stationary point strictly contained in* Ω. *Let* $\Omega = (0, 1)^3$, $T = 1$, $\gamma = x_0 = (0.5, 0.5, 0.5)$. *We set* $y =$

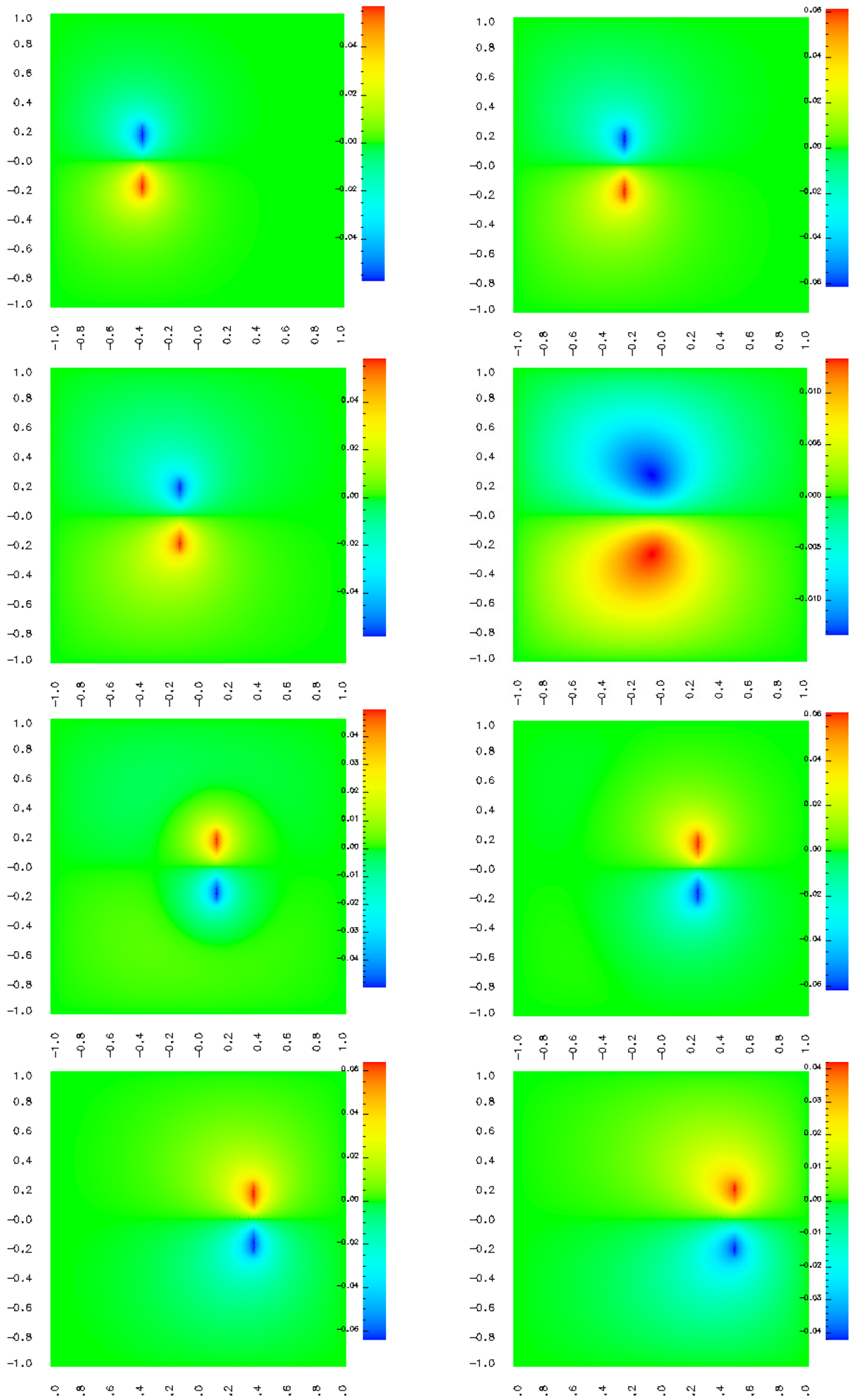

FIGURE 2. The computed state $y_{h\tau}$ on fine mesh with 263169 Dofs and 64 time steps for Example 5.2. The subplots show the solutions of the state at different time $t = \frac{1}{16}, \frac{1}{8}, \frac{3}{16}, \frac{1}{4}, \frac{5}{16}, \frac{3}{8}, \frac{7}{16}, \frac{1}{2}$ (from top left to bottom right).

Table 4. Errors of control u, state y and adjoint state z for Example 5.3 with respect to space and time.

Dof	N	$\|u-u_{h\tau}\|_{L^2(0,T)}$	Rate	$\|y-y_{h\tau}\|_{L^2(\Omega_T)}$	Rate	$\|z-z_{h\tau}\|_{L^2(\Omega_T)}$	Rate
175	8	0.281920287108	\	0.112081236271	\	0.247646970005	\
1085	32	0.143024589103	0.9790	0.083390837752	0.4266	0.065471735028	1.9193
7577	128	0.055364072472	1.3692	0.059371782409	0.4901	0.016788036277	1.9634
56497	512	0.018597982354	1.5738	0.042058290409	0.4974	0.004240873359	1.9850

$\frac{1}{2\pi}\frac{1}{|x-x_0|}(e^t-e)$, $p=-2\sin(\pi x_1)\sin(\pi x_2)\sin(\pi x_3)(e^t-e)$, $u=2(e^t-e)$, $a=-\infty$, $b=+\infty$ *and* $\alpha=1$.

Similar to Example 5.1 we show in Table 4 the convergence orders for the L^2-norms of the control, state and adjoint state. The time step is set to be $O(h^2)$. We can observe nearly $O(h^{\frac{3}{2}})$ order of convergence for the optimal control which implies that our error estimates in 3D can be improved. However, only $O(h^{\frac{1}{2}})$ order of convergence can be observed for the L^2-norm of the state, which is caused by the low regularity of the state equation due to the Dirac measure and is in agreement with our theoretical results. We note that the convergence order for the control is higher than the one $O(h)$ we expect, see Remark 4.3. We think the exact solution is too regular so that the superconvergence phenomena may happen.

Appendix A. Proof of Theorem 3.4

Proof of Theorem 3.4. We use duality argument to prove this theorem. The proof of this theorem follows the idea of [18], see also [8].

At first we derive the following stability estimates for numerical scheme (3.28). Let $y_{h\tau}^i\in V_0^h,\ i=1,2,\cdots,N$ be the solutions of fully discrete scheme (3.28). Then there exists a constant C independent of h,τ and the data u such that

$$\sum_{i=1}^N\|y_{h\tau}^i-y_{h\tau}^{i-1}\|_{0,\Omega}^2+\tau_N\|y_{h\tau}^N\|_{1,\Omega}^2\leqslant C\tau h^{-n}(\|u\|_U^2+\|y_0\|_{0,\Omega}^2) \tag{A.1}$$

for $k=0$ and $n=2$ or 3, and

$$\sum_{i=1}^N\|y_{h\tau}^i-y_{h\tau}^{i-1}\|_{0,\Omega}^2+\tau_N\|y_{h\tau}^N\|_{1,\Omega}^2\leqslant C\tau h^{1-\frac{6}{\sigma}}(\|u\|_U^2+\|y_0\|_{0,\Omega}^2) \tag{A.2}$$

for $k=1$ and $n=3$. The proof is a slight modification of the proof in [18], here we include the proof for completeness.

Let $w_h=\tau_i(y_{h\tau}^i-y_{h\tau}^{i-1})$ in (3.28) we get

$$(y_{h\tau}^i-y_{h\tau}^{i-1},y_{h\tau}^i-y_{h\tau}^{i-1})+\tau_i(\nabla y_{h\tau}^i,\nabla(y_{h\tau}^i-y_{h\tau}^{i-1}))\quad=\quad\tau_i\langle u\delta_{\gamma(t)},y_{h\tau}^i-y_{h\tau}^{i-1}\rangle_{I_i}.$$

Since $y_{h\tau}^i - y_{h\tau}^{i-1}$ is piecewise constant on each time interval I_i w.r.t time, thus we have for $k=0$ and $n=2$ or 3 that

$$\begin{aligned}
&\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{0,\Omega}^2 + \tau_i\|y_{h\tau}^i\|_{1,\Omega}^2\\
\leqslant\ & \tau_i(\nabla y_{h\tau}^i, \nabla y_{h\tau}^{i-1}) + \int_{I_i}\sum_{j=1}^m u_j(t)(y_{h\tau}^i - y_{h\tau}^{i-1})(\gamma_j(t))dt\\
\leqslant\ & \frac{1}{2}\tau_i\|y_{h\tau}^i\|_{1,\Omega}^2 + \frac{1}{2}\tau_i\|y_{h\tau}^{i-1}\|_{1,\Omega}^2 + \int_{I_i}\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{L^\infty(\Omega)}\sum_{j=1}^m |u_j(t)|dt\\
\leqslant\ & \frac{1}{2}\tau_i\|y_{h\tau}^i\|_{1,\Omega}^2 + \frac{1}{2}\tau_i\|y_{h\tau}^{i-1}\|_{1,\Omega}^2 + Ch^n\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{0,\infty,\Omega}^2 + C\tau_i h^{-n}\|u\|_{L^2(I_i;\mathbb{R}^m)}^2\\
\leqslant\ & \frac{1}{2}\tau_i\|y_{h\tau}^i\|_{1,\Omega}^2 + \frac{1}{2}\tau_i\|y_{h\tau}^{i-1}\|_{1,\Omega}^2 + C\tau_i h^{-n}\|u\|_{L^2(I_i;\mathbb{R}^m)}^2 + \frac{1}{2}\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{0,\Omega}^2, \qquad \text{(A.3)}
\end{aligned}$$

and for $k=1$ and $n=3$ that

$$\begin{aligned}
&\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{0,\Omega}^2 + \tau_i\|y_{h\tau}^i\|_{1,\Omega}^2\\
\leqslant\ & \tau_i(\nabla y_{h\tau}^i, \nabla y_{h\tau}^{i-1}) + \int_{I_i}(\int_{\gamma(t)} u(x,t)(y_{h\tau}^i - y_{h\tau}^{i-1})dx)dt\\
\leqslant\ & \frac{1}{2}\tau_i\|y_{h\tau}^i\|_{1,\Omega}^2 + \frac{1}{2}\tau_i\|y_{h\tau}^{i-1}\|_{1,\Omega}^2 + \int_{I_i}\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{L^2(\gamma(t))}\|u(x,t)\|_{L^2(\gamma(t))}dt\\
\leqslant\ & \frac{1}{2}\tau_i\|y_{h\tau}^i\|_{1,\Omega}^2 + \frac{1}{2}\tau_i\|y_{h\tau}^{i-1}\|_{1,\Omega}^2 + C\int_{I_i}\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{W_0^{1,\sigma'}(\Omega)}\|u(x,t)\|_{L^2(\gamma(t))}dt\\
\leqslant\ & \frac{1}{2}\tau_i\|y_{h\tau}^i\|_{1,\Omega}^2 + \frac{1}{2}\tau_i\|y_{h\tau}^{i-1}\|_{1,\Omega}^2 + Ch^{\frac{6}{\sigma}-1}\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{W_0^{1,\sigma'}(\Omega)}^2 + C\tau_i h^{1-\frac{6}{\sigma}}\|u(x,t)\|_{L^2(I_i;L^2(\gamma(t)))}^2\\
\leqslant\ & \frac{1}{2}\tau_i\|y_{h\tau}^i\|_{1,\Omega}^2 + \frac{1}{2}\tau_i\|y_{h\tau}^{i-1}\|_{1,\Omega}^2 + C\tau_i h^{1-\frac{6}{\sigma}}\|u(x,t)\|_{L^2(I_i;L^2(\gamma(t)))}^2 + \frac{1}{2}\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{0,\Omega}^2. \qquad \text{(A.4)}
\end{aligned}$$

In the above estimates we have used the following inverse estimates

$$\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{0,\infty,\Omega} \quad\leqslant\quad Ch^{-\frac{n}{2}}\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{0,\Omega}$$

and

$$\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{W_0^{1,\sigma'}(\Omega)} \quad\leqslant\quad Ch^{\frac{1}{2}-\frac{3}{\sigma}}\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{0,\Omega}.$$

Summing the above equations over i from 1 to N and using the inverse estimate we get

$$\begin{aligned}
\sum_{i=1}^N\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{0,\Omega}^2 + \tau\|y_{h\tau}^N\|_{1,\Omega}^2 \quad&\leqslant\quad \tau\|y_{h\tau}^0\|_{1,\Omega}^2 + C\tau h^{-n}\sum_{i=1}^N\|u(x,t)\|_{L^2(I_i;\mathbb{R}^m)}^2\\
&\leqslant\quad C\tau h^{-2}\|\mathcal{P}_h y_0\|_{0,\Omega}^2 + C\tau h^{-n}\|u(x,t)\|_{L^2(I;\mathbb{R}^m)}^2\\
&\leqslant\quad C\tau h^{-n}(\|u(x,t)\|_{L^2(I;\mathbb{R}^m)}^2 + \|y_0\|_{0,\Omega}^2) \qquad \text{(A.5)}
\end{aligned}$$

for $k=0$ and $n=2$ or 3, and

$$\begin{aligned}
\sum_{i=1}^N\|y_{h\tau}^i - y_{h\tau}^{i-1}\|_{0,\Omega}^2 + \tau\|y_{h\tau}^N\|_{1,\Omega}^2 \quad&\leqslant\quad \tau\|y_{h\tau}^0\|_{1,\Omega}^2 + C\tau h^{1-\frac{6}{\sigma}}\sum_{i=1}^N\|u(x,t)\|_{L^2(I_i;L^2(\gamma(t)))}^2\\
&\leqslant\quad C\tau h^{-2}\|\mathcal{P}_h y_0\|_{0,\Omega}^2 + C\tau h^{1-\frac{6}{\sigma}}\|u(x,t)\|_{L^2(I;L^2(\gamma(t)))}^2\\
&\leqslant\quad C\tau h^{1-\frac{6}{\sigma}}(\|y_0\|_{0,\Omega}^2 + \|u(x,t)\|_{L^2(I;L^2(\gamma(t)))}^2) \qquad \text{(A.6)}
\end{aligned}$$

for $k=1$ and $n=3$. This gives the results.

Let ψ be the solution of problem (2.1) with $f \in L^2(I;L^2(\Omega))$. Note that $\psi = 0$ on $\partial\Omega$, $\psi^N = \psi(T) = 0$, it follows from (2.6) that

$$\begin{aligned}
&\int_{\Omega_T}(y-y_{h\tau})f\,dxdt = \int_I\int_\Omega (y-y_{h\tau})(-\partial_t\psi + \mathcal{A}^*\psi)dxdt \\
= \;& \langle\partial_t y,\psi\rangle_I + (\nabla y,\nabla\psi)_I + (y_0,\psi(\cdot,0)) + \sum_{i=1}^N\int_{I_i}((y_{h\tau}^i,\partial_t\psi) - (\nabla y_{h\tau}^i,\nabla\psi))dt \\
= \;& \langle u\delta_{\gamma(t)},\psi\rangle_I + (y_0,\psi(\cdot,0)) + \sum_{i=1}^N\int_{I_i}(\tau_i^{-1}(y_{h\tau}^i,\psi^i-\psi^{i-1}) - (\nabla y_{h\tau}^i,\nabla\psi))dt \\
= \;& -\sum_{i=1}^N\int_{I_i}(\tau_i^{-1}(y_{h\tau}^i - y_{h\tau}^{i-1},\psi^{i-1}) + (\nabla y_{h\tau}^i,\nabla\psi))dt + \langle u\delta_{\gamma(t)},\psi\rangle_I + (y_0 - y_{h\tau}^0,\psi(\cdot,0)).
\end{aligned}$$

Note that from (3.28) we have

$$\sum_{i=1}^N(\tau_i^{-1}(y_{h\tau}^i - y_{h\tau}^{i-1},\pi_\tau\mathcal{R}_h\psi) + (\nabla y_{h\tau}^i,\nabla\pi_\tau\mathcal{R}_h\psi)) = \sum_{i=1}^N\langle u\delta_{\gamma(t)},\pi_\tau\mathcal{R}_h\psi\rangle_{I_i},$$

where $\mathcal{R}_h$ is the Ritz-projection operator. Furthermore,

$$\int_{I_i}(\nabla y_{h\tau}^i,\nabla(\pi_\tau\psi - \pi_\tau\mathcal{R}_h\psi))dt = 0. \tag{A.7}$$

Thus,

$$\begin{aligned}
\int_{\Omega_T}(y-y_{h\tau})f\,dxdt \;=\;& \langle u\delta_{\gamma(t)},\psi\rangle_I - \sum_{i=1}^N\int_{I_i}\langle u,\pi_\tau\mathcal{R}_h\psi\rangle_{I_i} \\
&+\sum_{i=1}^N\int_{I_i}(\tau_i^{-1}(y_{h\tau}^i - y_{h\tau}^{i-1},\pi_\tau\mathcal{R}_h\psi) + (\nabla y_{h\tau}^i,\nabla\pi_\tau\mathcal{R}_h\psi))dt \\
&-\sum_{i=1}^N\int_{I_i}(\tau_i^{-1}(y_{h\tau}^i - y_{h\tau}^{i-1},\psi^{i-1}) + (\nabla y_{h\tau}^i,\nabla\pi_\tau\psi))dt + (y_0 - y_{h\tau}^0,\psi(\cdot,0)) \\
=\;& -\sum_{i=1}^N\int_{I_i}\tau_i^{-1}(y_{h\tau}^i - y_{h\tau}^{i-1},\psi^{i-1} - \pi_\tau\mathcal{R}_h\psi)dt \\
&+\Big(\langle u\delta_{\gamma(t)},\psi\rangle_I - \sum_{i=1}^N\int_{I_i}\langle u\delta_{\gamma(t)},\pi_\tau\mathcal{R}_h\psi\rangle_{I_i}\Big) + (y_0 - y_{h\tau}^0,\psi(\cdot,0)) \\
=:\;& \widetilde{E}_1 + \widetilde{E}_2 + \widetilde{E}_3.
\end{aligned}\tag{A.8}$$

From (3.10) it is obvious that

$$|\widetilde{E}_3| = |(y_0 - y_{h\tau}^0,\psi(\cdot,0))| \leqslant Ch\|y_0\|_{0,\Omega}\|\psi(\cdot,0)\|_{1,\Omega}. \tag{A.9}$$

Now it remains to estimate $\widetilde{E}_1$ and $\widetilde{E}_2$. At first we consider the case $k = 0$ and $n = 2$ or 3. We can deduce that

$$\begin{aligned}
|\widetilde{E}_2| \;=\;& \Big|\langle u\delta_{\gamma(t)},\psi\rangle_I - \sum_{i=1}^N\int_{I_i}\langle u\delta_{\gamma(t)},\pi_\tau\mathcal{R}_h\psi\rangle_{I_i}\Big| \\
=\;& \Big|\sum_{i=1}^N\int_{I_i}\Big(\sum_{j=1}^m u(\gamma_j(t),t)(\psi - \pi_\tau\mathcal{R}_h\psi)(\gamma_j(t),t)\Big)dt\Big| \\
\leqslant\;& C\|u\|_U\|\psi - \pi_\tau\mathcal{R}_h\psi\|_{L^2(I;L^\infty(\Omega))}.
\end{aligned}\tag{A.10}$$

Standard error estimates (3.12) yield

$$\|\psi - \mathcal{R}_h\psi\|_{L^2(I;L^\infty(\Omega))} \leqslant Ch^{2-\frac{n}{2}}\|\psi\|_{L^2(I;H^2(\Omega))}. \tag{A.11}$$

Moreover, from the inverse inequality we have (see [8])

$$\begin{aligned}
\|\mathcal{R}_h\psi - \pi_\tau\mathcal{R}_h\psi\|^2_{L^2(I;L^\infty(\Omega))} &\leqslant \sum_{i=1}^N \frac{1}{\tau_i}\int_{I_i}\int_{I_i}\|\mathcal{R}_h\psi(t) - \mathcal{R}_h\psi(s)\|^2_{L^\infty(\Omega)}dsdt \\
&\leqslant Ch^{-n}\tau^{-1}\sum_{i=1}^N\int_{I_i}\int_{I_i}\|\mathcal{R}_h\psi(t) - \mathcal{R}_h\psi(s)\|^2_{L^2(\Omega)}dsdt \\
&\leqslant Ch^{-n}\tau^{-1}\sum_{i=1}^N\Big(\tau_i\int_{I_i}\|\mathcal{R}_h\psi(t) - \psi(t)\|^2_{L^2(\Omega)}dt + \tau_i\int_{I_i}\|\psi(s) - \mathcal{R}_h\psi(s)\|^2_{L^2(\Omega)}ds \\
&\quad + \int_{I_i}\int_{I_i}\|\psi(t) - \psi(s)\|^2_{L^2(\Omega)}dsdt\Big) \\
&\leqslant Ch^{-n}h^4\|\psi\|^2_{L^2(I;H^2(\Omega))} + Ch^{-n}\tau^{-1}\Big(\sum_{i=1}^N\int_{I_i}\int_{I_i}\|\int_{I_i}\partial_t\psi(r)dr\|^2_{L^2(\Omega)}dsdt\Big) \\
&\leqslant Ch^{-n}h^4\|\psi\|^2_{L^2(I;H^2(\Omega))} + Ch^{-n}\tau^2\|\partial_t\psi\|^2_{L^2(I;L^2(\Omega))}. \qquad \text{(A.12)}
\end{aligned}$$

The triangle inequality together with Lemma 2.1 implies

$$|\widetilde{E}_2| \leqslant C(h^{-\frac{n}{2}}\tau + h^{2-\frac{n}{2}})\|u\|_U\|f\|_{L^2(I;L^2(\Omega))}. \tag{A.13}$$

The Cauchy-Schwarz inequality, the stability (A.1) and the standard error estimate yield (see [18])

$$\begin{aligned}
|\widetilde{E}_1| &\leqslant \Big(\sum_{i=1}^N\|y^i_{h\tau} - y^{i-1}_{h\tau}\|^2_{0,\Omega}\Big)^{\frac{1}{2}}\Big(\sum_{i=1}^N\|\psi^{i-1} - \pi_\tau\mathcal{R}_h\psi\|^2_{0,\Omega}\Big)^{\frac{1}{2}} \\
&\leqslant C(h^2\tau^{-\frac{1}{2}} + \tau^{\frac{1}{2}})\|\psi\|_{L^2(I;H^2(\Omega))\cap H^1(I;L^2(\Omega))}h^{-\frac{n}{2}}\tau^{\frac{1}{2}}(\|y_0\|_{0,\Omega} + \|u\|_U) \\
&\leqslant C(h^{2-\frac{n}{2}} + h^{-\frac{n}{2}}\tau)\|f\|_{L^2(I;L^2(\Omega))}(\|y_0\|_{0,\Omega} + \|u\|_U). \qquad \text{(A.14)}
\end{aligned}$$

Then from Lemma 2.1, (A.8), (A.9), (A.13) and (A.14) we have

$$\begin{aligned}
\|y - y_{h\tau}\|_{L^2(I;L^2(\Omega))} &= \sup_{f\in L^2(I;L^2(\Omega)),f\neq 0}\frac{(f, y - y_{h\tau})_{\Omega_T}}{\|f\|_{L^2(I;L^2(\Omega))}} \\
&\leqslant C(h^{2-\frac{n}{2}} + h^{-\frac{n}{2}}\tau)(\|y_0\|_{0,\Omega} + \|u\|_U),
\end{aligned}$$

which completes the proof of (3.29).

Then we consider the case when $k = 1$ and $n = 3$. We can deduce from the trace inequality (2.7) that

$$\begin{aligned}
|\widetilde{E}_2| &= \Big|\langle u\delta_{\gamma(t)}, \psi\rangle_I - \sum_{i=1}^N\int_{I_i}\langle u\delta_{\gamma(t)}, \pi_\tau\mathcal{R}_h\psi\rangle_{I_i}\Big| \\
&= \Big|\sum_{i=1}^N\int_{I_i}\Big(\int_{\gamma(t)}u(x,t)(\psi - \pi_\tau\mathcal{R}_h\psi)(x,t)dx\Big)dt\Big| \\
&\leqslant C\|u\|_{L^2(I;L^2(\gamma(t)))}\|\psi - \pi_\tau\mathcal{R}_h\psi\|_{L^2(I;L^2(\gamma(t)))} \\
&\leqslant C\|u\|_{L^2(I;L^2(\gamma(t)))}\|\psi - \pi_\tau\mathcal{R}_h\psi\|_{L^2(I;W_0^{1,\sigma'}(\Omega))}. \qquad \text{(A.15)}
\end{aligned}$$

Standard error estimates ([4, Sec.8.5]) yield

$$\|\psi - \mathcal{R}_h\psi\|_{L^2(I;W_0^{1,\sigma'}(\Omega))} \leqslant Ch^{\frac{5}{2}-\frac{3}{\sigma}}\|\psi\|_{L^2(I;H^2(\Omega))}. \tag{A.16}$$

Moreover, from the inverse inequality we have

$$\|\mathcal{R}_h\psi - \pi_\tau\mathcal{R}_h\psi\|^2_{L^2(I;W_0^{1,\sigma'}(\Omega))} \leqslant \sum_{i=1}^N \frac{1}{\tau_i}\int_{I_i}\int_{I_i} \|\mathcal{R}_h\psi(t) - \mathcal{R}_h\psi(s)\|^2_{W_0^{1,\sigma'}(\Omega)} dsdt$$

$$\leqslant \quad Ch^{1-\frac{6}{\sigma}}\tau^{-1}\sum_{i=1}^N \int_{I_i}\int_{I_i} \|\mathcal{R}_h\psi(t) - \mathcal{R}_h\psi(s)\|^2_{L^2(\Omega)} dsdt$$

$$\leqslant \quad Ch^{1-\frac{6}{\sigma}}\tau^{-1}\sum_{i=1}^N \Big(\tau_i\int_{I_i} \|\mathcal{R}_h\psi(t) - \psi(t)\|^2_{L^2(\Omega)} dt + \tau_i\int_{I_i} \|\psi(s) - \mathcal{R}_h\psi(s)\|^2_{L^2(\Omega)} ds$$

$$+ \int_{I_i}\int_{I_i} \|\psi(t) - \psi(s)\|^2_{L^2(\Omega)} dsdt\Big)$$

$$\leqslant \quad Ch^{1-\frac{6}{\sigma}}h^4\|\psi\|^2_{L^2(I;H^2(\Omega))} + Ch^{1-\frac{6}{\sigma}}\tau^{-1}\Big(\sum_{i=1}^N \int_{I_i}\int_{I_i} \|\int_{I_i}\partial_t\psi(r)dr\|^2_{L^2(\Omega)} dsdt\Big)$$

$$\leqslant \quad Ch^{1-\frac{6}{\sigma}}h^4\|\psi\|^2_{L^2(I;H^2(\Omega))} + Ch^{1-\frac{6}{\sigma}}\tau^2\|\partial_t\psi\|^2_{L^2(I;L^2(\Omega))}. \tag{A.17}$$

The triangle inequality together with Lemma 2.1 gives

$$|\widetilde{E}_2| \leqslant C(h^{\frac{1}{2}-\frac{3}{\sigma}}\tau + h^{\frac{5}{2}-\frac{3}{\sigma}})\|u\|_U\|f\|_{L^2(I;L^2(\Omega))}. \tag{A.18}$$

The Cauchy-Schwarz inequality, the stability (A.2) and the standard error estimate yield (see [18])

$$|\widetilde{E}_1| \quad \leqslant \quad \Big(\sum_{i=1}^N \|y^i_{h\tau} - y^{i-1}_{h\tau}\|^2_{0,\Omega}\Big)^{\frac{1}{2}}\Big(\sum_{i=1}^N \|\psi^{i-1} - \pi_\tau\mathcal{R}_h\psi\|^2_{0,\Omega}\Big)^{\frac{1}{2}}$$

$$\leqslant \quad C(h^2\tau^{-\frac{1}{2}} + \tau^{\frac{1}{2}})\|\psi\|_{L^2(I;H^2(\Omega))\cap H^1(I;L^2(\Omega))}h^{\frac{1}{2}-\frac{3}{\sigma}}\tau^{\frac{1}{2}}(\|y_0\|_{0,\Omega} + \|u\|_{L^2(I;L^2(\gamma(t)))})$$

$$\leqslant \quad C(h^{\frac{5}{2}-\frac{3}{\sigma}} + h^{\frac{1}{2}-\frac{3}{\sigma}}\tau)\|f\|_{L^2(I;L^2(\Omega))}(\|y_0\|_{0,\Omega} + \|u\|_U). \tag{A.19}$$

Then from Lemma 2.1, (A.8), (A.9), (A.18)and (A.19) we have

$$\|y - y_{h\tau}\|_{L^2(I;L^2(\Omega))} \quad = \quad \sup_{f\in L^2(I;L^2(\Omega)), f\neq 0} \frac{(f, y - y_{h\tau})_{\Omega_T}}{\|f\|_{L^2(I;L^2(\Omega))}}$$

$$\leqslant \quad C(h^{\frac{5}{2}-\frac{3}{\sigma}} + h^{\frac{1}{2}-\frac{3}{\sigma}}\tau)(\|y_0\|_{0,\Omega} + \|u\|_U),$$

this completes the proof of (3.30).

Acknowledgements

The authors would like to thank two anonymous referees for their careful reviews and many helpful suggestions which helped to improve the paper greatly. The first author was supported by the National Basic Research Program of China under grant 2012CB821204, the National Natural Science Foundation of China under grant 91530204 and 11201464, and the scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry. The second author acknowledged the support of the National Natural Science Foundation of China under grant 91530204 and 11571356. Moreover, we would like to thank prof. Gengsheng Wang from Wuhan University, China for the helpful discussions.

References

[1] Robert A. Adams, John J.F. Fournier, *Sobolev Spaces, second edition*, Elsevier, Singapore, 2009.

[2] M. Andrle, F. Ben Belgacem and A. El Badia, *Identification of moving pointwise sources in an advection-dispersion-reaction equation*, Inverse Problems, 27 (2011), 025007.

[3] C. Bernardi, *Optimal finite-element interpolation on curved domains*, SIAM J. Numer. Anal., 25 (1989), pp. 1212–1240.

[4] Susanne C. Brenner and L. Ridgway Scott, *The Mathematical Theory of Finite Element Methods*, Third Edition, Springer, New York, 2008.

[5] E. Casas, *L^2 estimates for the finite element method for the Dirichlet problem with singular data*, Numer. Math., 47 (1985), pp. 627–632.

[6] E. Casas, *Pontryagin's principle for state-constrained boundary control problems of semilinear parabolic equations*, SIAM J. Control Optim., 35 (1997), pp. 1297–1327.

[7] E. Casas, C. Clason and K. Kunisch, *Approximation of elliptic control problems in measure spaces with sparse solutions*, SIAM J. Control Optim., 50 (2012), pp. 1735–1752.

[8] E. Casas, C. Clason and K. Kunisch, *Parabolic control problems in measure spaces with sparse solutions*, SIAM J. Control Optim., 51 (2013), pp. 28–63.

[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), pp. 1585–1611.

[10] C. Castro and E. Zuazua, *Unique continuation and control for the heat equation from an oscillating lower dimensional manifold*, SIAM J. Control Optim., 43 (2004), pp. 1400–1434.

[11] K. Chrysafinos and E. Karatzas, *Error estimates for discontinuous Galerkin time-stepping schemes for Robin boundary control problems constrained to parabolic PDEs*, SIAM J. Numer. Anal., 52 (2014), no. 6, pp. 2837–2862.

[12] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, Class. Appl. Math., SIAM, Philadelphia, 2002.

[13] A. Demlow, J. Guzmán and A. H. Schatz, *Local energy estimates for the finite element method on sharply varying grids*, Math. Comput., 80 (2011), pp. 1–9.

[14] J. Elschner, J. Rehberg and G. Schmidt, *Optimal regularity for elliptic transmission problems including $\mathcal{C}^1$ interfaces*, Interfaces Free Bound., 9 (2007), pp. 233–252.

[15] L. C. Evans, *Partial Differential Equations*, Grad. Stud. Math., 19, American Mathematical Society, Providence, RI, 1998.

[16] F. S. Falk, *Approximation of a class of optimal control problems with order of convergence estimates*, J. Math. Anal. Appl., 44 (1973), pp. 28–47.

[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), pp. 79–104.

[18] W. Gong, *Error estimates for finite element approximations of parabolic equations with measure data*, Math. Comput., 82 (2013), no. 281, pp. 69–98.

[19] 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), no. 1, pp. 97–119.

[20] 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), no. 3, pp. 941–967.

[21] W. Gong, G. S. Wang and N. N. Yan, *Approximations of elliptic optimal control problems with controls acting on a lower dimensional manifold*, SIAM J. Control Optim., 52 (2014), no. 3, pp. 2008–2035.

[22] W. Gong and N. N. Yan, *Mixed finite element method for Dirichlet boundary control problem governed by elliptic PDEs*, SIAM J. Control Optim., 49 (2011), pp. 984–1014.

[23] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Pitman, Boston, 1985.

[24] M. Hinze, *A varitional discretization concept in control constrained optimization: The linear-quadratic case*, Comput. Optim. Appl., 30 (2005), pp. 45–63.

[25] M. Hinze, R. Pinnau, M. Ulbrich and S. Ulbrich, *Optimization with PDE Constraints*, MMTA 23, Springer, 2009.

[26] D. Jerison and C. E. Kenig, *The inhomogeneous Dirichlet problem in Lipschitz domains*, J. Funct. Anal., 130 (1995), pp. 161–219.

[27] A. Khapalov, *Mobile point controls versus locally distributed ones for the controllability of the semilinear parabolic equation*, SIAM J. Control Optim., 40 (2001), pp. 231–252.

[28] K. Kunisch, K. Pieper and B. Vexler, *Measure valued directional sparsity for parabolic optimal control problems*, SIAM J. Control Optim., 52 (2014), no. 5, pp. 3078–3108.

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

[30] D. Leykekhman and B. Vexler, *Optimal a priori error estimates of parabolic optimal control problems with pointwise control*, SIAM J. Numer. Anal., 51 (2013), no. 5, pp. 2797–2821.

[31] R. Li and W. B. Liu, *AFEPack* available online at *http://dsec.pku.edu.cn/∼rli/*.

[32] L. Ling, M. Yamamoto, Y. C. Hon and T. Takeuchi, *Identification of source locations in two-dimensional heat equations*, Inverse Problems, 22 (2006), pp. 1289–1305.

[33] J. L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, Springer-Verlag, Berlin, 1971.

[34] J. L. Lions and E. Magenes, *Non-homogeneous Boundary Value Problems and Applications* Vol. I-II, Springer-Verlag, Berlin, 1972.

[35] W. B. Liu and N. N. Yan, *A posteriori error estimates for distributed convex optimal control problems*, Adv. Comput. Math., 15 (2001), no. 1-4, pp. 285–309.

[36] W. B. Liu and N. N. Yan, *A posteriori error estimates for optimal control problems governed by parabolic equations*, Numer. Math., 93 (2003), pp. 497–521.
[37] W. B. Liu and N. N. Yan, *Adaptive Finite Element Methods for Optimal Control Governed by PDEs*, Beijing, Science press, 2008.
[38] K. Malanowski, *Convergence of approximations vs. regularity of solutions for convex, control constrained optimal-control problems*, Appl. Math. Optim., 8 (1981), pp. 69–95.
[39] R. S. McNight and W. E. Bosarge, *The Ritz-Galerkin procedure for parabolic control problems*, SIAM J. Control Optim., 11 (1973), pp. 510–524.
[40] 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), pp. 1150–1177.
[41] 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), pp. 1301–1329.
[42] P. A. Nguyen and J. P. Raymond, *Control localized on thin structures for semilinear parabolic equations*, In: Nonlinear Partial Differential Equations and Their Applications, Collège de France Seminar, vol. XIV, Paris, 1997/1998. Stud. Math. Appl., vol. 31, pp. 591–645. North-Holland, Amsterdam (2002).
[43] P. A. Nguyen and J. P. Raymond, *Control problems for convection-diffusion equations with control localized on manifolds*, ESAIM: Control Optim. Calc. Var., 6 (2001), pp. 467–488.
[44] A. Rösch, *Error estimates for parabolic optimal control problems with control constraints*, Zeitschrift für Analysis und ihre Anwendungen (ZAA), 23 (2004), pp. 353–376.
[45] A. H. Schatz and L. B. Wahlbin, *Interior maximum norm estimates for finite element methods*, Math. Comput., 31 (1977), pp. 414–442.
[46] R. Scott, *Optimal L^∞ estimates for the finite element method on irregular meshes*, Math. Comput., 30 (1976), pp. 681–697.
[47] V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, Springer-Verlag, Berlin, 2006.
[48] R. Winther, *Error estimates for a Galerkin approximation of a parabolic control problem*, Ann. Math. Pura. App., 117 (1978), no. 4, pp. 173–206.