Finite Element Error Estimation for Parabolic Optimal Control Problems with Pointwise Observations

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

[1] *School of Mathematics, Sichuan University, Chengdu 610064, China*
[2] *LSEC, Institute of Computational Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences & School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100190, China*

Abstract. In this paper, we consider parabolic distributed control problems with cost functional of pointwise observation type either in space or in time. First, we show the well-posedness of the optimization problems and derive the first order optimality systems, where the adjoint state can be expressed as the linear combination of solutions to two backward parabolic equations that involve the Dirac delta distribution as source either in space or in time. Second, we use a space-time finite element method to discretize the control problems, where the state variable is approximated by piecewise constant functions in time and continuous piecewise linear polynomials in space, and the control variable is discretized by following the variational discretization concept. We obtain a priori error estimates for the control and state variables with order $\mathcal{O}(k^{\frac{1}{2}} + h)$ up to a logarithmic factor under the L^2-norm. Finally, we perform several numerical experiments to support our theoretical results.

1. Introduction

Let $\Omega \subset \mathbb{R}^n$ ($n = 2, 3$) be a convex polygonal or polyhedron domain, and let $T > 0$ be a constant. We consider an optimal control problem of parabolic type where the cost functional involves pointwise values of the state variable either in space or in time. The model of controlled system is characterized by the parabolic equation

---

　　　　165

$$\begin{aligned}
\partial_t y - \Delta y &= \mathcal{B}u &&\text{in } (0, T) \times \Omega, \\
y &= 0 &&\text{on } (0, T) \times \partial\Omega, \\
y(0) &= 0 &&\text{in } \Omega,
\end{aligned} \qquad (1.1)$$

where y and u are the state and control variables, respectively. For convenience, we denotes by $I = (0, T)$ the time interval and by $U = L^2(I \times \omega)$ the control space, where the subdomain $\omega \subseteq \Omega$ with positive measure is called the control domain. We impose additional pointwise constraints on the control variable u and define the admissible control set as follows:

$$U_{ad} := \left\{ u \in U : u_a \leq u(t, x) \leq u_b, \text{ a.e., } (t, x) \in I \times \omega \right\}$$

with $-\infty < u_a < u_b < +\infty$. The control operator $\mathcal{B}$ maps U to $L^2(I \times \Omega)$ and is a linear bounded operator. For example, if $\omega = \Omega$, then we can take $\mathcal{B}$ as the identity operator; otherwise, $\mathcal{B}$ can be defined as a zero extension operator.

The cost functional of pointwise tracking type can be defined as

$$\mathcal{J}_\beta(y, u) := (2 - \beta) J_1(y, u) + (\beta - 1) J_2(y, u) + \frac{\alpha}{2} \int_0^T \int_\omega |u(t, x)|^2 dx dt. \qquad (1.2)$$

Here J_1 and J_2 denote respectively the spatial and time observations of state, defined by

$$J_1(y, u) = \frac{1}{2} \sum_{i=1}^{N_1} \int_0^T \left| y(x^i, t) - y_{d_S}^i(t) \right|^2 dt,$$

$$J_2(y, u) = \frac{1}{2} \sum_{i=1}^{N_2} \int_\Omega \left| y(t^i) - y_{d_T}^i \right|^2 dx + \frac{1}{2} \int_\Omega \left(y(T) - y_T \right)^2 dx,$$

where $N_1, N_2 \in N_+$, $y_T \in L^2(\Omega)$, $\alpha > 0$ is a regularization parameter and N_+ denotes the set of positive integers. The sets $\{x^i, i = 1, \ldots, N_1\} \subset \text{Int}(\Omega)$ and $\{t^i, i = 1, \ldots, N_2\} \subset (0, T)$ are respectively called the set of spatial observation points and the set of time observation points, and $\{y_{d_S}^i \in L^2(\Omega), i = 1, \ldots, N_1\}$ and $\{y_{d_T}^i \in L^2(0, T), i = 1, \ldots, N_2\}$ are respectively called the spatial observations and the time observations. The parameter $\beta \in [1, 2]$ is the weight between the pointwise spatial observations and pointwise time observations. Specifically, $\beta = 1$ refers to the case of pure spatial observation and $\beta = 2$ refers to the case of pure time observation, while $\beta \in (1, 2)$ refers to the case with both observations and weights the importance of two observations.

With the above defined cost functional, our parameter-dependent optimal control problem reads: Find $(\bar{y}, \bar{u}) \in X \times U_{ad}$ such that

$$\mathcal{J}_\beta(\bar{y}, \bar{u}) \leq \mathcal{J}_\beta(y, u), \quad \forall (y, u) \in X \times U_{ad} \quad \text{subject to} \quad (1.1), \qquad (1.3)$$

where X is the state space given in Section 3.

The optimization problem (1.3) with cost functional of pointwise type serves as a model problem for several applications, such as the parameter identification or inverse problems with finitely many pointwise measurements (cf. [35]), and optimal control problems where the target states on some spatial or time points are of particular interest or can be measured by sensors [2,4,9]. We also refer the reader to [29] for the derivation of the necessary optimality conditions, and to [2,4,5,7,9,15,35,39] for the numerical analysis of this kind of problems.

There are a lot of publications concerning the analysis and approximation to optimal control of parabolic equations; see, e.g., the references [8,12,25,31,33,34], where the cost functionals are of distributed type, so that the state and adjoint equations are all regular. As for parabolic control problems with pointwise observations, the adjoint equations involve Dirac distributions and thus exhibit low regularity (cf. [8]). Note that such problems are closely related to parabolic control problems with pointwise control (cf. [18,19,26–28]), where instead the state equations involve the Dirac distributions.

In this paper, we consider a space-time finite element approximation to the parabolic control problems with cost functional of pointwise observation type either in space or in time. Although there are already some works on the error estimates of elliptic or Stokes control problems with pointwise tracking (cf. [2,4,5,7,9,15]), to the authors' best knowledge, there are no such results for the parabolic control problems with pointwise tracking. For the discretization of the state and adjoint equations, we use a piecewise constant discontinuous Galerkin scheme (the DG(0) method) for the temporal discretization, and the standard linear finite element method (the CG(1) method) for the spatial discretization. This is a special case of the so-called DG(r)-CG(s) method; see, e.g., [33] for the error estimation of such kinds of discretization schemes for parabolic equations. For the discretization of the control variable, we apply the variational discretization concept proposed in [22].

Our numerical analysis for the considered parabolic control problems is strongly related to the error estimation of finite element approximation to the parabolic equations with rough right-hand side terms. We are faced with two types of backward parabolic equations with Dirac distributions, i.e., the one with spatial Dirac measure and the other with temporal Dirac measure. For the first type, the error estimate of order $\mathcal{O}(k^{\frac{1}{2}} + h^{2-\frac{n}{2}})$ was derived in [17,18] for n-dimensions under the norm $\| \cdot \|_{L^2(I,L^2(\Omega))}$ and a mesh condition $k = \mathcal{O}(h^n)$, where h and k denote the spatial mesh size and the time step size, respectively. We mention that the order $\mathcal{O}(k + h^2)$ for two-dimensions was obtained in [26] under the norm $\| \cdot \|_{L^2(I,L^1(\Omega))}$, and the order $\mathcal{O}(k^{\frac{1}{2}} + h)$ for three-dimensions was obtained in [27] under the norm $\| \cdot \|_{L^2(I,L^{\frac{3}{2}}(\Omega))}$, both up to a logarithmic term. For the second type, the error estimate of order $\mathcal{O}(k^{\frac{1}{2}} + h)$ was derived in [17] under the norm $\| \cdot \|_{L^2(I,L^2(\Omega))}$ and a mesh condition $k = \mathcal{O}(h^n)$.

The goal of this paper is to provide a finite element analysis of the optimal control problem (1.3). To this end, we introduce two adjoint equations, one is with Dirac measure in space as the external force term and the other is with Dirac measure in time as the force term, to study respectively the pointwise observations in space and in time

so as to obtain the first order necessary conditions (cf. [23]); see Section 3 for details. We also need a priori error estimates for the finite element discretizations of the two adjoint equations under the norms $\|\cdot\|_{L^2(I,L^1(\Omega))}$ and $\|\cdot\|_{L^2(\Omega,L^1(I))}$, respectively. The former estimate in $\|\cdot\|_{L^2(I,L^1(\Omega))}$-norm can be found in [26] for two dimensions, and will be extended to three dimensions in Section 3; see Lemma 2.1. The latter estimate in $\|\cdot\|_{L^2(\Omega,L^1(I))}$-norm will be derived by using an estimate due to [32] under the norm $\|\cdot\|_{L^\infty(I,L^2(\Omega))}$ for parabolic equations with low regularity of solutions. In conclusion, the main result of this paper is the following error estimate between the optimal solution $\bar{u}$ of the continuous problem and the optimal solution $\bar{u}_{kh}$ of the discrete problem:

$$\|\bar{u} - \bar{u}_{kh}\|_{L^2(I,L^2(\omega))} \le C|\ln h| \left(\ln \frac{T}{k}\right)^{\frac{3}{4}} \left(k^{\frac{1}{2}} + h\right). \tag{1.4}$$

The rest of this paper is organized as follows. In Section 2 we discuss the regularity of the state equation, the local regularity of the parabolic equation and the well-posedness of two types of backward parabolic equations with rough right-hand sides. In Section 3, we give the optimal control problem and derive its first order necessary condition. In Section 4, we describe the finite element discretization of the optimal control problem, and give the discrete first order necessary condition. Section 5 mainly focuses on the error analysis for the optimal control and optimal state. And the last section is devoted to numerical experiments.

2. Preliminaries

Let the control space U be a Hilbert space with the inner product $(\cdot,\cdot)_U$. Introduce the following notations for inner products and norms on $L^2(I,L^2(\Omega))$ and $L^2(\Omega)$, respectively:

$$(u,v) := (u,v)_{L^2(\Omega)}, \qquad \|v\| := \|v\|_{L^2(\Omega)}, \qquad \forall u,v \in L^2(\Omega),$$
$$(u,v)_I := (u,v)_{L^2(I,L^2(\Omega))}, \quad \|v\|_I := \|v\|_{L^2(I,L^2(\Omega))}, \quad \forall u,v \in L^2(I,L^2(\Omega)).$$

In order to define the weak solution to the state equation (1.1), for given $\varphi \in L^2(I,H^{-1}(\Omega))$ we introduce an auxiliary problem

$$\begin{aligned} \partial_t \psi - \Delta \psi &= \varphi && \text{in } I \times \Omega, \\ \psi &= 0 && \text{on } I \times \partial\Omega, \\ \psi(0) &= 0 && \text{in } \Omega. \end{aligned} \tag{2.1}$$

In order to find the weak solution of (2.1) we reformulate it as follows: Find a function $\psi \in L^2(I,H_0^1(\Omega)) \cap H^1(I,H^{-1}(\Omega))$ such that

$$\begin{aligned} (\partial_t \psi, \phi)_{H^{-1},H_0^1} + (\nabla\psi, \nabla\phi)_I &= (\varphi, \phi)_I, \quad \forall \phi \in L^2\big(I, H_0^1(\Omega)\big), \\ \psi(0) &= 0, \end{aligned} \tag{2.2}$$

where $(\cdot, \cdot)_{H^{-1}, H_0^1}$ denotes the duality pairing between $L^2(I, H^{-1}(\Omega))$ and $L^2(I, H_0^1(\Omega))$.

We have the following results on the existence, uniqueness and regularity of solution to Eq. (2.1).

Proposition 2.1. *For given* $\varphi \in L^2(I, H^{-1}(\Omega))$*, there exists a unique solution,* $\psi \in L^2(I, H_0^1(\Omega)) \cap H^1(I, H^{-1}(\Omega))$*, to problem* (2.1)*. Moreover, if* $\varphi \in L^2(I, L^2(\Omega))$*, then* ψ *has the improved regularity*

$$\psi \in L^2\big(I, H^2(\Omega) \cap H_0^1(\Omega)\big) \cap H^1\big(I, L^2(\Omega)\big) \cap C\big(\bar{I}, H_0^1(\Omega)\big) \tag{2.3}$$

and there holds the stability estimate

$$\|\psi\|_{C(\bar{I}, H_0^1(\Omega))} + \|\partial_t \psi\|_I + \|\nabla^2 \psi\|_I \le C\|\varphi\|_I, \tag{2.4}$$

where $\|\cdot\|_{C(\bar{I}, H_0^1(\Omega))}$ *denotes the norm of the space* $C(\bar{I}, H_0^1(\Omega))$*.*

Proof. The proof of the existence and uniqueness of solution has been given in [29]. The improved regularity and the stability estimate are classical; see, e.g., [14]. The regularity $\psi \in C(\bar{I}, H_0^1(\Omega))$ can be obtained from the fact that $L^2(I, H^2(\Omega) \cap H_0^1(\Omega)) \cap H^1(I, L^2(\Omega)) \hookrightarrow C(\bar{I}, H_0^1(\Omega))$. $\square$

In addition, we have the following improved local regularity of the solution to (2.1).

Lemma 2.1. *Assume that* $\tilde{\Omega} \subset \Omega$ *is an open subdomain such that* $\text{dist}(\tilde{\Omega}, \partial\Omega) > 0$ *and that there exists a smooth open subdomain* Ω_0 *such that* $\tilde{\Omega} \subset\subset \Omega_0 \subset\subset \Omega$*, and assume that* $\varphi \in L^r(I, L^2(\Omega))$ *and* $\varphi|_{\Omega_0} \in L^r(I, L^s(\Omega_0))$ *for some* $1 < r < \infty$ *and* $2 \le s < \infty$*. Then the solution* ψ *to* (2.1) *belongs to* $W^{1,r}(I, L^s(\tilde{\Omega})) \cap L^r(I, W^{2,s}(\tilde{\Omega}))$ *and the following estimate holds:*

$$\|\psi\|_{L^r(I, W^{2,s}(\tilde{\Omega}))} + \|\partial_t \psi\|_{L^r(I, L^s(\tilde{\Omega}))} \le C_r^4 C_s^2 \big(\|\varphi\|_{L^r(I, L^s(\Omega_0))} + \|\varphi\|_{L^r(I, L^2(\Omega))}\big), \tag{2.5}$$

where C_r*,* C_s *are positive constants depending only on* r *and* s*, and satisfying* $C_r \le \frac{Cr^2}{r-1}$ *for some* $C > 0$ *and* $C_s \approx s$ *for* $s \to +\infty$*.*

Proof. The idea of proof for this lemma follows from [27, Lemma 2.4], [26, Lemma 2.2] and [4, Lemma 4.2]. Let $\tilde{\Omega}_1 \subset \Omega$ be an open smooth subdomain such that $\tilde{\Omega} \subset\subset \tilde{\Omega}_1 \subset\subset \Omega_0$. We consider a smooth cut-off function $\tilde{\omega}$ with the following properties:

$$\begin{aligned} \tilde{\omega}(x) &\in [0, 1], \quad \forall x \in \Omega, \\ \tilde{\omega}(x) &= 1, \qquad \forall x \in \tilde{\Omega}_1, \\ \tilde{\omega}(x) &= 0, \qquad \forall x \in \Omega \setminus \Omega_0, \end{aligned}$$

and set $\tilde{\psi} = \psi\tilde{\omega}$. Therefore, $\tilde{\psi}$ satisfies the following equation:

$$\begin{aligned} \partial_t \tilde{\psi} - \Delta\tilde{\psi} &= \tilde{\varphi} \quad && \text{in } I \times \Omega_0, \\ \tilde{\psi} &= 0 && \text{on } I \times \partial\Omega_0, \\ \tilde{\psi}(0) &= 0 && \text{in } \Omega_0, \end{aligned} \tag{2.6}$$

where $\tilde{\varphi} = \tilde{\omega}\varphi - \Delta\tilde{\omega}\psi - 2\nabla\tilde{\omega} \cdot \nabla\psi$. By using the continuous embedding $H^2(\Omega) \hookrightarrow W^{1,s}(\Omega)$ for $2 \le s \le 6$ and [27, Lemma 2.1], there exist positive constants C and C_r, with $C_r \le \frac{Cr^2}{r-1}$, such that the following estimate holds:

$$\|\tilde{\varphi}\|_{L^r(I,L^s(\Omega_0))} \le C\big(\|\varphi\|_{L^r(I,L^s(\Omega_0))} + \|\psi\|_{L^r(I,L^s(\Omega_0))} + \|\nabla\psi\|_{L^r(I,L^s(\Omega_0))}\big)$$
$$\le C_r\big(\|\varphi\|_{L^r(I,L^s(\Omega_0))} + \|\varphi\|_{L^r(I,L^2(\Omega_0))}\big).$$

The maximum parabolic regularity yields that for any $\tilde{\varphi} \in L^r(I,L^s(\Omega_0))$, Eq. (2.6) admits a unique solution $\tilde{\psi}$ satisfying $\partial_t\tilde{\psi}, \Delta\tilde{\psi} \in L^r(I,L^s(\Omega_0))$. Moreover, there holds

$$\|\partial_t\tilde{\psi}\|_{L^r(I,L^s(\Omega_0))} + \|\Delta\tilde{\psi}\|_{L^r(I,L^s(\Omega_0))} \le C_r\|\tilde{\varphi}\|_{L^r(I,L^s(\Omega_0))}$$

for $2 \le s < \infty$ and $1 < r < \infty$ (cf. [3, 13]). Noting that Ω_0 has a smooth boundary, we also have

$$\|\tilde{\psi}\|_{L^r(I,W^{2,s}(\Omega_0))} \le C_s\|\Delta\tilde{\psi}\|_{L^r(I,L^s(\Omega_0))}$$

for $2 \le s < \infty$ and $C_s \approx s$ for $s \to +\infty$ (cf. [16]). Observing that $\tilde{\psi} = \psi$ in $\tilde{\Omega}_1$, we obtain the desired estimate for ψ as follows:

$$\|\psi\|_{L^r(I,W^{2,s}(\tilde{\Omega}_1))} + \|\partial_t\psi\|_{L^r(I,L^s(\tilde{\Omega}_1))} \le C_r^2 C_s\big(\|\varphi\|_{L^r(I,L^s(\Omega_0))} + \|\varphi\|_{L^r(I,L^2(\Omega_0))}\big)$$

for $2 \le s \le 6$ and $1 < r < \infty$.

For $s > 6$, we repeat the previous steps with another smooth cut-off function

$$\omega(x) = 1, \quad \forall x \in \tilde{\Omega}_2,$$
$$\omega(x) = 0, \quad \forall x \in \Omega\backslash\tilde{\Omega}_1,$$

and use the continuous embedding $W^{2,6}(\tilde{\Omega}_1) \hookrightarrow W^{1,s}(\tilde{\Omega}_1)$ for any $2 \le s < \infty$, where $\tilde{\Omega}_2 \subset \Omega$ is a smooth open subdomain such that $\tilde{\Omega} \subset\subset \tilde{\Omega}_2 \subset\subset \tilde{\Omega}_1$. By repeating the above procedure with $\tilde{\Omega}_1$ replaced by $\tilde{\Omega}_2$, we can obtain the desired conclusion. $\square$

We denote by $C_0(\Omega)$ the space of continuous functions defined in $\bar{\Omega}$ that vanish on $\partial\Omega$. $C_0(\Omega)$ endowed with the supremum-norm $\|\cdot\|_\infty$ is a Banach space. The dual space of $C_0(\Omega)$ is identified with the space $M(\Omega)$ of real and regular Borel measures in Ω (cf. [36]). The norm in $M(\Omega)$ can be defined as

$$\|\mu\|_{M(\Omega)} := \sup\left\{\int_\Omega v d\mu : v \in C_0(\Omega), \|v\|_\infty \le 1\right\}$$

for each $\mu \in M(\Omega)$. Therefore, $\|\mu\|_{M(\Omega)}$ is the total variation of measure μ and the space $M(\Omega)$ is a Banach space under the norm $\|\cdot\|_{M(\Omega)}$.

The space $M(\Omega)$ is not separable, so we need to distinguish weakly and strongly measurability of functions $u : I \to M(\Omega)$. Hereafter, we use $L^2(I, M(\Omega))$ to denote the space of weakly measurable functions u defined in I and valued in $M(\Omega)$ such that the norm

$$\|u\|_{L^2(I,M(\Omega))} := \left(\int_I \|u(t)\|_{M(\Omega)}^2 dt\right)^{\frac{1}{2}}$$

is finite. Then $L^2(I, M(\Omega))$ is a Banach space under the norm $\| \cdot \|_{L^2(I,M(\Omega))}$ and can be identified with the dual space of $L^2(I, C_0(\Omega))$. The space $L^2(I, C_0(\Omega))$ consists of all measurable functions defined in I and valued in $C_0(\Omega)$, and the definition of the norm of $L^2(I, C_0(\Omega))$ is similar to that of $L^2(I, M(\Omega))$ with $\| \cdot \|_{M(\Omega)}$ replaced by $\| \cdot \|_\infty$. The duality pairing between $L^2(I, M(\Omega))$ and $L^2(I, C_0(\Omega))$ can be given by

$$\langle u, y \rangle_{L^2(I,M(\Omega)),L^2(I,C_0(\Omega))} = \int_I \langle u(t), y(t) \rangle_{M(\Omega),C_0(\Omega)} dt$$

for any $u \in L^2(I, M(\Omega))$ and $y \in L^2(I, C_0(\Omega))$, where $\langle \cdot, \cdot \rangle_{M(\Omega),C_0(\Omega)}$ denotes the duality pairing between $M(\Omega)$ and $C_0(\Omega)$.

For any given $\mu \in L^2(I, M(\Omega))$, let us introduce the following backward heat equation:

$$\begin{aligned}
-\partial_t p - \Delta p &= \mu &&\text{in } (0,T) \times \Omega, \\
p &= 0 &&\text{on } (0,T) \times \partial\Omega, \\
p(T) &= 0 &&\text{in } \Omega.
\end{aligned} \tag{2.7}$$

To define the weak solution of this equation, we use the method of transposition (cf. [30]).

Definition 2.1. *For any $\mu \in L^2(I, M(\Omega))$, we call a function $p \in L^2(I, L^2(\Omega))$ the very weak solution of Eq. (2.7), if p satisfies*

$$(p, \varphi)_I = \int_I \langle \mu(t), \psi(t) \rangle_{M(\Omega),C_0(\Omega)} dt, \quad \forall \varphi \in L^2(I, L^2(\Omega)), \tag{2.8}$$

where $\psi \in L^2(I, H^2(\Omega) \cap H_0^1(\Omega)) \cap H^1(I, L^2(\Omega))$ is the solution to Eq. (2.1) with right-hand side φ.

Remark 2.1. It is well known that $H^2(\Omega) \cap H_0^1(\Omega) \subset C_0(\Omega)$ ($n \leq 3$) and the inclusion is continuous (cf. [1]). Therefore, $L^2(I, H^2(\Omega) \cap H_0^1(\Omega)) \hookrightarrow L^2(I, C_0(\Omega))$ and Definition 2.1 is meaningful. For $1 < s < \frac{n}{n-1}$, we have $W_0^{1,s'}(\Omega) \hookrightarrow C_0(\Omega)$, and thus $\mu \in L^2(I, W^{-1,s}(\Omega))$. For $\frac{2n}{2+n} \leq s < \frac{n}{n-1}$, we have $H^2(\Omega) \cap H_0^1(\Omega) \hookrightarrow W_0^{1,s'}(\Omega)$ so that the right-hand side of (2.8) can be viewed as the duality pairing between $L^2(I, W^{-1,s}(\Omega))$ and $L^2(I, W_0^{1,s'}(\Omega))$ for s, s' satisfying $\frac{1}{s} + \frac{1}{s'} = 1$. This fact will be used in the proof of Theorem 2.1.

The existence, uniqueness and regularity of the very weak solution of Eq. (2.7) are given by the following theorem.

Theorem 2.1. *For any given $\mu \in L^2(I, M(\Omega))$, there exists a unique solution $p \in L^2(I, L^2(\Omega))$ of Eq. (2.7) in the very weak sense, and $p \in L^2(I, W_0^{1,s}(\Omega))$ ($1 < s < \frac{n}{n-1}$) with*

$$\|p\|_{L^2(I,W_0^{1,s}(\Omega))} \leq C \|\mu\|_{L^2(I,M(\Omega))},$$

where $C > 0$ is a constant independent of p. In addition, there exists a constant $1 < s_0 < \frac{3}{2}$ such that $\partial_t p \in L^2(I, W^{-1,s}(\Omega))$ for $s_0 < s < \frac{n}{n-1}$, and the identity

$$-\int_I \langle \partial_t p, \psi \rangle_{W^{-1,s}, W_0^{1,s'}} \, dt + \int_I (\nabla p, \nabla \psi)_{L^s, L^{s'}} dt = \int_I \langle \mu(t), \psi(t) \rangle_{M(\Omega), C_0(\Omega)} dt$$

holds for any $\psi \in L^2(I, W_0^{1,s'}(\Omega))$. Here $\langle \cdot, \cdot \rangle_{W^{-1,s}, W_0^{1,s'}}$ denotes the duality pairing between $W^{-1,s}(\Omega)$ and $W_0^{1,s'}(\Omega)$, $(\cdot, \cdot)_{L^s, L^{s'}}$ denotes the duality pairing between $L^s(\Omega)^n$ and $L^{s'}(\Omega)^n$, and s and s' satisfy $\frac{1}{s} + \frac{1}{s'} = 1$.

Proof. The idea of proof follows from [8, 17, 27, 28], and we sketch it for completeness. Let $\{\mu_k\}_{k=1}^\infty \subset C(\bar{I} \times \bar{\Omega})$ be such that $\mu_k \overset{\star}{\rightharpoonup} \mu$ in $L^2(I, M(\Omega))$ and $\|\mu_k\|_{L^2(I, L^1(\Omega))} \leq \|\mu\|_{L^2(I, M(\Omega))}$, $\forall k \in \mathbb{N}_+$. Let $p_k \in L^2(I, H^2(\Omega) \cap H_0^1(\Omega)) \cap H^1(I, L^2(\Omega))$ be the solution of the following equation:

$$
\begin{aligned}
-\partial_t p_k - \Delta p_k = \mu_k & \quad \text{in } (0, T) \times \Omega, \\
p_k = 0 & \quad \text{on } (0, T) \times \partial\Omega, \\
p_k(T) = 0 & \quad \text{in } \Omega.
\end{aligned}
$$

For any $\varphi \in \mathscr{D}(I \times \Omega)$, let $\psi \in L^2(I, C_0(\Omega))$ be the solution of (2.1), where $\mathscr{D}(I \times \Omega)$ denotes the set of all infinitely differentiable functions on $I \times \Omega$ with compact support. For $1 < s < \frac{n}{n-1}$, we have

$$
\begin{aligned}
\int_I \int_\Omega p_k \varphi \, dx dt &= \int_I \int_\Omega p_k (\partial_t \psi - \Delta \psi) \, dx dt \\
&= \int_I \int_\Omega (-\partial_t p_k - \Delta p_k) \psi \, dx dt = \int_I \int_\Omega \mu_k \psi \, dx dt \\
&\leq \|\mu_k\|_{L^2(I, M(\Omega))} \cdot \|\psi\|_{L^2(I, C(\bar{\Omega}))} \\
&\leq C \|\mu_k\|_{L^2(I, L^1(\Omega))} \cdot \|\psi\|_{L^2(I, W^{1,s'}(\Omega))}.
\end{aligned}
$$

Now we use the maximal regularity of the heat equation for further estimate. Since Ω is convex with a Lipschitz boundary, there exists a constant $\hat{s}$ with $\hat{s} > 4$ if $n = 2$, and $\hat{s} > 3$ if $n = 3$, such that $\Delta : W_0^{1,s}(\Omega) \to W^{-1,s}(\Omega)$ is an isomorphism for each $\hat{s}' < s < \hat{s}$, where $\frac{1}{\hat{s}'} + \frac{1}{\hat{s}} = 1$ (cf. [24]). Furthermore, using [13, Theorem 5.4] we obtain that, for every $\hat{s}' < s' < \hat{s}$, there exists a constant $C > 0$ such that

$$
\begin{aligned}
\int_I \int_\Omega p_k \varphi \, dx dt &\leq C \|\mu_k\|_{L^2(I, L^1(\Omega))} \cdot \|\psi\|_{L^2(I, W^{1,s'}(\Omega))} \\
&\leq C \|\mu_k\|_{L^2(I, L^1(\Omega))} \cdot \|\varphi\|_{L^2(I, W^{-1,s'}(\Omega))}.
\end{aligned}
$$

Since $\mathscr{D}(I \times \Omega)$ is dense in $L^2(I, W^{-1,s'}(\Omega))$ and

$$\left(L^2(I, W_0^{1,s}(\Omega)) \right)^* = L^2\left(I, W^{-1,s'}(\Omega) \right),$$

one obtains

$$\|p_k\|_{L^2(I,W_0^{1,s}(\Omega))} \leq C\|\mu\|_{L^2(I,M(\Omega))}, \quad \forall k \in \mathbb{N}_+. \qquad (2.9)$$

Therefore, $\{p_k\}_{k\in\mathbb{N}_+} \subset L^2(I,W_0^{1,s}(\Omega))$ is bounded. By the reflexivity of $L^2(I,W_0^{1,s}(\Omega))$ we extract a subsequence, still denoted by $\{p_k\}_{k\in\mathbb{N}_+}$ such that $p_k \rightharpoonup p$ in $L^2(I,W_0^{1,s}(\Omega))$ for some $p \in L^2(I,W_0^{1,s}(\Omega))$. Now we check that p is a solution to (2.7). Note that $\mu_k \overset{\star}{\rightharpoonup} \mu$, and p_k satisfies

$$\int_I \int_\Omega p_k \varphi \, dx dt = \int_I \int_\Omega \mu_k \psi \, dx dt, \quad \forall \varphi \in \mathscr{D}(I \times \Omega), \qquad (2.10)$$

where $(\psi,\varphi) \in L^2(I,H^2(\Omega) \cap H_0^1(\Omega)) \cap H^1(I,L^2(\Omega)) \times L^2(I,L^2(\Omega))$ satisfies Eq. (2.1). Passing to the limit in (2.10), we then obtain the identity (2.8) for any $\varphi \in \mathscr{D}(I \times \Omega)$. Because $\mathscr{D}(I \times \Omega)$ is dense in $L^2(I,L^2(\Omega))$ and $(\partial_t - \Delta)$ is an isomorphism from $L^2(I,H^2(\Omega) \cap H_0^1(\Omega)) \cap H^1(I,L^2(\Omega))$ to $L^2(I,L^2(\Omega))$, we arrive at (2.8). Therefore, p is a solution to (2.7), and the estimate for p can be obtained by Eq. (2.9) for any $\hat{s}' < s < \frac{n}{n-1}$.

When

$$\frac{2n}{n+2} \leq s < \frac{n}{n-1},$$

$W_0^{1,s}(\Omega) \hookrightarrow L^2(\Omega)$ is dense and continuous. For any $\hat{s}' < s < \frac{n}{n-1}$, the solution of (2.7) belongs to $L^2(I,W_0^{1,s}(\Omega))$ and $\Delta : W_0^{1,s}(\Omega) \to W^{-1,s}(\Omega)$ is an isomorphism and $\mu \in L^2(I,W^{-1,s}(\Omega))$, therefore, $\partial_t p \in L^2(I,W^{-1,s}(\Omega))$. Let $s_0 = \max\{\hat{s}', \frac{2n}{n+2}\}$, then $p \in L^2(I,W_0^{1,s}(\Omega)) \cap L^2(I,L^2(\Omega))$ and $\partial_t p \in L^2(I,W^{-1,s}(\Omega))$ for $s_0 < s < \frac{n}{n-1}$.

When

$$s_0 < s < \frac{n}{n-1},$$

set

$$W := \Big\{ \psi \in L^2\big(I,H^2(\Omega) \cap H_0^1(\Omega)\big) \cap H^1\big(I,L^2(\Omega)\big) :$$
$$\psi \text{ satisfies } (2.1) \text{ for some } \varphi \in L^2\big(I,L^2(\Omega)\big) \Big\}.$$

Taking $\varphi = \partial_t \psi - \Delta\psi$ in (2.8) for any $\psi \in W$ and integrating by parts (cf. [37]), we obtain

$$-\int_I \langle \partial_t p, \psi \rangle_{W^{-1,s},W_0^{1,s'}} dt + \int_I (\nabla p, \nabla \psi)_{L^s,L^{s'}} dt$$
$$= \int_I \langle \mu(t), \psi(t) \rangle_{M(\Omega),C_0(\Omega)} dt, \quad \forall \psi \in W, \qquad (2.11)$$

which, together with the fact that W is dense in $L^2(I,W_0^{1,s'}(\Omega))$ and $L^2(I,M(\Omega)) \subset L^2(I,W^{-1,s}(\Omega))$, implies the desired identity of this theorem.

Remark 2.2. Using integration by parts (cf. [37]) for the first term in identity (2.11), we obtain the following identity for any $\psi \in \tilde{W}$:

$$\int_I \langle p, \partial_t \psi \rangle_{W_0^{1,s}, W^{-1,s'}} dt + \int_I (\nabla p, \nabla \psi)_{L^s, L^{s'}} dt = \int_I \langle \mu(t), \psi(t) \rangle_{M(\Omega), C_0(\Omega)} dt, \quad (2.12)$$

where
$$\tilde{W} = \left\{ v \in L^2(I, W_0^{1,s'}(\Omega)) : \partial_t v \in L^2(I, W^{-1,s'}(\Omega)), \, v(0) = 0 \right\}.$$

Replacing (2.8) in Definition 2.1 by the identity (2.11) or (2.12), we obtain the equivalent definition of solution to (2.7) with higher regularity.

For any given $g \in C(\bar{I}, L^2(\Omega))$ and $\hat{t} \in I$, we consider the following backward in time heat equation:

$$\begin{aligned}
-\partial_t p - \Delta p &= g(t) \otimes \delta(t - \hat{t}) && \text{in } (0, T) \times \Omega, \\
p &= 0 && \text{on } (0, T) \times \partial\Omega, \qquad (2.13)\\
p(T) &= 0 && \text{in } \Omega,
\end{aligned}$$

where $\delta(t - \hat{t})$ is the Dirac delta distribution concentrated on $\hat{t}$, and $g(t) \otimes \delta(t - t_i)$ is a distribution defined by

$$\psi \to \int_\Omega g(x, \hat{t}) \psi(x, \hat{t}) dx, \quad \forall \psi \in \mathscr{D}(I \times \Omega).$$

Similarly, we can define the solution of (2.13) by the transposition method.

Definition 2.2. *We call a function $p \in L^2(I, L^2(\Omega))$ the very weak solution of Eq. (2.13), if it satisfies the following identity:*

$$(p, \varphi)_I = \int_\Omega g(\hat{t}) \psi(\hat{t}) dx, \quad \forall \varphi \in L^2(I, L^2(\Omega)), \qquad (2.14)$$

where $\psi \in C(\bar{I}, L^2(\Omega))$ satisfies (2.1) for given right-hand side φ.

The following proposition concerning the well-posedness of (2.13) is taken from [17], which is very useful in the finite element error estimation for the adjoint equation.

Proposition 2.2. *System (2.13) has a unique solution $p \in L^2(I, H_0^1(\Omega)) \cap L^\infty(I, L^2(\Omega))$. Moreover,*

$$(p, \partial_t v)_{H_0^1, H^{-1}} + (\nabla p, \nabla v)_I = \int_\Omega g(\hat{t}) v(\hat{t}) dx, \quad \forall v \in W(I)$$

and

$$\|p\|_{L^2(I, H_0^1(\Omega))} + \|p\|_{L^\infty(I, L^2(\Omega))} \le C \|g\|_{L^\infty(I, L^2(\Omega))}$$

for some constant $C > 0$ independent of p, where

$$W(I) := \left\{ L^2(I, H_0^1(\Omega)) \cap H^1(I, H^{-1}(\Omega)), \, v(0) = 0 \right\}.$$

3. Optimal control problems

In this section we study the optimal control problem. We denote the state space by $X = L^2(I, C(\bar{\Omega})) \cap C(\bar{I}, L^2(\Omega))$, where $C(\bar{\Omega})$ denotes the space of continuous functions in $\bar{\Omega}$. By using Proposition 2.1 we can see that the state equation (1.1) admits a unique solution $y(u) \in X$ for any given control $u \in U$. Therefore, we can define two mappings $S_i : L^2(I, L^2(\omega)) \to A_i$ such that $S_i u = y(u)$ for $i = 1, 2$, where $A_1 = L^2(I, C(\bar{\Omega}))$ and $A_2 = C(\bar{I}, L^2(\Omega))$.

Using the mappings S_1, S_2 we introduce the reduced cost functional $\hat{\mathcal{J}}_\beta : U \to \mathbb{R}$ as

$$\hat{\mathcal{J}}_\beta(u) := (2 - \beta)J_1(S_1 u, u) + (\beta - 1)J_2(S_2 u, u) + \frac{\alpha}{2}\|u\|_U^2.$$

The optimal control problem (1.3) can be equivalently written as

$$\min_{u \in U_{ad}} \hat{\mathcal{J}}_\beta(u). \tag{3.1}$$

By using standard arguments (cf. [29]) we can prove that the optimization problem (3.1) admits a unique solution $\bar{u} \in U_{ad}$ for any $\beta \in [1, 2]$. The first order necessary (also sufficient) optimality condition of the optimization problem (3.1) at $\bar{u}$ reads as follows:

$$\hat{\mathcal{J}}_\beta'(\bar{u})(v - \bar{u}) \geq 0, \quad \forall v \in U_{ad}, \tag{3.2}$$

where $\hat{\mathcal{J}}_\beta'(\bar{u})$ denotes the Fréchet derivative of $\hat{\mathcal{J}}_\beta$ at $\bar{u}$ that takes the form

$$\hat{\mathcal{J}}_\beta'(u)(v) = (2 - \beta)\hat{J}_1'(u)(v) + (\beta - 1)\hat{J}_2'(u)(v) + \alpha(u, v)_U, \quad \forall v \in U$$

for given $u \in U$. Here $\hat{J}_1'(u)$ and $\hat{J}_2'(u)$ are given by

$$\hat{J}_1'(u)(v) = \sum_{i=1}^{N_1} \int_0^T \big(y(u)(x^i, t) - y_{d_S}^i(t)\big)\tilde{y}_1(v)(x^i, t)dt, \quad \forall v \in U,$$

$$\hat{J}_2'(u)(v) = \sum_{i=1}^{N_2+1} \int_\Omega \big(y(u)(x, t^i) - y_{d_T}^i(x)\big)\tilde{y}_2(v)(x, t^i)dx, \quad \forall v \in U$$

with $t^{N_2+1} = T$ and $y_{d_T}^{N_2+1} = y_T$. Moreover, $\tilde{y}_i(v) = S_i'(u)v \in A_i$ is the solution to (1.1) with right-hand side $\mathcal{B}v$, and is independent of u. In other words, the necessary optimality condition (3.2) is equivalent to the following variational inequality:

$$(2 - \beta)\hat{J}_1'(\bar{u})(v - \bar{u}) + (\beta - 1)\hat{J}_2'(\bar{u})(v - \bar{u}) + \alpha(\bar{u}, v - \bar{u})_U \geq 0, \quad \forall v \in U_{ad}. \tag{3.3}$$

Particularly, the above variational inequalities for the cases of $\beta = 1$ and $\beta = 2$ have the following expressions: For any $v \in U_{ad}$ there holds

$$\sum_{i=1}^{N_1} \int_0^T \big(y(\bar{u})(x^i, t) - y_{d_S}^i(t)\big)\tilde{y}(v - \bar{u})(x^i, t)dt$$

$$+ \alpha \int_0^T \int_\omega \bar{u}(t, x)\big(v(t, x) - \bar{u}(t, x)\big)dxdt \geq 0 \tag{3.4}$$

for $\beta = 1$, and

$$\sum_{i=1}^{N_2+1} \int_\Omega \big(y(\bar{u})(x,t^i) - y_{d_T}^i(x)\big)\tilde{y}(v-\bar{u})(x,t^i)dx$$

$$+ \alpha \int_0^T \int_\omega \bar{u}(t,x)(v-\bar{u})(t,x)dxdt \geq 0 \qquad (3.5)$$

for $\beta = 2$.

Remark 3.1. Multiply the variational inequalities (3.4) and (3.5) by $2 - \beta$ and $\beta - 1$, respectively, and add them up, then we obtain the variational inequality (3.3). In other words, we can obtain the first order optimality condition for the case of $\beta \in (1,2)$ by using the linear combination of the cases for $\beta = 1$ and $\beta = 2$, which implies the relationship between the adjoint state variable of the case $\beta \in (1,2)$ and that of the case $\beta = 1$ or $\beta = 2$.

However, the above form of first order optimality condition is not friendly for numerical approximation. In order to better characterize the optimal solution $\bar{u}$ to the optimal control problem (3.1) and its first order necessary optimality condition (3.2), we will introduce two backward in time heat equations, the so-called adjoint state equations. The first one reads

$$- \partial_t p^1(x,t) - \Delta p^1(x,t)$$

$$= \sum_{i=1}^{N_1} \big(y(u)(x^i,t) - y_{d_S}^i(t)\big) \otimes \delta(x - x^i) \quad \text{in } (0,T) \times \Omega,$$

$$p^1(x,t) = 0 \qquad\qquad\qquad\qquad \text{on } (0,T) \times \partial\Omega,$$

$$p^1(x,T) = 0 \qquad\qquad\qquad\qquad \text{in } \Omega,$$

(3.6)

where $\delta(x-x^i)$ is the Dirac delta distribution concentrated on x^i, and $\sum_{i=1}^{N_1}(y(u)(x^i,t) - y_{d_S}^i(t)) \otimes \delta(x - x^i)$ is a distribution defined by

$$\psi \to \sum_{i=1}^{N_1} \int_0^T \big(y(u)(x^i,t) - y_{d_S}^i(t)\big)\psi(x^i,t)dt, \quad \forall \psi \in \mathscr{D}(I \times \Omega),$$

corresponding to the case $\beta = 1$. The above equation involves singular right-hand side and does not admit a classical variational solution, therefore, we should use the method of transposition to define the solution to (3.6) (cf. [30]), see Definition 2.1. By Theorem 2.1 we can infer the existence of a unique solution.

The second adjoint state equation reads

$$- \partial_t p^2(x,t) - \Delta p^2(x,t)$$

$$= \sum_{i=1}^{N_2} \big(y(u)(x,t^i) - y_{d_T}^i(x)\big) \otimes \delta(t - t^i) \quad \text{in } (0,T) \times \Omega, \qquad (3.7a)$$

$$p^2(x,t) = 0 \qquad \text{on } (0,T) \times \partial\Omega, \tag{3.7b}$$

$$p^2(T) = y(u)(T) - y_T \quad \text{in } \Omega, \tag{3.7c}$$

where $\delta(t - t^i)$ is the Dirac delta function concentrated on t^i, and $\sum_{i=1}^{N_2}(y(u)(x,t^i) - y_{d_T}^i(x)) \otimes \delta(t - t^i)$ is a distribution defined by

$$\psi \to \sum_{i=1}^{N_2} \int_\Omega \big(y(u)(x,t^i) - y_d^i(x)\big)\psi(x,t^i)dx, \quad \forall \psi \in \mathscr{D}(I \times \Omega), \tag{3.8}$$

corresponding to the case $\beta = 2$.

The existence and uniqueness of the very weak solution to (3.7) can be obtained by Definition 2.2 and Proposition 2.2. In fact, for given $y(u)(T) - y_T \in L^2(\Omega)$ (cf. [30]), let $p^0 \in L^2(I, H_0^1(\Omega)) \cap L^\infty(I, L^2(\Omega))$ satisfy the following equation:

$$\begin{aligned}
-\partial_t p^0(x,t) - \Delta p^0(x,t) &= 0 \quad \text{in } (0,T) \times \Omega, \\
p^0 &= 0 \qquad \text{on } (0,T) \times \partial\Omega, \\
p^0(T) &= y(u)(T) - y_T \qquad \text{in } \Omega.
\end{aligned}$$

Let $\tilde{p}^i$ be the solution of (2.13) with the right-hand side replaced by $(y(u)(x,t^i) - y_{d_T}^i(x)) \otimes \delta(t - t^i)$. By Proposition 2.2, we know that $\tilde{p}^i$, $i = 1, \ldots, N_2$, are well-defined. Now let $p^2 = p^0 + \sum_{i=1}^{N_2} \tilde{p}^i$. From the superposition principle, it follows that p^2 is the unique solution to (3.7). To emphasize the dependence of p^1, p^2 on u, we denote by $p^1(u)$ and $p^2(u)$ the solutions to (3.6) and (3.7), respectively. From the optimality condition (3.2) we see that the directional derivative of $\hat{\mathcal{J}}_\beta(\bar{u})$ along $v - \bar{u}$ is non-negative for all $v \in U_{ad}$. By using the solutions to adjoint equations (3.6) and (3.7) we can give an explicit expression for $\hat{\mathcal{J}}_\beta'(u)$.

Let us introduce the adjoint variable $p(u) \in L^2(I, L^2(\Omega))$, with

$$p(u) := (2 - \beta)p^1(u) + (\beta - 1)p^2(u). \tag{3.9}$$

Then we can obtain an explicit expression for $\hat{\mathcal{J}}_\beta'(u)$, viz.

$$\begin{aligned}
\hat{\mathcal{J}}_\beta'(u)(v) &= \int_0^T \int_\omega \chi_\omega(x)p(u)(t,x)v(t,x)dxdt \\
&\quad + \alpha \int_0^T \int_\omega u(t,x)v(t,x)dxdt, \quad \forall v \in U. \tag{3.10}
\end{aligned}$$

In fact, taking $\varphi = \mathcal{B}v$ in Eq. (2.8), and letting ψ be the solution of (2.1), we have $\psi = \tilde{y}(v)$. Taking $\mu = \sum_{i=1}^{N_1}(y(u)(x^i,t) - y_{d_S}^i(t)) \otimes \delta(x - x^i)$ in (2.7) and using (2.8), we get

$$\int_I \int_\omega \chi_\omega p^1(u)vdxdt = \sum_{i=1}^{N_1} \int_\Omega \big(y(u)(x^i,t) - y_{d_S}^i(t)\big)\tilde{y}(v)(x^i)dt = \hat{J}_1'(u)(v).$$

Similarly, by using Definition 2.2 and (2.14) we can derive

$$
\begin{aligned}
\int_I \int_\omega \chi_\omega p^2(u) v \, dx dt &= \sum_{i=1}^{N_2} \int_\Omega \big(y(u)(x,t^i) - y_{d_T}^i(x)\big) \tilde{y}(v)(t^i) dx \\
&\quad + \int_\Omega \big(y(u)(x,T) - y_T(x)\big) \tilde{y}(v)(T) dx \\
&= \hat{J}_2'(u)(v).
\end{aligned}
$$

Multiplying the first identity by $2 - \beta$ and the second one by $\beta - 1$, then adding them up, we obtain (3.10).

By using (3.10), a more convenient expression for (3.2) is given as follows:

$$
\int_I \int_\omega \big(\chi_\omega p(\bar{u})(t,x) + \alpha \bar{u}(t,x)\big)\big(v(t,x) - \bar{u}(t,x)\big) dx dt \geq 0, \quad \forall v \in U_{ad}. \tag{3.11}
$$

The necessary optimality condition (3.11) can be equivalently formulated via the pointwise projection

$$
P_{U_{ad}} : U \to U_{ad}, \quad P_{U_{ad}}(z)(t,x) = \max\big(u_a, \min(u_b, z(t,x))\big) \quad \text{a.e.} \quad (t,x) \in I \times \omega.
$$

That is,

$$
\bar{u} = P_{U_{ad}}\left(-\frac{1}{\alpha} \chi_\omega p(\bar{u})\right). \tag{3.12}
$$

Now we collect the results on the first order optimality system for the optimization problem.

Theorem 3.1. *A control $\bar{u} \in U_{ad}$ with associated state $\bar{y} \in X$ is an optimal pair of the optimal control problem* (1.3) *if and only if there exist $\bar{p}^1 \in L^2(I, W_0^{1,s}(\Omega))$, $\bar{p}^2 \in L^2(I, H_0^1(\Omega))$ and an adjoint state $\bar{p} \in L^2(I, L^2(\Omega))$ such that*

$$
(\partial_t \bar{y}, v)_{H^{-1},H^1} + (\nabla \bar{y}, \nabla v)_I = (\mathcal{B}\bar{u}, v)_I, \quad \forall v \in L^2\big(I, H_0^1(\Omega)\big), \quad \bar{y}(0) = 0,
$$

$$
-\big(\partial_t \bar{p}^1, \psi\big)_{W^{-1,s},W^{1,s'}} + \int_I (\nabla \bar{p}^1, \nabla \psi)_{L^s,L^{s'}} dt
$$

$$
= \sum_{i=1}^{N_1} \int_0^T \big(\bar{y}(x^i) - y_{d_S}^i, \psi(x^i)\big) dt, \qquad \forall \psi \in L^2\big(I, W_0^{1,s'}(\Omega)\big), \quad \bar{p}^1(T) = 0,
$$

$$
\big(\bar{p}^2, \partial_t \psi\big)_{H^1,H^{-1}} + \big(\nabla \bar{p}^2, \nabla \psi\big)_I
$$

$$
= \sum_{i=1}^{N_2} \int_\Omega \big(\bar{y}(t^i) - y_{d_T}^i\big)\psi(t^i) dx + \int_\Omega \big(\bar{y}(T) - y_T\big)\psi(T) dx, \quad \forall \psi \in W(I),
$$

$$
(\chi_\omega \bar{p} + \alpha \bar{u}, v - \bar{u})_U \geq 0, \quad \forall v \in U_{ad}, \quad \bar{p} = (2 - \beta)\bar{p}^1 + (\beta - 1)\bar{p}^2.
$$

Here, $(\cdot, \cdot)_{H^{-1},H^1}$ is the dual relation between $L^2(I, H^{-1}(\Omega))$ and $L^2(I, H_0^1(\Omega))$, and $(\cdot, \cdot)_{W^{-1,s},W^{1,s'}}$ is the dual relation between $L^2(I, W^{-1,s'}(\Omega))$ and $L^2(I, W_0^{1,s}(\Omega))$, and $s_0 < s < \frac{n}{n-1}$ for some $1 < s_0 < \frac{3}{2}$.

Remark 3.2. Due to the linearity of the governing state equation, we can incorporate an additional nonzero right-hand side $f \in L^2(I, L^2(\Omega))$ and an initial data y_0 in the state equation (1.1), by requiring $y_0 \in H_0^1(\Omega)$ when $\beta < 2$ and $y_0 \in L^2(\Omega)$ when $\beta = 2$, and modifying the desired observations. Moreover, the above setup is a model problem, where, for simplicity, we choose the heat equation as the state equation. However, all our results can be extended directly to more general self-adjoint elliptic operator of second order with smooth coefficients (instead of $-\Delta$) on the left-hand side of (1.1).

4. Finite element discretization for the control problem

In this section we consider the finite element discretization of the control problem. To begin with, we divide $\bar{I} = [0, T]$ into subintervals $I_m = (t_{m-1}, t_m]$ of length $k_m = t_m - t_{m-1}$, where $0 = t_0 < t_1 < \cdots < t_M = T$. The maximal time step is denoted by $k = \max_{1 \leq m \leq M} k_m$. We impose the following conditions (cf. [27]) which are valid for a large class of time grids:

(i) There exist constants $c, \gamma > 0$, independent of k, such that

$$\min_{1 \leq m \leq M} k_m \geq ck^\gamma.$$

(ii) There exists a constant $\kappa > 0$, independent of k, such that for $m = 1, \ldots, M - 1$,

$$\kappa^{-1} \leq \frac{k_m}{k_{m+1}} \leq \kappa.$$

(iii) The maximal time step size satisfies $k \leq \frac{1}{4} \min\{T, 1\}$.

Now we define the time semi-discrete space consisting of all piecewise constant functions as

$$X_k^0 = \left\{ z_k \in L^2\big(I, H_0^1(\Omega)\big) : z_k|_{I_m} \in P_0\big(I_m, H_0^1(\Omega)\big), \ m = 1, \ldots, M \right\},$$

where $P_0(I_m, H_0^1(\Omega))$ denotes the function space of constants on I_m valued in $H_0^1(\Omega)$, $m = 1, \ldots, M$. We will need the following notations for functions in space X_k^0:

$$v_m = v_m^- := \lim_{t \to 0^+} v(t_m - t), \quad v_{m+1} = v_m^+ := \lim_{t \to 0^+} v(t_m + t), \quad [v]_m := v_m^+ - v_m^-.$$

Let $\mathscr{T}_h = \{\tau\}$ be a family of quasi-uniform and shape regular partitions of Ω into n-simplex τ with diameter h_τ, and denote the mesh size of $\mathscr{T}_h$ by $h = \max_{\tau \in \mathscr{T}_h} h_\tau$. We assume further that there exists a constant $c > 0$, independent of h and τ, such that $\mathrm{diam}(\tau) \leq h \leq c|\tau|^{\frac{1}{n}}$. Let

$$V_h := \left\{ v_h \in C(\bar{\Omega}) : v_h|_\tau \in P_1(\tau), \forall \tau \in \mathscr{T}_h \right\}$$

be the usual continuous piecewise linear finite element space, where $P_1(\tau)$ denotes the linear polynomials on τ. We set $V_h^0 = V_h \cap H_0^1(\Omega)$.

Let π_h be the usual nodal Lagrange interpolation (cf. [6]) from $C_0(\Omega)$ to V_h or the Clément-type interpolation (cf. [11]) from $L^1(\Omega)$ to V_h. Let $P_h : L^2(\Omega) \to V_h$ be the L^2-projection operator defined by

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

and let $R_h : H_0^1(\Omega) \to V_h^0$ be the Ritz projection operator defined by

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

Then, the following error estimates for the Ritz projection and the inverse estimates are classical (cf. [6, 10]).

Lemma 4.1. *Let R_h be the Ritz projection operator defined above. Then there hold*

$$\begin{aligned}
\|v_h\|_{L^\infty(\Omega)} &\leq Ch^{-\frac{n}{2}} \|v_h\|, & \forall v_h \in V_h, \\
\|v - R_h v\|_{L^\infty(\Omega)} &\leq Ch^{2-\frac{n}{2}} \|\nabla^2 v\|, & \forall v \in H_0^1(\Omega) \cap H^2(\Omega), \\
\|v - R_h v\|_{H^1(\Omega)} &\leq Ch \|\nabla^2 v\|, & \forall v \in H_0^1(\Omega) \cap H^2(\Omega), \\
\|v - R_h v\| &\leq Ch \|\nabla(v - R_h v)\|, & \forall v \in H_0^1(\Omega).
\end{aligned}$$

Proof. The first estimate is the standard inverse estimate for finite element functions (cf. [6]), the second one is the uniform estimate for Ritz projection operator and we refer to [6, 10]. The third and the last ones are standard error estimates for Ritz projection in H^1 and L^2 norms which can also be found in [6, 10]. $\square$

Remark 4.1. The application of the operators R_h, P_h and π_h to time-dependent arguments has to be understood pointwisely in time, and Lemma 4.1 holds for time-dependent case with the corresponding estimates under the space-time norms.

In order to define the fully discrete approximation, we need to introduce the space-time finite element space

$$X_{k,h}^{0,1} := \left\{ v_{kh} \in X_k^0 : v_{kh}|_{I_m} \in P_0(I_m, V_h^0), \ m = 1, \dots, M \right\}$$

and the bilinear form (cf. [33])

$$B(v, w) := \sum_{m=1}^M \langle \partial_t v, w \rangle_{I_m \times \Omega} + (\nabla v, \nabla w)_I + \sum_{m=2}^M \left([v]_{m-1}, w_{m-1}^+ \right) + (v_0^+, w_0^+) \quad (4.1)$$

for $v, w \in X_k^0$ or $v, w \in X_{k,h}^{0,1}$. Here $\langle \cdot, \cdot \rangle_{I_m \times \Omega}$ denotes the duality pairing between $L^2(I_m, H^{-1}(\Omega))$ and $L^2(I_m, H_0^1(\Omega))$. By integration by parts we obtain an equivalent dual expression of B as follows (cf. [33]):

$$B(v, w) = -\sum_{m=1}^M \langle v, \partial_t w \rangle_{I_m \times \Omega} + (\nabla v, \nabla w)_I - \sum_{m=1}^{M-1} \left(v_m^-, [w]_m \right) + \left(v_M^-, w_M^- \right). \quad (4.2)$$

For given control $u \in U$, the DG(0)-CG(1) approximation $y_{kh} \in X_{k,h}^{0,1}$ of the state equation (1.1) is defined by (cf. [38])

$$B(y_{kh}, v_{kh}) = (\mathcal{B}u, v_{kh})_I, \quad \forall v_{kh} \in X_{k,h}^{0,1}. \tag{4.3}$$

It is not difficult to verify the following Galerkin orthogonality relation

$$B(y - y_{kh}, v_{kh}) = 0, \quad \forall v_{kh} \in X_{k,h}^{0,1}. \tag{4.4}$$

Now we give the discretized optimal control problem

$$\min \mathcal{J}_\beta(y_{kh}, u) : (y_{kh}, u) \in X_{k,h}^{0,1} \times U_{ad} \quad \text{subject to} \quad (4.3). \tag{4.5}$$

By standard arguments, it is easy to prove that the discrete optimal control problem (4.5) has a unique solution. Note that by following the variational discretization concept (cf. [22]) the admissible control set is not discretized explicitly in (4.5). It can be checked that the discrete control is constant on each I_m but may not belong to the finite element space V_h.

Similar to the continuous case, the discrete state equation (4.3) defines a mapping $u \in U_{ad} \to y_{kh} \in X_{k,h}^{0,1}$ from the control space to the discrete state space. To emphasize the dependence of the mapping on u, we write $y_{kh} = y_{kh}(u)$. Define the reduced discrete cost functional $\hat{\mathcal{J}}_{\beta,kh} : U \to \mathbb{R}$ as

$$\hat{\mathcal{J}}_{\beta,kh}(u) := \mathcal{J}_\beta(y_{kh}(u), u).$$

Then the optimal control problem (4.5) is reduced to the following optimization problem:

$$\min \hat{\mathcal{J}}_{\beta,kh}(u), \quad u \in U_{ad}. \tag{4.6}$$

We denote by $\bar{u}_{kh}$ the solution to (4.6) and we have the following first order necessary condition:

$$\hat{\mathcal{J}}'_{\beta,kh}(\bar{u}_{kh})(v - \bar{u}_{kh}) \geq 0, \quad \forall v \in U_{ad}, \tag{4.7}$$

where $\hat{\mathcal{J}}'_{\beta,kh}$ denotes the Fréchet derivative of $\hat{\mathcal{J}}_{\beta,kh}$ at $\bar{u}_{kh}$. Due to the convexity of the optimization problem (4.6), it is easy to verify that the condition (4.7) is also sufficient for (4.6).

Given $u, v \in U$, let $\tilde{y}_{kh}(v) = y'_{kh}(u)v$ be the directional derivative of $y_{kh}(u)$ at v, then it is easy to check that $\tilde{y}_{kh}(v)$ satisfies

$$B(\tilde{y}_{kh}(v), \varphi_{kh}) = (\mathcal{B}v, \varphi_{kh})_I, \quad \forall \varphi_{kh} \in X_{k,h}^{0,1}, \tag{4.8}$$

and that $\hat{\mathcal{J}}'_{\beta,kh}(u)$ has the following form:

$$\hat{\mathcal{J}}'_{\beta,kh}(u)v = (2 - \beta)\hat{J}'_{1,kh}(u)v + (\beta - 1)\hat{J}'_{2,kh}(u)v + \alpha(u, v)_U, \tag{4.9}$$

where

$$\hat{J}'_{1,kh}(u)v = \sum_{i=1}^{N_1} \int_0^T \left(y_{kh}(u)(x^i) - y_{d_S}^i \right) \tilde{y}_{kh}(v)(x^i) dt,$$

$$\hat{J}'_{2,kh}(u)v = \sum_{i=1}^{N_2} \int_\Omega \big(y_{kh}(u)(t^i) - y^i_{d_T}\big)\tilde{y}_{kh}(v)(t^i)dx$$

$$+ \int_\Omega \big(y_{kh}(u)(T) - y_T\big)\tilde{y}_{kh,M}(v)dx.$$

We denote by $\bar{y}_{kh}$ the optimal discrete state of the optimal control problem (4.5), i.e., $\bar{y}_{kh}$ is the solution of (4.3) with the right-hand side u replaced by $\bar{u}_{kh}$. Similar to the continuous level, in what follows we use the adjoint argument for $\hat{\mathcal{J}}'_{\beta,kh}$ to derive an explicit expression for (4.7). We firstly introduce the discrete variables $p^i_{kh}(u) \in X^{0,1}_{k,h}$, $i = 1, 2$, defined respectively by

$$B\big(\varphi_{kh}, p^1_{kh}(u)\big) = \sum_{i=1}^{N_1} \int_0^T \big(y_{kh}(u)(x^i) - y^i_{d_S}\big)\varphi_{kh}(x^i)dt, \tag{4.10}$$

$$B\big(\varphi_{kh}, p^2_{kh}(u)\big) = \sum_{i=1}^{N_2} \big(y_{kh}(u)(t^i) - y^i_{d_T}, \varphi_{kh}(t^i)\big) + \big(y_{kh}(u)(T) - y_T, \varphi_{kh,M}\big) \tag{4.11}$$

for any $\varphi_{kh} \in X^{0,1}_{k,h}$. We note that $p^i_{kh}(u)$ are well-defined by the standard theory (cf. [38]). Then we define the discrete adjoint state $p_{kh}(u) \in X^{0,1}_{k,h}$ as follows:

$$p_{kh}(u) := (2 - \beta)p^1_{kh}(u) + (\beta - 1)p^2_{kh}(u). \tag{4.12}$$

Taking $\varphi_{kh} = \tilde{y}_{kh}(v)$ in (4.10) and (4.11), multiplying (4.10) and (4.11) by $2 - \beta$ and $\beta - 1$, respectively, and adding them up, we get

$$B\big(\tilde{y}_{kh}(v), p_{kh}(u)\big) = (2 - \beta)\hat{J}'_{1,kh}(u)v + (\beta - 1)\hat{J}'_{2,kh}(u)v.$$

Once again, by taking $\varphi_{kh} = p_{kh}(u)$ in (4.8) we arrive at

$$B\big(\tilde{y}_{kh}(v), p_{kh}(u)\big) = \big(\mathcal{B}v, p_{kh}(u)\big).$$

Eventually, we obtain the expression of $\hat{\mathcal{J}}'_{\beta,kh}$, i.e.,

$$\hat{\mathcal{J}}'_{\beta,kh}(v) = \int_0^T \int_\omega \big(\chi_\omega(x)p_{kh}(u)(t,x) + \alpha u(t,x)\big)v(t,x)dxdt, \quad \forall v \in U. \tag{4.13}$$

Using this form, we also obtain the explicit expression of (4.7), i.e.,

$$\int_0^T \int_\omega \big(\chi_\omega(x)\bar{p}_{kh}(t,x) + \alpha\bar{u}_{kh}(t,x)\big)\big(v(t,x) - \bar{u}_{kh}\big)dxdt \geq 0, \quad \forall v \in U_{ad}, \tag{4.14}$$

where $\bar{p}_{kh} = p_{kh}(\bar{u}_{kh})$ is the so-called discrete adjoint state. In the following we write $\bar{p}^1_{kh} = p^1_{kh}(\bar{u}_{kh})$ and $\bar{p}^2_{kh} = p^2_{kh}(\bar{u}_{kh})$.

Once again, we use the projection operator $P_{U_{ad}}$ to write $\bar{u}_{kh}$ as the projection of the adjoint

$$\bar{u}_{kh} = P_{U_{ad}}\left(-\frac{1}{\alpha}\chi_\omega\bar{p}_{kh}\right),$$

or equivalently,

$$\bar{u}_{kh}(t,x) = \max\left(u_a, \min\left(u_b, -\frac{1}{\alpha}\bar{p}_{kh}(t,x)\right)\right) \quad \text{a.e.} \quad (t,x) \in (0,T) \times \omega. \quad (4.15)$$

From the above discussions it follows the following conclusion on the first order optimality system for the discretized optimal control problem (4.5).

Theorem 4.1. *The pair of approximate state and control, $(\bar{y}_{kh}, \bar{u}_{kh}) \in X_{k,h}^{0,1} \times U_{ad}$, is the optimal pair of problem* (4.5) *if and only if there exist two discrete variables $\bar{p}_{kh}^1, \bar{p}_{kh}^2 \in X_{k,h}^{0,1}$ and a discrete adjoint state $\bar{p}_{kh} \in X_{k,h}^{0,1}$ such that*

$$B(\bar{y}_{kh}, v_{kh}) = (\mathcal{B}\bar{u}_{kh}, v_{kh})_I, \qquad\qquad \forall v_{kh} \in X_{k,h}^{0,1},$$

$$B(\varphi_{kh}, \bar{p}_{kh}^1) = \sum_{i=1}^{N_1}\int_0^T\left(\bar{y}_{kh}(x^i) - y_{d_S}^i\right)\varphi_{kh}(x^i)dt, \quad \forall \varphi_{kh} \in X_{k,h}^{0,1},$$

$$B\left(\varphi_{kh}, \bar{p}_{kh}^2\right) = \sum_{i=1}^{N_2}\int_\Omega\left(\bar{y}_{kh}(t^i) - y_{d_T}^i\right)\varphi_{kh}(t^i)dx$$
$$+ \int_\Omega(\bar{y}_{kh}(T) - y_T)\varphi_{kh,M}dx, \qquad \forall \varphi_{kh} \in X_{k,h}^{0,1},$$

$$\bar{p}_{kh} = (2-\beta)\bar{p}_{kh}^1 + (\beta-1)\bar{p}_{kh}^2,$$

$$\int_0^T\int_\omega\left(\chi_\omega\bar{p}_{kh}(t,x) + \alpha\bar{u}_{kh}(t,x)\right)\left(v(t,x) - \bar{u}_{kh}(t,x)\right)dxdt \geq 0, \quad \forall v \in U_{ad}.$$

Remark 4.2. In our time discretization scheme we shall avoid the situation that the time observation points t^i $(i = 1, \ldots, N_2)$ and the time grid points coincide, this is because we use discontinuous test functions in time for the adjoint equation. Therefore, the Petrov-Galerkin scheme employing continuous piecewise linear states and piecewise constant test functions, proposed in [12, 34], could be a better choice when using a general time partition. We refer to [21] for the application of this scheme to space-time sparse control problems. In case that some time grid points and time observation points coincide, for example, at t^m, we may define $\int_I v d\delta(t - t^m) := v(t^m)$ for any piecewise-constant-in-time function v.

In the following we study the stability of the semi-discrete and fully discrete solutions. Let ψ be the solution of the auxiliary problem (2.1). Its time semi-discrete approximation reads: Find $\psi_k \in X_k^0$ such that

$$B(\psi_k, \phi_k) = (\varphi, \phi_k)_I, \quad \forall \phi_k \in X_k^0. \quad (4.16)$$

This problem admits a unique solution $\psi_k \in X_k^0$ (cf. [38]) and the following conclusion holds.

Lemma 4.2. *Let ψ_k be the solution of* (4.16). *Then $\psi_k \in X_k^0 \cap L^2(I, H^2(\Omega) \cap H_0^1(\Omega))$ and*

$$\|\nabla^2 \psi_k\|_I \leq C \|\Delta \psi_k\|_I, \tag{4.17}$$

$$\|\nabla \psi_k\|_I + \|\Delta \psi_k\|_I + \|\psi_k\|_{L^2(I,L^\infty(\Omega))} \leq C \|\varphi\|_{L^2(I,L^2(\Omega))}, \tag{4.18}$$

where C is a positive constant independent of k and φ.

Proof. Note that on each time interval I_m, the solution ψ_k of problem (4.16) satisfies

$$(\nabla \psi_k, \nabla \phi_k)_{I_m} + ([\psi_k]_{m-1}, \phi_{k,m-1}^+) = (\varphi, \phi_k)_{I_m}, \quad m = 1, \dots, M, \tag{4.19}$$

where $\psi_{k,0} = 0$. The above formulation can be equivalently written as

$$(\nabla \psi_{k,m}, \nabla \phi_{k,m}) + \frac{1}{k_m}(\psi_{k,m}, \phi_{k,m})$$
$$= \left(\bar{\varphi} + \frac{1}{k_m}\psi_{k,m-1}, \phi_{k,m}\right), \quad \forall \phi_{k,m} \in H_0^1(\Omega), \tag{4.20}$$

where $\bar{\varphi}|_{I_m}$ is the mean value of φ on I_m. Since Ω is a convex polygon or polyhedron and $\bar{\varphi} + \frac{1}{k_m}\psi_{k,m-1} \in L^2(\Omega)$, by the theory of elliptic regularity (cf. [20]) the solution of above equation satisfies $\psi_{k,m} \in H^2(\Omega) \cap H_0^1(\Omega)$ for $m = 1, \dots, M$, so that the first estimate holds.

As for the second estimate, the result

$$\|\nabla \psi_k\|_I + \|\Delta \psi_k\|_I \leq C \|\varphi\|_{L^2(I,L^2(\Omega))}$$

can be found in [33, Theorems 4.1 and 4.3], which, together with the embedding $H^2(\Omega) \hookrightarrow C(\bar{\Omega})$, also yields

$$\|\psi_k\|_{L^2(I,L^\infty(\Omega))} \leq C \|\varphi\|_{L^2(I,L^2(\Omega))}.$$

This completes the proof.   □

Consider the following fully discrete finite element scheme for the auxiliary problem (2.1): Find $\psi_{kh} \in X_{k,h}^{0,1}$ such that

$$B(\psi_{kh}, \phi_{kh}) = (\varphi, \phi_{kh})_I, \quad \forall \phi_{kh} \in X_{k,h}^{0,1}. \tag{4.21}$$

Let $\tilde{\Omega}$ be an open subset of Ω and let Ω_0 be a smooth open subdomain of Ω, such that $\tilde{\Omega} \subset\subset \Omega_0 \subset\subset \Omega$. Let $\{x^i\}_{i=1}^{N_1} \subset \tilde{\Omega}$ and $d = \min_{1 \leq i \leq N_1} \text{dist}(x^i, \partial\tilde{\Omega}) > 0$. The following lemma concerning the interior estimation is taken from [27, 28]. This result is useful for the a priori error estimation of the optimal control problems.

Lemma 4.3. *Let ψ and ψ_{kh} satisfy (2.1) and (4.21), respectively. There exists a constant $h_0 > 0$, independent of h, k and ψ, such that for any $2 \leq s \leq \infty$, $i = 1, \ldots, N_1$, and $h \leq h_0$, there hold*

$$\int_0^T |(\psi - \psi_{kh})(t, x^i)|^2 dt$$

$$\leq C|\text{ln}h|^2 \inf_{\chi \in X_{k,h}^{0,1}} \left\{ \|\psi - \chi\|_{L^2(I, L^\infty(\tilde{\Omega}))}^2 + h^{-\frac{4}{s}} \|\pi_k \psi - \chi\|_{L^2(I, L^s(\tilde{\Omega}))}^2 + \|\psi - \chi\|_{L^2(I, L^2(\Omega))}^2 \right.$$

$$\left. + \|\pi_k \psi - \chi\|_{L^2(I, L^2(\Omega))}^2 + h^2 \|\nabla(\psi - \chi)\|_{L^2(I, L^2(\Omega))}^2 \right\}$$

for $n = 2$, and

$$\int_0^T |(\psi - \psi_{kh})(t, x^i)|^2 dt$$

$$\leq C \left(\ln \frac{T}{k} \right)^2 |\text{ln}h|^2 \inf_{\chi \in X_{k,h}^{0,1}} \left\{ \|\psi - \chi\|_{L^2(I, L^\infty(\tilde{\Omega}))}^2 + h^{-\frac{6}{s}} \|\pi_k \psi - \chi\|_{L^2(I, L^s(\tilde{\Omega}))}^2 \right.$$

$$+ \|\psi - \chi\|_{L^2(I, L^2(\Omega))}^2 + \|\pi_k \psi - \chi\|_{L^2(I, L^2(\Omega))}^2$$

$$\left. + h^2 \|\nabla(\psi - \chi)\|_{L^2(I, L^2(\Omega))}^2 \right\}$$

for $n = 3$. Here the constant $C > 0$ depends only on d and h_0.

From [32, Corollary 5.5 and 5.11] we have the following lemma.

Lemma 4.4. *For given $\varphi \in L^\infty(I, L^2(\Omega))$, let ψ and ψ_{kh} be the solutions to (2.1) and (4.21), respectively. Then there exists a constant $C > 0$, independent of h, k and ψ, such that*

$$\|\psi - \psi_{kh}\|_{L^\infty(I, L^2(\Omega))} \leq C \left(\ln \frac{T}{k} \right)^{\frac{1}{2}} \left(k + h^2 \left(\ln \frac{T}{k} \right)^{\frac{1}{2}} \right) \|\varphi\|_{L^\infty(I, L^2(\Omega))}. \qquad (4.22)$$

5. Error estimates for the control problem

For given $u \in U_{ad}$, let $y(u)$ be the solution of the state equation (1.1). We define the auxiliary variables $\hat{p}_{kh}^i(y) \in X_{k,h}^{0,1}$, $i = 1, 2$, by

$$B\big(\varphi_{kh}, \hat{p}_{kh}^1(y)\big) = \sum_{i=1}^{N_1} \big(y(u)(x^i) - y_{d_S}^i, \varphi_{kh}(x^i)\big)_{L^2(I)}, \quad \forall \varphi_{kh} \in X_{k,h}^{0,1}, \qquad (5.1)$$

$$B\big(\varphi_{kh}, \hat{p}_{kh}^2(y)\big) = \sum_{i=1}^{N_2} \big(y(u)(t^i) - y_{d_T}^i, \varphi_{kh}(t^i)\big)$$

$$+ \big(y(u)(T) - y_T, \varphi_{kh,M}\big), \qquad \forall \varphi_{kh} \in X_{k,h}^{0,1}, \qquad (5.2)$$

respectively, and set

$$\hat{p}_{kh}(y) := (2 - \beta)\hat{p}^1_{kh}(y) + (\beta - 1)\hat{p}^2_{kh}(y). \tag{5.3}$$

In addition, we introduce another auxiliary variable $y_{kh}(u) \in X^{0,1}_{k,h}$ that solves

$$B\big(y_{kh}(u), \varphi_{kh}\big) = (\mathcal{B}u, \varphi_{kh})_I, \quad \forall \varphi_{kh} \in X^{0,1}_{k,h}. \tag{5.4}$$

Lemma 5.1. *The second order derivative of the cost functional $\hat{\mathcal{J}}_{\beta,kh}$ at any $u \in U_{ad}$ is independent of u, i.e., $\hat{\mathcal{J}}''_{\beta,kh}(u)$ is a constant operator for any $u \in U_{ad}$. Moreover, it holds*

$$\hat{\mathcal{J}}''_{\beta,kh}(u)(v, v) \geq \alpha \|v\|^2_{L^2(I,L^2(\omega))}, \quad \forall\, u, v \in U_{ad}. \tag{5.5}$$

A straightforward calculation implies the above lemma. In the following we show the stability of the fully discrete scheme (4.21).

Proposition 5.1. *For $\varphi \in L^\infty(I, L^2(\Omega))$, let ψ_{kh} be the solution of (4.21). Then there exist sufficiently small constants h_0, $k_0 > 0$ such that the stability estimates*

$$\|\psi_{kh}\|_{L^\infty(I,L^2(\Omega))} \leq C \left| \ln\frac{T}{k} \right| \|\varphi\|_{L^\infty(I,L^2(\Omega))}, \tag{5.6}$$

$$\|\psi_{kh}\|_{L^2(I,L^2(\Omega))} + \|\psi_{kh}\|_{L^2(I,L^\infty(\Omega))} \leq C\|\varphi\|_{L^2(I,L^2(\Omega))} \tag{5.7}$$

hold for $h \leq h_0$, $k \leq k_0$, where $C > 0$ is a constant independent of k and h.

Proof. Applying Proposition 2.1 and Lemma 4.4 under the conditions $k \leq k_0$ and $h \leq h_0$ for some $k_0 > 0$ and $h_0 > 0$, we obtain

$$\|\psi_{kh}\|_{L^\infty(I,L^2(\Omega))} \leq \|\psi_{kh} - \psi\|_{L^\infty(I,L^2(\Omega))} + \|\psi\|_{L^\infty(I,L^2(\Omega))}$$
$$\leq C \left| \ln\frac{T}{k} \right| \|\varphi\|_{L^\infty(I,L^2(\Omega))},$$

i.e., the first estimate holds.

The estimate for $\|\psi_{kh}\|_{L^2(I,L^2(\Omega))}$ follows similarly by using the results of [33]. The thing left is to prove estimate $\|\psi_{kh}\|_{L^2(I,L^\infty(\Omega))}$. In fact, we have

$$\|\psi_{kh}\|_{L^2(I,L^\infty(\Omega))}$$
$$\leq \|\psi_{kh} - R_h\psi_k\|_{L^2(I,L^\infty(\Omega))} + \|R_h\psi_k - \psi_k\|_{L^2(I,L^\infty(\Omega))} + \|\psi_k\|_{L^2(I,L^\infty(\Omega))}$$
$$\leq Ch^{-\frac{n}{2}}\|\psi_{kh} - R_h\psi_k\|_{L^2(I,L^2(\Omega))} + Ch^{2-\frac{n}{2}}\|\nabla^2\psi_k\|_{L^2(I,L^2(\Omega))} + \|\psi_k\|_{L^2(I,L^\infty(\Omega))}$$
$$\leq Ch^{-\frac{n}{2}}\big(\|\psi_{kh} - \psi_k\|_{L^2(I,L^2(\Omega))} + \|\psi_k - R_h\psi_k\|_{L^2(I,L^2(\Omega))}\big)$$
$$\qquad + Ch^{2-\frac{n}{2}}\|\nabla^2\psi_k\|_{L^2(I,L^2(\Omega))} + \|\psi_k\|_{L^2(I,L^\infty(\Omega))}$$
$$\leq Ch^{-\frac{n}{2}}\big(\|\psi_{kh} - \psi_k\|_{L^2(I,L^2(\Omega))} + Ch\|\nabla(\psi_k - R_h\psi_k)\|_{L^2(I,L^2(\Omega))}\big)$$
$$\qquad + Ch^{2-\frac{n}{2}}\|\nabla^2\psi_k\|_{L^2(I,L^2(\Omega))} + \|\psi_k\|_{L^2(I,L^\infty(\Omega))}$$

$$\leq Ch^{2-\frac{n}{2}} \|\nabla^2 \psi_k\|_{L^2(I,L^2(\Omega))} + \|\psi_k\|_{L^2(I,L^\infty(\Omega))}$$

$$\leq Ch^{2-\frac{n}{2}} \|\Delta \psi_k\|_{L^2(I,L^2(\Omega))} + \|\psi_k\|_{L^2(I,L^\infty(\Omega))}$$

$$\leq Ch^{2-\frac{n}{2}} \|\varphi\|_{L^2(I,L^2(\Omega))} + C\|\varphi\|_{L^2(I,L^2(\Omega))}$$

$$\leq C\|\varphi\|_{L^2(I,L^2(\Omega))},$$

where we used Lemmas 4.1 and 4.2 when $h \leq h_0$ for some $h_0 > 0$. $\qquad\square$

We also need a stability result for the fully discrete solution of (5.2).

Proposition 5.2. *Let $\hat{p}_{kh}^2(y)$ be the solution of problem* (5.2). *Then there holds*

$$\left\|\hat{p}_{kh}^2(y)\right\|_{L^\infty(I,L^2(\Omega))} \leq \sum_{i=1}^{N_2} \left\|y(u)(t^i) - y_d^i\right\| + \|y(u)(T) - y_T\|. \tag{5.8}$$

Proof. It is easy to see that

$$\left\|\hat{p}_{kh}^2(y)\right\|_{L^\infty(I,L^2(\Omega))} = \max_{1\leq m\leq M} \left\|\hat{p}_{kh}^2(y)\right\|_{L^\infty(I_m,L^2(\Omega))} = \max_{1\leq m\leq M} \left\|\hat{p}_{kh,m}^2(y)\right\|.$$

Therefore, we only need to estimate $\|\hat{p}_{kh,m}^2(y)\|$ for $m = 1, \ldots, M$ in two steps.

Step 1. We firstly consider the case $m = M$. Introduce the following auxiliary problem:

$$\begin{aligned}
\partial_t \psi - \Delta \psi &= 0 && \text{in } I_M \times \Omega, \\
\psi &= 0 && \text{on } I_M \times \partial\Omega, \\
\psi(t_{M-1}) &= \hat{p}_{kh,M}^2(y) && \text{in } \Omega,
\end{aligned}$$

where $I_M = (t_{M-1}, t_M]$. Its discrete approximation $\psi_{kh} \in X_{k,h}^{0,1}(I_M)$ is defined by

$$\begin{aligned}
(\nabla \psi_{kh}, \nabla \varphi_{kh})_{I_M} &+ (\psi_{kh,M}, \varphi_{kh,M-1}^+) \\
&= (\hat{p}_{kh,M}^2(y), \varphi_{kh,M-1}^+), \quad \forall \varphi_{kh} \in X_{k,h}^{0,1}(I_M),
\end{aligned} \tag{5.9}$$

where

$$X_{k,h}^{0,1}(I_M) := \left\{v_{kh}|_{I_M} : v_{kh} \in X_{k,h}^{0,1}\right\}.$$

Note that on the time interval I_M the solution of problem (5.2) satisfies

$$\begin{aligned}
\left(\nabla \varphi_{kh}, \nabla \hat{p}_{kh}^2(y)\right)_{I_M} &- \left(\varphi_{kh,M}, [\hat{p}_{kh}^2(y)]_M\right) \\
&= \sum_{i=1}^{N_2} \left((y(u)(t^i) - y_d^i)\chi_{I_M}(t^i), \varphi_{kh,M}\right)
\end{aligned} \tag{5.10}$$

for any $\varphi_{kh} \in X_{k,h}^{0,1}(I_M)$, where χ_{I_M} is the characteristic function of I_M. With $\hat{p}_{kh,M+1}^2(y) = y(u)(T) - y_T$, the above equation is equivalent to the following one:

$$\begin{aligned}
\left(\nabla \varphi_{kh}, \nabla \hat{p}_{kh}^2(y)\right)_{I_M} &+ \left(\varphi_{kh,M}, \hat{p}_{kh,M}^2(y)\right) \\
&= \sum_{i=1}^{N_2} \left((y(u)(t^i) - y_d^i)\chi_{I_M}(t^i), \varphi_{kh,M}\right) + \left(\varphi_{kh,M}, y(u)(T) - y_T\right).
\end{aligned} \tag{5.11}$$

We take $\varphi_{kh} = \hat{p}_{kh}^2(y)|_{I_M} = \hat{p}_{kh,M}^2(y)$ in (5.9) and $\varphi_{kh} = \psi_{kh}$ in (5.11) to obtain

$$
\begin{aligned}
\left\|\hat{p}_{kh,M}^2(y)\right\|^2 &= \left(\hat{p}_{kh,M}^2(y), \hat{p}_{kh,M-1}^{2+}(y)\right) \\
&= \left(\nabla\psi_{kh}, \nabla\hat{p}_{kh}^2(y)\right)_{I_M} + \left(\psi_{kh,M}, \hat{p}_{kh,M-1}^{2+}(y)\right) \\
&= \left(\nabla\psi_{kh}, \nabla\hat{p}_{kh}^2(y)\right)_{I_M} + \left(\psi_{kh,M}, \hat{p}_{kh,M}^2(y)\right) \\
&= \sum_{i=1}^{N_2} \left(\left(y(u)(t^i) - y_d^i\right)\chi_{I_M}(t^i), \psi_{kh,M}\right) + \left(\psi_{kh,M}, y(u)(T) - y_T\right) \\
&\leq \left(\sum_{i=1}^{N_2} \left\|(y(u)(t^i) - y_d^i)\chi_{I_M}(t^i)\right\| + \|y(u)(T) - y_T\|\right) \|\psi_{kh,M}\| \\
&\leq \left(\sum_{i=1}^{N_2} \left\|(y(u)(t^i) - y_d^i)\chi_{I_M}(t^i)\right\| + \|y(u)(T) - y_T\|\right) \left\|\hat{p}_{kh,M}^2(y)\right\|,
\end{aligned}
$$

where we have used $\|\psi_{kh,M}\| \leq \|\hat{p}_{kh,M}^2(y)\|$, which is a direct consequence of (5.9) by setting $\varphi_{kh} = \psi_{kh}$ and using Schwarz's inequality. Therefore, we obtain

$$
\left\|\hat{p}_{kh,M}^2(y)\right\| \leq \sum_{i=1}^{N_2} \left\|(y(u)(t^i) - y_d^i)\chi_{I_M}(t^i)\right\| + \|y(u)(T) - y_T\|. \tag{5.12}
$$

Step 2. Now we consider the case $m < M$. Introduce the auxiliary problem

$$
\begin{aligned}
\partial_t\psi - \Delta\psi &= 0 && \text{in } I_m \times \Omega, \\
\psi &= 0 && \text{on } I_m \times \partial\Omega, \\
\psi(t_{m-1}) &= \hat{p}_{kh,m}^2(y) && \text{in } \Omega,
\end{aligned}
$$

where $I_m = (t_{m-1}, t_m]$. Similar to Step 1, the analog of Eq. (5.9) becomes

$$
(\nabla\psi_{kh}, \nabla\varphi_{kh})_{I_m} + \left(\psi_{kh,m}, \varphi_{kh,m-1}^+\right) = \left(\hat{p}_{kh,m}^2(y), \varphi_{kh,m-1}^+\right), \quad \forall\varphi_{kh} \in X_{k,h}^{0,1}(I_m),
$$

and the analog of Eq. (5.11) becomes

$$
\begin{aligned}
&\left(\nabla\varphi_{kh}, \nabla\hat{p}_{kh}^2(y)\right)_{I_m} + (\varphi_{kh,m}, \hat{p}_{kh,m}^2(y)) \\
&= \sum_{i=1}^{N_2} \left((y(u)(t^i) - y_d^i)\chi_{I_m}(t^i), \varphi_{kh,m}\right) + \left(\varphi_{kh,m}, \hat{p}_{kh,m+1}^2(y)\right).
\end{aligned}
$$

Similarly, we can obtain

$$
\left\|\hat{p}_{kh,m}^2(y)\right\| \leq \sum_{i=1}^{N_2} \left\|(y(u)(t^i) - y_d^i)\chi_{I_m}(t^i)\right\| + \left\|\hat{p}_{kh,m+1}^2(y)\right\|.
$$

By induction for m and combining the result of Step 1, we obtain

$$\left\|\hat{p}_{kh,m}^2(y)\right\| \leq \sum_{m=1}^{M} \sum_{i=1}^{N_2} \left\|\left(y(u)(t^i) - y_d^i\right)\chi_{I_m}(t^i)\right\| + \left\|y(u)(T) - y_T\right\|$$

$$\leq \sum_{i=1}^{N_2} \left\|y(u)(t^i) - y_d^i\right\| + \left\|y(u)(T) - y_T\right\|$$

for any $1 \leq m \leq M$. Then we complete the proof. $\qquad\square$

Now we are ready to give some finite element error estimates. The first estimate is devoted to the backward parabolic equation with Dirac measure in space under the norm $\|\cdot\|_{L^2(I;L^1(\Omega))}$.

Lemma 5.2. *For given $u \in U_{ad}$, let $p^1(u)$ be the solution of (3.6) and $\hat{p}_{kh}^1(y)$ be the solution of (5.1). Then there exists a constant $C > 0$ independent of k and h, such that the following estimates:*

$$\left\|p^1(u) - \hat{p}_{kh}^1(y)\right\|_{L^2(I,L^1(\Omega))}$$

$$\leq C \left(\sum_{i=1}^{N_1} \left\|y(u)(x^i) - y_{d_S}^i\right\|_{L^2(I)}\right) |\ln h|^2 \left(k + h^2\right) \tag{5.13}$$

hold for $n = 2$, and

$$\left\|p^1(u) - \hat{p}_{kh}^1(y)\right\|_{L^2(I,L^1(\Omega))}$$

$$\leq C \left(\sum_{i=1}^{N_1} \left\|y(u)(x^i) - y_{d_S}^i\right\|_{L^2(I)}\right) \ln \frac{T}{k}|\ln h|^2 \left(k + h^2\right) \tag{5.14}$$

hold for $n = 3$.

Proof. The proof of the first assertion follows from [26–28], and the second one can be proved similarly. For the sake of completeness, we give the proof for the second assertion.

Let $e = p^1(u) - \hat{p}_{kh}^1(y)$, and ψ be the solution to problem (2.1) with right-hand side $\varphi(\cdot,\cdot) = \mathrm{sgn}(e(\cdot,\cdot))\|e(\cdot)\|_{L^1(\Omega)}$ where $\mathrm{sgn}(\cdot)$ is the sign function. The fully discrete solution $\psi_{kh} \in X_{k,h}^{0,1}$ of (2.1) is defined as

$$B(\psi_{kh}, v_{kh}) = (\varphi, v_{kh})_I, \quad \forall v_{kh} \in X_{k,h}^{0,1}.$$

Then from the orthogonality we obtain

$$\|e\|_{L^2(I,L^1(\Omega))}^2 = (e, \varphi)_{L^2(I,L^2(\Omega))} = B\left(\psi, p^1(u) - \hat{p}_{kh}^1(y)\right)$$

$$= \sum_{i=1}^{N_1} \left(y(u)(x^i) - y_{d_S}^i, \psi(x^i) - \psi_{kh}(x^i)\right)_{L^2(I)}$$

$$\leq \left(\sum_{i=1}^{N_1} \left\|y(u)(x^i) - y_{d_S}^i\right\|_{L^2(I)}\right) \sum_{i=1}^{N_1} \left\|\psi(x^i) - \psi_{kh}(x^i)\right\|_{L^2(I)}.$$

Using the second inequality of Lemma 4.3 and taking $\chi = \pi_h \pi_k \psi$, we obtain

$$\int_I \left\| \psi(t, x^i) - \psi_{kh}(t, x^i) \right\|^2 dt \le C \left(\ln \frac{T}{k} \right)^2 |\ln h|^2 (I_1 + I_2 + I_3), \quad i = 1, \ldots, N_1.$$

Here π_k is the semi-discrete L^2 projection operator from $L^2(I, H_0^1(\Omega))$ to X_k^0, and we refer to [27] for its definition and properties. Using the fact that Clément interpolation is stable with respect to L^s norm (cf. [11]), and from the interpolation estimate and the inverse estimate we conclude

$$
\begin{aligned}
I_1 &= \|\psi - \pi_h \pi_k \psi\|_{L^2(I, L^\infty(\tilde{\Omega}))}^2 + h^{-\frac{6}{s}} \|\pi_k \psi - \pi_h \pi_k \psi\|_{L^2(I, L^s(\tilde{\Omega}))}^2 \\
&\le C \left(\|\psi - \pi_h \psi\|_{L^2(I, L^\infty(\tilde{\Omega}))}^2 + h^{-\frac{6}{s}} \|\psi - \pi_k \psi\|_{L^2(I, L^s(\tilde{\Omega}))}^2 \right) \\
&\quad + C h^{-\frac{6}{s}} \left(\|\psi - \pi_h \psi\|_{L^2(I, L^s(\tilde{\Omega}))}^2 + \|\psi - \pi_k \psi\|_{L^2(I, L^s(\tilde{\Omega}))}^2 \right) \\
&\le C h^{-\frac{6}{s}} \left(k^2 + h^4 \right) \left(\|\psi\|_{L^2(I, W^{2,s}(\tilde{\Omega}))}^2 + \|\partial_t \psi\|_{L^2(I, L^s(\tilde{\Omega}))}^2 \right).
\end{aligned}
$$

By Lemma 2.1 and taking $s = |\ln h|$, we get

$$I_1 \le C |\ln h|^2 \left(k^2 + h^4 \right) \|e\|_{L^2(I, L^1(\Omega))}^2.$$

Applying the standard estimate and the estimate of [25, Lemma 3.13] for I_3 yields

$$
\begin{aligned}
I_3 &= h^2 \|\nabla(\psi - \pi_h \pi_k \psi)\|_{L^2(I, L^2(\Omega))}^2 \\
&\le C \left(h^2 \|\nabla(\psi - \pi_h \psi)\|_{L^2(I, L^2(\Omega))}^2 + h^2 \|\nabla(\psi - \pi_k \psi)\|_{L^2(I, L^2(\Omega))}^2 \right) \\
&\le C \left(h^4 + k h^2 \right) \left(\|\nabla^2 \psi\|_{L^2(I, L^2(\Omega))}^2 + \|\partial_t \psi\|_{L^2(I, L^2(\Omega))}^2 \right) \\
&\le C \left(h^4 + k^2 \right) \|e\|_{L^2(I, L^1(\Omega))}^2.
\end{aligned}
$$

We apply the standard estimate for I_2 to obtain

$$
\begin{aligned}
I_2 &\le C \left(h^4 + k^2 \right) \left(\|\nabla^2 \psi\|_{L^2(I, L^2(\Omega))}^2 + \|\partial_t \psi\|_{L^2(I, L^2(\Omega))}^2 \right) \\
&\le C \left(h^4 + k^2 \right) \|e\|_{L^2(I, L^1(\Omega))}^2.
\end{aligned}
$$

Combining the estimates for I_1, I_2 and I_3 completes the proof. $\square$

The second estimate is devoted to the backward parabolic equation with Dirac measure in space under the norm $\|\cdot\|_{L^2(I; L^2(\Omega))}$.

Lemma 5.3. *Given a control $u \in U_{ad}$, let $p_{kh}^1(u)$ be the solution of (4.10) and $\hat{p}_{kh}^1(y)$ be the solution of (5.1). Then the following estimates:*

$$\left\| p_{kh}^1(u) - \hat{p}_{kh}^1(y) \right\|_I \le C |\ln h|^2 \left(k + h^2 \right) \|u\|_{L^2(I, L^\infty(\omega))} \tag{5.15}$$

hold for $n = 2$, and

$$\left\| p_{kh}^1(u) - \hat{p}_{kh}^1(y) \right\|_I \le C \ln \frac{T}{k} |\ln h|^2 \left(k + h^2 \right) \|u\|_{L^2(I, L^\infty(\omega))} \tag{5.16}$$

hold for $n = 3$, where C is a positive constant independent of k and h.

Proof. Let $e_{kh} = p_{kh}^1(u) - \hat{p}_{kh}^1(y)$ and ψ be the solution to problem (2.1) with right-hand side e_{kh}. Let $\psi_{kh} \in X_{k,h}^{0,1}$ be the discretization of ψ satisfying

$$B(\psi - \psi_{kh}, v_{kh}) = 0, \quad \forall v_{kh} \in X_{k,h}^{0,1}.$$

Using the definitions of $p_{kh}^1(u)$ and $\hat{p}_{kh}^1(y)$ we obtain

$$\begin{aligned}
\left\| p_{kh}^1(u) - \hat{p}_{kh}^1(y) \right\|_I^2 &= B\left(\psi_{kh}, p_{kh}^1(u) - \hat{p}_{kh}^1(y)\right) \\
&= \sum_{i=1}^{N_1} \left(y_{kh}(u)(x^i) - y(u)(x^i), \psi_{kh}(x^i)\right)_{L^2(I)} \\
&\leq \sum_{i=1}^{N_1} \left\| y_{kh}(u)(x^i) - y(u)(x^i) \right\|_{L^2(I)} \|\psi_{kh}\|_{L^2(I, L^\infty(\Omega))} \\
&\leq C \sum_{i=1}^{N_1} \left\| y_{kh}(x^i) - y(x^i) \right\|_{L^2(I)} \|e_{kh}\|_I,
\end{aligned}$$

where we used Proposition 5.1. The rest of the proof is similar to that of the previous lemma. □

We also need the error estimates for the backward parabolic equation with Dirac measure in time.

Lemma 5.4. *For given $u \in U_{ad}$, let $p^2(u)$ be the solution of (3.7), $p_{kh}^2(u)$ be the solution of (4.11) and $\hat{p}_{kh}^2(y)$ be the solution of (5.2). Then there exists a constant $C > 0$, independent of k and h, such that*

$$\left\| p^2(u) - \hat{p}_{kh}^2(y) \right\|_{L^2(\Omega, L^1(I))}$$
$$\leq C \left| \ln\frac{T}{k} \right|^{\frac{1}{2}} \left(k + h^2 \left(\ln\frac{T}{k} \right)^{\frac{1}{2}} \right) \left(\sum_{i=1}^{N_2} \left\| y(u)(t^i) - y_{d_T}^i \right\| + \|y(u)(T) - y_T\| \right), \quad (5.17)$$

$$\left\| \hat{p}_{kh}^2(y) - p_{kh}^2(u) \right\|_I \leq C \left| \ln\frac{T}{k} \right| \left(k + h^2 \left(\ln\frac{T}{k} \right)^{\frac{1}{2}} \right) \|u\|_{L^\infty(I, L^2(\omega))}. \quad (5.18)$$

Proof. In order to obtain the above two estimates, we need to introduce an auxiliary problem. Given $\varphi \in L^2(I, L^2(\Omega))$, let ψ be the solution to problem (2.1) with right-hand side φ. Let ψ_{kh} be the discretization of ψ that is defined by

$$B(v_{kh}, \psi_{kh}) = (\varphi, v_{kh}), \quad \forall v_{kh} \in X_{k,h}^{0,1}.$$

By taking $\varphi = \text{sgn}(e_1(\cdot, \cdot)) \|e_1\|_{L^1(I)}$ with $e_1 = p^2(u) - \hat{p}_{kh}^2(y)$, we obtain

$$\left\| p^2(u) - \hat{p}_{kh}^2(y) \right\|_{L^2(\Omega, L^1(I))}^2$$
$$= \left(\varphi, p^2(u) - \hat{p}_{kh}^2(y)\right)_I = \left(\varphi, p^2(u)\right)_I - \left(\varphi, \hat{p}_{kh}^2(y)\right)_I$$

$$
= \left(p^2(u), \partial_t \psi - \Delta \psi\right)_I - B\left(\hat{p}_{kh}^2(y), \psi_{kh}\right)
$$

$$
= \left(p^2(u), \partial_t \psi\right)_I + \left(\nabla p^2(u), \nabla \psi\right)_I - \sum_{i=1}^{N_2} \left(y(u)(t^i) - y_{d_T}^i, \psi_{kh}(t^i)\right)
$$

$$
\quad - \left(y(u)(T) - y_T, \psi_{kh}(T)\right)
$$

$$
= \sum_{i=1}^{N_2} \left(y(u)(t^i) - y_{d_T}^i, \psi(t^i) - \psi_{kh}(t^i)\right) + \left(y(u)(T) - y_T, \psi(T) - \psi_{kh}(T)\right)
$$

$$
\leq \left(\sum_{i=1}^{N_2} \left\|y(u)(t^i) - y_{d_T}^i\right\| + \left\|y(u)(T) - y_T\right\|\right) \left\|\psi - \psi_{kh}\right\|_{L^\infty(I, L^2(\Omega))}
$$

$$
\leq C \left(\sum_{i=1}^{N_2} \left\|y(u)(t^i) - y_{d_T}^i\right\| + \left\|y(u)(T) - y_T\right\|\right) \left|\ln\frac{T}{k}\right|^{\frac{1}{2}} \left(k + h^2 \left(\ln\frac{T}{k}\right)^{\frac{1}{2}}\right) \left\|\varphi\right\|_{L^\infty(I, L^2(\Omega))}
$$

$$
\leq C \left(\sum_{i=1}^{N_2} \left\|y(u)(t^i) - y_{d_T}^i\right\| + \left\|y(u)(T) - y_T\right\|\right) \left|\ln\frac{T}{k}\right|^{\frac{1}{2}} \left(k + h^2 \left(\ln\frac{T}{k}\right)^{\frac{1}{2}}\right) \left\|e_1\right\|_{L^2(\Omega, L^1(I))}.
$$

In above estimates we used Propositions 2.1, 2.2 and Lemma 4.4.

Similarly, by taking $\varphi = e_2$ with $e_2 = \hat{p}_{kh}^2(y) - p_{kh}^2(u)$ we can also obtain

$$
\left\|\hat{p}_{kh}^2(y) - p_{kh}^2(u)\right\|_I^2
$$

$$
= \sum_{i=1}^{N_2} \left(y(u)(t^i) - y_{d_T}^i, \psi_{kh}(t^i)\right) + \left(y(u)(T) - y_T, \psi_{kh}(T)\right)
$$

$$
\quad - \sum_{i=1}^{N_2} \left(y_{kh}(u)(t^i) - y_{d_T}^i, \psi_{kh}(t^i)\right) - \left(y_{kh}(u)(T) - y_T, \psi_{kh}(T)\right)
$$

$$
= \sum_{i=1}^{N_2} \left(y(u)(t^i) - y_{kh}(u)(t^i), \psi_{kh}(t^i)\right) + \left(y(u)(T) - y_{kh}(u)(T), \psi_{kh}(T)\right)
$$

$$
\leq C \left\|y(u) - y_{kh}(u)\right\|_{L^\infty(I, L^2(\Omega))} \left\|\psi_{kh}\right\|_{L^\infty(I, L^2(\Omega))}
$$

$$
\leq C \left|\ln\frac{T}{k}\right| \left\|y(u) - y_{kh}(u)\right\|_{L^\infty(I, L^2(\Omega))} \left\|\hat{p}_{kh}^2(u) - p_{kh}^2(u)\right\|_{L^\infty(I, L^2(\Omega))}
$$

$$
\leq C \left|\ln\frac{T}{k}\right| \left\|y(u) - y_{kh}(u)\right\|_{L^\infty(I, L^2(\Omega))}^2
$$

$$
\leq C \left|\ln\frac{T}{k}\right|^2 \left(k + h^2 \left(\ln\frac{T}{k}\right)^{\frac{1}{2}}\right)^2 \left\|u\right\|_{L^\infty(I, L^2(\omega))}^2,
$$

where we have used Propositions 2.2, 5.2, 5.1, Lemma 4.4, and the pointwise constraints for the optimal control. Eventually, we obtain the second estimate. $\qquad\square$

Some simple calculations lead to the following stability result for the state equation.

Lemma 5.5. *Let $\bar{u} \in U_{ad}$ be the solution of the optimal control problem (1.3), and let $\bar{y} = y(\bar{u})$ be the corresponding optimal state. Then there exists a constant $C > 0$ such that*

$$\sum_{i=1}^{N_1} \left\| \bar{y}(x^i) - y_{d_S}^i \right\|_{L^2(I)} + \sum_{i=1}^{N_2} \left\| \bar{y}(t^i) - y_{d_T}^i \right\| + \left\| \bar{y}(T) - y_T \right\| + \left\| \bar{u} \right\|_{L^2(I,L^2(\omega))}$$

$$\leq C \left(\sum_{i=1}^{N_1} \left\| y_{d_S}^i \right\|_{L^2(I)} + \sum_{i=1}^{N_2} \left\| y_{d_T}^i \right\| + \left\| y_T \right\| \right). \tag{5.19}$$

Now we are ready to give our main result.

Theorem 5.1. *Let $\bar{u} \in U_{ad}$ and $\bar{y}$ be the solution of the optimal control problem (1.3) and the corresponding optimal state, respectively, and let $\bar{u}_{kh}$ and $\bar{y}_{kh}$ be the solution of discrete optimal control problem (4.5) and the corresponding discrete optimal state, respectively. Then there exists a constant $C > 0$, independent of k and h, such that*

$$\left\| \bar{u} - \bar{u}_{kh} \right\|_{L^2(I,L^2(\omega))} + \left\| \bar{y} - \bar{y}_{kh} \right\|_{L^2(I,L^2(\Omega))} \leq C \left| \ln h \right| \left| \ln \frac{T}{k} \right|^{\frac{3}{4}} \left(k^{\frac{1}{2}} + h \right). \tag{5.20}$$

Proof. Let $\bar{p}$ and $\bar{p}_{kh}$ be the optimal adjoint state of (1.3) and the optimal adjoint state of the discretized problem (4.5), respectively, and let $\hat{p}_{kh}(\bar{u})$ be defined in (5.3). Then we obtain

$$\alpha \left\| \bar{u} - \bar{u}_{kh} \right\|_{L^2(I,L^2(\omega))}^2 \leq \hat{\mathcal{J}}_{\beta,kh}''(\bar{u}_{kh})(\bar{u} - \bar{u}_{kh}, \bar{u} - \bar{u}_{kh})$$

$$= \hat{\mathcal{J}}_{\beta,kh}'(\bar{u})(\bar{u} - \bar{u}_{kh}) - \hat{\mathcal{J}}_{\beta,kh}'(\bar{u}_{kh})(\bar{u} - \bar{u}_{kh})$$

$$\leq \hat{\mathcal{J}}_{\beta,kh}'(\bar{u})(\bar{u} - \bar{u}_{kh}) - \hat{\mathcal{J}}_{\beta}'(\bar{u})(\bar{u} - \bar{u}_{kh})$$

$$= (\chi_\omega p_{kh}(\bar{u}), \bar{u} - \bar{u}_{kh})_{L^2(I,L^2(\omega))} - (\chi_\omega \bar{p}, \bar{u} - \bar{u}_{kh})_{L^2(I,L^2(\omega))}$$

$$= \left(\chi_\omega [p_{kh}(\bar{u}) - \hat{p}_{kh}(\bar{y})], \bar{u} - \bar{u}_{kh} \right)_{L^2(I,L^2(\omega))} + \left(\chi_\omega [\hat{p}_{kh}(\bar{y}) - \bar{p}], \bar{u} - \bar{u}_{kh} \right)_{L^2(I,L^2(\omega))}$$

$$= (2 - \beta) \left(\chi_\omega [p_{kh}^1(\bar{u}) - \hat{p}_{kh}^1(\bar{y})], \bar{u} - \bar{u}_{kh} \right)_{L^2(I,L^2(\omega))}$$

$$\quad + (\beta - 1) \left(\chi_\omega \left[p_{kh}^2(\bar{u}) - \hat{p}_{kh}^2(\bar{y}) \right], \bar{u} - \bar{u}_{kh} \right)_{L^2(I,L^2(\omega))}$$

$$\quad + (2 - \beta) \left(\chi_\omega \left[\hat{p}_{kh}^1(\bar{y}) - \bar{p}^1 \right], \bar{u} - \bar{u}_{kh} \right)_{L^2(I,L^2(\omega))}$$

$$\quad + (\beta - 1) \left(\chi_\omega \left[\hat{p}_{kh}^2(\bar{y}) - \bar{p}^2 \right], \bar{u} - \bar{u}_{kh} \right)_{L^2(I,L^2(\omega))}$$

$$\leq \left((2 - \beta) \left\| p_{kh}^1(\bar{u}) - \hat{p}_{kh}^1(\bar{y}) \right\|_I + (\beta - 1) \left\| p_{kh}^2(\bar{u}) - \hat{p}_{kh}^2(\bar{y}) \right\|_I \right) \left\| \bar{u} - \bar{u}_{kh} \right\|_{L^2(I,L^2(\omega))}$$

$$\quad + \left((2 - \beta) \left\| \hat{p}_{kh}^1(\bar{y}) - \bar{p}^1 \right\|_{L^2(I,L^1(\Omega))} + (\beta - 1) \left\| \hat{p}_{kh}^2(\bar{y}) - \bar{p}^2 \right\|_{L^2(\Omega,L^1(I))} \right)$$

$$\quad \times \left\| \bar{u} - \bar{u}_{kh} \right\|_{L^\infty(I \times \omega)},$$

where we have used the fact that

$$-\hat{\mathcal{J}}_{\beta,kh}'(\bar{u}_{kh})(\bar{u} - \bar{u}_{kh}) \leq 0 \leq -\hat{\mathcal{J}}_{\beta}'(\bar{u})(\bar{u} - \bar{u}_{kh})$$

and the estimates (5.3) and (4.12).

Due to the box constraints for the control we have $\bar{u}, \bar{u}_{kh} \in L^\infty(I \times \omega)$, and then from Lemmas 5.2-5.5 it follows

$$\|\bar{u} - \bar{u}_{kh}\|_{L^2(I,L^2(\omega))} \leq C \Big(\|p_{kh}^1(\bar{u}) - \hat{p}_{kh}^1(\bar{y})\|_I + \|p_{kh}^2(\bar{u}) - \hat{p}_{kh}^2(\bar{y})\|_I$$

$$+ \|\hat{p}_{kh}^1(\bar{y}) - \bar{p}^1\|_{L^2(I,L^1(\Omega))} + \|\hat{p}_{kh}^2(\bar{y}) - \bar{p}^2\|_{L^2(\Omega,L^1(I))} \Big)^{\frac{1}{2}}$$

$$\leq C \left| \ln \frac{T}{k} \right|^{\frac{3}{4}} |\ln h| \left(k^{\frac{1}{2}} + h \right).$$

In the following, we need to estimate the error between the discretized optimal state $\bar{y}_{kh}$ and the continuous one. Let $y_{kh}(\bar{u})$ be the solution of (5.4), we conclude from Proposition 5.1 that

$$\|\bar{y} - \bar{y}_{kh}\|_I \leq \|\bar{y} - y_{kh}(\bar{u})\|_I + \|y_{kh}(\bar{u}) - \bar{y}_{kh}\|_I$$

$$\leq C \left(\|\bar{y} - y_{kh}(\bar{u})\|_{L^2(I,L^2(\Omega))} + \|\bar{u} - \bar{u}_{kh}\|_{L^2(I,L^2(\omega))} \right)$$

$$\leq C \left| \ln \frac{T}{k} \right|^{\frac{3}{4}} |\ln h| \left(k^{\frac{1}{2}} + h \right),$$

where we have used Proposition 5.1 as well as an L^2-error estimate of finite element approximation to parabolic equations in [33]. □

6. Numerical results

In this section we provide two numerical examples for the parabolic optimal control problem (1.3) in two dimensions to verify the theoretical result in Theorem 5.1. The first example (with $\beta = 1$) corresponds to the problem with only spatial observations, while the second example (with $\beta = 2$) corresponds to the problem with only temporal observations.

For the sake of simplicity, we set $\Omega = (-1,1)^2$ and $I = (0,1)$. We choose the regularization parameter $\alpha = 0.8$, the control domain $\omega = \Omega$ and the control bounds $u_a = -0.4$, $u_b = 0.4$ in our numerical experiments, and use $N \times N$ uniform triangular spatial meshes (cf. Fig. 1) and M uniform temporal grids. We compute the errors of the
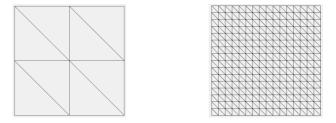

Figure 1: The spatial domain: 2×2 mesh (left) and 16×16 mesh (right).

state and control approximations, i.e. $\|\bar{y} - \bar{y}_{kh}\|_{L^2(I,L^2(\Omega))}$ and $\|\bar{u} - \bar{u}_{kh}\|_{L^2(I,L^2(\Omega))}$. Here we take the approximate solutions on the spatial and temporal meshes with $N = 64$ and $M = 128$ as our referential solutions for $(\bar{y}, \bar{u})$.

Example 6.1. We take $\beta = 1$ to investigate the order of convergence for the control and state variables corresponding to the case of spatial observations. Let $N_1 = 4$ and choose the spatial observation points $x^1 = (-\frac{1}{3}, -\frac{5}{7})$, $x^2 = (\frac{1}{2}, \frac{1}{2})$, $x^3 = (\frac{1}{2}, \frac{1}{3})$ and $x^4 = (-\frac{1}{2}, \frac{1}{2})$. The target spatial observations are chosen as $y^1_{d_S}(t) = e^t$, $y^2_{d_S}(t) = \sin t$, $y^3_{d_S}(t) = \cos t$ and $y^4_{d_S}(t) = 2t$.

We illustrate in Fig. 2 the profiles of the referential optimal state $\bar{y}$, adjoint state $\bar{p}$ and control $\bar{u}$ at time point $t = 0.5$, and can see that the optimal adjoint state variable is very singular, while the optimal state is smooth. This observation is consistent with our regularity results in Theorem 3.1.
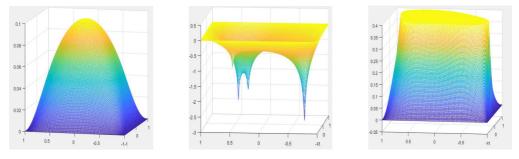

Figure 2: The profiles of $\bar{y}$, $\bar{p}$ and $\bar{u}$ at $t = 0.5$ for Example 6.1.

Tables 1 and 2 list error results of the state and control approximations to investigate the spatial accuracy with $k = \mathcal{O}(h^2)$ and the temporal accuracy with $h = \mathcal{O}(k)$, respectively. From Table 1 we can observe almost second order convergence for the state, and three halves convergence rate for the control which are better than our predicted result $\mathcal{O}(h|\ln h|)$. From Table 2 we can see first order convergence for the state and control variables, which is again better than our predicted result $\mathcal{O}(k^{1/2}|\ln \frac{T}{k}|^{3/4})$. This observation is consistent with the convergence behaviors of parabolic equations with right-hand side involving spatial Dirac measure (cf. [17, 18]).

Table 1: Convergence history of spatial discretization for Example 6.1: $M = N^2$.

N	$\|\bar{y} - \bar{y}_{kh}\|_{L^2(I,L^2(\Omega))}$	Rate	$\|\bar{u} - \bar{u}_{kh}\|_{L^2(I,L^2(\Omega))}$	Rate
2	0.0491	-	0.2054	-
4	0.0120	2.0327	0.1070	0.9408
8	0.0035	1.7776	0.0374	1.5165
16	0.0010	1.8074	0.0125	1.5811
32	2.9415e-04	1.7654	0.0033	1.9214

Table 2: Convergence history of temporal discretization for Example 6.1: $N = M$.

M	$\|\bar{y} - \bar{y}_{kh}\|_{L^2(I,L^2(\Omega))}$	Rate	$\|\bar{u} - \bar{u}_{kh}\|_{L^2(I,L^2(\Omega))}$	Rate
2	0.0306	-	0.1986	-
4	0.0172	0.8311	0.1429	0.4749
8	0.0086	1.0000	0.0764	0.9034
16	0.0043	1.0000	0.0423	0.8529
32	0.0018	1.2563	0.0189	1.1623

Example 6.2. We take $\beta = 2$ to examine the convergence behaviors of the control and state variables corresponding to the case of temporal observations. Let $N_2 = 3$ and choose the temporal observation points $t^1 = 0.3$, $t^2 = 0.4$ and $t^3 = 0.7$. The target temporal observations are chosen as

$$y_{d_T}^1(x_1, x_2) = -x_1(1 - x_2), \quad y_{d_T}^2(x_1, x_2) = -(x_1^2 + x_2^2)^{\frac{1}{2}} x_1,$$
$$y_{d_T}^3(x_1, x_2) = 2\cos x_1, \qquad y_T(x_1, x_2) = e^{x_1 + x_2}.$$

From Fig. 3 we find that the adjoint state $\bar{p}$ is smoother than the one for the case $\beta = 1$, as is consistent with the regularity result in Theorem 3.1.
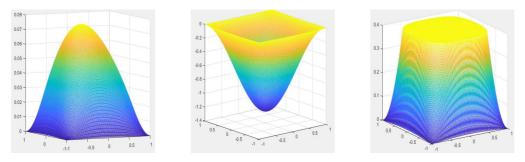

Figure 3: The profiles of $\bar{y}$, $\bar{p}$ and $\bar{u}$ at $t = 0.5$ for Example 6.2.

From Table 3 we find that the spatial accuracy is of second order for the state variable and is of first order for the control variable. The better convergence behaviour of the state than the predicted result is a common phenomenon for PDE-constrained

Table 3: Convergence history of spatial discretization for Example 6.2: $M = N^2$.

N	$\|\bar{y} - \bar{y}_{kh}\|_{L^2(I,L^2(\Omega))}$	Rate	$\|\bar{u} - \bar{u}_{kh}\|_{L^2(I,L^2(\Omega))}$	Rate
2	0.0746	-	0.4418	-
4	0.0199	1.9064	0.2701	0.7099
8	0.0052	1.9362	0.1360	0.9899
16	0.0013	2.0000	0.0640	1.0875
32	3.5966e-04	1.8538	0.0276	1.2134

Table 4: Convergence history of temporal discretization for Example 6.2: $N = M$.

M	$\|\bar{y} - \bar{y}_{kh}\|_{L^2(I, L^2(\Omega))}$	Rate	$\|\bar{u} - \bar{u}_{kh}\|_{L^2(I, L^2(\Omega))}$	Rate
2	0.0723	-	0.4831	-
4	0.0336	1.1055	0.3846	0.3290
8	0.0198	0.7630	0.2675	0.5238
16	0.0088	1.1699	0.1870	0.5165
32	0.0032	1.4594	0.1020	0.8745

optimal control problems (cf. [32, 33]), while the convergence order of the control is almost consistent with our theoretical result (cf. Theorem 5.1). From Table 4 we can also observe the first order convergence for the state variable and half an order for the control variable. The results in Tables 3 and 4 confirm the orders of convergence for the control variable for the control problems with pointwise observations in time.

Acknowledgments

References

[1] R.A. ADAMS, *Sobolev Spaces*, Academic Press, 1975.

[2] A. ALLENDES, F. FUICA, E. OTÁROLA AND D. QUERO, *An adaptive FEM for the pointwise tracking optimal control problem of the Stokes equations*, SIAM J. Sci. Comput. 41(5) (2019), A2967–A2998.

[3] A. ASHYRALYEV AND P.E. SOBOLEVSKII, *Well-Posedness of Parabolic Difference Equations*, Vol. 69, Springer, 1994.

[4] N. BEHRINGER, D. MEIDNER AND B. VEXLER, *Finite element error estimates for optimal control problems with pointwise tracking*, Pure Appl. Funct. Anal. 4(2) (2019), 177–204.

[5] A. BERMÚDEZ, P. GAMALLO AND R. RODRÍGUEZ, *Finite element methods in local active control of sound*, SIAM J. Control Optim. 43(2) (2004), 437–465.

[6] S. C. BRENNER AND L. R. SCOTT, *The Mathematical Theory of Finite Element Methods*, Springer, 2008.

[7] C. BRETT, A. DEDNER AND C. ELLIOTT, *Optimal control of elliptic PDEs at points*, IMA J. Numer. Anal. 36 (2016), 1015–1050.

[8] E. CASAS, *Pontryagin's principle for state-constrained boundary control problems of semilinear parabolic equations*, SIAM J. Control Optim. 35(4) (1997), 1297–1327.

[9] L. L. CHANG, W. GONG AND N. N. YAN, *Numerical analysis for the approximation of optimal control problems with pointwise observations*, Math. Methods Appl. Sci. 38(18) (2015), 4502–4520.

[10] P. G. CIARLET, *The Finite Element Methods for Elliptic Problems*, Vol. 4, North-Holland, 1978.

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

[12] N. VON DANIELS, M. HINZE AND M. VIERLING, *Crank-Nicolson time stepping and variational discretization of control-constrained parabolic optimal control problems*, SIAM J. Control Optim. 53(3) (2015), 1182–1198.

[13] J. ELSCHNER, J. REHBERG AND G. SCHMIDT, *Optimal regularity for elliptic transmission problems including C^1 interfaces*, Interfaces Free Bound. 9(2) (2007), 233–252.

[14] L. C. EVANS, *Partial Differential Equations*, Vol. 19, AMS, 2002.

[15] F. FUICA, E. OTÁROLA AND D. QUERO, *Error estimates for optimal control problems involving the Stokes system and Dirac measures*, Appl. Math. Optim. 84(2) (2021), 1717–1750.

[16] D. GILBARG AND N. S. TRUDINGER, *Elliptic Partial Differential Equations of Second Order*, Springer, 2001.

[17] W. GONG, *Error estimate for finite element approximations of parabolic equations with measure data*, Math. Comput. 82(281) (2013) 69–98.

[18] W. GONG, M. HINZE, AND Z. ZHOU, *A priori error analysis for finite element approximation of parabolic optimal control problems with pointwise control*, SIAM J. Control Optim. 52(1) (2014), 97–119.

[19] W. GONG AND N. N. YAN, *Finite element approximations of parabolic optimal control problems with controls acting on a lower dimensional manifold*, SIAM J. Numer. Anal. 54(2) (2016), 1229–1262.

[20] P. GRISVARD, *Elliptic Problems in Nonsmooth Domains*, Pitman Publishing Inc., 1985.

[21] E. HERBERG, M. HINZE AND H. SCHUMACHER, *Maximal discrete sparsity in parabolic optimal control with measures*, Math. Control Relat. Fields 10(4) (2020), 735–759.

[22] M. HINZE, *A variational discretization concept in control constrained optimization: the linear-quadratic case*, Comput. Optim. Appl. 30(1) (2005), 45–61.

[23] M. HINZE, R. PINNAU, M. ULBRICH AND S. ULBRICH, *Optimization with PDE Constraints*, Vol. 23, Springer, 2009.

[24] D. JERISON AND C. E. KENIG, *The inhomogeneous Dirichlet problem in Lipschitz domains*, J. Funct. Anal. 130(1) (1995), 161–219.

[25] K. KUNISCH, K. PIEPER AND B. VEXLER, *Measure valued directional sparsity for parabolic optimal control problems*, SIAM J. Control Optim. 52(5) (2014), 3078–3108.

[26] D. LEYKEKHMAN AND B. VEXLER, *Optimal a priori error estimate of parabolic optimal control problems with pointwise control*, SIAM J. Numer. Anal. 51(5) (2013), 2797–2821.

[27] D. LEYKEKHMAN AND B. VEXLER, *A priori error estimate for three dimensional parabolic optimal control problems with pointwise control*, SIAM J. Control Optim. 54(5) (2016), 2403–2435.

[28] D. LEYKEKHMAN AND B. VEXLER, *Optimal a priori error estimates of parabolic optimal control problems with a moving point control. Solvability, regularity, and optimal control of boundary value problems for PDEs*, 327–356, Springer INdAM Ser. 22, Springer, 2017.

[29] J. L. LIONS, *Optimal Control of Systems Governed by Partial Differential Equations*, Vol. 170, Springer, 1971.

[30] J. L. LIONS AND E. MAGENES, *Nonhomogeneous Boundary Value Problem and Applications*, Vol. 2, Springer, 1972.

[31] W. B. LIU AND N. N. YAN, *A posteriori error estimates for optimal control problems governed by parabolic equations*, Numer. Math. 93 (2003), 497–521.

[32] D. MEIDNER, R. RANNACHER AND B. VEXLER, *A priori error estimates for finite element*

discretizations of parabolic optimization problems with pointwise state constraints in time, SIAM J. Control Optim. 49(5) (2011), 1961–1997.

[33] D. MEIDNER AND B. VEXLER, *A priori error estimates for space-time finite element discretization of parabolic optimal control problems. Part I: problems without control constraints*, SIAM J. Control Optim. 47(3) (2008), 1150–1177.

[34] D. MEIDNER AND B. VEXLER, *A priori error analysis of the Petrov-Galerkin Crank-Nicolson scheme for parabolic optimal control problems*, SIAM J. Control Optim. 49(5) (2011), 2183–2211.

[35] R. RANNACHER AND B. VEXLER, *A priori error estimates for the finite element discretization of elliptic parameter identification problems with pointwise measurements*, SIAM J. Control Optim. 44(5) (2005), 1844–1863.

[36] E. RUDIN, *Real and Complex Analysis*, McGraw-Hill, 1966.

[37] R. E. SHOWALTER, *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*, Vol. 49, AMS, 1997.

[38] V. THOMÉE, *Galerkin Finite Element Methods for Parabolic Problems*, Springer Ser. Comput. Math. 25, Springer-Verlag, 2006.

[39] P. H. YOU AND J. X. ZHOU, *Constrained LQR problems in elliptic distributed control systems with point observations-on convergence rates*, SIAM J. Control Optim. 35(5) (1997), 1379–1754.