

## ADAPTIVE DISCONTINUOUS GALERKIN APPROXIMATION OF OPTIMAL CONTROL PROBLEMS GOVERNED BY TRANSIENT CONVECTION-DIFFUSION EQUATIONS\*

HAMDULLAH YÜCEL<sup>†</sup>, MARTIN STOLL<sup>‡</sup>, AND PETER BENNER<sup>§</sup>

**Abstract.** In this paper, we investigate a posteriori error estimates of a control-constrained optimal control problem governed by a time-dependent convection diffusion equation. The control constraints are handled by using the primal-dual active set algorithm as a semi-smooth Newton method and by adding a Moreau-Yosida-type penalty function to the cost functional. Residual-based error estimators are proposed for both approaches. The derived error estimators are used as error indicators to guide the mesh refinements. A symmetric interior penalty Galerkin method in space and a backward Euler method in time are applied in order to discretize the optimization problem. Numerical results are presented, which illustrate the performance of the proposed error estimators.

**Key words.** optimal control problem, a posteriori error estimate, discontinuous Galerkin method, convection diffusion equations

**AMS subject classifications.** 65N30, 65N50, 49J20, 65K10

**1. Introduction.** Optimal control problems (OCPs) governed by convection diffusion partial differential equations (PDEs) arise in environmental modeling, petroleum reservoir simulation, and in many other engineering applications [12, 13, 39]. Efficient numerical methods are essential for a successful application of such optimal control problems.

Several well-established techniques have been proposed to enhance stability and accuracy of the optimal control problems governed by the steady convection diffusion equation, e.g., the streamline upwind/Petrov Galerkin (SUPG) finite element method [11], the local projection stabilization [5], the edge stabilization [27, 51], and discontinuous Galerkin methods [32, 52, 53, 54, 55]. However, only few papers are published so far for unsteady optimal control problems governed by convection diffusion equations, e.g., the characteristic finite element method [14, 15], the streamline upwind/Petrov Galerkin (SUPG) finite element method [30], the local discontinuous Galerkin (LDG) method [57], the nonsymmetric interior penalty Galerkin (NIPG) method [45], and the symmetric interior penalty Galerkin (SIPG) method [2].

Adaptive finite element approximations are particularly attractive for the solution of optimal control problems governed by convection-dominated partial differential equations since the solution of the governing state PDE or the solution of the associated adjoint PDE may exhibit boundary and/or interior layers with small widths where their gradients change rapidly. When layers are not properly resolved, standard finite element discretizations lead to strong oscillations. One approach to improve the quality of a numerical solution is local mesh refinement around the layers, thereby achieving a desired residual error bound with as few degrees of freedom as possible. The literature in this area is huge, and we would like to refer to [1, 38, 48] and the references therein for more details. Adaptive finite element approximations use a posteriori error estimators or indicators to guide the mesh refinement procedure. A posteriori error estimates are computable quantities in terms of the discrete solutions without knowledge of the exact solutions. Only the area where the error estimator is

---

\*Received April 10, 2018. Accepted October 16, 2018. Published online on November 22, 2018. Recommended by Roland Herzog.

<sup>†</sup>Institute of Applied Mathematics, Middle East Technical University, 06800 Ankara, Turkey (yucelh@metu.edu.tr).

<sup>‡</sup>Technische Universität Chemnitz, Faculty of Mathematics, Reichenhainer Strasse 41, 09126 Chemnitz, Germany (martin.stoll@mathematik.tu-chemnitz.de).

<sup>§</sup>Computational Methods in Systems and Control Theory, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, 39106 Magdeburg, Germany (benner@mpi-magdeburg.mpg.de).

large will be refined so that a higher density of vertices is distributed over the area where the solutions are not well enough approximated. In this sense, the a posteriori error estimators are essential in designing algorithms to generate a mesh equidistributing the computational effort and optimizing the computation. The a posteriori error analysis of optimal control problems governed by parabolic equations is discussed in, e.g., [16, 20, 35, 36, 46]. For optimal control problems governed by time-dependent convection diffusion equations, the a posteriori error analyses are investigated by using a characteristic finite element discretization in [17] and by using edge stabilization in [56]. We also note that the error estimators proposed in [21, 27] do not contain a contribution of the residual in the control equation since the variational discretization concept proposed by Hinze [25] was used. The variational discrete optimal control can be understood as a discrete object, which is automatically discretized through a projection operator. Its structure depends on the discrete adjoint, the properties of the orthogonal projection, the Riesz isomorphism, and the control operator. Moreover, the numerical implementation of the variational discretization is not as straightforward as the full discretization; see [26, Chapter 3]. Although the variational discretization has an advantage in terms of accuracy, we believe it is still important to derive error estimates for the system with explicit discretized control. We will consider the transfer of our approach to the case of the variational control concept in a subsequent paper.

Here we investigate an a posteriori error analysis of optimal control problems governed by transient convection diffusion equations using the discontinuous Galerkin method in space and the backward Euler method in time. We apply a discontinuous Galerkin (DG) discretization for convection-dominated optimal control problems due to its better convergence behavior, local mass conservation, flexibility in approximating rough solutions on complicated meshes, and mesh adaptation. We would like to refer to [4, 23, 42] for details of the discontinuous Galerkin methods. For control-constrained optimal control problems, the resulting mapping is no longer smooth and not easy to differentiate. One has to tackle the semi-smoothness introduced by the box constraints either via the primal-dual active set method [6], which turns out to be a non-smooth Newton method, or via a regularization term such as the Moreau-Yosida regularization. Although the Moreau-Yosida regularization is a popular technique for optimal control problems with state constraints (see, e.g., [21, 22, 29, 52]), it is also used to regularize the control constraints in [40, 44]. For both strategies, we derive a posteriori error estimates in the  $L^2$ -norm for the control approximation with obstacle constraints. We need to mention that here we are not concerned with controlling the error contribution stemming from the regularization parameter. Therefore, we formally assume the Moreau-Yosida regularization parameter to be fixed in advance as done in [21, 52]. Our aim is only to understand the behavior of the a posteriori error estimate in connection with a Moreau-Yosida regularization. Then, we derive the residual-based error estimates for the regularized optimization problem up to the error involved by the Moreau-Yosida regularization. Further, by using  $L^2$ -duality techniques, some  $L^2$ -norm a posteriori estimates for the state and adjoint approximations are derived.

For time-dependent problems, adaptivity can be considered in both time and space. The spatial adaptivity can vary in time since the regions where refinement is necessary move in time. This feature of transient problems was successfully applied in [7, 19, 37, 49]. However, when there is a strong transport phenomenon, the optimal spatial mesh varies dramatically with time. Hence, there is a great need for space-time adaptivity. Such a general scheme requires recursive adaptations of the whole space-time mesh. In some optimal control problems, there exist two transport phenomena with opposite directions. This makes the computation of the problem more complex. Therefore, we focus on space adaptivity and use a fixed time step for the implicit Euler method in the analysis and in the numerical implementations.

The rest of the paper is organized as follows: in the next section, we introduce control-constrained optimal control problems governed by transient convection diffusion equations. We apply the symmetric interior penalty Galerkin (SIPG) method for the diffusion and the upwind discretization for the convection in order to discretize the optimization problem in space. The primal-dual active set strategy as a semi-smooth Newton method is also introduced to solve the optimality system. Some a posteriori error estimates for the primal-dual active set approach are derived in Section 3. The other approach to solve the control-constrained optimal control problem, the Moreau-Yosida regularization, is given in Section 4. Section 5 contains the numerical experiments to illustrate the performance of the proposed error estimators.

**2. Approximation schemes for the optimal control problem.** In this section, we introduce the discontinuous Galerkin finite element discretization in space and the backward Euler discretization in time for the approximation of the distributed linear-quadratic optimal control problems governed by unsteady convection diffusion PDEs.

We adopt the standard notations for Sobolev spaces on computational domains and their norms.  $\Omega$  and  $\Omega_U$  are bounded open sets in  $\mathbb{R}^2$  with Lipschitz boundaries  $\partial\Omega$  and  $\partial\Omega_U$ , respectively. Although adaptive finite element methods provide a real benefit on non-convex domains, for example such with reentrant corners in practical applications, we assume that  $\Omega$  and  $\Omega_U$  are convex polygons for simplicity. The inner products in  $L^2(\Omega_U)$  and  $L^2(\Omega)$  are denoted by  $(\cdot, \cdot)_U$  and  $(\cdot, \cdot)$ , respectively. Throughout the paper,  $C$  denotes a generic positive constant, and  $a \lesssim b$  means that  $a \leq Cb$  for some positive constant  $C$ . Further, we consider the Bochner spaces of functions mapping the time interval  $(0, T)$  to a Banach space  $V$  in which the norm  $\|\cdot\|_V$  is defined. For  $r \geq 1$ , we define

$$L^r(0, T; V) = \left\{ z : [0, T] \rightarrow V \text{ measurable} : \int_0^T \|z(\cdot)\|_V^r dt < \infty \right\}$$

with

$$\|z(\cdot)\|_{L^r(0, T; V)} = \begin{cases} \left( \int_0^T \|z(\cdot)\|_V^r dt \right)^{1/r}, & \text{if } 1 \leq r < \infty, \\ \text{ess sup}_{t \in (0, T)} \|z(\cdot)\|_V, & \text{if } r = \infty. \end{cases}$$

In this paper, we shall take the state space  $W = L^2(0, T; V)$  with  $V = H_0^1(\Omega)$  and the control space  $X = L^2(0, T; L^2(\Omega_U))$ . We are interested in the following distributed optimal control problem governed by a transient convection diffusion equation:

$$(2.1) \quad \min_{u \in U_{ad} \subseteq X} J(y, u) := \int_0^T \left( \frac{1}{2} \|y - y_d\|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \|u - u_d\|_{L^2(\Omega_U)}^2 \right) dt,$$

subject to

$$(2.2a) \quad \partial_t y - \epsilon \Delta y + \beta \cdot \nabla y = f + Bu, \quad x \in \Omega, \quad t \in (0, T],$$

$$(2.2b) \quad y(x, t) = 0, \quad x \in \partial\Omega, \quad t \in (0, T],$$

$$(2.2c) \quad y(x, 0) = y_0(x), \quad x \in \Omega,$$

where the closed convex admissible set of control constraints is given by

$$(2.3) \quad U_{ad} = \{u \in X : u_a \leq u \leq u_b, \text{ a.e. in } \Omega_U \times (0, T)\}$$

with the constant bounds  $u_a \leq u_b$ . The function  $u_d$ , called desired control, is a guideline for the control; see, e.g., [10]. Note that this formulation also allows for the special (and

most common) case  $u_d = 0$ , i.e., there is no a priori information on the optimal control.  $B$  is a linear continuous operator from  $L^2(\Omega_U)$  to  $L^2(\Omega)$  realizing the transition between  $\Omega_U$  and  $\Omega$ . Generally,  $\Omega_U$  can be a subset of  $\Omega$ . In the special case  $\Omega_U = \Omega$ ,  $B = I$  is the identity operator on  $L^2(\Omega)$ .

We make the following assumptions for the functions and parameters in the optimal control problem (2.1)–(2.3):

- (i) The source function  $f$ , the desired state  $y_d$ , and the desired control  $u_d$  satisfy the following regularity conditions:

$$f, y_d \in L^2(0, T; L^2(\Omega)) \quad \text{and} \quad u_d \in L^2(0, T; L^2(\Omega_U)).$$

- (ii) The initial condition is defined as  $y_0(x) \in V = H_0^1(\Omega)$ .

- (iii)  $\beta$  denotes a velocity field. It belongs to  $(W^{1,\infty}(\Omega))^2$  and satisfies the incompressibility condition, i.e.,  $\nabla \cdot \beta = 0$ . The diffusion parameter  $\epsilon$  is also taken as  $0 < \epsilon \ll 1$ .

Using the assumptions defined above, the following result on the regularity of the state solution can be stated.

**PROPOSITION 2.1** ([34]). *Under the assumptions defined above and for a given control  $u \in L^2(0, T; L^2(\Omega_U))$ , the state  $y$  satisfies the regularity condition*

$$y \in H^1(0, T; L^2(\Omega)) \cap W$$

and the weak formulation

$$\begin{aligned} (\partial_t y, v) + a(y, v) &= (f + Bu, v), \quad \forall v \in V, \\ y(x, 0) &= y_0, \end{aligned}$$

where the (bi)-linear forms are defined by

$$a(y, v) = \int_{\Omega} (\epsilon \nabla y \cdot \nabla v + \beta \cdot \nabla y v) dx, \quad (f, v) = \int_{\Omega} f v dx.$$

Then, the variational formulation corresponding to (2.1)–(2.3) can be written as

$$(2.4a) \quad \min_{u \in U_{ad}} J(y, u) := \int_0^T \left( \frac{1}{2} \|y - y_d\|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \|u - u_d\|_{L^2(\Omega_U)}^2 \right) dt$$

subject to

$$(2.4b) \quad (\partial_t y, v) + a(y, v) = (f + Bu, v), \quad \forall v \in V, \quad t \in (0, T],$$

$$(2.4c) \quad y(x, 0) = y_0,$$

$$(2.4d) \quad (y, u) \in H^1(0, T; L^2(\Omega)) \cap W \times U_{ad}.$$

It can be derived by standard techniques (see, e.g., [18] and [33]) that the control problem (2.4) has a unique solution  $(y, u)$  and that  $(y, u)$  is the solution of (2.4) if and only if there exists an adjoint  $p \in H^1(0, T; L^2(\Omega)) \cap W$  such that  $(y, u, p)$  satisfies the following optimality system for  $t \in (0, T]$ :

$$(2.5a) \quad (\partial_t y, v) + a(y, v) = (f + Bu, v), \quad \forall v \in V, \quad y(x, 0) = y_0,$$

$$(2.5b) \quad -(\partial_t p, \psi) + a(\psi, p) = (y - y_d, \psi), \quad \forall \psi \in V, \quad p(x, T) = 0,$$

$$(2.5c) \quad \int_0^T \left( \alpha(u - u_d) + B^* p, w - u \right)_U dt \geq 0, \quad \forall w \in U_{ad},$$

where  $B^*$  denotes the adjoint of  $B$ . From the second equation (2.5b), we deduce that the adjoint  $p$  satisfies the following transient convection diffusion equation:

$$(2.6a) \quad -\partial_t p - \epsilon \Delta p - \beta \cdot \nabla p = y - y_d, \quad x \in \Omega, \quad t \in (0, T],$$

$$(2.6b) \quad p(x, t) = 0, \quad x \in \partial\Omega, \quad t \in (0, T],$$

$$(2.6c) \quad p(x, T) = 0, \quad x \in \Omega.$$

The convection term in the adjoint equation (2.6) is the negative of the one in the state equation (2.2). As a consequence, errors in the solution can potentially propagate in both directions. Therefore, the numerical treatment of the state and adjoint systems together is more delicate.

**2.1. Discontinuous Galerkin (DG) scheme.** In the following, we construct the discontinuous Galerkin finite element scheme for the state equation (2.2).

Let  $\{\mathcal{T}_h\}_h$  be a family of shape-regular simplicial triangulations of  $\Omega$ . Each mesh  $\mathcal{T}_h$  consists of closed triangles such that  $\bar{\Omega} = \bigcup_{K \in \mathcal{T}_h} \bar{K}$  holds. We assume that the mesh is regular in the following sense: for different triangles  $K_i, K_j \in \mathcal{T}_h, i \neq j$ , the intersection  $K_i \cap K_j$  is either empty or a vertex or an edge, i.e., hanging nodes are not allowed. The diameter of an element  $K$  and the length of an edge  $E$  are denoted by  $h_K$  and  $h_E$ , respectively. Further, the maximum value of the element diameter is denoted by  $h = \max_{K \in \mathcal{T}_h} h_K$ .

We split the set of all edges  $\mathcal{E}_h$  into the set  $\mathcal{E}_h^0$  of interior edges and the set  $\mathcal{E}_h^\partial$  of boundary edges so that  $\mathcal{E}_h = \mathcal{E}_h^0 \cup \mathcal{E}_h^\partial$ . Let  $\mathbf{n}$  denote the unit outward normal to  $\partial\Omega$ . The inflow and outflow parts of  $\partial\Omega$  are denoted by  $\Gamma^-$  and  $\Gamma^+$ , respectively,

$$\Gamma^- = \{x \in \partial\Omega : \beta(x) \cdot \mathbf{n}(x) < 0\}, \quad \Gamma^+ = \{x \in \partial\Omega : \beta(x) \cdot \mathbf{n}(x) \geq 0\}.$$

Similarly, the inflow and outflow boundaries of an element  $K$  are defined by

$$\partial K^- = \{x \in \partial K : \beta(x) \cdot \mathbf{n}_K(x) < 0\}, \quad \partial K^+ = \{x \in \partial K : \beta(x) \cdot \mathbf{n}_K(x) \geq 0\},$$

where  $\mathbf{n}_K$  is the unit normal vector on the boundary  $\partial K$  of an element  $K$ .

Let the edge  $E$  be a common edge for two elements  $K$  and  $K^e$ . For a piecewise continuous scalar function  $y$ , there are two traces of  $y$  along  $E$ , denoted by  $y|_E$  from inside  $K$  and  $y^e|_E$  from inside  $K^e$ . The jump and average of  $y$  across the edge  $E$  are defined by

$$[[y]] = y|_E \mathbf{n}_K + y^e|_E \mathbf{n}_{K^e}, \quad \{\{y\}\} = \frac{1}{2}(y|_E + y^e|_E).$$

Similarly, for a piecewise continuous vector field  $\nabla y$ , the jump and average across an edge  $E$  are given by

$$[[\nabla y]] = \nabla y|_E \cdot \mathbf{n}_K + \nabla y^e|_E \cdot \mathbf{n}_{K^e}, \quad \{\{\nabla y\}\} = \frac{1}{2}(\nabla y|_E + \nabla y^e|_E).$$

For a boundary edge  $E \in K \cap \Gamma$ , we set  $\{\{\nabla y\}\} = \nabla y$  and  $[[y]] = y\mathbf{n}$ , where  $\mathbf{n}$  is the outward normal unit vector on  $\Gamma$ . In this paper, we only consider discontinuous linear finite element spaces to define the discrete spaces of the state and test functions

$$V_h = \{y \in L^2(\Omega) : y|_K \in \mathbb{P}^1(K), \quad \forall K \in \mathcal{T}_h\}.$$

**REMARK 2.2.** When the state equation (2.2) contains nonhomogeneous Dirichlet boundary conditions, the space of discrete states and the space of test functions can still be taken to be the same due to the weak treatment of boundary conditions for DG methods.

We now consider the discretization of the control variable. Let  $\{\mathcal{T}_h^U\}_h$  be also a family of shape-regular simplicial triangulations of  $\Omega_U$  such that  $\bar{\Omega}_U = \bigcup_{K_U \in \mathcal{T}_h^U} \bar{K}_U$  holds. For  $K_U^i, K_U^j \in \mathcal{T}_h^U, i \neq j$ , the intersection  $K_U^i \cap K_U^j$  is either empty or a vertex or an edge. The maximum diameter is defined by  $h_U = \max_{K_U \in \mathcal{T}_h^U} h_{K_U}$ , where  $h_{K_U}$  denotes the diameter of an element  $K_U$ . The discrete space of the control variable associated with  $\{\mathcal{T}_h^U\}_h$  is also a discontinuous linear finite element space

$$S_h = \{u \in L^2(\Omega_U) : u|_{K_U} \in \mathbb{P}^1(K_U), \quad \forall K_U \in \mathcal{T}_h^U\}.$$

Note that, given the reduced regularity of the control functions imposed by the control constraints, natural candidates for discretizing the control space would be the piecewise constant finite elements. Nevertheless, we use piecewise linear polynomials as is often done in the literature [2, 5, 43], resulting in the optimal convergence rate  $h_U^{3/2}$  in space for the control.

We can now give the DG discretizations of the state equation (2.2) in space for a fixed control  $u$ . The DG method proposed here is based on the upwind discretization of the convection term and on the SIPG discretization of the diffusion term. Recall that in discontinuous Galerkin methods we do not explicitly impose continuity constraints on the trial and test functions across the element interfaces. As a consequence, weak formulations include jump terms across interfaces, and penalty terms are typically added to control the jump terms. We refer to [4, 42] for a rigorous derivation of the following (bi-)linear forms applied to  $y_h \in H^1(0, T; W_h)$  for a fixed control  $u_h$  and  $\forall t \in (0, T]$ :

$$(\partial_t y_h, v_h) + a_h(y_h, v_h) = (f + B u_h, v_h), \quad \forall v_h \in V_h,$$

where

$$\begin{aligned} a_h(y, v) &= \sum_{K \in \mathcal{T}_h} \int_K \epsilon \nabla y \cdot \nabla v \, dx \\ &\quad - \sum_{E \in \mathcal{E}_h} \int_E \left( \{\{\epsilon \nabla y\}\} \cdot [v] + \{\{\epsilon \nabla v\}\} \cdot [y] \right) ds + \sum_{E \in \mathcal{E}_h} \frac{\sigma \epsilon}{h_E} \int_E [y] \cdot [v] \, ds \\ &\quad + \sum_{K \in \mathcal{T}_h} \int_K \beta \cdot \nabla y v \, dx + \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \setminus \Gamma} \beta \cdot \mathbf{n} (y^e - y) v \, ds - \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \cap \Gamma^-} \beta \cdot \mathbf{n} y v \, ds \end{aligned}$$

with the nonnegative real parameter  $\sigma$  being called the penalty parameter. We choose  $\sigma$  to be sufficiently large, independent of the mesh size  $h$ , and the diffusion coefficient  $\epsilon$  to ensure the stability of the DG discretization. However, it depends on the order of the finite element space and the position of the edge  $E$ . As a threshold, the value of the penalty parameter  $\sigma$  on the boundary edges  $E \in \mathcal{E}^\partial$  is twice the one on the interior edges  $E \in \mathcal{E}^0$  as described in [42, Section 2.7.1]. Further, large penalty parameters decrease the jumps across element interfaces, which can affect the numerical approximation. The DG approximation converges to the continuous Galerkin approximation as the penalty parameter goes to infinity; see, e.g., [8] for details.

**2.2. Primal-dual active set (PDAS) strategy.** We explain our first approach to solve the control-constrained optimal control problem (2.1)–(2.3), called the primal-dual active set (PDAS) strategy introduced in [6]. We first define the semi-discrete approximation of the

optimal control problem (2.4) as follows:

$$(2.7a) \quad \min_{u_h \in U_h^{ad}} \int_0^T \left( \frac{1}{2} \sum_{K \in \mathcal{T}_h} \|y_h - y_d\|_{L^2(K)}^2 + \frac{\alpha}{2} \sum_{K_U \in \mathcal{T}_h^U} \|u_h - u_d\|_{L^2(K_U)}^2 \right) dt,$$

subject to

$$(2.7b) \quad (\partial_t y_h, v_h) + a_h(y_h, v_h) = (f + Bu_h, v_h), \quad \forall v_h \in V_h, \quad t \in (0, T],$$

$$(2.7c) \quad y_h(x, 0) = y_h^0(x),$$

$$(2.7d) \quad (y_h, u_h) \in H^1(0, T; V_h) \times U_h^{ad},$$

where

$$(2.7e) \quad U_h^{ad} = \{u_h \in L^2(0, T; S_h) : u_a \leq u_h \leq u_b \text{ a.e. in } \Omega_U \times (0, T]\}$$

is a closed convex set in  $L^2(0, T; S_h)$ . For the ease of exposition, we also assume the condition  $U_h^{ad} \subset U_{ad} \cap L^2(0, T; S_h)$ .

Let  $J(\cdot)$  be a continuous functional in  $L^2(\Omega)$ . In addition, there exists at least one solution for the optimization problem (2.7) since the discrete state  $y_h$  can be bounded as done in [2, 45]. Then, it follows that the control problem (2.7) has a unique solution  $(y_h, u_h) \in H^1(0, T; V_h) \times U_h^{ad}$  (see, e.g., [33]) and that a pair  $(y_h, u_h)$  is the solution of (2.7) if and only if there is an adjoint  $p_h \in H^1(0, T; V_h)$  such that the triple  $(y_h, u_h, p_h)$  satisfies the following optimality system:

$$\begin{aligned} (\partial_t y_h, v_h) + a_h(y_h, v_h) &= (f + Bu_h, v_h), & \forall v_h \in V_h, \\ y_h(x, 0) &= y_h^0, \\ -(\partial_t p_h, \psi_h) + a_h(\psi_h, p_h) &= (y_h - y_d, \psi_h), & \forall \psi_h \in V_h, \\ p_h(x, T) &= 0, \\ \int_0^T (\alpha(u_h - u_d) + B^* p_h, w_h - u_h)_U dt &\geq 0, & \forall w_h \in U_h^{ad}. \end{aligned}$$

We now consider fully-discrete approximations of the optimal control problem (2.1)–(2.3) using the backward Euler scheme in time and the discontinuous Galerkin discretization in space. Let  $N_T$  be a positive integer. The discrete time interval  $\bar{I} = [0, T]$  is defined as

$$0 = t_0 < t_1 < \dots < t_{N_T-1} < t_{N_T} = T$$

with size  $k_n = t_n - t_{n-1}$ , for  $n = 1, \dots, N_T$ , and  $k = \max_{n=1, \dots, N_T} k_n$ . Then, the fully-discrete approximation scheme of the semi-discrete problem (2.7) is

$$(2.8a) \quad \min_{u_{h,n} \in U_{h,n}^{ad}} \sum_{n=1}^{N_T} k_n \left( \frac{1}{2} \sum_{K \in \mathcal{T}_h} \|y_{h,n} - y_n^d\|_{L^2(K)}^2 + \frac{\alpha}{2} \sum_{K_U \in \mathcal{T}_h^U} \|u_{h,n} - u_n^d\|_{L^2(K_U)}^2 \right),$$

subject to

$$(2.8b) \quad \left( \frac{y_{h,n} - y_{h,n-1}}{k_n}, v \right) + a_h(y_{h,n}, v) = (f_n + Bu_{h,n}, v), \quad \forall v \in V_h,$$

$$(2.8c) \quad y_{h,0}(x, 0) = y_h^0(x),$$

where

$$(2.8d) \quad U_{h,n}^{ad} = \{u_{h,n} \in S_h : u_a \leq u_{h,n} \leq u_b \text{ a.e. in } \Omega_U\}, \quad \text{for } n = 1, 2, \dots, N_T.$$

The fully discretized state equation (2.8b) has a unique solution  $y_{h,n}$  for each control variable  $u_{h,n}$  due to the boundedness of the solution; see [2, Lemma 6]. The cost functional (2.8a) is also convex in  $y_{h,n}$  and is strictly convex in  $u_{h,n}$ . We then conclude that the fully discretized control problem (2.8) has a unique solution  $(Y_{h,n}, U_{h,n}) \in V_h \times U_{h,n}^{ad}$ ,  $n = 1, 2, \dots, N_T$ , and  $(Y_{h,n}, U_{h,n})$ ,  $n = 1, 2, \dots, N_T$ , is the solution of (2.8) if and only if there is an adjoint  $P_{h,n-1} \in V_h$ ,  $i = 1, 2, \dots, N_T$ , and  $(Y_{h,n}, U_{h,n}, P_{h,n-1}) \in V_h \times U_{h,n}^{ad} \times V_h$  satisfies the following optimality system [26, 47]:

$$(2.9a) \quad \begin{aligned} \left( \frac{Y_{h,n} - Y_{h,n-1}}{k_n}, v \right) + a_h(Y_{h,n}, v) &= (f_n + BU_{h,n}, v), & \forall v \in V_h, \\ Y_{h,0} &= y_h^0, & n = 1, 2, \dots, N_T, \end{aligned}$$

$$(2.9b) \quad \begin{aligned} \left( \frac{P_{h,n-1} - P_{h,n}}{k_n}, q \right) + a_h(q, P_{h,n-1}) &= (Y_{h,n} - y_n^d, q), & \forall q \in V_h, \\ P_{h,T} &= 0, & n = N_T, \dots, 2, 1, \end{aligned}$$

$$(2.9c) \quad (\alpha(U_{h,n} - u_n^d) + B^*P_{h,n-1}, w - U_{h,n})_U \geq 0, \quad \begin{aligned} &\forall w \in U_{h,n}^{ad}, \\ &n = 1, 2, \dots, N_T. \end{aligned}$$

In the direction of time, for  $n = 1, 2, \dots, N_T$ , we define

$$\begin{aligned} Y_h|_{(t_{n-1}, t_n]} &= ((t_n - t)Y_{h,n-1} + (t - t_{n-1})Y_{h,n}) / k_n, \\ P_h|_{(t_{n-1}, t_n]} &= ((t_n - t)P_{h,n-1} + (t - t_{n-1})P_{h,n}) / k_n, \\ U_h|_{(t_{n-1}, t_n]} &= U_{h,n} \end{aligned}$$

as done in [36]. Let  $\hat{w}(x, t)|_{t \in (t_{n-1}, t_n]} = w(x, t_n)$  and  $\tilde{w}(x, t)|_{t \in (t_{n-1}, t_n]} = w(x, t_{n-1})$  for any function  $w \in C(0, T; L^2(\Omega))$ . Then, the optimality system (2.9) can be restated as

$$(2.10a) \quad \begin{aligned} \left( \frac{\partial Y_h}{\partial t}, v \right) + a_h(\hat{Y}_h, v) &= (\hat{f} + BU_h, v), & \forall v \in V_h, t \in (t_{n-1}, t_n], \\ Y_h(x, 0) &= y_h^0(x), & n = 1, 2, \dots, N_T, \end{aligned}$$

$$(2.10b) \quad \begin{aligned} - \left( \frac{\partial P_h}{\partial t}, q \right) + a_h(q, \tilde{P}_h) &= (\hat{Y}_h - \hat{y}_d, q), & \forall q \in V_h, t \in (t_{n-1}, t_n], \\ P_h(x, T) &= 0, & n = N_T, \dots, 2, 1, \end{aligned}$$

$$(2.10c) \quad (\alpha(U_h - \hat{u}_d) + B^*\tilde{P}_h, w - U_h)_U \geq 0, \quad \begin{aligned} &\forall w \in U_{h,n}^{ad}, t \in (t_{n-1}, t_n], \\ &n = 1, 2, \dots, N_T. \end{aligned}$$

We solve the optimality system (2.9) by using the primal-dual active set (PDAS) algorithm as a semi-smooth Newton method [6]. To use this approach, we first need to define the active sets

$$\begin{aligned} \mathcal{A}_-^n &= \bigcup \{x \in K_U : -B^*P_{h,n-1} - \alpha(u_a - u_n^d) < 0, \forall K_U \in \mathcal{T}_h^U\}, \\ \mathcal{A}_+^n &= \bigcup \{x \in K_U : -B^*P_{h,n-1} - \alpha(u_b - u_n^d) > 0, \forall K_U \in \mathcal{T}_h^U\}, \end{aligned}$$



and the inactive set  $\mathcal{I}^n = \mathcal{T}_h^U \setminus (\mathcal{A}_-^n \cup \mathcal{A}_+^n)$  for each time step  $t_n$ . For  $n = 1, 2, \dots, N_T$ , the discretized optimality system (2.9) is equivalent to

$$(2.11a) \quad (\mathcal{M} + k_n \mathcal{K}) Y_n - \mathcal{M} Y_{n-1} = \ell(f_n) + \mathcal{M} B U_n,$$

$$(2.11b) \quad (\mathcal{M} + k_n \mathcal{K}^T) P_{n-1} - \mathcal{M} P_n = \mathcal{M} Y_n - \ell(y_n^d),$$

$$(2.11c) \quad \alpha \mathcal{M}_U U_n - \alpha \chi_{\mathcal{I}^n} \ell(u_n^d) + \mathcal{M}_U \chi_{\mathcal{I}^{n-1}} B^* P_{n-1} = \alpha \mathcal{M}_U (\chi_{\mathcal{A}_-^n} u_a + \chi_{\mathcal{A}_+^n} u_b),$$

where  $\mathcal{K}$  is the stiffness matrix corresponding to  $a_h(\cdot, \cdot)$  and  $\mathcal{M}$  and  $\mathcal{M}_U$  are the mass matrices on the domain  $\Omega$  and  $\Omega_U$ , respectively.  $\chi_{\mathcal{A}_-^n}$ ,  $\chi_{\mathcal{A}_+^n}$ , and  $\chi_{\mathcal{I}^n}$  denote the characteristic functions of  $\mathcal{A}_-^n$ ,  $\mathcal{A}_+^n$ , and  $\mathcal{I}^n$ , respectively. The characteristic functions correspond to vectors consisting of 0 and 1. Also,  $\ell(z) = \int_{\Omega} z v \, dx$  with  $v \in V_h$ . At each time step, we then apply the active set algorithm described in Algorithm 1 for the iteration number  $m$ .

---

**Algorithm 1** Active set algorithm.

---

Choose initial values for  $y^{(\cdot,0)}$ ,  $u^{(\cdot,0)}$ , and  $p^{(\cdot,0)}$ .

Set the active sets  $\mathcal{A}_-^{(\cdot,0)}$ ,  $\mathcal{A}_+^{(\cdot,0)}$  and inactive set  $\mathcal{I}^{(\cdot,0)}$ .

**for**  $m = 1, 2, \dots$  **do**

Solve (2.11) and update the active sets  $\mathcal{A}_-^{(\cdot,m)}$ ,  $\mathcal{A}_+^{(\cdot,m)}$  and the inactive set  $\mathcal{I}^{(\cdot,m)}$  for all time steps.

**if**  $\mathcal{A}_-^{(\cdot,m)} = \mathcal{A}_-^{(\cdot,m+1)}$ ,  $\mathcal{A}_+^{(\cdot,m)} = \mathcal{A}_+^{(\cdot,m+1)}$ , and  $\mathcal{I}^{(\cdot,m)} = \mathcal{I}^{(\cdot,m+1)}$  **then**

STOP.

**end if**

**end for**

---

**3. A posteriori error estimates.** We analyse the a posteriori error estimates of the optimal control problem governed by the transient convection diffusion equation discretized by the symmetric interior penalty Galerkin scheme in space and the backward Euler scheme in time. In general, the a posteriori error analysis of unsteady optimal control problems is more complicated than the one of steady optimal control problems due to the fact that the properties of the time variable and its discretization are quite different from those of the space variables. Thus, different approaches are needed to handle the two groups of variables and their interactions.

In order to avoid over-estimation and hence derive a sharp estimator for the control, at each time level  $t_n$ ,  $n = 1, 2, \dots, N_T$ , we divide  $\Omega_U$  as follows:

$$(3.1a) \quad \Omega_U^{n,a} = \{x \in \Omega_U : (B^* P_h)(x, t_{n-1}) > \alpha(u_d(x, t_n) - u_a), U_{h,n} = u_a\},$$

$$(3.1b) \quad \Omega_U^{n,a+} = \{x \in \Omega_U : (B^* P_h)(x, t_{n-1}) > \alpha(u_d(x, t_n) - u_a), U_{h,n} > u_a\},$$

$$(3.1c) \quad \Omega_U^{n,b} = \{x \in \Omega_U : (B^* P_h)(x, t_{n-1}) < \alpha(u_d(x, t_n) - u_b), U_{h,n} = u_b\},$$

$$(3.1d) \quad \Omega_U^{n,b-} = \{x \in \Omega_U : (B^* P_h)(x, t_{n-1}) < \alpha(u_d(x, t_n) - u_b), U_{h,n} < u_b\},$$

$$(3.1e) \quad \Omega_U^{n,0} = \{x \in \Omega_U : \alpha(u_d(x, t_n) - u_b) \leq (B^* P_h)(x, t_{n-1}) \leq \alpha(u_d(x, t_n) - u_a)\}.$$

It is clear that the intersection of the above sets is empty, i.e.,  $\Omega_U^{n,i} \cap \Omega_U^{n,j} = \emptyset$ , for  $i \neq j$ ,  $i, j \in \{0, a, a+, b, b-\}$ , and

$$\Omega_U = \Omega_U^{n,0} \cup \Omega_U^{n,a} \cup \Omega_U^{n,a+} \cup \Omega_U^{n,b} \cup \Omega_U^{n,b-}.$$

To simplify the notation, set

$$\Omega_U^{n,*} = \Omega_U^{n,0} \cup \Omega_U^{n,a+} \cup \Omega_U^{n,b-}.$$

In the following lemma, we derive an estimate for the control variable by making a connection with the adjoint variable.

LEMMA 3.1. *Let  $(y, u, p)$  and  $(Y_h, U_h, P_h)$  be the solutions of (2.5) and (2.10), respectively. Then, we have the following estimate*

$$\|u - U_h\|_{L^2(0,T;L^2(\Omega_U))}^2 \leq C \left( \eta_u^2 + \|\tilde{P}_h - p(U_h)\|_{L^2(0,T;L^2(\Omega))}^2 \right),$$

where

$$\eta_u^2 = \sum_{n=1}^{N_T} \int_{t_{n-1}}^{t_n} \int_{\Omega_U^{n,*}} (\alpha(U_h - \hat{u}_d) + B^* \tilde{P}_h)^2 dx dt + \|\alpha(u_d - \hat{u}_d)\|_{L^2(0,T;L^2(\Omega_U))}^2,$$

and the auxiliary solutions, i.e.,  $y(U_h), p(U_h) \in H^1(0, T; L^2(\Omega)) \cap W$ , are defined as follows:

$$(3.2a) \quad \begin{aligned} \left( \frac{\partial}{\partial t} y(U_h), w \right) + a(y(U_h), w) &= (f + BU_h, w), & \forall w \in V, \\ y(U_h)(x, t)|_{\partial\Omega} &= 0, \quad y(U_h)(x, 0) = y_0(x), & x \in \Omega, \end{aligned}$$

$$(3.2b) \quad \begin{aligned} - \left( \frac{\partial}{\partial t} p(U_h), q \right) + a(q, p(U_h)) &= (y(U_h) - y_d, q), & \forall q \in V, \\ p(U_h)(x, t)|_{\partial\Omega} &= 0, \quad p(U_h)(x, T) = 0, & x \in \Omega. \end{aligned}$$

*Proof.* The inequality (2.5c) gives us

$$(3.3) \quad \begin{aligned} & \alpha \|u - U_h\|_{L^2(0,T;L^2(\Omega_U))}^2 \\ &= \int_0^T (\alpha u, u - U_h)_U dt - \int_0^T (\alpha U_h, u - U_h)_U dt \\ &\leq \int_0^T (\alpha u_d - B^* p, u - U_h)_U dt - \int_0^T (\alpha U_h, u - U_h)_U dt \\ &= \underbrace{\int_0^T (\alpha(U_h - \hat{u}_d) + B^* \tilde{P}_h, U_h - u)_U dt}_{M_1} + \underbrace{\int_0^T (\alpha(u_d - \hat{u}_d), u - U_h)_U dt}_{M_2} \\ &\quad + \underbrace{\int_0^T (B^*(\tilde{P}_h - p(U_h)), u - U_h)_U dt}_{M_3} + \underbrace{\int_0^T (B^*(p(U_h) - p), u - U_h)_U dt}_{M_4}. \end{aligned}$$

We first derive an estimate of  $M_1$  for any  $t \in (t_{i-1}, t_i]$ ,

$$(3.4) \quad \begin{aligned} & (\alpha(U_h - \hat{u}_d) + B^* \tilde{P}_h, U_h - u)_U \\ &= \int_{\Omega_U^{n,*}} (\alpha(U_h - \hat{u}_d) + B^* \tilde{P}_h)(U_h - u) dx \\ &\quad + \int_{\Omega_U^{n,a} \cup \Omega_U^{n,b}} (\alpha(U_h - \hat{u}_d) + B^* \tilde{P}_h)(U_h - u) dx. \end{aligned}$$

By the definitions of  $\Omega_U^{n,a}$  and  $\Omega_U^{n,b}$  in (3.1), we have

$$\begin{aligned}
 & \int_{\Omega_U^{n,a} \cup \Omega_U^{n,b}} (\alpha(U_h - \hat{u}_d) + B^* \tilde{P}_h)(U_h - u) \, dx \\
 (3.5) \quad &= \int_{\Omega_U^{n,a}} \underbrace{(\alpha(u_a - \hat{u}_d) + B^* \tilde{P}_h)}_{>0} \underbrace{(u_a - u)}_{\leq 0} \, dx \\
 &+ \int_{\Omega_U^{n,b}} \underbrace{(\alpha(u_b - \hat{u}_d) + B^* \tilde{P}_h)}_{<0} \underbrace{(u_b - u)}_{\geq 0} \, dx \leq 0.
 \end{aligned}$$

Then, with the help of (3.4), (3.5), and Young's inequality with the parameter  $\gamma > 0$ , we obtain

$$\begin{aligned}
 (3.6) \quad M_1 &\leq \int_0^T (\alpha(U_h - \hat{u}_d) + B^* \tilde{P}_h, U_h - u)_{\Omega_U^{n,*}} \, dt \\
 &\leq \frac{1}{2\gamma} \eta_u^2 + \frac{\gamma}{2} \|u - U_h\|_{L^2(0,T;L^2(\Omega_U))}^2.
 \end{aligned}$$

Next, we estimate  $M_2$  and  $M_3$  by invoking again Young's inequality with the parameter  $\gamma > 0$ ,

$$\begin{aligned}
 (3.7) \quad M_2 &= \int_0^T (\alpha(u_d - \hat{u}_d), u - U_h)_U \, dt \\
 &\leq \frac{1}{2\gamma} \|\alpha(u_d - \hat{u}_d)\|_{L^2(0,T;L^2(\Omega_U))}^2 + \frac{\gamma}{2} \|u - U_h\|_{L^2(0,T;L^2(\Omega_U))}^2,
 \end{aligned}$$

$$\begin{aligned}
 (3.8) \quad M_3 &= \int_0^T (B^*(\tilde{P}_h - p(U_h)), u - U_h)_U \, dt \\
 &\leq \frac{1}{2\gamma} \int_0^T \|B^*(\tilde{P}_h - p(U_h))\|_{L^2(\Omega_U)}^2 \, dt + \frac{\gamma}{2} \int_0^T \|u - U_h\|_{L^2(\Omega_U)}^2 \, dt \\
 &\leq \frac{1}{2\gamma} \|\tilde{P}_h - p(U_h)\|_{L^2(0,T;L^2(\Omega))}^2 + \frac{\gamma}{2} \|u - U_h\|_{L^2(0,T;L^2(\Omega_U))}^2.
 \end{aligned}$$

Finally, the auxiliary equations in (3.2) yield

$$\begin{aligned}
 M_4 &= \int_0^T (p(U_h) - p, B(u - U_h)) \, dt \\
 &= \int_0^T (\partial_t(y - y(U_h)), p(U_h) - p) \, dt + \int_0^T (a(y - y(U_h)), p(U_h) - p) \, dt \\
 &= \int_0^T (\partial_t(y - y(U_h)), p(U_h) - p) \, dt + \int_0^T (\partial_t(p(U_h) - p), y - y(U_h)) \, dt \\
 &\quad + \int_0^T (y(U_h) - y, y - y(U_h)) \, dt.
 \end{aligned}$$

An application of integration by parts on the time derivatives using that  $(y - y(U_h))|_{t=0} = 0$  and  $(p(U_h) - p)|_{t=T} = 0$  yields

$$(3.9) \quad M_4 = \int_0^T (y(U_h) - y, y - y(U_h)) \, dt \leq 0.$$

By inserting the estimates (3.6)–(3.9) of  $M_1 - M_4$  into (3.3) and setting the arbitrary positive constant  $\gamma = \alpha/2$ , we obtain the desired result.  $\square$

Before deriving error estimates for the state and adjoint equations, we need the following results for the Lagrange interpolation operator  $\Pi_h$  and the trace inequality.

LEMMA 3.2 ([9]). *Let  $\Pi_h$  be the standard Lagrange interpolation operator. For  $m = 0, 1$ ,  $q > 1$ , and  $v \in W^{2,q}(\Omega)$ , there exists a positive constant  $C$  such that*

$$|v - \Pi_h v|_{W^{m,q}(\Omega)} \leq Ch^{2-m} |v|_{W^{2,q}(\Omega)}.$$

LEMMA 3.3 ([31]). *For all  $v \in W^{1,q}(\Omega)$ ,  $1 \leq q < \infty$ ,*

$$\|v\|_{W^{0,q}(\partial K)} \leq C \left( h_K^{-1/q} \|v\|_{W^{0,q}(K)} + h_K^{1-1/q} |v|_{W^{1,q}(K)} \right).$$

Moreover, we have the following inequalities, derived in [3],

$$\|v\|_{L^2(E)}^2 \leq Ch_E^{-1} \|v\|_{L^2(K)}^2, \quad \|\mathbf{n}_E \cdot \nabla v\|_{L^2(E)}^2 \leq Ch_E^{-1} \|\nabla v\|_{L^2(K)}^2,$$

where the constant  $C$  depends on the shape regularity of the mesh. Then, the above inequalities yield the following estimation

$$(3.10) \quad \sum_{E \in \mathcal{E}_h} |h_E| \|\{\!\{ \nabla v \}\!\}\|_{L^2(E)}^2 \leq C \sum_{K \in \mathcal{T}_h} \|\nabla v\|_{L^2(K)}^2, \quad \forall v \in V_h.$$

We finally introduce the following stability results derived in [28] for the convection diffusion equations.

LEMMA 3.4 ([28]). *Assume that  $\Omega$  is a convex domain. Let  $\phi$  and  $\psi$  be the solutions of the dual (forward and backward) problems (3.11) and (3.12), respectively. Then, for given  $F \in L^2(0, T; L^2(\Omega))$ ,*

$$\begin{aligned} \|v\|_{L^\infty(0, T; L^2(\Omega))} &\lesssim \|F\|_{L^2(0, T; L^2(\Omega))}, \\ \|\nabla v\|_{L^2(0, T; L^2(\Omega))} &\lesssim \|F\|_{L^2(0, T; L^2(\Omega))}, \\ \|\Delta v\|_{L^2(0, T; L^2(\Omega))} &\lesssim \|F\|_{L^2(0, T; L^2(\Omega))}, \\ \|v_t\|_{L^2(0, T; L^2(\Omega))} &\lesssim \|F\|_{L^2(0, T; L^2(\Omega))}, \end{aligned}$$

where  $v \in \{\phi, \psi\}$  satisfies

$$(3.11a) \quad \phi_t - \epsilon \Delta \phi + \beta \cdot \nabla \phi = F, \quad (x, t) \in \Omega \times (0, T],$$

$$(3.11b) \quad \phi(x, t) = 0, \quad (x, t) \in \partial\Omega \times [0, T],$$

$$(3.11c) \quad \phi(x, 0) = 0, \quad x \in \Omega$$

or

$$(3.12a) \quad -\psi_t - \epsilon \Delta \psi - \beta \cdot \nabla \psi = F, \quad (x, t) \in \Omega \times (0, T],$$

$$(3.12b) \quad \psi(x, t) = 0, \quad (x, t) \in \partial\Omega \times [0, T],$$

$$(3.12c) \quad \psi(x, 0) = 0, \quad x \in \Omega.$$

Now, we turn to an estimation of the error  $\|P_h - p(U_h)\|_{L^2(0, T; L^2(\Omega))}^2$ .

LEMMA 3.5. *Let  $(y, u, p)$  and  $(Y_h, U_h, P_h)$  be the solutions of (2.5) and (2.10), respectively. The auxiliary solutions  $y(U_h)$  and  $p(U_h)$  are defined by the system (3.2). Assume that  $\Omega$  is a convex domain. Then,*

$$\|P_h - p(U_h)\|_{L^2(0, T; L^2(\Omega))}^2 \leq C \left( \|Y_h - y(U_h)\|_{L^2(0, T; L^2(\Omega))}^2 + \sum_{i=1}^7 \eta_i^2 \right),$$

where

$$\begin{aligned}
 \eta_1^2 &= \int_0^T \sum_{K \in \mathcal{T}_h} h_K^4 \int_K \left( \widehat{Y}_h - \widehat{y}_d + \frac{\partial P_h}{\partial t} + \epsilon \Delta \widetilde{P}_h + \beta_h \cdot \nabla \widetilde{P}_h \right)^2 dx dt, \\
 \eta_2^2 &= \int_0^T \sum_{E \in \mathcal{E}_h} h_E^3 \int_E \left[ \left[ \epsilon \nabla \widetilde{P}_h \right] \right]^2 ds dt, \\
 \eta_3^2 &= \|Y_h - \widehat{Y}_h\|_{L^2(0,T;L^2(\Omega))}^2 + \|\widehat{y}_d - y_d\|_{L^2(0,T;L^2(\Omega))}^2, \\
 \eta_4^2 &= \int_0^T \sum_{E \in \mathcal{E}_h} h_E \int_E \left[ \left[ \widetilde{P}_h \right] \right]^2 ds dt, \\
 \eta_5^2 &= \int_0^T \int_{\Omega} (|\epsilon \nabla (P_h - \widetilde{P}_h)|^2 + |\beta \cdot \nabla (P_h - \widetilde{P}_h)|^2) dx dt, \\
 \eta_6^2 &= \int_0^T \sum_{K \in \mathcal{T}_h} \int_{\partial K^+ \setminus \Gamma} h_E^3 (\beta \cdot \mathbf{n}_E \left[ \left[ \widetilde{P}_h \right] \right])^2 ds dt \\
 &\quad + \int_0^T \sum_{K \in \mathcal{T}_h} \int_{\partial K^+ \cap \Gamma^+} h_E^3 (\beta \cdot \mathbf{n}_E \widetilde{P}_h)^2 ds dt, \\
 \eta_7^2 &= \int_0^T \sum_{E \in \mathcal{E}_h} h_E^3 \int_E \left( (\epsilon + \beta \cdot \mathbf{n}_E) \left[ \left[ P_h \right] \right] \right)^2 ds dt.
 \end{aligned}$$

*Proof.* Let  $\phi$  be solution of (3.11) with  $F = P_h - p(U_h)$ . Let  $\phi_I = \Pi_h \phi$  be the Lagrange interpolation of  $\phi$  defined as in Lemma 3.2. Then, by using the adjoint equation (2.10b), the auxiliary equation (3.2), and the dual problem (3.11), we obtain

$$\begin{aligned}
 &\|P_h - p(U_h)\|_{L^2(0,T;L^2(\Omega))}^2 \\
 &= \int_0^T (P_h - p(U_h), F) dt \\
 &= \int_0^T (P_h - p(U_h), \phi_t - \epsilon \Delta \phi + \beta \cdot \nabla \phi) dt \\
 &= \int_0^T \left( -\left( \frac{\partial}{\partial t} (P_h - p(U_h)), \phi \right) + a(\phi, P_h - p(U_h)) \right) dt \\
 &\quad + \int_0^T \sum_{K \in \mathcal{T}_h} \int_{\partial K} \epsilon (p(U_h) - P_h) (\nabla \phi \cdot \mathbf{n}) ds dt \\
 &= \int_0^T \left( -\left( \frac{\partial}{\partial t} (P_h - p(U_h)), \phi - \phi_I \right) + a(\phi - \phi_I, \widetilde{P}_h - p(U_h)) \right) dt \\
 &\quad + \int_0^T \left( -\left( \frac{\partial}{\partial t} (P_h - p(U_h)), \phi_I \right) + a(\phi_I, \widetilde{P}_h - p(U_h)) + a(\phi, P_h - \widetilde{P}_h) \right) dt \\
 &\quad + \int_0^T \sum_{K \in \mathcal{T}_h} \int_{\partial K} \epsilon (p(U_h) - P_h) (\nabla \phi \cdot \mathbf{n}) ds dt \\
 &= \int_0^T \left( \left( -\frac{\partial P_h}{\partial t} - y(U_h) + y_d, \phi - \phi_I \right) + a(\phi, \widetilde{P}_h) \right) dt
 \end{aligned}$$

$$\begin{aligned}
 & + \int_0^T \sum_{K \in \mathcal{T}_h} \int_{\partial K} \epsilon (p(U_h) - P_h) (\nabla \phi \cdot \mathbf{n}) \, ds \, dt \\
 & + \int_0^T \left( -a_h(\phi_I, \tilde{P}_h) + (\hat{Y}_h - \hat{y}_d, \phi_I) - (y(U_h) - y_d, \phi_I) + a(\phi, P_h - \tilde{P}_h) \right) dt.
 \end{aligned}$$

Integrating by parts, we obtain

(3.13)

$$\begin{aligned}
 & \|P_h - p(U_h)\|_{L^2(0,T;L^2(\Omega))}^2 \\
 & = \underbrace{\int_0^T \left( -\frac{\partial P_h}{\partial t} - \epsilon \Delta \tilde{P}_h - \beta \cdot \nabla \tilde{P}_h - \hat{Y}_h + \hat{y}_d, \phi - \phi_I \right) dt}_{I_1} \\
 & + \underbrace{\int_0^T \sum_{K \in \mathcal{T}_h} \int_{\partial K} (\epsilon \nabla \tilde{P}_h \cdot \mathbf{n})(\phi - \phi_I) \, ds \, dt}_{I_2} \\
 & + \underbrace{\int_0^T (\hat{Y}_h - y(U_h) + y_d - \hat{y}_d, \phi) \, dt}_{I_3} \\
 & + \underbrace{\int_0^T \sum_{E \in \mathcal{E}_h} \int_E \left( \{\{\epsilon \nabla \tilde{P}_h\}\} \cdot [\phi_I] + \{\{\epsilon \nabla \phi_I\}\} \cdot [[\tilde{P}_h]] \right) ds \, dt}_{I_4} \\
 & - \underbrace{\int_0^T \sum_{E \in \mathcal{E}_h} \frac{\epsilon \sigma}{h_E} \int_E [[\tilde{P}_h]] [\phi_I] \, ds \, dt}_{I_5} \\
 & + \underbrace{\int_0^T \int_{\Omega} (\epsilon \nabla (P_h - \tilde{P}_h) \nabla \phi - \beta \cdot \nabla (P_h - \tilde{P}_h) \phi) \, dx \, dt}_{I_6} \\
 & + \underbrace{\int_0^T \left( \sum_{K \in \mathcal{T}_h} \int_{\partial K^+ \setminus \Gamma} \beta \cdot \mathbf{n} (\tilde{P}_h - \tilde{P}_h^e) \phi_I \, ds + \sum_{K \in \mathcal{T}_h} \int_{\partial K^+ \cap \Gamma^+} \beta \cdot \mathbf{n} \tilde{P}_h \phi_I \, ds \right) dt}_{I_7} \\
 & + \underbrace{\int_0^T \sum_{K \in \mathcal{T}_h} \int_{\partial K} (\beta \cdot \mathbf{n}) P_h \phi \, ds \, dt}_{I_8} \\
 & + \underbrace{\int_0^T \sum_{K \in \mathcal{T}_h} \int_{\partial K} \epsilon (p(U_h) - P_h) (\nabla \phi \cdot \mathbf{n}) \, ds \, dt}_{I_9}.
 \end{aligned}$$

We now estimate the terms on the right-hand side of (3.13) term by term. To estimate the first term on the right-hand side of (3.13), we use Lemma 3.2 and Lemma 3.4 with Young's

inequality such that

$$\begin{aligned}
 (3.14) \quad I_1 &\leq C(\gamma) \int_0^T \sum_{K \in \mathcal{T}_h} h_K^4 \int_K \left( \widehat{Y}_h - \widehat{y}_d + \frac{\partial P_h}{\partial t} + \epsilon \Delta \widetilde{P}_h + \beta \cdot \nabla \widetilde{P}_h \right)^2 dx dt \\
 &\quad + C\gamma \int_0^T |\phi|_{H^2(\Omega)}^2 dt \\
 &\leq C(\gamma) \eta_1^2 + C\gamma \|P_h - p(U_h)\|_{L^2(0,T;L^2(\Omega))}^2.
 \end{aligned}$$

Next, if we rewrite  $I_2$  in terms of the jump of  $\nabla \widetilde{P}_h$  and use Lemmas 3.2–3.4, we obtain

$$\begin{aligned}
 (3.15) \quad I_2 &= \int_0^T \sum_{E \in \mathcal{E}_h} \int_E \left[ [\epsilon \nabla \widetilde{P}_h] \right] (\phi - \phi_I) ds dt \\
 &\leq C(\gamma) \int_0^T \sum_{E \in \mathcal{E}_h} h_E^3 \int_E \left[ [\epsilon \nabla \widetilde{P}_h] \right]^2 ds dt + C\gamma \int_0^T |\phi|_{H^2(\Omega)}^2 dt \\
 &\leq C(\gamma) \eta_2^2 + C\gamma \|P_h - p(U_h)\|_{L^2(0,T;L^2(\Omega))}^2.
 \end{aligned}$$

Then, Lemma 3.4, Young's inequality, and the triangle inequality give us

$$\begin{aligned}
 (3.16) \quad I_3 &\leq C(\gamma) \left( \|\widehat{Y}_h - y(U_h)\|_{L^2(0,T;L^2(\Omega))}^2 + \|\widehat{y}_d - y_d\|_{L^2(0,T;L^2(\Omega))}^2 \right) \\
 &\quad + C\gamma \|\phi\|_{L^2(0,T;L^2(\Omega))}^2 \\
 &\leq C(\gamma) \left( \eta_3^2 + \|Y_h - y(U_h)\|_{L^2(0,T;L^2(\Omega))}^2 \right) + C\gamma \|P_h - p(U_h)\|_{L^2(0,T;L^2(\Omega))}^2.
 \end{aligned}$$

Similarly, using Young's inequality, the triangle inequality with respect to  $\phi_I$ , Lemma 3.3, Lemma 3.4, and the inequality in (3.10), we obtain

$$(3.17) \quad I_4 \leq C(\gamma) (\eta_2^2 + \eta_4^2) + C\gamma \|P_h - p(U_h)\|_{L^2(0,T;L^2(\Omega))}^2,$$

$$(3.18) \quad I_5 \leq C(\gamma) \eta_4^2 + C\gamma \|P_h - p(U_h)\|_{L^2(0,T;L^2(\Omega))}^2,$$

$$\begin{aligned}
 (3.19) \quad I_6 &\leq C(\gamma) \int_0^T \int_{\Omega} (|\epsilon \nabla (P_h - \widetilde{P}_h)|^2 + |\beta \cdot \nabla (P_h - \widetilde{P}_h)|^2) dx dt \\
 &\quad + C\gamma \int_0^T (|\nabla \phi|_{L^2(\Omega)}^2 + |\phi|_{L^2(\Omega)}^2) dt
 \end{aligned}$$

$$(3.20) \quad \leq C(\gamma) \eta_5^2 + C\gamma \|P_h - p(U_h)\|_{L^2(0,T;L^2(\Omega))}^2,$$

$$(3.21) \quad I_7 \leq C(\gamma) \eta_6^2 + C\gamma \|P_h - p(U_h)\|_{L^2(0,T;L^2(\Omega))}^2.$$

Finally, rewriting  $I_8$  and  $I_9$  in terms of the jump operator and using Lemma 3.3 and Lemma 3.4 with  $\llbracket p(U_h) \rrbracket = 0$ , we have

$$(3.22) \quad I_8 + I_9 \leq C(\gamma) \eta_7^2 + C\gamma \|P_h - p(U_h)\|_{L^2(0,T;L^2(\Omega))}^2.$$

By inserting (3.14)–(3.22) into (3.13) with  $\gamma$  being small enough, the desired result is obtained.  $\square$

Next, we need to find an estimate for  $\|Y_h - y(U_h)\|_{L^2(0,T;L^2(\Omega))}$ .

**LEMMA 3.6.** *Let  $(y, u, p)$  and  $(Y_h, U_h, P_h)$  be the solutions of (2.5) and (2.10), respectively. The auxiliary solutions  $y(U_h)$  and  $p(U_h)$  are defined by the system (3.2). Assume that  $\Omega$  is a convex domain, then,*

$$\|Y_h - y(U_h)\|_{L^2(0,T;L^2(\Omega))}^2 \leq C \sum_{i=8}^{15} \eta_i^2,$$

where

$$\begin{aligned}
 \eta_8^2 &= \int_0^T \sum_{K \in \mathcal{T}_h} h_K^4 \int_K \left( \widehat{f}_h - BU_h - \frac{\partial Y_h}{\partial t} + \epsilon \Delta \widehat{Y}_h - \beta_h \cdot \nabla \widehat{Y}_h \right)^2 dx dt, \\
 \eta_9^2 &= \int_0^T \sum_{E \in \mathcal{E}_h} h_E^3 \int_E \left[ \left[ \epsilon \nabla \widehat{Y}_h \right] \right]^2 ds dt, \\
 \eta_{10}^2 &= \int_0^T \sum_{E \in \mathcal{E}_h} h_E \int_E \left[ \left[ \widehat{Y}_h \right] \right]^2 ds dt, \\
 \eta_{11}^2 &= \int_0^T \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \setminus \Gamma} h_E^3 \left( \beta \cdot \mathbf{n}_E \left[ \left[ \widehat{Y}_h \right] \right] \right)^2 ds dt \\
 &\quad + \int_0^T \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \cap \Gamma^-} h_E^3 \left( \beta \cdot \mathbf{n}_E \widehat{Y}_h \right)^2 ds dt, \\
 \eta_{12}^2 &= \|\widehat{f} - f\|_{L^2(0,T;L^2(\Omega))}^2 + \|\widetilde{P}_h - P_h\|_{L^2(0,T;L^2(\Omega))}^2, \\
 \eta_{13}^2 &= \int_0^T \int_{\Omega} (|\epsilon \nabla (Y_h - \widehat{Y}_h)|^2 + |\beta \cdot \nabla (Y_h - \widehat{Y}_h)|^2) dx dt, \\
 \eta_{14}^2 &= \|Y_h(x, 0) - y_0(x)\|_{L^2(\Omega)}^2, \\
 \eta_{15}^2 &= \int_0^T \sum_{E \in \mathcal{E}_h} h_E^3 \int_E \left( (\epsilon + \beta \cdot \mathbf{n}_E) \left[ \left[ Y_h \right] \right] \right)^2 ds dt.
 \end{aligned}$$

*Proof.* Similar as before, let  $\psi$  be the solution of (3.12) with  $F = Y_h - y(U_h)$ . Let  $\psi_I = \Pi_h \psi$  be the Lagrange interpolation of  $\psi$  defined as in Lemma 3.2. Then, we conclude from (2.10a), (3.2), and (3.12) that

$$\begin{aligned}
 &\|Y_h - y(U_h)\|_{L^2(0,T;L^2(\Omega))}^2 \\
 &= \int_0^T (Y_h - y(U_h), F) dt = \int_0^T (Y_h - y(U_h), -\psi_t - \epsilon \Delta \psi - \beta \cdot \nabla \psi) dt \\
 &= \int_0^T \left( \frac{\partial Y_h}{\partial t} - \epsilon \Delta \widehat{Y}_h + \beta \cdot \nabla \widehat{Y}_h - \widehat{f}_h - BU_h, \psi - \psi_I \right) dt \\
 &\quad + \int_0^T \sum_{K \in \mathcal{T}_h} \int_{\partial K} (\epsilon \nabla \widehat{Y}_h \cdot \mathbf{n}) (\psi - \psi_I) ds dt \\
 &\quad + \int_0^T \sum_{E \in \mathcal{E}_h} \int_E \left( \left\{ \left\{ \epsilon \nabla \psi_I \right\} \right\} \cdot \left[ \left[ \widehat{Y}_h \right] \right] + \left\{ \left\{ \epsilon \nabla \widehat{Y}_h \right\} \right\} \cdot \left[ \left[ \psi_I \right] \right] \right) ds dt \\
 &\quad - \int_0^T \sum_{E \in \mathcal{E}_h} \frac{\epsilon \sigma}{h_E} \int_E \left[ \left[ \widehat{Y}_h \right] \right] \left[ \left[ \psi_I \right] \right] ds dt - \int_0^T \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \setminus \Gamma} (\beta \cdot \mathbf{n}) (\widehat{Y}_h^e - \widehat{Y}_h) \psi_I ds dt \\
 &\quad + \int_0^T \sum_{K \in \mathcal{T}_h} \int_{\partial K^- \cap \Gamma^-} (\beta \cdot \mathbf{n}) \widehat{Y}_h \psi_I ds dt \\
 &\quad + \int_0^T (\widehat{f}_h - f, \psi) dt + \int_0^T a(Y_h - \widehat{Y}_h, \psi) dt + ((Y_h - y(U_h))(x, 0), \psi(x, 0))
 \end{aligned}$$



$$+ \int_0^T \sum_{K \in \mathcal{T}_h} \int_{\partial K} (y(U_h) - Y_h) (\epsilon(\nabla \psi \cdot \mathbf{n}) + (\beta \cdot \mathbf{n}) \psi).$$

Applying the same arguments as done in (3.14)–(3.22), the desired result is obtained.  $\square$

From Lemmas 3.1, 3.5, and 3.6, the following a posteriori error estimate is derived.

**THEOREM 3.7.** *Let  $(y, u, p)$  and  $(Y_h, U_h, P_h)$  be the solutions of (2.5) and (2.10), respectively. The auxiliary solutions  $y(U_h)$  and  $p(U_h)$  are defined by the system (3.2). Assume that  $\Omega$  is a convex domain. Then,*

$$(3.23) \quad \begin{aligned} & \|u - U_h\|_{L^2(0,T;L^2(\Omega_U))}^2 + \|y - Y_h\|_{L^2(0,T;L^2(\Omega))}^2 + \|p - P_h\|_{L^2(0,T;L^2(\Omega))}^2 \\ & \leq C \left( \eta_u^2 + \sum_{i=1}^{15} \eta_i^2 \right). \end{aligned}$$

*Proof.* It follows from (2.5) and (3.2) that

$$(3.24a) \quad \|y(U_h) - y\|_{L^2(0,T;L^2(\Omega))}^2 \leq C \|u - U_h\|_{L^2(0,T;L^2(\Omega_U))}^2,$$

$$(3.24b) \quad \|p(U_h) - p\|_{L^2(0,T;L^2(\Omega))}^2 \leq C \|y(U_h) - y\|_{L^2(0,T;L^2(\Omega))}^2.$$

Lemma 3.1, 3.5, and 3.6 yield

$$(3.25) \quad \begin{aligned} & \|u - U_h\|_{L^2(0,T;L^2(\Omega_U))}^2 \leq C \left( \eta_u^2 + \|\tilde{P}_h - p(U_h)\|_{L^2(0,T;L^2(\Omega))}^2 \right) \\ & \leq C \left( \eta_u^2 + \|\tilde{P}_h - P_h\|_{L^2(0,T;L^2(\Omega))}^2 + \|P_h - p(U_h)\|_{L^2(0,T;L^2(\Omega))}^2 \right) \\ & \leq C \left( \eta_u^2 + \sum_{i=1}^{15} \eta_i^2 \right). \end{aligned}$$

Then, the desired result is obtained by applying the triangle inequality and using the inequalities (3.24)–(3.25) with Lemmas 3.1, 3.5, and 3.6.  $\square$

**4. Moreau-Yosida regularization.** Moreau-Yosida regularization is a popular technique for optimal control problems with state constraints. Some recent progress in this area has been summarised in [21, 22, 29, 52] and the references cited therein. However, it also provides challenges for the control-constrained case; see, e.g., [40, 44].

We penalize the control constraint, i.e.,  $u_a \leq u \leq u_b$ , with a Moreau-Yosida-based regularization by modifying the objective functional  $J(y, u)$  in (2.1). Now, we wish to solve the following problem:

$$(4.1) \quad \begin{aligned} & \min \int_0^T \left( \frac{1}{2} \|y - y_d\|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \|u - u_d\|_{L^2(\Omega_U)}^2 \right) dt \\ & + \frac{1}{2\delta} \int_0^T \left( \|\max\{0, u - u_b\}\|_{L^2(\Omega_U)}^2 + \|\min\{0, u - u_a\}\|_{L^2(\Omega_U)}^2 \right) dt, \end{aligned}$$

subject to

$$\begin{aligned} \partial_t y - \epsilon \Delta y + \beta \cdot \nabla y &= f + Bu, & x \in \Omega, & t \in (0, T], \\ y(x, t) &= 0, & x \in \partial\Omega, & t \in (0, T], \\ y(x, 0) &= y_0(x), & x \in \Omega, & \end{aligned}$$

where  $\delta$  is the Moreau-Yosida regularization parameter. The min- and max-expressions in the regularized objective functional arise from regularizing the indicator function corresponding to the set of admissible controls.

The unconstrained optimal control problem (4.1) has a unique solution  $(y, u) \in W \times X$  if and only if there is an adjoint  $p \in W$  such that  $(y, u, p)$  satisfies the following system for  $t \in (0, T]$ :

$$(4.2a) \quad \begin{aligned} (\partial_t y, v) + a(y, v) &= (f + Bu, v), & \forall v \in V, \\ y(x, 0) &= y_0, \end{aligned}$$

$$(4.2b) \quad \begin{aligned} -\partial_t p, \psi) + a(\psi, p) &= (y - y_d, \psi), & \forall \psi \in V, \\ p(x, T) &= 0, \end{aligned}$$

$$(4.2c) \quad \int_0^T (\alpha(u - u_d) + B^*p + \sigma) dt = 0,$$

where  $\sigma = \frac{1}{\delta} (\max\{0, u - u_b\} + \min\{0, u - u_a\})$ . Then, the fully discretized optimality system of the regularized optimal control problem (4.1) is written as

$$(4.3a) \quad \begin{aligned} \left( \frac{Y_{h,n} - Y_{h,n-1}}{k_n}, v \right) + a_h(Y_{h,n}, v) &= (f_n + BU_{h,n}, v), & \forall v \in V_h, \\ Y_{h,0} &= y_h^0, & n = 1, 2, \dots, N_T, \end{aligned}$$

$$(4.3b) \quad \begin{aligned} \left( \frac{P_{h,n-1} - P_{h,n}}{k_n}, q \right) + a_h(q, P_{h,n-1}) &= (Y_{h,n} - y_n^d, q), & \forall q \in V_h, \\ P_{h,T} &= 0, & n = N_T, \dots, 2, 1, \end{aligned}$$

$$(4.3c) \quad \alpha(U_{h,n} - u_n^d) + B^*P_{h,n-1} + \sigma_{h,n} = 0, \quad n = 1, 2, \dots, N_T,$$

where  $\sigma_{h,n} = \frac{1}{\delta} (\max\{0, U_{h,n} - u_b\} + \min\{0, U_{h,n} - u_a\})$ . As in the previous section, we restate the optimality system (4.3) as follows:

$$(4.4a) \quad \begin{aligned} \left( \frac{\partial Y_h}{\partial t}, v \right) + a_h(\widehat{Y}_h, v) &= (\widehat{f}_h + BU_h, v), & \forall v \in V_h, t \in (t_{n-1}, t_n], \\ Y_h(x, 0) &= y_h^0(x), & n = 1, 2, \dots, N_T, \end{aligned}$$

$$(4.4b) \quad \begin{aligned} \left( \frac{\partial P_h}{\partial t}, q \right) + a_h(q, \widetilde{P}_h) &= (\widehat{Y}_h - \widehat{y}_d, q), & \forall q \in V_h, t \in (t_{n-1}, t_n], \\ P_h(x, T) &= 0, & n = N_T, \dots, 2, 1, \end{aligned}$$

$$(4.4c) \quad \alpha(U_h - \widehat{u}_d) + B^*\widetilde{P}_h + \widehat{\sigma}_h = 0, \quad t \in (t_{n-1}, t_n], \quad n = 1, 2, \dots, N_T$$

with  $\widehat{\sigma}_h = \frac{1}{\delta} (\max\{0, U_h - u_b\} + \min\{0, U_h - u_a\})$ .

The optimality system (4.3) of the Moreau-Yosida approach leads to the following linear system for  $n = 1, \dots, N_T$ :

$$(4.5a) \quad (\mathcal{M} + k_n \mathcal{K}) Y_n - \mathcal{M} Y_{n-1} = \ell(f_n) + \mathcal{M} B U_n,$$

$$(4.5b) \quad (k_n \mathcal{K}^T + \mathcal{M}) P_{n-1} - \mathcal{M} P_n = \mathcal{M} Y_n - \ell(y_n^d),$$

$$(4.5c) \quad \begin{aligned} \left( \alpha \mathcal{M}_U + \frac{1}{\delta} \chi_{\mathcal{A}_n} \mathcal{M}_U \chi_{\mathcal{A}_n} \right) U_n - \alpha \ell(u_n^d) + \mathcal{M}_U B^* P_{n-1} \\ = \frac{1}{\delta} \left( \chi_{\mathcal{A}_n^a} \mathcal{M}_U \chi_{\mathcal{A}_n^a} u_a + \chi_{\mathcal{A}_n^b} \mathcal{M}_U \chi_{\mathcal{A}_n^b} u_a \right), \end{aligned}$$

where

$$\begin{aligned}\mathcal{A}_n^a &= \bigcup \{x \in K_U : U_h - u_a < 0, K_U \in \mathcal{T}_h^U\}, \\ \mathcal{A}_n^b &= \bigcup \{x \in K_U : U_h - u_b > 0, K_U \in \mathcal{T}_h^U\}, \\ \mathcal{A}_n &= \mathcal{A}_n^a \cup \mathcal{A}_n^b.\end{aligned}$$

We combine the Moreau-Yosida regularization approach with a semi-smooth Newton solver. For a fixed regularization parameter we develop a residual-based a posteriori error estimate. We note that in this work we are not interested in the error involved by the regularization. Our main aim is to understand the behavior of the error estimate in connection with the Moreau-Yosida regularization of the control constraints. Now, we derive an a posteriori error estimate for the Moreau-Yosida-regularized optimization problem (4.1) up to the error contribution stemming from the regularization.

LEMMA 4.1. *Let  $(y, u, p)$  and  $(Y_h, U_h, P_h)$  be the solutions of (4.2) and (4.4), respectively. Then, we have the following estimate*

$$(4.6) \quad \begin{aligned}\|u - U_h\|_{L^2(0,T;L^2(\Omega_U))}^2 &\leq C \left( (n_u^M)^2 + \|\tilde{P}_h - p(U_h)\|_{L^2(0,T;L^2(\Omega))}^2 \right. \\ &\quad \left. + \|\hat{\sigma}_h - \sigma\|_{L^2(0,T;L^2(\Omega))}^2 \right),\end{aligned}$$

where

$$\begin{aligned}(n_u^M)^2 &= \sum_{n=1}^{N_T} \int_{t_{n-1}}^{t_n} \int_{\Omega_U} \left( \alpha(U_h - \hat{u}_d) + B^* \tilde{P}_h + \frac{1}{\delta} (\chi_{\mathcal{A}_n^a}(U_h - u_a) + \chi_{\mathcal{A}_n^b}(U_h - u_b)) \right)^2 dx dt \\ &\quad + \|\alpha(u_d - \hat{u}_d)\|_{L^2(0,T;L^2(\Omega_U))}^2\end{aligned}$$

and the auxiliary functions, i.e.,  $y(U_h)$  and  $p(U_h)$ , are defined as in (3.2).

*Proof.* By using the inequalities (4.2c), (4.3c), and (4.4c), we obtain

$$\begin{aligned}\alpha \|u - U_h\|_{L^2(0,T;L^2(\Omega_U))}^2 &= \int_0^T (\alpha u_d - B^* p - \sigma, u - U_h)_U dt - \int_0^T (\alpha U_h, u - U_h)_U dt \\ &= \underbrace{\int_0^T (\alpha(U_h - \hat{u}_d) + B^* \tilde{P}_h + \hat{\sigma}_h, U_h - u)_U dt}_{M_1} + \underbrace{\int_0^T (\hat{\sigma}_h - \sigma, u - U_h)_U dt}_{M_2} \\ &\quad + \underbrace{\int_0^T (B^*(\tilde{P}_h - p(U_h)), u - U_h)_U dt}_{M_3} + \underbrace{\int_0^T (B^*(p(U_h) - p), u - U_h)_U dt}_{M_4} \\ &\quad + \underbrace{\int_0^T (\alpha(u_d - \hat{u}_d), u - U_h)_U dt}_{M_5}.\end{aligned}$$

Then, by following the procedure in Lemma 3.1, we can derive the desired result in (4.6).  $\square$

Similarly, we have the following a posteriori error estimate for the regularized optimization problem (4.1) from Lemma 3.5, 3.6, and 4.1.

**THEOREM 4.2.** *Let  $(y, u, p)$  and  $(Y_h, U_h, P_h)$  be the solutions of (2.5) and (4.4), respectively. The auxiliary solutions  $y(U_h)$  and  $p(U_h)$  are defined in the system (3.2). Assume that  $\Omega$  is a convex domain, then,*

$$(4.7) \quad \begin{aligned} & \|u - U_h\|_{L^2(0,T;L^2(\Omega_U))}^2 + \|y - Y_h\|_{L^2(0,T;L^2(\Omega))}^2 + \|p - P_h\|_{L^2(0,T;L^2(\Omega))}^2 \\ & \leq C \left( (\eta_u^M)^2 + \sum_{i=1}^{15} \eta_i^2 + \|\hat{\sigma}_h - \sigma\|_{L^2(0,T;L^2(\Omega))}^2 \right). \end{aligned}$$

**5. Numerical Implementation.** In this section, we present some numerical results to demonstrate the performance of the estimators proposed in Sections 3 and 4. The initial guess for the control variable is equal to zero for all discretization levels in Algorithm 1. The penalty parameter within SIPG is chosen as  $\sigma = 6$  on the interior edges and 12 on the boundary edges as taken in [42]. The Moreau-Yosida regularization parameter  $\delta$  is equal to  $10^{-6}$ . We use uniform time steps, and the time-step size is  $k = 1/50$ . Further, we take  $\Omega = \Omega_U$  and  $B = I$ . Our adaptive strategy is briefly described in Algorithm 2.

---

**Algorithm 2** Adaptive Algorithm.

**(Input)** Given an initial mesh partition  $\mathcal{T}_h$ , a refinement parameter  $\theta$ , and a tolerance parameter  $Tol$ .

**Step 1. (Solve)** Solve the optimality system (2.11) obtained by the primal-dual active set (PDAS) algorithm on the current mesh or the optimality system (4.5) obtained by the Moreau-Yosida regularization.

**Step 2. (Estimate)** Calculate the local error indicators on each element  $K$ , and then sum them over the whole space-time domain.

**Step 3. (Mark)** The edges and elements for the refinement are specified by using the a posteriori error indicator and by choosing subsets  $\mathcal{M}_K \subset \mathcal{T}_h$  such that the following bulk criterion is satisfied for the given marking parameter  $\theta$ :

$$\theta \sum_{K \in \mathcal{T}_h} (\eta_K)^2 \leq \sum_{K \in \mathcal{M}_K} (\eta_K)^2,$$

where  $\sum_{K \in \mathcal{T}_h} (\eta_K)^2$  is corresponding to the total error indicator in (3.23) for the PDAS approach and in (4.7) for the Moreau-Yosida regularization, respectively.

**Step 4. (Refine)** The marked elements are refined by longest edge bisection, where the elements of the marked edges are refined by bisection.

**Step 5.** Return to Step 1 on the new mesh to update the solutions, until the error estimators are less than the given tolerance value  $Tol$ .

---

**5.1. Example 1.** We first consider the following example given in [17] of the transport of a rotating Gaussian pulse with only a lower bound, i.e.,  $u_a = 0$ . Fu et al. use this example in their analysis of the norm-residual-based estimator in combination with a characteristic finite element approximation. The problem data are given by

$$\Omega = [-0.5, 0.5]^2, \quad T = 1, \quad \epsilon = 10^{-4}, \quad \beta = (-x_2, x_1)^T, \quad \text{and} \quad \alpha = 1.$$

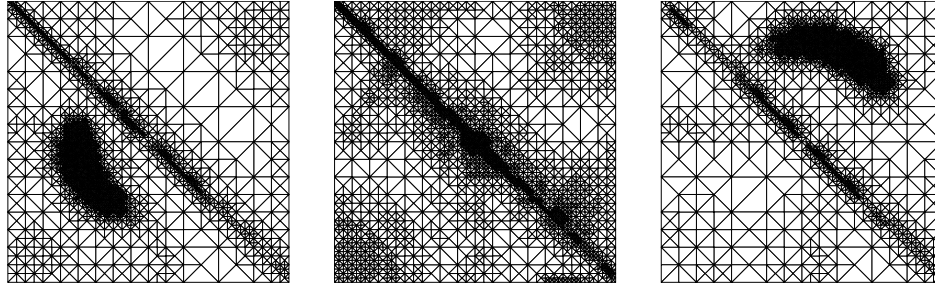


FIG. 5.1. *Example 5.1:* The adaptively refined meshes for different values of  $(x_0, y_0)$ , i.e., left:  $(-0.25, 0)$ , middle:  $(0, 0)$ , right:  $(0.25, 0.25)$ , at  $t = 1$  using the primal-dual active set strategy. The number of refinement steps and vertices are  $(9,4224)$ ,  $(10,3615)$ , and  $(9,4728)$  (from left to right) with an adaptive parameter  $\theta = 0.35$ .

The corresponding analytical solutions are given by

$$\begin{aligned}
 y(x, t) &= \frac{2\sigma_0^2}{2\sigma_0^2 + 4t\epsilon} \exp\left(-\frac{(\bar{x}_1 - x_0)^2 + (\bar{x}_2 - y_0)^2}{2\sigma_0^2 + 4t\epsilon}\right), \\
 p(x, t) &= 0, \\
 z(x, t) &= \begin{cases} 1/2, & x_1 + x_2 > 0, \\ 0, & x_1 + x_2 \leq 0, \end{cases} \\
 u_d(x, t) &= \sin(\pi t/2) \sin(\pi x_1) \sin(\pi x_2) + z(x, t), \\
 u(x, t) &= \max\left(0, u_d - \frac{p}{\alpha}\right),
 \end{aligned}$$

with a varying center point  $(x_0, y_0)$ , the standard deviation  $\sigma_0^2 = 0.0447$ , and  $\bar{x}_1 = x_1 \cos(t) + x_2 \sin(t)$ ,  $\bar{x}_2 = x_2 \cos(t) - x_1 \sin(t)$ . The source  $f$  and the desired state functions are taken as  $f = -u$  and  $y_d = y$ , respectively.

The optimal control problem exhibits a strong jump (discontinuity) introduced by the desired control  $u_d(x, t)$ . Figure 5.1 shows that a high density of vertices are distributed along  $x_1 + x_2 = 0$ . By using the a posteriori error indicators of the control variable, we pick out the discontinuity caused by the desired control  $u_d(x, t)$  and construct an adaptive mesh to obtain a better accuracy for the control with an adaptive parameter  $\theta = 0.35$ .

As the state  $y$  exhibits a different regularity, we make experiments for different values of the center point such as  $(x_0, y_0) \in \{(-0.25, 0), (0, 0), (0.25, 0.25)\}$ . For all cases, we obtain a higher density of vertices in the neighborhood of  $(\bar{x}_1, \bar{x}_2) = (x_0, y_0)$  as is observed in Figure 5.1. Table 5.1 provides the percentage of each component of the total estimator obtained by the PDAS approach for  $(x_0, y_0) = (0, 0)$  in each adaptive step. We observe that the refinement process is dominated by the contribution of the control, including the desired control  $u_d(x, t)$ .

Figure 5.2 displays the summation of the  $L^2(0, T; L^2(\Omega))$ -errors for each time step for the state, adjoint, and control variables at  $(x_0, y_0) = (0, 0)$ , obtained using the primal-dual active set strategy and the Moreau-Yosida regularization. For both approaches, the errors on adaptively refined meshes are decreasing faster than the errors on uniformly refined meshes. Although we do not present any theoretical results, the numerical results show that the Moreau-Yosida approach better captures the properties of the control compared to the PDAS approach

TABLE 5.1

*Example 5.1:* The contributions of each components of the error estimator obtained by the PDAS approach for  $(x_0, y_0) = (0, 0)$  in terms of percentage in each adaptive step.

	1	2	3	4	5	6	7	8	9	10
$\eta_u$	55.08	69.18	69.45	73.62	75.75	78.47	78.77	78.18	78.55	77.45
$\eta_1$	0.401	0.198	0.183	0.094	0.080	0.026	0.016	0.007	0.004	0.002
$\eta_2$	0.007	0.004	0.003	0.002	0.001	0.001	0.000	0.000	0.000	0.000
$\eta_3$	0.291	0.328	0.358	0.365	0.402	0.322	0.296	0.262	0.258	0.258
$\eta_4$	9.317	3.614	3.308	2.031	1.254	0.678	0.472	0.279	0.153	0.096
$\eta_5$	0.378	0.430	0.521	0.615	0.738	0.581	0.505	0.423	0.355	0.277
$\eta_6$	0.036	0.018	0.017	0.010	0.009	0.003	0.002	0.001	0.000	0.000
$\eta_7$	0.104	0.051	0.047	0.025	0.021	0.008	0.005	0.002	0.001	0.001
$\eta_8$	1.585	1.455	1.288	0.750	0.617	0.312	0.185	0.071	0.049	0.025
$\eta_9$	0.023	0.016	0.009	0.010	0.007	0.006	0.005	0.005	0.003	0.003
$\eta_{10}$	22.25	17.32	17.32	13.06	9.906	7.037	5.515	4.283	2.959	2.373
$\eta_{11}$	0.235	0.186	0.174	0.112	0.113	0.055	0.027	0.010	0.007	0.004
$\eta_{12}$	1.747	2.420	2.804	3.672	4.662	5.691	7.107	9.177	10.97	13.11
$\eta_{13}$	1.364	1.837	2.338	3.470	4.842	5.393	5.967	6.207	5.889	5.689
$\eta_{14}$	6.862	2.667	1.895	1.982	1.762	1.340	1.083	1.075	0.787	0.712
$\eta_{15}$	0.320	0.281	0.296	0.190	0.173	0.075	0.048	0.019	0.012	0.007

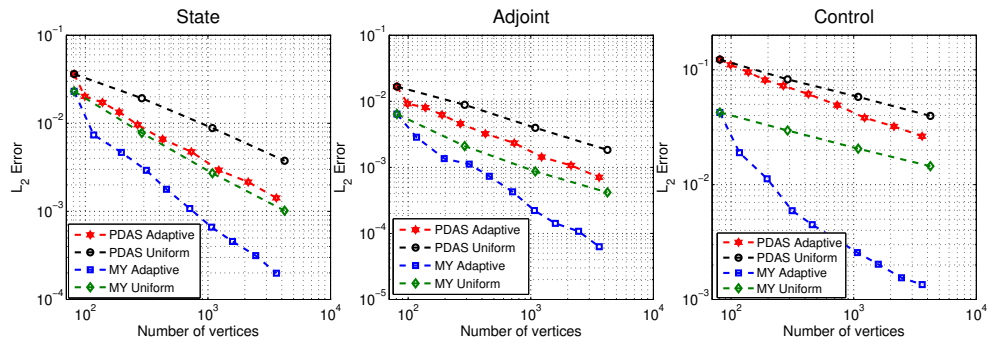


FIG. 5.2. *Example 5.1:* The global errors of the state, adjoint, and control in the  $L^2(0, T; L^2(\Omega))$ -norm with  $(x_0, y_0) = (0, 0)$ .

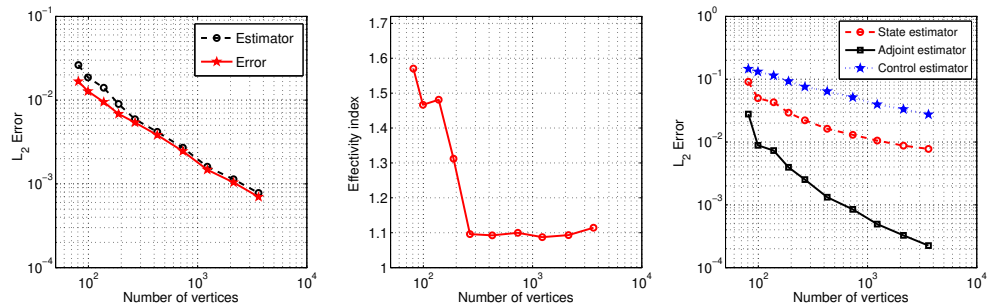


FIG. 5.3. *Example 5.1:* The convergence behavior for the a posteriori error estimate obtained by the PDAS approach for  $(x_0, y_0) = (0, 0)$ .

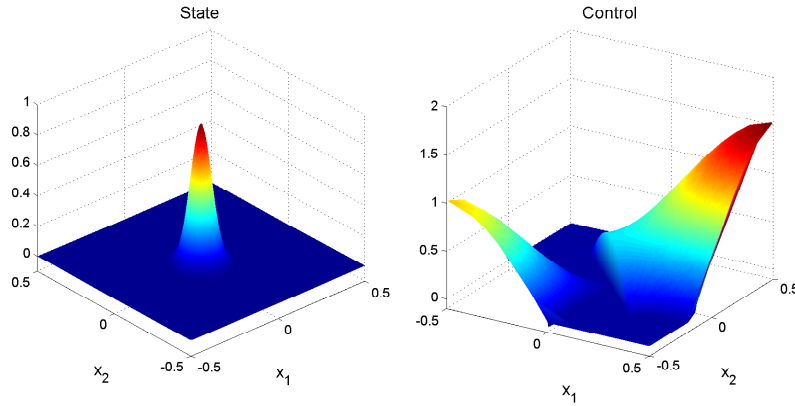


FIG. 5.4. Example 5.1: The computed state (left) and control (right) on an adaptively refined mesh with 3,618 vertices by using the Moreau-Yosida regularization for  $(x_0, y_0) = (0, 0)$  at  $t = 1$  after 10 refinement steps.

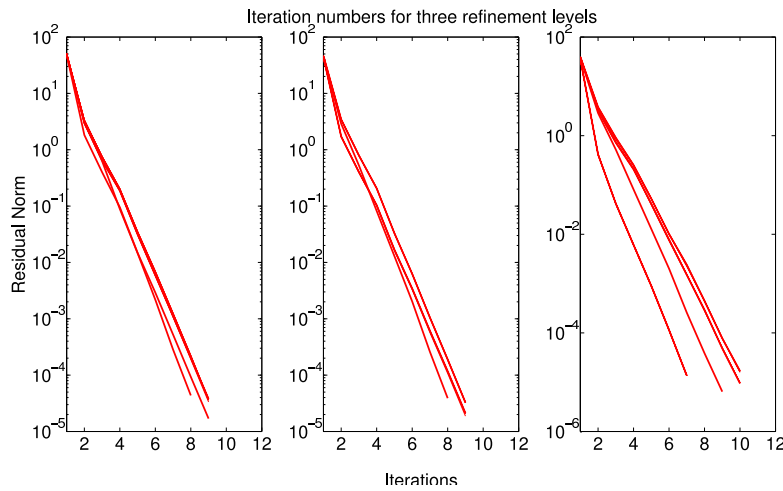


FIG. 5.5. Example 5.1: The GMRES iterations for three different refinement levels. A block-triangular preconditioner was used, and the stopping criterion is set to  $10^{-4}$  for the relative preconditioned residual.

as indicated in Figure 5.2. This observation shows that a further understanding is required. We display the performance of the indicator obtained by the PDAS approach for  $(x_0, y_0) = (0, 0)$  in each adaptive step in Figure 5.3. In the left plot of Figure 5.3, we plot the  $L^2$ -error and the value of the estimator in (3.23) against the number of vertices in each refinement step. The middle plot in Figure 5.3 displays the ratio of the estimator and the  $L^2$ -error. On the other hand, the right one in Figure 5.3 shows that the control variable dominates the refinement process as can be seen from Table 5.1. Figure 5.4 displays the computed solutions on an adaptively refined mesh with 3,618 vertices by using the Moreau-Yosida regularization for  $(x_0, y_0) = (0, 0)$  at  $t = 1$  after 10 refinement steps.

Additionally, our approach is also amendable by efficient preconditioning strategies such as the ones given in [40, 41], where an iterative method of Krylov subspace-type is combined

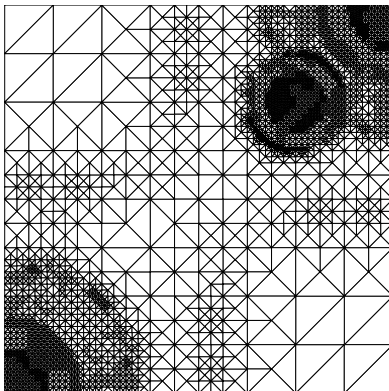


FIG. 5.6. Example 5.2: The adaptively refined mesh with 4.417 vertices at  $t = 0.5$  after 6 refinement steps with  $\theta = 0.55$  by using the primal-dual active set strategy.

with efficient and robust Schur complement approaches. Figure 5.5 displays the iteration numbers of GMRES with a block-triangular preconditioner for three consecutive stages of refinement and the associated systems within the Newton method. In Figure 5.5, each line corresponds to a nonlinear iteration.

**5.2. Example 2.** We set up our second example according to

$$\Omega = [-1, 1]^2, \quad T = 0.5, \quad \epsilon = 10^{-5}, \quad \text{and} \quad \beta = (2, 3)^T, \quad \text{and} \quad \alpha = 0.1.$$

The source function  $f$  and the desired state  $y_d$  are computed by using the following analytical solutions:

$$\begin{aligned}
 y(x, t) &= 16 \sin(\pi t) x_1 (1 - x_1) x_2 (1 - x_2) \\
 &\quad \times \left( \frac{1}{2} + \frac{1}{\pi} \arctan \left[ \frac{2}{\sqrt{\epsilon}} \left( \frac{1}{16} - \left( x_1 - \frac{1}{2} \right)^2 - \left( x_2 - \frac{1}{2} \right)^2 \right) \right] \right), \\
 p(x, t) &= 0, \\
 u_d(x, t) &= \sin(\pi t) \sin\left(\frac{\pi}{2} x_1\right) \sin\left(\frac{\pi}{2} x_2\right), \\
 u(x, t) &= \max \left( 0, \min \left( 0.5, u_d - \frac{p}{\alpha} \right) \right).
 \end{aligned}$$

The optimal state exhibits an interior layer depending on the diffusion parameter  $\epsilon$ . Also, it involves a hump changing its height in the course of the time. Figure 5.6 indicates a high density of vertices being distributed along the interior layer and the contact set. It again demonstrates that the proposed error indicators work well.

The global  $L^2(0, T; L^2(\Omega))$ -errors of the state, adjoint, and the control variables, obtained using both approaches, are given in Figure 5.7. We here only present the results of the primal-dual active set strategy on the uniform meshes for the state and adjoint since the results for both approaches are quite similar.

The performance of the indicator obtained by the PDAS approach in each adaptive step is displayed in Figure 5.8. Although the effectivity index is not close to one as in the previous example, it does not oscillate after a few iterations. The right plot in Figure 5.8 and Table 5.2 show that the refinement process is dominated by the contribution of the state.



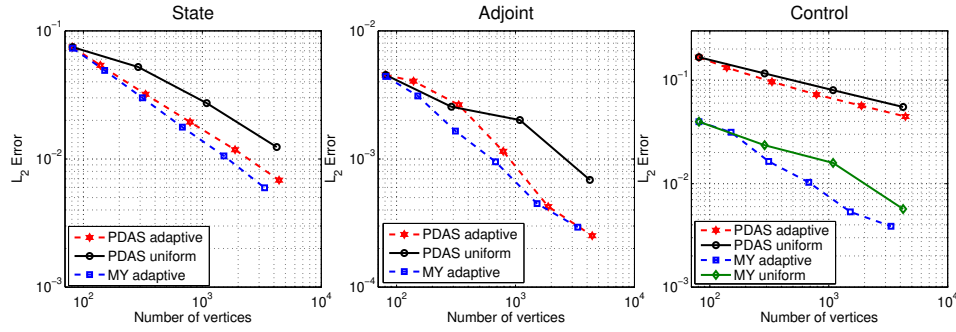


FIG. 5.7. Example 5.2: The global errors of the state, adjoint, and control in the  $L^2(0, T; L^2(\Omega))$ -norm.

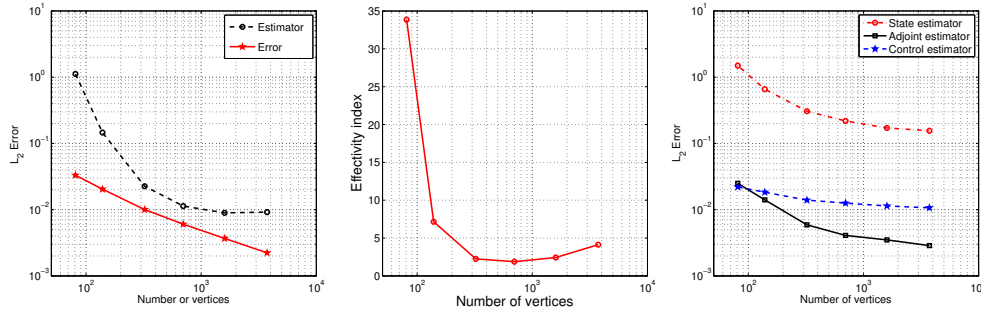


FIG. 5.8. Example 5.2: The convergence behavior for the a posteriori error estimate obtained by the PDAS approach.

TABLE 5.2

Example 5.2: The contributions of each components of the error estimator obtained by the PDAS approach in terms of percentage in each adaptive step.

	1	2	3	4	5	6
$\eta_u$	1.342	2.290	3.434	3.825	3.634	3.538
$\eta_1$	0.621	0.476	0.247	0.100	0.027	0.013
$\eta_2$	0.000	0.000	0.000	0.000	0.000	0.000
$\eta_3$	0.262	0.528	1.004	1.282	1.349	1.393
$\eta_4$	0.447	0.577	0.435	0.245	0.094	0.053
$\eta_5$	0.170	0.359	0.437	0.380	0.271	0.177
$\eta_6$	0.055	0.065	0.038	0.016	0.005	0.002
$\eta_7$	0.166	0.163	0.091	0.038	0.011	0.005
$\eta_8$	62.59	43.21	19.85	7.733	3.078	1.029
$\eta_9$	0.001	0.001	0.001	0.001	0.001	0.000
$\eta_{10}$	12.38	14.93	14.80	12.28	8.507	5.150
$\eta_{11}$	0.235	0.186	2.433	1.267	0.490	0.199
$\eta_{12}$	8.231	17.48	32.99	41.19	44.67	46.05
$\eta_{13}$	3.055	9.160	21.22	30.15	37.21	42.14
$\eta_{14}$	0.000	0.000	0.000	0.000	0.000	0.000
$\eta_{15}$	6.690	5.914	3.024	1.495	0.656	0.242

**6. Conclusions.** We discuss optimal control problems governed by transient convection diffusion equations, discretized by the symmetric interior penalty Galerkin (SIPG) method in space and the backward Euler method in time. In order to handle control constraints, we apply the primal-dual active set strategy and a Moreau-Yosida-based regularization. For both approaches, we propose error estimators to guide the mesh refinement. Numerical results show that a substantial amount of computational work can be saved by using efficient adaptive meshes for both approaches. In addition, we observe that although an additional error arises by the regularization of the box constraints, the Moreau-Yosida technique captures the errors of the control better than the PDAS strategy for both numerical examples. However, this observation requires a careful theoretical investigation. In this work we have tried to understand the behavior of the a posteriori error estimate in connection with a Moreau-Yosida regularization by fixing the regularization parameter. The aim of balancing the errors arising from the discretization and the Moreau-Yosida regularization can be carried out combining techniques from [24] and [50]. Further, instead of a fixed time, the adaptivity both in time and space will be addressed in future work for transient optimal control problems with suitable computation techniques.

**Acknowledgments.** The authors would like to express their sincere thanks to the referees for their most valuable suggestions.

## REFERENCES

- [1] M. AINSWORTH AND J. T. ODEN, *A Posteriori Error Estimation in Finite Element Analysis*, Wiley, New York, 2000.
- [2] T. AKMAN, H. YÜCEL, AND B. KARASÖZEN, *A priori error analysis of the upwind symmetric interior penalty Galerkin (SIPG) method for the optimal control problems governed by unsteady convection diffusion equations*, *Comput. Optim. Appl.*, 57 (2014), pp. 703–729.
- [3] D. N. ARNOLD, *An interior penalty finite element method with discontinuous elements*, *SIAM J. Numer. Anal.*, 19 (1982), pp. 742–760.
- [4] D. N. ARNOLD, F. BREZZI, B. COCKBURN, AND L. D. MARINI, *Unified analysis of discontinuous Galerkin methods for elliptic problems*, *SIAM J. Numer. Anal.*, 39 (2002), pp. 1749–1779.
- [5] R. BECKER AND B. VEXLER, *Optimal control of the convection-diffusion equation using stabilized finite element methods*, *Numer. Math.*, 106 (2007), pp. 349–367.
- [6] M. BERGOUNIOUX, K. ITO, AND K. KUNISCH, *Primal-dual strategy for constrained optimal control problems*, *SIAM J. Control Optim.*, 37 (1999), pp. 1176–1194.
- [7] A. CAGNIANI, E. H. GEORGOULIS, AND S. METCALFE, *Adaptive discontinuous Galerkin methods for nonstationary convection-diffusion problem*, *IMA J. Numer. Anal.*, 34 (2014), pp. 1578–1597.
- [8] A. CANGIANI, J. CHAPMAN, E. H. GEORGOULIS, AND M. JENSEN, *On local super-penalization of interior penalty discontinuous Galerkin methods*, *Int. J. Numer. Anal. Mod.*, 11 (2014), pp. 478–495.
- [9] P. G. CIARLET, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, 1978.
- [10] D. CLEVER, J. LANG, S. ULBRICH, AND J. C. ZIEMS, *Combination of an adaptive multilevel SQP method and a space-time adaptive PDAE solver for optimal control problems*, *Procedia Comput. Sci.*, 1 (2010), pp. 1435–1443.
- [11] S. S. COLLIS AND M. HEINKENSCHLOSS, *Analysis of the streamline upwind/Petrov Galerkin method applied to the solution of optimal control problems*, Tech. Rep. TR02–01, Department of Computational and Applied Mathematics, Rice University, Houston, 2002.
- [12] L. DEDÈ, S. MICHELETTI, AND S. PEROTTO, *Anisotropic error control for environmental applications*, *Appl. Numer. Math.*, 58 (2008), pp. 1320–1339.
- [13] H. J. S. FERNANDO, S. M. LEE, J. ANDERSON, M. PRINCEVAC, E. PARDYJAK, AND S. GROSSMAN-CLARKE, *Urban fluid mechanics: Air circulation and contaminant dispersion in cities*, *Environmental Fluid Mechanics*, 1 (2001), pp. 107–164.
- [14] H. FU, *A characteristic finite element method for optimal control problems governed by convection-diffusion equations*, *J. Comput. Appl. Math.*, 235 (2010), pp. 825–836.
- [15] H. FU AND H. RUI, *A priori error estimates for optimal control problems governed by transient advection-diffusion equations*, *J. Sci. Comput.*, 38 (2009), pp. 290–315.
- [16] ———, *A priori and a posteriori error estimates for the method of lumped masses for parabolic optimal control problems*, *Int. J. Comput. Math.*, 88 (2011), pp. 2798–2823.

- [17] ———, *Adaptive characteristic finite element approximation of convection-diffusion optimal control problems*, Numer. Methods Partial Differential Equations, 29 (2012), pp. 979–998.
- [18] A. V. FURSIKOV, *Optimal Control of Distributed Systems. Theory and Applications*, Amer. Math. Soc., Providence, 2000.
- [19] E. H. GEORGIOULIS, O. LAKKIS, AND J. M. VIRTANEN, *A posteriori error control for discontinuous Galerkin methods for parabolic problems*, SIAM J. Numer. Anal., 49 (2011), pp. 427–458.
- [20] W. GONG AND N. YAN, *A posteriori error estimate for boundary control problems governed by the parabolic partial differential equations*, J. Comput. Math., 27 (2009), pp. 68–88.
- [21] A. GÜNTHER, M. HINZE, AND M. H. TBER, *A posteriori error representations for elliptic optimal control problems with control and state constraints*, in Constrained Optimization and Optimal Control for Partial Differential Equations, G. Leugering, S. Engell, A. Griewank, M. Hinze, R. Rannacher, V. Schulz, M. Ulbrich, and S. Ulbrich, eds., vol. 160 of Internat. Ser. Numer. Math., Springer, Basel, 2012, pp. 303–317.
- [22] R. HERZOG AND E. W. SACHS, *Preconditioned conjugate gradient method for optimal control problems with control and state constraints*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2291–2317.
- [23] J. S. HESTHAVEN AND T. WARBURTON, *Nodal Discontinuous Galerkin Methods*, Springer, New York, 2008.
- [24] M. HINTERMÜLLER AND M. HINZE, *Moreau-Yosida regularization in state constrained elliptic control problems: error estimates and parameter adjustment*, SIAM J. Numer. Anal., 47 (2009), pp. 1666–1683.
- [25] M. HINZE, *A variational discretization concept in control constrained optimization: the linear-quadratic case*, Comput. Optim. Appl., 30 (2005), pp. 45–63.
- [26] M. HINZE, R. PINNAU, M. ULBRICH, AND S. ULBRICH, *Optimization with PDE Constraints*, Springer, New York, 2009.
- [27] M. HINZE, N. YAN, AND Z. ZHOU, *Variational discretization for optimal control governed by convection dominated diffusion equations*, J. Comp. Math., 27 (2009), pp. 237–253.
- [28] P. HOUSTON AND E. SÜLI, *Adaptive Lagrange-Galerkin methods for unsteady convection-diffusion problems*, Math. Comp., 70 (2000), pp. 77–106.
- [29] K. ITO AND K. KUNISCH, *Semi-smooth Newton methods for state-constrained optimal control problems*, Systems Control Lett., 50 (2003), pp. 221–228.
- [30] Z. KANAR SEYMEN, H. YÜCEL, AND B. KARASÖZEN, *Distributed optimal control of time-dependent diffusion-convection-reaction equations using space-time discretization*, J. Comput. Appl. Math., 261 (2014), pp. 146–157.
- [31] A. KUFNER, O. JOHN, AND S. FUCIK, *Function Spaces*, Nordhoff, Leyden, 1977.
- [32] D. LEYKEKHMAN AND M. HEINKENSCHLOSS, *Local error analysis of discontinuous Galerkin methods for advection-dominated elliptic linear-quadratic optimal control problems*, SIAM J. Numer. Anal., 50 (2012), pp. 2012–2038.
- [33] J.-L. LIONS, *Optimal Control of Systems Governed by Partial Differential Equations*, Springer, Berlin, 1971.
- [34] J.-L. LIONS AND E. MAGENES, *Non-Homogeneous Boundary Value Problems and Applications. Vol. I*, Springer, New York, 1972.
- [35] W. LIU, H. MA, T. TANG, AND N. YAN, *A posteriori error estimates for discontinuous Galerkin time-stepping method for optimal control problems governed by parabolic equations*, SIAM J. Numer. Anal., 42 (2004), pp. 1032–1061.
- [36] W. LIU AND N. YAN, *A posteriori error estimates for optimal control problems governed by parabolic equations*, Numer. Math., 93 (2003), pp. 497–521.
- [37] C. MAKRIDAKIS AND R. H. NOCHETTO, *Elliptic reconstruction and a posteriori error estimates for parabolic problems*, SIAM J. Numer. Anal., 41 (2003), pp. 1585–1594.
- [38] R. H. NOCHETTO, K. G. SIEBERT, AND A. VEESER, *Theory of adaptive finite element methods: an introduction*, in Multiscale, Nonlinear and Adaptive Approximation, R. A. DeVore and A. Kunoth, eds., Springer, Berlin, 2009, pp. 409–542.
- [39] D. PARRA-GUEVARA AND Y. N. SKIBA, *On optimal solution of an inverse air pollution problem: theory and numerical approach*, Math. Comput. Modelling, 43 (2006), pp. 766–778.
- [40] J. W. PEARSON AND M. STOLL, *Fast iterative solution of reaction-diffusion control problems arising from chemical processes*, SIAM J. Sci. Comput., 35 (2013), pp. B987–B1009.
- [41] J. W. PEARSON, M. STOLL, AND A. J. WATHEN, *Regularization-robust preconditioners for time-dependent PDE-constrained optimization problems*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 1126–1152.
- [42] B. RIVIÈRE, *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations. Theory and Implementation*, SIAM, Philadelphia, 2008.
- [43] A. RÖSCH, *Error estimates for linear-quadratic control problems with control constraints*, Optim. Methods Softw., 21 (2006), pp. 121–134.
- [44] M. STOLL AND A. WATHEN, *Precondining for partial differential equation constrained optimization with control constraints*, Numer. Linear Algebra Appl., 19 (2012), pp. 53–71.
- [45] T. SUN, *Discontinuous Galerkin finite element method with interior penalties for convection diffusion optimal control problem*, Int. J. Numer. Anal. Model., 7 (2010), pp. 87–107.

- [46] T. SUN, L. GE, AND W. LIU, *Equivalent a posteriori error estimates for a constrained optimal control problem governed by parabolic equations*, Int. J. Numer. Anal. Model., 10 (2013), pp. 1–23.
- [47] F. TRÖLTZSCH, *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*, Amer. Math. Soc., Providence, 2010.
- [48] R. VERFÜRTH, *A Review of A Posteriori Error Estimation and Adaptive Mesh–Refinement Techniques*, Wiley, Chichester, 1996.
- [49] ———, *Robust a posteriori error estimates for nonstationary convection-diffusion equations*, SIAM J. Numer. Anal., 43 (2005), pp. 1783–1802.
- [50] W. WOLLNER, *A posteriori error estimates for a finite element discretization of interior point methods for an elliptic optimization problem with state constraints*, Comput. Optim. Appl., 47 (2010), pp. 133–159.
- [51] N. YAN AND Z. ZHOU, *A priori and a posteriori error analysis of edge stabilization Galerkin method for the optimal control problem governed by convection-dominated diffusion equation*, J. Comput. Appl. Math., 223 (2009), pp. 198–217.
- [52] H. YÜCEL AND P. BENNER, *Adaptive discontinuous Galerkin methods for state constrained optimal control problems governed by convection diffusion equations*, Comput. Optim. Appl., 62 (2015), pp. 291–321.
- [53] H. YÜCEL, M. HEINKENSCHLOSS, AND B. KARASÖZEN, *An adaptive discontinuous Galerkin method for convection dominated distributed optimal control problems*, Tech Report, Department of Computational and Applied Mathematics, Rice University, Houston, 2012.
- [54] ———, *Distributed optimal control of diffusion-convection-reaction equations using discontinuous Galerkin methods*, in Numerical Mathematics and Advanced Applications 2011, A. Cangiani, R. L. Davidchack, E. Georgoulis, A. N. Gorban, J. Levesley, and M. V. Tretyakov, eds., Springer, Heidelberg, 2013, pp. 389–397.
- [55] H. YÜCEL AND B. KARASÖZEN, *Adaptive symmetric interior penalty Galerkin (SIPG) method for optimal control of convection diffusion equations with control constraints*, Optimization, 63 (2014), pp. 145–166.
- [56] Z. ZHOU AND H. FU, *A posteriori error estimates for continuous interior penalty Galerkin approximation of transient convection diffusion optimal control problems*, Bound. Value Probl., 2014 (2014), Art. 2014:207 (19 pages).
- [57] Z. ZHOU AND N. YAN, *The local discontinuous Galerkin method for optimal control problem governed by convection diffusion equations*, Int. J. Numer. Anal. Model., 7 (2010), pp. 681–699.