ADAPTIVE MULTILEVEL CORRECTION METHOD FOR FINITE ELEMENT APPROXIMATIONS OF ELLIPTIC OPTIMAL CONTROL PROBLEMS

WEI GONG$^{\diamond}$, HEHU XIE[†], AND NINGNING YAN[‡]

Abstract: In this paper we propose an adaptive multilevel correction scheme to solve optimal control problems discretized with finite element method. Different from the classical adaptive finite element method (AFEM for short) applied to optimal control which requires the solution of the optimization problem on new finite element space after each mesh refinement, with our approach we only need to solve two linear boundary value problems on current refined mesh and an optimization problem on a very low dimensional space. The linear boundary value problems can be solved with well-established multigrid method designed for elliptic equation and the optimization problems are of small scale corresponding to the space built with the coarsest space plus two enriched bases. Our approach can achieve the similar accuracy with standard AFEM but greatly reduces the computational cost. Numerical experiments demonstrate the efficiency of our proposed algorithm.

Keywords: Optimal control problems, elliptic equation, control constraints, a posteriori error estimates, adaptive finite element method, multilevel correction method

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

1. Introduction

As a typical application of PDE-constrained optimization, optimal control problem (OCP for short) plays an increasingly important role in modern scientific community. Contributed to the pioneer work of Lions (see [22]), optimal controls of PDE system became a hot research topic in the decades and had positive impacts on the development of some related fields like optimization and numerical analysis. A lot of achievements have been made in these directions.

Due to the PDE constraints of optimal control problems, discretization methods for PDE are indispensable to solve this kind of problems together with the optimization algorithm. For the optimization algorithms to solve PDE-constrained optimization in both finite dimensional space and Banach space we refer to [14] and [16] for more details. As a mainstream discretization method in the community of numerical analysis, finite element method became very popular in the numerical solutions of optimal control problems. A priori and a posteriori error estimates for the finite element approximations to different kind of optimal control problems are summarized in [16] and [26], respectively.

Adaptive finite element method, aiming at generating a sequence of optimal triangulations by refining those elements where the errors are relatively large as the local error estimators indicate, is very efficient to reduce the computational cost while achieving satisfactory accuracy. Adaptive finite element method was firstly proposed in [2] by Babuška and Rheinboldt and by now became a well-developed algorithm for which the convergence

Date: Today is September 27, 2016.

$^{\diamond}$ NCMIS, LSEC, Institute of Computational Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China (wgong@lsec.cc.ac.cn).

[†]NCMIS, LSEC, Institute of Computational Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China (hhxie@lsec.cc.ac.cn).

[‡]NCMIS, LSEC, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China (ynn@amss.ac.cn).

and optimal computational complexity have been well established, see [6], [9], [27] and so on. For the derivation of reliable and efficient a posteriori error estimators of different kind we refer to [28]. The application of adaptive finite element method to optimal control problems was contributed to Liu, Yan ([23]) and Becker and coauthors ([3]). Since then, a lot of works can be found in, e.g., [13], [17], [19], [24] and the references therein. For the convergence analysis of AFEM for optimal control problems we mention the works [10] and [12] which treated the full control discretization case and the variational control discretization case, respectively.

Although the adaptive finite element method has promising accuracy applied to optimal control problems, one has to solve an optimization problem after each mesh refinement, which may be very costly when the number of Dofs is large. In this paper, we propose an adaptive multilevel correction method to solve optimal control problems with finite element method. Different from the classical approach which solves the optimization problem on the new finite element space after each mesh refinement, with our approach we only need to solve two linear boundary value problems (BVPs for short) on the refined mesh and an optimization problem on the coarsest mesh from which we start the adaptive algorithm enriched with two bases corresponding to the solutions of two linear BVPs. The linear boundary value problems can be solved efficiently with well-established multigrid method designed for elliptic equation and the optimization problems are of small scale corresponding to the coarsest finite element space plus two bases, this greatly reduces the computational cost but achieves the similar accuracy with the standard AFEM for optimal control problems, which is proved in this paper.

The adaptive multilevel correction method proposed in this paper is a combination of the multilevel correction scheme for optimal controls of elliptic equation proposed in [11] and the adaptive FEM, which originated from the multilevel correction method proposed in [20], [21], [30] and [31] for eigenvalue problems. For the related two-grid method and adaptive correction method for elliptic equation we refer to [32] and [33]. Here we should also comment on the existing multigrid methods for solving PDE-constrained optimization problems. Roughly speaking, there are three kind of multigrid methods for PDE-constrained optimization: the direct (so-called one-shot) multigrid method where the optimization problem is implemented within the hierarchy of grid levels, the use of multigrid schemes as inner solvers within an outer optimization loop and the MG/OPT algorithm where the multigrid method defines the outer solver ([18]), for an overview we refer the readers to [5]. We remark that the approach proposed in this paper is totally different from the above mentioned multigrid methods to solve the OCP. Our method solves the OCP from coarsest mesh to finest mesh while other methods solve the OCP from finest mesh to coarsest mesh.

The success of our proposed adaptive multilevel correction method for solving OCPs lies in the fact that the solutions of two linear BVPs on current refined mesh with sources from solutions on previous mesh give better approximations of the state and adjoint state (compare step (6) in Algorithm 3.1) measured in L^2-norm, due to the Aubin-Nitsche technique. Then the necessary information from the finer mesh is contained in the solutions of two BVPs and the multilevel correction algorithm can preserve the same accuracy with the direct method asymptotically, by solving an optimization problem on the coarsest space enriched with two bases, we refer to [11] for a priori error analysis. The ideas were previously used for two-grid method ([32]) and applied later on to nonlinear problems and eigenvalue problems. The recent advance by Xie et al. ([21, 30]) makes it possible to generalize the two-grid method to multilevel version by introducing the coarsest space.

The structure of this paper is as follows: In section 2 we present the optimal control problem as well as its finite element approximation. In section 3 we formulate our adaptive multilevel correction method to solve the OCP, a posteriori error estimates are also derived for this kind of adaptive method. Convergence results of the algorithm are presented in

Section 4. Section 5 is devoted to some numerical experiments to illustrate the efficiency of our proposed algorithm.

Let $\Omega \subset \mathbb{R}^d$, $d = 2, 3$ be a bounded polygonal or polyhedral domain. Let $\|\cdot\|_{m,s,\Omega}$ and $\|\cdot\|_{m,\Omega}$ be the usual norms of the Sobolev spaces $W^{m,s}(\Omega)$ and $H^m(\Omega)$ respectively. Let $|\cdot|_{m,s,\Omega}$ and $|\cdot|_{m,\Omega}$ be the usual seminorms of the above-mentioned two spaces respectively.

2. Finite element method for optimal control problem

In this section we will introduce the general formulation of linear-quadratic optimal control problems governed by elliptic equations. The theoretical aspects including the existence and uniqueness of solution and the first order optimality conditions will be presented. Moreover, we will introduce a finite element approximation to the control problems.

Consider the following controlled equation:

$$\begin{cases} Ly = u & \text{in } \Omega, \\ y = 0 & \text{on } \partial\Omega, \end{cases} \tag{2.1}$$

where L is a linear second order elliptic operator of the following type

$$Ly := -\sum_{i,j=1}^{d} \frac{\partial}{\partial x_j}(a_{ij}\frac{\partial y}{\partial x_i}) + cy.$$

Here $a_{ij} \in W^{1,\infty}(\Omega)$ $(i, j = 1, \cdots, d)$ is symmetric, positive definite and $0 \leqslant c < \infty$. Thus, L is self-adjoint and we denote the adjoint operator $L^* = L$. We denote $A = (a_{ij})_{d\times d}$ and A^* its adjoint. We use the standard notations

$$a(y, v) := \int_\Omega (\sum_{i,j=1}^{d} a_{ij}\frac{\partial y}{\partial x_i}\frac{\partial v}{\partial x_j} + cyv)dx, \qquad \forall y, v \in H_0^1(\Omega),$$

$$(y, v) := \int_\Omega yvdx, \qquad \forall y, v \in L^2(\Omega).$$

Note that the bilinear form $a(\cdot,\cdot)$ induces a norm which is denoted by $\|v\|_{a,\Omega} := \sqrt{a(v, v)}$.

The linear-quadratic optimal control problem considered in this paper reads:

$$\min_{u\in U_{ad}} \quad J(y, u) = \frac{1}{2}\|y - y_d\|_{0,\Omega}^2 + \frac{\alpha}{2}\|u\|_{0,\Omega}^2 \quad \text{subject to} \;\; (2.1). \tag{2.2}$$

The set of admissible control is of bilateral type:

$$U_{ad} := \Big\{u \in L^2(\Omega) : \;\; a \leqslant u(x) \leqslant b \;\; \text{a.e. in } \Omega\Big\} \tag{2.3}$$

with $a < b$ two constants.

Since the above optimization problem is linear and strictly convex, there exists a unique solution $u \in U_{ad}$ by standard arguments (see [22, Chap.2, Thm.1.2]). Moreover, by introducing the adjoint state p, the optimal solution can be characterized by the following necessary and sufficient (first order) optimality condition

$$\begin{cases} a(y, v) = (u, v), & \forall v \in H_0^1(\Omega), \\ a(w, p) = (y - y_d, w), & \forall w \in H_0^1(\Omega), \\ (\alpha u + p, v - u) \geqslant 0, & \forall v \in U_{ad}. \end{cases} \tag{2.4}$$

Hereafter, we call u, y and p the optimal control, state and adjoint state, respectively.

With the set of admissible control (2.3) we can get the pointwise representation of the optimal control u through the adjoint state p

$$u = P_{U_{ad}}\big\{-\frac{1}{\alpha}p\big\}, \tag{2.5}$$

where $P_{U_{ad}}$ is the orthogonal projection operator onto U_{ad}.

Let $\mathcal{T}_h$ be a regular triangulation of Ω such that $\bar{\Omega} = \cup_{\tau\in\mathcal{T}_h}\bar{\tau}$. Throughout the paper we denote $\mathcal{E}_h$ the set of interior faces (edges or sides) of $\mathcal{T}_h$. On $\mathcal{T}_h$ we construct the piecewise linear and continuous finite element space V_h such that $V_h \subset C(\bar{\Omega}) \cap H_0^1(\Omega)$.

In this paper, we use piecewise linear finite element to approximate the state y, and variational discretization for the optimal control u (see [15]). Then we can define the finite element approximation to the optimal control problem (2.2) as follows:

$$\min_{u\in U_{ad}} J_h(y_h(u),u) = \frac{1}{2}\|y_h(u)-y_d\|_{0,\Omega}^2 + \frac{\alpha}{2}\|u\|_{0,\Omega}^2 \tag{2.6}$$

subject to

$$y_h(u)\in V_h: \quad a(y_h(u),v_h) = (u,v_h), \qquad \forall v_h\in V_h. \tag{2.7}$$

Similar to the infinite dimensional problem (2.2), the above discretized optimization problem also admits a unique solution $u_h \in U_{ad}$. Moreover, the discretized first order necessary and sufficient optimality condition can be stated as follows:

$$\left\{\begin{array}{ll} a(y_h,v_h) = (u_h,v_h), & \forall v_h\in V_h,\\ a(w_h,p_h) = (y_h-y_d,w_h), & \forall w_h\in V_h,\\ (\alpha u_h + p_h, v_h-u_h)\geqslant 0, & \forall v_h\in U_{ad}, \end{array}\right. \tag{2.8}$$

where $p_h \in V_h$ is the discrete adjoint state. Similar to the continuous case we have

$$u_h = P_{U_{ad}}\big\{-\frac{1}{\alpha}p_h\big\}. \tag{2.9}$$

We denote $S: L^2(\Omega)\to H_0^1(\Omega)$ the control-to-state mapping with S^* its adjoint. Then we can write $y = Su$ and $p = S^*(y-y_d)$. For the discretized state equation we also define $S_h : L^2(\Omega)\to V_h$ as the discrete solution operator such that $y_h = S_h u_h$ and S_h^* the associated discrete adjoint solution operator for the adjoint state equation with $p_h = S_h^*(y_h - y_d)$.

For the following purpose, we firstly introduce some notations. For each element $T\in\mathcal{T}_h$, we define the local error indicators $\eta_{y,h}(u_h,y_h,T)$ and $\eta_{p,h}(y_h,p_h,T)$ by

$$\eta_{y,h}^2(u_h,y_h,T) := h_T^2\|u_h - Ly_h\|_{0,T}^2 + \sum_{E\in\mathcal{E}_h, E\subset\partial T} h_E\|[A\nabla y_h]\cdot n_E\|_{0,E}^2, \tag{2.10}$$

$$\eta_{p,h}^2(y_h,p_h,T) := h_T^2\|y_h - y_d - L^*p_h\|_{0,T}^2 + \sum_{E\in\mathcal{E}_h, E\subset\partial T} h_E\|[A^*\nabla p_h]\cdot n_E\|_{0,E}^2, \tag{2.11}$$

where $[A\nabla y_h]\cdot n_E$ denotes the jump of $A\nabla y_h$ across the common side E of elements T^+ and T^-, n_E denotes the outward normal oriented to T^-. Then on a subset $\omega\subset\Omega$, we define the error estimators $\eta_{y,h}(u_h,y_h,\omega)$ and $\eta_{p,h}(y_h,p_h,\omega)$ by

$$\eta_{y,h}(u_h,y_h,\omega) := \Big(\sum_{T\in\mathcal{T}_h, T\subset\omega}\eta_{y,h}^2(u_h,y_h,T)\Big)^{\frac{1}{2}}, \tag{2.12}$$

$$\eta_{p,h}(y_h,p_h,\omega) := \Big(\sum_{T\in\mathcal{T}_h, T\subset\omega}\eta_{p,h}^2(y_h,p_h,T)\Big)^{\frac{1}{2}}. \tag{2.13}$$

Thus, $\eta_{y,h}(u_h,y_h,\Omega)$ and $\eta_{p,h}(y_h,p_h,\Omega)$ constitute the error estimators for the state equation and the adjoint state equation on Ω with respect to $\mathcal{T}_h$.

With the above defined error estimators we have the following properties whose proof can be found in [6, 31].

Lemma 2.1. *For the error indicator $\eta_{v,h}(g,v_h,\omega)$ with $g = u_h, v = y$ or $g = y_h, v = p$ there hold*

$$\eta_{v,h}(g,v_h+w_h,\omega) \leqslant \eta_{v,h}(g,v_h,\omega) + \eta_{v,h}(g,w_h,\omega) \quad \forall v_h,w_h\in V_h, \tag{2.14}$$

$$\eta_{v,h}(g,v_h,\Omega) \leqslant C_R\|v_h\|_{a,\Omega} \quad \forall v_h\in V_h. \tag{2.15}$$

Now standard a posterior error estimates for elliptic equation give the following upper bounds (see, e.g., [28]) which show the reliability of the error estimators.

Lemma 2.2. *Let S and S_h be the continuous and discrete solution operators defined above. Then the following a posteriori error estimates hold*

$$\|Su_h - S_h u_h\|_{a,\Omega}^2 \leqslant \tilde{C}_1 \eta_{y,h}^2(u_h, S_h u_h, \Omega), \tag{2.16}$$

$$\|S^*(y_h - y_d) - S_h^*(y_h - y_d)\|_{a,\Omega}^2 \leqslant \tilde{C}_1 \eta_{p,h}^2(y_h, S_h^*(y_h - y_d), \Omega). \tag{2.17}$$

For $f \in L^2(\Omega)$ we define the data oscillation (see [27]) by

$$\operatorname{osc}(f, \mathcal{T}_h) := \Big(\sum_{T\in\mathcal{T}_h} \operatorname{osc}^2(f,T)\Big)^{\frac{1}{2}}, \quad \operatorname{osc}^2(f,T) := \|h_T(f - f_T)\|_{0,T}^2, \tag{2.18}$$

where f_T denotes the average of f on element T. Then we can also derive the following global a posteriori error lower bounds, i.e., the global efficiency of the error estimators.

Lemma 2.3. *Let S and S_h be the continuous and discrete solution operators defined above. Then the following a posteriori error lower bounds hold*

$$\tilde{C}_2 \eta_{y,h}^2(u_h, S_h u_h, \Omega) \leqslant \|Su_h - S_h u_h\|_{a,\Omega}^2 + \tilde{C}_3 osc^2(u_h - LS_h u_h, \mathcal{T}_h), \tag{2.19}$$

$$\begin{aligned}\tilde{C}_2 \eta_{p,h}^2(y_h, S_h^*(y_h - y_d), \Omega) \leqslant\ & \|S^*(y_h - y_d) - S_h^*(y_h - y_d)\|_{a,\Omega}^2 \\ & + \tilde{C}_3 osc^2(y_h - y_d - L^* S_h^*(y_h - y_d), \mathcal{T}_h).\end{aligned} \tag{2.20}$$

3. Adaptive multilevel correction method for optimal control problems

The adaptive finite element procedure consists of the following loops

$$\text{SOLVE} \rightarrow \text{ESTIMATE} \rightarrow \text{MARK} \rightarrow \text{REFINE}.$$

The ESTIMATE step is based on the a posteriori error indicators which will be derived in the following, while the step REFINE can be done by using iterative or recursive bisection of elements with the minimal refinement condition. In this section, we propose a type of adaptive multilevel correction method for the optimal control problem (2.6)-(2.7) which corresponds to the SOLVE step of the adaptive procedure. In the loop of adaptive finite element method, solving the optimization problem on the refined mesh after each REFINE module is transformed to the solutions of two linear boundary value problems on current mesh and the solution of one optimization problem on the coarsest finite element space. The algorithm is described as follows:

Algorithm 3.1. *An adaptive multilevel correction method for optimal control problem:*

(1) *Given a coarse mesh $\mathcal{T}_{h_0}$ with mesh size h_0 and construct the finite element space V_{h_0}.*
(2) *Refine the mesh $\mathcal{T}_{h_0}$ to obtain an initial mesh $\mathcal{T}_{h_1}$ by regular refinement and construct the finite element space V_{h_1}. Set $k = 1$ and solve the following optimal control problem*

$$\min_{u_{h_1}\in U_{ad},\ y_{h_1}\in V_{h_1}} \quad J(y_{h_1}, u_{h_1}) = \frac{1}{2}\|y_{h_1} - y_d\|_{0,\Omega}^2 + \frac{\alpha}{2}\|u_{h_1}\|_{0,\Omega}^2$$

subject to

$$a(y_{h_1}, v_{h_1}) = (u_{h_1}, v_{h_1}), \qquad \forall v_{h_1} \in V_{h_1}.$$

(3) *Compute the local error indicators $\eta_{h_k}((u_{h_k}, y_{h_k}, p_{h_k}), T)$.*
(4) *Construct $\tilde{\mathcal{T}}_{h_k} \subset \mathcal{T}_{h_k}$ by the marking algorithm.*
(5) *Refine $\tilde{\mathcal{T}}_{h_k}$ to get a new conforming mesh $\mathcal{T}_{h_{k+1}}$.*

(6) *Solve two BVPs on $\mathcal{T}_{h_{k+1}}$ for the discrete solutions $y^*_{h_{k+1}} \in V_{h_{k+1}}$ such that*

$$a(y^*_{h_{k+1}}, v_{h_{k+1}}) = (u_{h_k}, v_{h_{k+1}}), \qquad \forall v_{h_{k+1}} \in V_{h_{k+1}}$$

and $p^*_{h_{k+1}} \in V_{h_{k+1}}$ such that

$$a(v_{h_{k+1}}, p^*_{h_{k+1}}) = (y_{h_k} - y_d, v_{h_{k+1}}), \qquad \forall v_{h_{k+1}} \in V_{h_{k+1}}.$$

(7) *Construct a new finite element space $V_{h_0,h_{k+1}} := V_{h_0} + \mathrm{span}\{y^*_{h_{k+1}}\} + \mathrm{span}\{p^*_{h_{k+1}}\}$ and solve the following optimal control problem:*

$$\min_{u_{h_{k+1}} \in U_{ad},\ y_{h_{k+1}} \in V_{h_0,h_{k+1}}} J(y_{h_{k+1}}, u_{h_{k+1}}) = \frac{1}{2}\|y_{h_{k+1}} - y_d\|^2_{0,\Omega} + \frac{\alpha}{2}\|u_{h_{k+1}}\|^2_{0,\Omega}$$

subject to

$$a(y_{h_{k+1}}, v_{h_0,h_{k+1}}) = (u_{h_{k+1}}, v_{h_0,h_{k+1}}), \qquad \forall v_{h_0,h_{k+1}} \in V_{h_0,h_{k+1}}.$$

(8) *Set $k = k + 1$ and go to Step (3).*

Remark 3.2. *A number of remarks are in order. Firstly, step (6) in Algorithm 3.1 can be viewed as one gradient step with initial guess from previous iteration. The solutions of two BVPs on current refined mesh give better approximations of the state and adjoint state variables, which contain necessary information of the finer mesh so that the multilevel correction algorithm can preserve the same accuracy with the direct optimization method asymptotically, we refer to* [11] *for a priori error analysis. Secondly, the stiffness and mass matrices of the state equation in step (7) of Algorithm 3.1 are still sparse except for the last two rows and columns. Since the matrix is of small size, it can be solved efficiently with even direct method. Thirdly, in our numerical experiments we use projected gradient method (*[15]*) to solve the resulting optimization problem in step (7). The incorporation of semi-smooth Newton method (*[14]*) introduces difficulty because the definition of active sets should be posed on fine mesh which may be costly. Lastly, it seems from the proof of Theorem 3.3 in* [11] *that the coarset mesh should be appropriately chosen according to the regularization parameter α. That is to say, if α is small the coarsest mesh size should be chosen also small to guarantee the convergence of the algorithm.*

Remark 3.3. *In this remark we intend to explain the computational complexity of Algorithm 3.1. There are several places where the evaluation of integral on fine mesh should be done, for example, when evaluating the right hand side of the state equation contributed from the control. Therefore, the total computational complexity can not really reduced to the scale of the coarsest grid $\mathcal{T}_{h_0}$. The savings of computational time come from the solving of the state and adjoint state equations during each optimization step, as the dimension of the governing state equation is greatly reduced.*

Now we are in a position to derive a posteriori error estimator for the optimal control problems solved by adaptive Algorithm 3.1. To begin with, we define the following quantity

$$\gamma(h) = \sup_{f \in L^2(\Omega), \|f\|_{0,\Omega}=1} \inf_{v_h \in V_h} \|L^{-1}f - v_h\|_{a,\Omega}.$$

Note that $\gamma(h) \to 0$ as $h \to 0$.

At first, we establish some relationships between the boundary value approximations and the optimal control approximations. In the following of this paper, we set $H = h_k$ and $h = h_{k+1}$. Let $V_{\tilde{H}} := V_{h_0,h_k}$ and $V_{\tilde{h}} := V_{h_0,h_{k+1}}$ be the enriched finite element spaces defined in the step (7) of Algorithm 3.1. We also denote $S_{\tilde{H}} : L^2(\Omega) \to V_{\tilde{H}}$ the solution operator of the state equation with $S^*_{\tilde{H}}$ its adjoint. Let $y^h = Su_h$ and $p^h = S^*(y_h - y_d)$, then it is clear that y_h and p_h are the standard finite element approximations of y^h and p^h in $V_{\tilde{h}}$, i.e., $y_h = S_{\tilde{h}}u_h$ and $p_h = S^*_{\tilde{h}}(y_h - y_d)$.

Theorem 3.4. *Let $h, H \in (0, h_0]$ and $(u, y, p) \in U_{ad} \times H_0^1(\Omega) \times H_0^1(\Omega)$ be the solution of problem (2.2) and $(u_h, y_h, p_h) \in U_{ad} \times V_h \times V_h$ be the solution sequence produced by Algorithm 3.1. Then the following properties hold*

$$\begin{aligned} \|y - y_h\|_{a,\Omega} &= \|y^h - S_h u_h\|_{a,\Omega} + O(\gamma(h_0))(\|y - y_H\|_{a,\Omega} + \|p - p_H\|_{a,\Omega} \\ &\quad + \|y - y_h\|_{a,\Omega} + \|p - p_h\|_{a,\Omega}), \end{aligned} \tag{3.1}$$

$$\begin{aligned} \|p - p_h\|_{a,\Omega} &= \|p^h - S_h^*(y_h - y_d)\|_{a,\Omega} + O(\gamma(h_0))(\|y - y_H\|_{a,\Omega} + \|p - p_H\|_{a,\Omega} \\ &\quad + \|y - y_h\|_{a,\Omega} + \|p - p_h\|_{a,\Omega}) \end{aligned} \tag{3.2}$$

provided $h_0 \ll 1$.

Proof. Note that we have the splitting:

$$y - y_h = y - y^h + y^h - S_h u_h + S_h u_h - S_h u_H + S_h u_H - y_h,$$
$$p - p_h = p - p^h + p^h - S_h^*(y_h - y_d) + S_h^*(y_h - y_d) - S_h^*(y_H - y_d) + S_h^*(y_H - y_d) - p_h.$$

From the stability of elliptic equation we can derive

$$\|y - y^h\|_{a,\Omega} \leqslant C\|u - u_h\|_{0,\Omega}. \tag{3.3}$$

Similarly, we have

$$\|p - p^h\|_{a,\Omega} \leqslant C\|y - y_h\|_{0,\Omega}.$$

In the following we estimate $\|y - y_h\|_{0,\Omega}$. Let $\psi \in H_0^1(\Omega)$ be the solution of the following auxiliary problem

$$\begin{cases} L^*\psi = y - y_h & \text{in } \Omega, \\ \psi = 0 & \text{on } \partial\Omega. \end{cases} \tag{3.4}$$

Let $\psi_{\tilde{h}} \in V_{\tilde{h}}$ be the finite element approximation of ψ. Then we can conclude from the standard Aubin-Nitsche technique that

$$\begin{aligned} \|y - y_h\|_{0,\Omega}^2 &= a(y - y_h, \psi) \\ &= a(y - y_h, \psi - \psi_{\tilde{h}}) + a(y - y_h, \psi_{\tilde{h}}) \\ &= a(y - y_h, \psi - \psi_{\tilde{h}}) + (u - u_h, \psi_{\tilde{h}} - \psi) + (u - u_h, \psi) \\ &\leqslant \tilde{C}\Big(\gamma(h_0)\|y - y_h\|_{a,\Omega} + \|u - u_h\|_{0,\Omega}\Big)\|y - y_h\|_{0,\Omega}, \end{aligned}$$

which in turn implies

$$\|y - y_h\|_{0,\Omega} \leqslant \tilde{C}\gamma(h_0)\|y - y_h\|_{a,\Omega} + C\|u - u_h\|_{0,\Omega}.$$

Therefore, we can obtain

$$\|p - p^h\|_{a,\Omega} \leqslant \tilde{C}\gamma(h_0)\|y - y_h\|_{a,\Omega} + C\|u - u_h\|_{0,\Omega}. \tag{3.5}$$

From the discrete stability of finite element solutions we have

$$\|S_h u_h - S_h u_H\|_{a,\Omega} \leqslant C\|u - u_h\|_{0,\Omega} + C\|u - u_H\|_{0,\Omega}. \tag{3.6}$$

Similarly, we have

$$\begin{aligned} &\|S_h^*(y_h - y_d) - S_h^*(y_H - y_d)\|_{a,\Omega} \leqslant C\|y - y_h\|_{0,\Omega} + C\|y - y_H\|_{0,\Omega} \\ &\leqslant \tilde{C}\gamma(h_0)(\|y - y_H\|_{a,\Omega} + \|y - y_h\|_{a,\Omega}) + C(\|u - u_H\|_{0,\Omega} + \|u - u_h\|_{0,\Omega}). \end{aligned} \tag{3.7}$$

Note that $S_h u_H = y_{\tilde{h}}^*$ and $S_h^*(y_H - y_d) = p_{\tilde{h}}^*$, thus there hold $S_h u_H - y_h \in V_{\tilde{h}}$ and $S_h^*(y_H - y_d) - p_h \in V_{\tilde{h}}$. It follows from $V_{\tilde{h}} \subset V_h$ that

$$a(S_h u_H - y_h, S_h u_H - y_h) = (u_H - u_h, S_h u_H - y_h),$$

which yields

$$\|S_h u_H - y_h\|_{a,\Omega} \leqslant C\|u - u_H\|_{0,\Omega} + C\|u - u_h\|_{0,\Omega}. \tag{3.8}$$

Similarly, we can prove

$$\begin{aligned}\|S_h^*(y_H - y_d) - p_h\|_{a,\Omega} &\leqslant \tilde{C}\gamma(h_0)(\|y - y_H\|_{a,\Omega} + \|y - y_h\|_{a,\Omega})\\ &\quad + C(\|u - u_H\|_{0,\Omega} + \|u - u_h\|_{0,\Omega}).\end{aligned} \tag{3.9}$$

Now it remains to estimate $\|u - u_H\|_{0,\Omega}$ and $\|u - u_h\|_{0,\Omega}$. Note that the first order optimality condition of the control problem in step (7) of Algorithm 3.1 is the same as (2.8) except V_h being replaced by $V_{\tilde{h}}$. Setting $v = u_h$ in the third equation of (2.4) and $v_h = u$ in the third equation of (2.8) we are led to

$$(\alpha u + p, u_h - u) \geqslant 0, \quad (\alpha u_h + p_h, u - u_h) \geqslant 0.$$

Adding the above two inequalities together and noticing that $p_h = S_{\tilde{h}}^*(S_{\tilde{h}}u_h - y_d)$, we obtain

$$\begin{aligned}&\alpha\|u - u_h\|_{0,\Omega}^2 \leqslant (p_h - p, u - u_h)\\ = \ &(p_h - S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d), u - u_h) + (S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d) - p, u - u_h)\\ = \ &(S_{\tilde{h}}u_h - S_{\tilde{h}}u, S_{\tilde{h}}(u - u_h)) + (S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d) - p, u - u_h)\\ \leqslant \ &(S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d) - p, u - u_h).\end{aligned}$$

Note that $p = S^*(Su - y_d)$, it follows from the ε-Young inequality that

$$\alpha\|u - u_h\|_{0,\Omega}^2 \leqslant C\|S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d) - S^*(Su - y_d)\|_{0,\Omega}^2. \tag{3.10}$$

To estimate the above term we use again the Aubin-Nitsche technique. Let $\phi \in H_0^1(\Omega)$ be the solution of equation (2.1) with right hand side $S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d) - S^*(Su - y_d)$ and $\phi_{\tilde{h}} \in V_{\tilde{h}}$ its finite element approximation. Then we have

$$\begin{aligned}&\|S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d) - S^*(Su - y_d)\|_{0,\Omega}^2 = a(S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d) - S^*(Su - y_d), \phi)\\ = \ &a(S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d) - S^*(Su - y_d), \phi - \phi_{\tilde{h}}) + a(S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d) - S^*(Su - y_d), \phi_{\tilde{h}})\\ = \ &a(S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d) - S^*(Su - y_d), \phi - \phi_{\tilde{h}}) + (S_{\tilde{h}}u - Su, \phi_{\tilde{h}}).\end{aligned}$$

Note that $V_{h_0} \subset V_{\tilde{h}}$, it follows from the standard finite element error estimate that

$$\begin{aligned}&a(S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d) - S^*(Su - y_d), \phi - \phi_{\tilde{h}})\\ \leqslant \ &\tilde{C}\gamma(h_0)\|S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d) - S^*(Su - y_d)\|_{0,\Omega}\|S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d) - S^*(Su - y_d)\|_{a,\Omega}\end{aligned} \tag{3.11}$$

and

$$\begin{aligned}&(S_{\tilde{h}}u - Su, \phi_{\tilde{h}}) = (S_{\tilde{h}}u - Su, \phi_{\tilde{h}} - \phi) + (S_{\tilde{h}}u - Su, \phi)\\ \leqslant \ &\tilde{C}\gamma(h_0)\|S_{\tilde{h}}u - Su\|_{a,\Omega}\|S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d) - S^*(Su - y_d)\|_{0,\Omega}.\end{aligned} \tag{3.12}$$

Combining the above estimates we are led to

$$\begin{aligned}&\|S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d) - S^*(Su - y_d)\|_{0,\Omega}\\ \leqslant \ &\tilde{C}\gamma(h_0)(\|S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d) - S^*(Su - y_d)\|_{a,\Omega} + \|S_{\tilde{h}}u - Su\|_{a,\Omega}).\end{aligned} \tag{3.13}$$

It follows from (3.10) and (3.13) that

$$\begin{aligned}\|u - u_h\|_{0,\Omega} &\lesssim \gamma(h_0)(\|S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d) - S^*(Su - y_d)\|_{a,\Omega} + \|S_{\tilde{h}}u - Su\|_{a,\Omega})\\ &\lesssim \gamma(h_0)(\|p_h - p\|_{a,\Omega} + \|S_{\tilde{h}}^*(S_{\tilde{h}}u - y_d) - S_{\tilde{h}}^*(S_{\tilde{h}}u_h - y_d)\|_{a,\Omega} + \|S_{\tilde{h}}u - Su\|_{a,\Omega})\\ &\lesssim \gamma(h_0)(\|p_h - p\|_{a,\Omega} + \|S_{\tilde{h}}u - S_{\tilde{h}}u_h\|_{a,\Omega} + \|S_{\tilde{h}}u - Su\|_{a,\Omega})\\ &\lesssim \gamma(h_0)(\|p_h - p\|_{a,\Omega} + \|S_{\tilde{h}}u_h - Su\|_{a,\Omega} + \|S_{\tilde{h}}u - S_{\tilde{h}}u_h\|_{a,\Omega})\\ &\lesssim \gamma(h_0)(\|p_h - p\|_{a,\Omega} + \|y_h - y\|_{a,\Omega} + \|u - u_h\|_{0,\Omega}).\end{aligned} \tag{3.14}$$

If $h_0 \ll 1$ then $\gamma(h_0) \ll 1$ for all $h \in (0, h_0)$, and we arrive at

$$\|u - u_h\|_{0,\Omega} \lesssim \gamma(h_0)(\|p_h - p\|_{a,\Omega} + \|y_h - y\|_{a,\Omega}). \tag{3.15}$$

Similarly, we can prove that

$$\|u - u_H\|_{0,\Omega} \lesssim \gamma(h_0)(\|p - p_H\|_{a,\Omega} + \|y - y_H\|_{a,\Omega}). \tag{3.16}$$

Inserting the above estimates into (3.3) and (3.5)-(3.9) we can conclude from the splitting of $y - y_h$ and $p - p_h$ the desired results (3.1)-(3.2). This completes the proof. $\square$

To derive a posteriori error estimates for the optimal control problem solved by Algorithm 3.1 we define the norm

$$\|(y,p)\|_a^2 = a(y,y) + a(p,p).$$

For ease of exposition we also define the following quantities:

$$\begin{aligned}\eta_h^2((u_h,y_h,p_h),T) &= \eta_{y,h}^2(u_h,y_h,T) + \eta_{p,h}^2(y_h,p_h,T),\\ \mathrm{osc}^2((u_h,y_h,p_h),T) &= \mathrm{osc}^2(u_h - Ly_h,T) + \mathrm{osc}^2(y_h - y_d - L^*p_h,T),\end{aligned}$$

and the straightforward modifications for $\eta_h^2((u_h,y_h,p_h),\Omega)$ and $\mathrm{osc}^2((u_h,y_h,p_h),\mathcal{T}_h)$.

Now we state the following a posteriori error estimates for the finite element approximation of the optimal control problem solved by the adaptive Algorithm 3.1.

Theorem 3.5. *Let $h \in (0,h_0]$ and $(u,y,p) \in U_{ad} \times H_0^1(\Omega) \times H_0^1(\Omega)$ be the solution of problem (2.2) and $(u_h,y_h,p_h) \in U_{ad} \times V_{\tilde{h}} \times V_{\tilde{h}}$ be the sequence produced by Algorithm 3.1. Then the following a posteriori error estimates hold*

$$\|(y-y_h,p-p_h)\|_a^2 \leqslant C_1\eta_h^2((u_h,y_h,p_h),\Omega) + O(\gamma^2(h_0))\|(y-y_H,p-p_H)\|_a^2 \tag{3.17}$$

provided $h_0 \ll 1$. Moreover, there holds the global lower bound

$$\begin{aligned}C_2\eta_h^2((u_h,y_h,p_h),\Omega) \leqslant\ & \|(y-y_h,p-p_h)\|_a^2 + C_3\, osc^2((u_h,y_h,p_h),\mathcal{T}_h)\\ & + O(\gamma^2(h_0))\|(y-y_H,p-p_H)\|_a^2.\end{aligned} \tag{3.18}$$

Proof. From Lemmas 2.1 and 2.2, (3.6), (3.8), (3.15) and (3.16) we have

$$\begin{aligned}\|y^h - S_hu_h\|_{a,\Omega}^2 &\leqslant \tilde{C}_1\eta_{y,h}^2(u_h,S_hu_h,\Omega)\\ &\leqslant 2\tilde{C}_1\eta_{y,h}^2(u_h,y_h,\Omega) + 2\tilde{C}_1\eta_{y,h}^2(u_h,S_hu_h - y_h,\Omega)\\ &\leqslant 2\tilde{C}_1\eta_{y,h}^2(u_h,y_h,\Omega) + 2\tilde{C}_1C_R^2\|S_hu_h - y_h\|_{a,\Omega}^2\\ &\leqslant 2\tilde{C}_1\eta_{y,h}^2(u_h,y_h,\Omega) + 2\tilde{C}_1C_R^2(\|S_hu_h - S_hu_H\|_{a,\Omega}^2 + \|S_hu_H - y_h\|_{a,\Omega}^2)\\ &\leqslant 2\tilde{C}_1\eta_{y,h}^2(u_h,y_h,\Omega) + 2\tilde{C}^2\tilde{C}_1C_R^2\gamma^2(h_0)(\|y-y_H\|_{a,\Omega}^2 + \|y-y_h\|_{a,\Omega}^2\\ &\quad + \|p-p_H\|_{a,\Omega}^2 + \|p-p_h\|_{a,\Omega}^2).\end{aligned} \tag{3.19}$$

Similarly, we can derive from (3.7), (3.9), (3.15) and (3.16) that

$$\begin{aligned}&\|p^h - S_h^*(y_h-y_d)\|_{a,\Omega}^2 \leqslant \tilde{C}_1\eta_{p,h}^2(y_h,S_h^*(y_h-y_d),\Omega)\\ \leqslant\ & 2\tilde{C}_1\eta_{p,h}^2(y_h,p_h,\Omega) + 2\tilde{C}_1\eta_{y,h}^2(y_h,S_h^*(y_h-y_d)-p_h,\Omega)\\ \leqslant\ & 2\tilde{C}_1\eta_{p,h}^2(y_h,p_h,\Omega) + 2\tilde{C}_1C_R^2\|S_h^*(y_h-y_d)-p_h\|_{a,\Omega}^2\\ \leqslant\ & 2\tilde{C}_1\eta_{p,h}^2(y_h,p_h,\Omega) + 2\tilde{C}_1C_R^2(\|S_h^*(y_h-y_d)-S_h^*(y_H-y_d)\|_{a,\Omega}^2\\ & + \|S_h^*(y_H-y_d)-p_h\|_{a,\Omega}^2)\\ \leqslant\ & 2\tilde{C}_1\eta_{p,h}^2(y_h,p_h,\Omega) + 2\tilde{C}^2\tilde{C}_1C_R^2\gamma^2(h_0)(\|y-y_H\|_{a,\Omega}^2 + \|y-y_h\|_{a,\Omega}^2\\ & + \|p-p_H\|_{a,\Omega}^2 + \|p-p_h\|_{a,\Omega}^2).\end{aligned} \tag{3.20}$$

Combing the above estimates and Theorem 3.4 yields (3.17) with

$$C_1 = \frac{4\tilde{C}_1}{1 - 4\tilde{C}^2\gamma^2(h_0)(1+2\tilde{C}_1C_R^2)}. \tag{3.21}$$

Lemma 2.2 in [12] says that there exists a constant C_* depending on A, the mesh regularity constant and coefficient c such that

$$\mathrm{osc}(Lv,\mathcal{T}_h) \leqslant C_*\|v\|_{a,\Omega}, \quad \mathrm{osc}(L^*v,\mathcal{T}_h) \leqslant C_*\|v\|_{a,\Omega} \quad \forall v \in V_h. \tag{3.22}$$

This together with Lemma 2.3, Theorem 3.4, (3.6), (3.8), (3.15) and (3.16) implies that

$$
\begin{aligned}
&\tilde{C}_2\eta^2_{y,h}(u_h, y_h, \Omega) - 4\tilde{C}_3\mathrm{osc}^2(u_h - Ly_h, \mathcal{T}_h)\\
\leqslant\ & 2\tilde{C}_2\eta^2_{y,h}(u_h, S_hu_h, \Omega) - 2\tilde{C}_3\mathrm{osc}^2(u_h - LS_hu_h, \mathcal{T}_h)\\
&+4(C_*^2 + C_R^2)(\tilde{C}_2 + \tilde{C}_3)\|S_hu_h - y_h\|^2_{a,\Omega}\\
\leqslant\ & 2\|y^h - S_hu_h\|^2_{a,\Omega} + 4(C_*^2 + C_R^2)(\tilde{C}_2 + \tilde{C}_3)\|S_hu_h - y_h\|^2_{a,\Omega}\\
\leqslant\ & 4\|y - y_h\|^2_{a,\Omega} + 4\tilde{C}^2\gamma^2(h_0)(1 + (C_*^2 + C_R^2)(\tilde{C}_2 + \tilde{C}_3))(\|y - y_H\|^2_{a,\Omega} + \|p - p_H\|^2_{a,\Omega}\\
&+\|y - y_h\|^2_{a,\Omega} + \|p - p_h\|^2_{a,\Omega}).
\end{aligned} \tag{3.23}
$$

We can also derive that

$$
\begin{aligned}
&\tilde{C}_2\eta^2_{p,h}(y_h, p_h, \Omega) - 4\tilde{C}_3\mathrm{osc}^2(y_h - y_d - L^*p_h, \mathcal{T}_h)\\
\leqslant\ & 2\tilde{C}_2\eta^2_{p,h}(y_h, S_h^*(y_h - y_d), \Omega) - 2\tilde{C}_3\mathrm{osc}^2(y_h - y_d - L^*S_h^*(y_h - y_d), \mathcal{T}_h)\\
&+4(C_*^2 + C_R^2)(\tilde{C}_2 + \tilde{C}_3)\|S_h^*(y_h - y_d) - p_h\|^2_{a,\Omega}\\
\leqslant\ & 2\|p^h - S_h^*(y_h - y_d)\|^2_{a,\Omega} + 4(C_*^2 + C_R^2)(\tilde{C}_2 + \tilde{C}_3)\|S_h^*(y_h - y_d) - p_h\|^2_{a,\Omega}\\
\leqslant\ & 4\|p - p_h\|^2_{a,\Omega} + 4\tilde{C}^2\gamma^2(h_0)(1 + (C_*^2 + C_R^2)(\tilde{C}_2 + \tilde{C}_3))(\|y - y_H\|^2_{a,\Omega} + \|p - p_H\|^2_{a,\Omega}\\
&+\|y - y_h\|^2_{a,\Omega} + \|p - p_h\|^2_{a,\Omega}).
\end{aligned} \tag{3.24}
$$

Combing the above estimates yields (3.18) with

$$
\begin{aligned}
C_2 &= \frac{\tilde{C}_2}{4 + 4\tilde{C}^2\gamma^2(h_0)(1 + (C_*^2 + C_R^2)(\tilde{C}_2 + \tilde{C}_3))},\\
C_3 &= \frac{4\tilde{C}_3}{4 + 4\tilde{C}^2\gamma^2(h_0)(1 + (C_*^2 + C_R^2)(\tilde{C}_2 + \tilde{C}_3))}.
\end{aligned}
$$

□

There are several alternatives for MARK procedure like Max strategy or Dörfler's strategy ([9]) while we adopt the later one. Note that there are two error indicators $\eta_{y,h}(u_h, y_h, T)$ and $\eta_{p,h}(y_h, p_h, T)$ contributed to the state approximation and adjoint state approximation, respectively. We combine the two estimators as the error indicator of optimal control problems. The marking algorithm is described as follows

Algorithm 3.6. *Dörfler's marking strategy for OCPs*

1. *Given the parameter* $0 < \theta < 1$*;*
2. *Construct a minimal subset* $\tilde{\mathcal{T}}_h \subset \mathcal{T}_h$ *such that*

$$
\sum_{T\in\tilde{\mathcal{T}}_h} \eta_h^2((u_h, y_h, p_h), T) \geqslant \theta\eta_h^2((u_h, y_h, p_h), \Omega).
$$

3. *Mark all the elements in* $\tilde{\mathcal{T}}_h$.

4. Convergence of adaptive multilevel correction method for optimal control problems

In this section we intend to prove the convergence of the adaptive multilevel correction Algorithm 3.1. For the proof we follow the idea of [12] and use some results of [6, 7, 8]. Following Theorem 3.4, we will firstly establish certain relationships between the two level approximations, which will be used in our convergence analysis.

Theorem 4.1. *Let* $h, H \in (0, h_0]$ *and* $(u, y, p) \in U_{ad} \times H_0^1(\Omega) \times H_0^1(\Omega)$ *be the solution of problem (2.2). Assume that* $(u_h, y_h, p_h) \in U_{ad} \times V_{\tilde{h}} \times V_{\tilde{h}}$ *and* $(u_H, y_H, p_H) \in U_{ad} \times V_{\tilde{H}} \times V_{\tilde{H}}$

are produced by Algorithm 3.1, respectively. Define $y^H := Su_H$ *and* $p^H := S^*(y_H - y_d)$. *Then the following properties hold*

$$\begin{aligned} \|y-y_h\|_{a,\Omega} &= \|y^H - S_h u_H\|_{a,\Omega} + O(\gamma(h_0))\big(\|y-y_h\|_{a,\Omega} + \|y-y_H\|_{a,\Omega} \\ &\quad + \|p-p_h\|_{a,\Omega} + \|p-p_H\|_{a,\Omega}\big), \end{aligned} \tag{4.1}$$

$$\begin{aligned} \|p-p_h\|_{a,\Omega} &= \|p^H - S_h^*(y_H - y_d)\|_{a,\Omega} + O(\gamma(h_0))\big(\|y-y_h\|_{a,\Omega} + \|y-y_H\|_{a,\Omega} \\ &\quad + \|p-p_h\|_{a,\Omega} + \|p-p_H\|_{a,\Omega}\big), \end{aligned} \tag{4.2}$$

$$\begin{aligned} osc(u_h - Ly_h, \mathcal{T}_h) &= osc(u_H - LS_h u_H, \mathcal{T}_h) + O(\gamma(h_0))\big(\|y-y_h\|_{a,\Omega} + \|p-p_h\|_{a,\Omega} \\ &\quad + \|y-y_H\|_{a,\Omega} + \|p-p_H\|_{a,\Omega}\big), \end{aligned} \tag{4.3}$$

$$\begin{aligned} osc(y_h - y_d - L^* p_h, \mathcal{T}_h) &= osc(y_H - y_d - L^* S_h^*(y_H - y_d), \mathcal{T}_h) + O(\gamma(h_0))\big(\|y-y_h\|_{a,\Omega} \\ &\quad + \|p-p_h\|_{a,\Omega} + \|y-y_H\|_{a,\Omega} + \|p-p_H\|_{a,\Omega}\big) \end{aligned} \tag{4.4}$$

and

$$\begin{aligned} \eta_{y,h}(u_h, y_h, \Omega) &= \eta_{y,h}(u_H, S_h u_H, \Omega) + O(\gamma(h_0))\big(\|y-y_h\|_{a,\Omega} + \|y-y_H\|_{a,\Omega} \\ &\quad + \|p-p_h\|_{a,\Omega} + \|p-p_H\|_{a,\Omega}\big), \end{aligned} \tag{4.5}$$

$$\begin{aligned} \eta_{p,h}(y_h, p_h, \Omega) &= \eta_{p,h}(y_H, S_h^*(y_H - y_d), \Omega) + O(\gamma(h_0))\big(\|y-y_h\|_{a,\Omega} + \|y-y_H\|_{a,\Omega} \\ &\quad + \|p-p_h\|_{a,\Omega} + \|p-p_H\|_{a,\Omega}\big) \end{aligned} \tag{4.6}$$

provided $h_0 \ll 1$.

Proof. The proof is quite similar to the proof of Theorem 4.1 in [12] and Theorem 3.4, we omit it here. □

Now we are ready to prove the error reduction for the sum of the energy errors and the scaled error estimators of the state y and the adjoint state p plus some additional terms, between two consecutive adaptive loops.

Theorem 4.2. *Let* $(u, y, p) \in U_{ad} \times H_0^1(\Omega) \times H_0^1(\Omega)$ *be the solution of problem (2.2) and* $(u_{h_k}, y_{h_k}, p_{h_k}) \in U_{ad} \times V_{h_0,h_k} \times V_{h_0,h_k}$ *be a sequence of solutions produced by Algorithm 3.1. Then there exist constants* $\gamma, \beta_0 > 0$ *and* $\beta \in (0,1)$ *depending only on the shape regularity of meshes and the parameter* θ *used by Algorithm 3.6, such that for any two consecutive iterates* k *and* $k+1$*, we have*

$$\begin{aligned} &\|(y-y_{h_{k+1}}, p-p_{h_{k+1}})\|_a^2 + \gamma\eta_{h_{k+1}}^2((u_{h_{k+1}}, y_{h_{k+1}}, p_{h_{k+1}}), \Omega) \\ \leqslant\ &\beta^2\Big(\|(y-y_{h_k}, p-p_{h_k})\|_a^2 + \gamma\eta_{h_k}^2((u_{h_k}, y_{h_k}, p_{h_k}), \Omega)\Big) \\ &+\beta_0^2\gamma^2(h_0)\|(y-y_{h_{k-1}}, p-p_{h_{k-1}})\|, \quad \text{for } k \geqslant 2 \end{aligned} \tag{4.7}$$

and

$$\begin{aligned} &\|(y-y_{h_2}, p-p_{h_2})\|_a^2 + \gamma\eta_{h_2}^2((u_{h_2}, y_{h_2}, p_{h_2}), \Omega) \\ \leqslant\ &\beta^2\Big(\|(y-y_{h_1}, p-p_{h_1})\|_a^2 + \gamma\eta_{h_1}^2((u_{h_1}, y_{h_1}, p_{h_1}), \Omega)\Big) \end{aligned} \tag{4.8}$$

provided $h_0 \ll 1$.

Proof. At first we prove the case $k \geqslant 2$. For convenience, we use (u_h, y_h, p_h), (u_H, y_H, p_H) and $(u_{H_{-1}}, y_{H_{-1}}, p_{H_{-1}})$ to denote $(u_{h_{k+1}}, y_{h_{k+1}}, p_{h_{k+1}})$, $(u_{h_k}, y_{h_k}, p_{h_k})$ and $(u_{h_{k-1}}, y_{h_{k-1}}, p_{h_{k-1}})$, respectively. So it suffices to prove that

$$\begin{aligned} &\|(y-y_h, p-p_h)\|_a^2 + \gamma\eta_h^2((u_h, y_h, p_h), \Omega) \\ \leqslant\ &\beta^2\Big(\|(y-y_H, p-p_H)\|_a^2 + \gamma\eta_H^2((u_H, y_H, p_H), \Omega)\Big) \\ &+\beta_0^2\gamma^2(h_0)\|(y-y_{H_{-1}}, p-p_{H_{-1}})\| \end{aligned} \tag{4.9}$$

holds for $\gamma, \beta_0 > 0$ and $\beta \in (0,1)$.

Recall that $y^H := Su_H$, $p^H := S^*(y_H - y_d)$ and $S_h u_H$, $S_h^*(y_H - y_d)$ are their finite element approximations in V_h. So we conclude from Theorem 2.6 in [12] that there exist constants $\tilde{\gamma}$ and $\tilde{\beta} \in (0,1)$ satisfying (see also [7, Theorem 2.4])

$$\begin{aligned} & \|(y^H - S_h u_H, p^H - S_h^*(y_H - y_d))\|_a^2 + \tilde{\gamma}\big(\eta_{y,h}^2(u_H, S_h u_H, \Omega) \\ & +\eta_{p,h}^2(y_H, S_h^*(y_H - y_d), \Omega)\big) \\ \leqslant\ & \tilde{\beta}^2\Big(\|(y^H - S_H u_H, p^H - S_H^*(y_H - y_d))\|_a^2 \\ & +\tilde{\gamma}\big(\eta_{y,H}^2(u_H, S_H u_H, \Omega) + \eta_{p,H}^2(y_H, S_H^*(y_H - y_d), \Omega)\big)\Big). \end{aligned} \tag{4.10}$$

We note that the above result is the standard error reduction property of AFEM for elliptic boundary value problems consisting of the state and adjoint state equations.

It follows from (4.1)-(4.2) and (4.5)-(4.6) that there exists a constant $\tilde{C}_4 > 0$ such that

$$\begin{aligned} & \|(y - y_h, p - p_h)\|_a^2 + \tilde{\gamma}\eta_h^2((u_h, y_h, p_h), \Omega) \\ \leqslant\ & (1+\delta_1)\|(y^H - S_h u_H, p^H - S_h^*(y_H - y_d))\|_a^2 \\ & +(1+\delta_1)\tilde{\gamma}\big(\eta_{y,h}^2(u_H, S_h u_H, \Omega) + \eta_{p,h}^2(y_H, S_h^*(y_H - y_d), \Omega)\big) \\ & +\tilde{C}_4(1+\delta_1^{-1})\gamma^2(h_0)\Big(\|(y - y_h, p - p_h)\|_a^2 + \|(y - y_H, p - p_H)\|_a^2\Big) \\ & +\tilde{C}_4(1+\delta_1^{-1})\gamma^2(h_0)\tilde{\gamma}\Big(\|(y - y_h, p - p_h)\|_a^2 + \|(y - y_H, p - p_H)\|_a^2\Big), \end{aligned}$$

where the δ_1-Young inequality is used and $\delta_1 \in (0,1)$ satisfies

$$(1+\delta_1)\tilde{\beta}^2 < 1. \tag{4.11}$$

Thus, there exists a positive constant $\tilde{C}_5$ depending on $\tilde{C}_4$ and $\tilde{\gamma}$ such that

$$\begin{aligned} & \|(y - y_h, p - p_h)\|_a^2 + \tilde{\gamma}\eta_h^2((u_h, y_h, p_h), \Omega) \\ \leqslant\ & (1+\delta_1)\Big(\|(y^H - S_h u_H, p^H - S_h^*(y_H - y_d))\|_a^2 \\ & +\tilde{\gamma}\big(\eta_{y,h}^2(u_H, S_h u_H, \Omega) + \eta_{p,h}^2(y_H, S_h^*(y_H - y_d), \Omega)\big)\Big) \\ & +\tilde{C}_5\delta_1^{-1}\gamma^2(h_0)\Big(\|(y - y_h, p - p_h)\|_a^2 + \|(y - y_H, p - p_H)\|_a^2\Big). \end{aligned} \tag{4.12}$$

We combine the estimates (4.10) and (4.12) to derive

$$\begin{aligned} & \|(y - y_h, p - p_h)\|_a^2 + \tilde{\gamma}\eta_h^2((u_h, y_h, p_h), \Omega) \\ \leqslant\ & (1+\delta_1)\tilde{\beta}^2\Big(\|(y^H - S_H u_H, p^H - S_H^*(y_H - y_d))\|_a^2 \\ & +\tilde{\gamma}\big(\eta_{y,H}^2(u_H, S_H u_H, \Omega) + \eta_{p,H}^2(y_H, S_H^*(y_H - y_d), \Omega)\big)\Big) \\ & +\tilde{C}_5\delta_1^{-1}\gamma^2(h_0)\Big(\|(y - y_h, p - p_h)\|_a^2 + \|(y - y_H, p - p_H)\|_a^2\Big). \end{aligned} \tag{4.13}$$

Similar to the proof of (3.6)-(3.9) and (3.15) we can derive that

$$\begin{aligned} & \|S_H u_H - y_H\|_{a,\Omega}^2 + \|S_H^*(y_H - y_d) - p_H\|_{a,\Omega}^2 \\ \leqslant\ & \|S_H u_H - S_H u_{H_{-1}}\|_{a,\Omega}^2 + \|S_H u_{H_{-1}} - y_H\|_{a,\Omega}^2 \\ & +\|S_H^*(y_H - y_d) - S_H^*(y_{H_{-1}} - y_d)\|_{a,\Omega}^2 + \|S_H^*(y_{H_{-1}} - y_d) - p_H\|_{a,\Omega}^2 \\ \leqslant\ & \tilde{C}\gamma^2(h_0)(\|y - y_H\|_{a,\Omega}^2 + \|y - y_{H_{-1}}\|_{a,\Omega}^2 + \|p - p_H\|_{a,\Omega}^2 + \|p - p_{H_{-1}}\|_{a,\Omega}^2). \end{aligned}$$

Using Theorem 3.4 and Lemma 2.1 we arrive at

$$\begin{aligned} & \|(y^H - S_H u_H, p^H - S_H^*(y_H - y_d))\|_a^2 \\ & +\tilde{\gamma}\big(\eta_{y,H}^2(u_H, S_H u_H, \Omega) + \eta_{p,H}^2(y_H, S_H^*(y_H - y_d), \Omega)\big) \\ \leqslant\ & (1+\delta_2)\|(y - y_H, p - p_H)\|_a^2 + (1+\delta_2)\tilde{\gamma}\big(\eta_{y,H}^2(u_H, y_H, \Omega) + \eta_{p,H}^2(y_H, p_H, \Omega)\big) \end{aligned}$$

$$
\begin{aligned}
& +(1+\delta_2^{-1})C_R^2\tilde{\gamma}(\|S_Hu_H-y_H\|_{a,\Omega}^2+\|S_H^*(y_H-y_d)-p_H\|_{a,\Omega}^2) \\
& +\tilde{C}_6(1+\delta_2^{-1})\gamma^2(h_0)\Big(\|(y-y_H,p-p_H)\|_a^2+\|(y-y_{H_{-1}},p-p_{H_{-1}})\|_a^2\Big) \\
\leqslant\ & (1+\delta_2)\|(y-y_H,p-p_H)\|_a^2+(1+\delta_2)\tilde{\gamma}\eta_H^2((u_H,y_H,p_H),\Omega) \\
& +\tilde{C}_6(1+\delta_2^{-1})\gamma^2(h_0)\Big(\|(y-y_H,p-p_H)\|_a^2+\|(y-y_{H_{-1}},p-p_{H_{-1}})\|_a^2\Big) \\
& +\tilde{C}C_R^2(1+\delta_2^{-1})\gamma^2(h_0)\tilde{\gamma}\Big(\|(y-y_H,p-p_H)\|_a^2+\|(y-y_{H_{-1}},p-p_{H_{-1}})\|_a^2\Big),
\end{aligned}
$$

where the δ_2-Young inequality is used with $\delta_2\in(0,1)$ satisfying

$$
(1+\delta_1)(1+\delta_2+\tilde{C}_7\delta_2^{-1}\gamma^2(h_0))\tilde{\beta}^2<1. \tag{4.14}
$$

Then we have

$$
\begin{aligned}
& \|(y^H-S_Hu_H,p^H-S_H^*(y_H-y_d))\|_a^2 \\
& +\tilde{\gamma}\big(\eta_{y,H}^2(u_H,S_Hu_H,\Omega)+\eta_{p,H}^2(y_H,S_H^*(y_H-y_d),\Omega)\big) \\
\leqslant\ & (1+\delta_2)\Big(\|(y-y_H,p-p_H)\|_a^2+\tilde{\gamma}\eta_H^2((u_H,y_H,p_H),\Omega)\Big) \\
& +\tilde{C}_7\delta_2^{-1}\gamma^2(h_0)\Big(\|(y-y_H,p-p_H)\|_a^2+\|(y-y_{H_{-1}},p-p_{H_{-1}})\|_a^2\Big),
\end{aligned} \tag{4.15}
$$

where $\tilde{C}_7$ depends on $\tilde{C}_6$, $\tilde{C}$, C_R and $\tilde{\gamma}$. It follows from (4.13) and (4.15) that

$$
\begin{aligned}
& \|(y-y_h,p-p_h)\|_a^2+\tilde{\gamma}\eta_h^2((u_h,y_h,p_h),\Omega) \\
\leqslant\ & (1+\delta_1)\tilde{\beta}^2\Big((1+\delta_2)\big(\|(y-y_H,p-p_H)\|_a^2+\tilde{\gamma}\eta_H^2((u_H,y_H,p_H),\Omega)\big) \\
& +\tilde{C}_7\delta_2^{-1}\gamma^2(h_0)\big(\|(y-y_H,p-p_H)\|_a^2+\|(y-y_{H_{-1}},p-p_{H_{-1}})\|_a^2\big)\Big) \\
& +\tilde{C}_5\delta_1^{-1}\gamma^2(h_0)\Big(\|(y-y_h,p-p_h)\|_a^2+\|(y-y_H,p-p_H)\|_a^2\Big),
\end{aligned}
$$

and thus

$$
\begin{aligned}
& (1-\tilde{C}_5\delta_1^{-1}\gamma^2(h_0))\|(y-y_h,p-p_h)\|_a^2+\tilde{\gamma}\eta_h^2((u_h,y_h,p_h),\Omega) \\
\leqslant\ & \big((1+\delta_1)(1+\delta_2+\tilde{C}_7\delta_2^{-1}\gamma^2(h_0))\tilde{\beta}^2+\tilde{C}_5\delta_1^{-1}\gamma^2(h_0)\big)\|(y-y_H,p-p_H)\|_a^2 \\
& +(1+\delta_1)(1+\delta_2)\tilde{\gamma}\tilde{\beta}^2\eta_H^2((u_H,y_H,p_H),\Omega) \\
& +(1+\delta_1)\tilde{\beta}^2\tilde{C}_7\delta_2^{-1}\gamma^2(h_0)\|(y-y_{H_{-1}},p-p_{H_{-1}})\|_a^2.
\end{aligned} \tag{4.16}
$$

This gives

$$
\begin{aligned}
& \|(y-y_h,p-p_h)\|_a^2+\frac{\tilde{\gamma}}{1-\tilde{C}_5\delta_1^{-1}\gamma^2(h_0)}\eta_h^2((u_h,y_h,p_h),\Omega) \\
\leqslant\ & \frac{(1+\delta_1)(1+\delta_2+\tilde{C}_7\delta_2^{-1}\gamma^2(h_0))\tilde{\beta}^2+\tilde{C}_5\delta_1^{-1}\gamma^2(h_0)}{1-\tilde{C}_5\delta_1^{-1}\gamma^2(h_0)}\|(y-y_H,p-p_H)\|_a^2 \\
& +\frac{(1+\delta_1)(1+\delta_2)\tilde{\gamma}\tilde{\beta}^2}{1-\tilde{C}_5\delta_1^{-1}\gamma^2(h_0)}\eta_H^2((u_H,y_H,p_H),\Omega) \\
& +\frac{(1+\delta_1)\tilde{\beta}^2\tilde{C}_7\delta_2^{-1}\gamma^2(h_0)}{1-\tilde{C}_5\delta_1^{-1}\gamma^2(h_0)}\|(y-y_{H_{-1}},p-p_{H_{-1}})\|_a^2.
\end{aligned} \tag{4.17}
$$

Since $\gamma(h_0)\ll 1$ provided that $h_0\ll 1$, we can define the constant β as

$$
\beta:=\Big(\frac{(1+\delta_1)(1+\delta_2+\tilde{C}_7\delta_2^{-1}\gamma^2(h_0))\tilde{\beta}^2+\tilde{C}_5\delta_1^{-1}\gamma^2(h_0)}{1-\tilde{C}_5\delta_1^{-1}\gamma^2(h_0)}\Big)^{\frac{1}{2}}, \tag{4.18}
$$

which satisfies $\beta\in(0,1)$ if $h_0\ll 1$. Then

$$
\|(y-y_h,p-p_h)\|_a^2+\frac{\tilde{\gamma}}{1-\tilde{C}_5\delta_1^{-1}\gamma^2(h_0)}\eta_h^2((u_h,y_h,p_h),\Omega)\leqslant\beta^2\Big(\|(y-y_H,p-p_H)\|_a^2
$$

$$+\frac{(1+\delta_1)(1+\delta_2)\tilde{\gamma}\tilde{\beta}^2}{(1+\delta_1)(1+\delta_2+\tilde{C}_7\delta_2^{-1}\gamma^2(h_0))\tilde{\beta}^2+\tilde{C}_5\delta_1^{-1}\gamma^2(h_0)}\eta_H^2((u_H,y_H,p_H),\Omega)\Big)$$

$$+\frac{(1+\delta_1)\tilde{\beta}^2\tilde{C}_7\delta_2^{-1}\gamma^2(h_0)}{1-\tilde{C}_5\delta_1^{-1}\gamma^2(h_0)}\|(y-y_{H_{-1}},p-p_{H_{-1}})\|_a^2. \tag{4.19}$$

Now we choose

$$\gamma:=\frac{\tilde{\gamma}}{1-\tilde{C}_5\delta_1^{-1}\gamma^2(h_0)}, \tag{4.20}$$

it is obvious that

$$\frac{(1+\delta_1)(1+\delta_2)\tilde{\gamma}\tilde{\beta}^2}{(1+\delta_1)(1+\delta_2+\tilde{C}_7\delta_2^{-1}\gamma^2(h_0))\tilde{\beta}^2+\tilde{C}_5\delta_1^{-1}\gamma^2(h_0)}<\gamma.$$

We set

$$\beta_0=\Big(\frac{(1+\delta_1)\tilde{\beta}^2\tilde{C}_7\delta_2^{-1}}{1-\tilde{C}_5\delta_1^{-1}\gamma^2(h_0)}\Big)^{\frac{1}{2}}, \tag{4.21}$$

this completes the proof of (4.9). The proof of (4.8) is very similar and we omit it here. □

Now we are ready to give the final convergence result.

Theorem 4.3. *Let $(u,y,p)\in U_{ad}\times H_0^1(\Omega)\times H_0^1(\Omega)$ be the solution of problem (2.2) and $(u_{h_k},y_{h_k},p_{h_k})\in U_{ad}\times V_{h_0,h_k}\times V_{h_0,h_k}$ be a sequence of solutions produced by Algorithm 3.1. Then there exist constants $\varrho>0$ and $\bar{\beta}\in(0,1)$ depending only on the shape regularity of meshes, β and the parameter θ used by Algorithm 3.6, such that for any two consecutive iterates k and $k+1$, we have*

$$E_{h_{k+1}}^2+\varrho^2\gamma^2(h_0)E_{h_k}^2 \leqslant \bar{\beta}^2(E_{h_k}^2+\varrho^2\gamma^2(h_0)E_{h_{k-1}}^2),\quad \textit{for } k\geqslant 2, \tag{4.22}$$

$$E_{h_2}^2 \leqslant \bar{\beta}^2E_{h_1}^2 \tag{4.23}$$

provided that $h_0\ll 1$, where $E_{h_k}^2=\|(y-y_{h_k},p-p_{h_k})\|_a^2+\gamma\eta_{h_k}^2((u_{h_k},y_{h_k},p_{h_k}),\Omega)$. Then the adaptive Algorithm 3.1 converges with a linear rate $\bar{\beta}$, i.e., the k-th iteration solution $(u_{h_k},y_{h_k},p_{h_k})$ of Algorithm 3.1 has the following property:

$$E_{h_k}^2+\varrho^2\gamma^2(h_0)E_{h_{k-1}}^2 \leqslant C_0\bar{\beta}^{2(k-1)}, \tag{4.24}$$

where $C_0=\|(y-y_{h_1},p-p_{h_1})\|_a^2+\gamma\eta_{h_1}^2((u_{h_1},y_{h_1},p_{h_1}),\Omega)$ and $k\geqslant 2$.

Proof. From Theorem 4.2 we have

$$E_{h_{k+1}}^2 \leqslant \beta^2E_{h_k}^2+\beta_0^2\gamma^2(h_0)\|(y-y_{h_{k-1}},p-p_{h_{k-1}})\|_a^2,\quad \text{for } k\geqslant 2, \tag{4.25}$$

$$E_{h_2}^2 \leqslant \beta^2E_{h_1}^2. \tag{4.26}$$

Let $\bar{\beta}$ and ϱ satisfy the following properties

$$\bar{\beta}^2-\varrho^2\gamma^2(h_0)=\beta^2,$$
$$\bar{\beta}^2\varrho^2=\beta_0^2.$$

It is clear that the solutions can be written as

$$\bar{\beta}^2=\frac{\beta^2+\sqrt{\beta^4+4\beta_0^2\gamma^2(h_0)}}{2},\quad \varrho^2=\frac{2\beta_0^2}{\beta^2+\sqrt{\beta^4+4\beta_0^2\gamma^2(h_0)}}.$$

Note that $\beta<1$, if $h_0\ll 1$ we can conclude that $\bar{\beta}<1$. Thus, (4.22) and (4.23) are the direct consequences of (4.25)-(4.26) with the chosen ϱ and $\bar{\beta}$. The proof of (4.24) is obvious by simple calculation. □

We also would like to analyze the complexity of Algorithm 3.1. Similar to [6] and [8], for our purpose to analyze the complexity of AFEM for optimal control problems we need to introduce a function approximation class as follows

$$\mathcal{A}_\gamma^s := \Big\{(y,p,y_d) \in H_0^1(\Omega)\times H_0^1(\Omega)\times L^2(\Omega): \quad |(y,p,y_d)|_{s,\gamma} < +\infty\Big\},$$

where $\gamma > 0$ is some constant and

$$|(y,p,y_d)|_{s,\gamma} = \sup_{\varepsilon>0}\varepsilon \inf_{\mathcal{T}\subset\mathcal{T}_{h_1}:\ \inf(\|(y-y_\mathcal{T},p-p_\mathcal{T})\|_a^2+(\gamma+1)\mathrm{osc}^2((u_\mathcal{T},y_\mathcal{T},p_\mathcal{T}),\mathcal{T}))^{1/2}\leqslant\varepsilon}(\#\mathcal{T}-\#\mathcal{T}_{h_1})^s.$$

Here $\mathcal{T}\subset\mathcal{T}_{h_1}$ means $\mathcal{T}$ is a refinement of $\mathcal{T}_{h_1}$, $y_\mathcal{T}$ and $p_\mathcal{T}$ are elements of the finite element space corresponding to the partition $\mathcal{T}$. It is seen from the definition that $\mathcal{A}_\gamma^s = \mathcal{A}_1^s$ for all $\gamma > 0$, thus we use $\mathcal{A}^s$ throughout the paper with corresponding norm $|\cdot|_s$. So $\mathcal{A}^s$ is the class of functions that can be approximated with a given tolerance ε by continuous piecewise linear polynomial functions over a partition $\mathcal{T}$ with number of degrees of freedom $\#\mathcal{T}-\#\mathcal{T}_{h_1}\lesssim\varepsilon^{-1/s}|v|_s^{1/s}$.

To begin with, we assume the initial mesh size h_0 is small enough such that

$$\gamma(h_0)\|(y-y_{h_{k-1}},p-p_{h_{k-1}})\|_a^2 \leqslant \|(y-y_{h_k},p-p_{h_k})\|_a^2. \tag{4.27}$$

With the above assumption we can conclude from Theorem 4.2 that

$$\begin{aligned} &\|(y-y_{h_{k+1}},p-p_{h_{k+1}})\|_a^2+\gamma\eta_{h_{k+1}}^2((u_{h_{k+1}},y_{h_{k+1}},p_{h_{k+1}}),\Omega)\\ \leqslant\ & \delta^2\Big(\|(y-y_{h_k},p-p_{h_k})\|_a^2+\gamma\eta_{h_k}^2((u_{h_k},y_{h_k},p_{h_k}),\Omega)\Big)\end{aligned}$$

with $\delta^2=\beta^2+\beta_0^2\gamma(h_0)$ when h_0 is small enough. We note that the above error reduction property of Algorithm 3.1 is the same as Theorem 4.2 in [12] for standard AFEM. By using the similar technique we can prove that Algorithm 3.1 possesses optimal complexity for the state and adjoint state approximations. The details can be found in [12].

Theorem 4.4. *Let $(u,y,p)\in U_{ad}\times H_0^1(\Omega)\times H_0^1(\Omega)$ be the solution of problem (2.2) and $(u_{h_n},y_{h_n},p_{h_n})\in U_{ad}\times V_{h_n}\times V_{h_n}$ be a sequence of solutions of problem (2.6)-(2.7) corresponding to a sequence of finite element spaces V_{h_n} with partitions $\mathcal{T}_{h_n}$ produced by Algorithm 3.1. Then the n-th iterate solution (y_{h_n},p_{h_n}) of Algorithm 3.1 satisfies the optimal bound*

$$\|(y-y_{h_n},p-p_{h_n})\|_a^2+\gamma osc^2((u_{h_n},y_{h_n},p_{h_n}),\mathcal{T}_{h_n})\lesssim(\#\mathcal{T}_{h_n}-\#\mathcal{T}_{h_1})^{-2s}, \tag{4.28}$$

where the hidden constant depends on the exact solution (u,y,p) and θ, C_1, C_2, C_3 and γ.

Remark 4.5. *In Theorems 4.3 and 4.4 we have proved the convergence and quasi-optimality of the adaptive algorithm for the state and adjoint state approximations. We note that Theorem 4.3 also implies the convergence of $\|u-u_{h_k}\|_{0,\Omega}$ in view of the estimate (3.15), namely, for the k-th iterate solution u_{h_k} $(k\geqslant 2)$ of Algorithm 3.1 there holds*

$$\|u-u_{h_k}\|_{0,\Omega}^2\lesssim C_0\bar\beta^{2(k-1)}. \tag{4.29}$$

Moreover, the control variable can also be included into the complexity analysis of AFEM for optimal control problems to obtain

$$\|u-u_{h_k}\|_{0,\Omega}^2\lesssim(\#\mathcal{T}_{h_k}-\#\mathcal{T}_{h_1})^{-2s}. \tag{4.30}$$

However, as pointed out in [12], the above results are sub-optimal for the optimal control, which can also be observed from the numerical results in Section 5. To prove the optimality of AFEM for control variable it seems that we need to work with AFEM based on L^2-norm error estimators, this becomes more clear if compared with the optimal a priori error estimates, see, e.g., [15].

5. Numerical Examples

In the final section we carry out some numerical experiments to demonstrate the efficiency of our proposed adaptive multilevel correction algorithm and to validate our theoretical results.

Example 5.1. *We consider an example defined on the L-shaped domain* $\Omega = (-1,1)^2\backslash([0,1]\times(-1,0])$. *We set* $L = -\Delta$ *with homogeneous Dirichlet boundary condition. We take the exact solutions as*

$$\begin{aligned} y(r,\vartheta) &= (r^2\cos^2(\vartheta)-1)(r^2\sin^2(\vartheta)-1)r^\lambda\sin(\lambda\vartheta), \\ p(r,\vartheta) &= \alpha(r^2\cos^2(\vartheta)-1)(r^2\sin^2(\vartheta)-1)r^\lambda\sin(\lambda\vartheta), \\ u(r,\vartheta) &= P_{U_{ad}}(-\frac{p}{\alpha}) \end{aligned}$$

with $\lambda = \frac{2}{3}$, *where* (r,ϑ) *denotes the polar coordinate. We set* $\alpha = 0.1$, $a = -0.3$ *and* $b = 0.5$. *We assume the additional right hand side* f *for the state equation.*

We refer to [1] for a similar example, here the modifications are made to preserve homogeneous Dirichlet boundary conditions on the L-shaped domain. For our computations we set the tolerance of the stopping rule of projected gradient method as $1.0e-8$. We use $P_{U_{ad}}(-\frac{p^*_{h_{k+1}}}{\alpha})$ as the initial guess for the solution of the optimization problem in step (7) of Algorithm 3.1 while $p^*_{h_{k+1}}$ is the solution of the BVP in step (6). We note that only three iterations are needed for the solutions of the coarse optimization problem.

We give the numerical results for the optimal control approximation by Algorithm 3.1 with parameter $\theta = 0.2$ and $\theta = 0.4$, respectively. We note that the adaptive algorithm with smaller θ yields more optimal mesh distribution while increases the adaptive loops. In Figure 1 we plot the profiles of the numerically computed optimal state and control with $\theta = 0.2$, while the profile of the adjoint state is similar to the state with a scaling parameter α. We present in Figure 2 the meshes by Algorithm 3.1 after 10 and 15 adaptive iterations with $\theta = 0.2$. We can see that there are more grid points around the reentrant corner where the singularities located. In Figure 3 we also illustrate the active sets of the continuous solution, the discrete solutions with variational control discretization and piecewise linear control discretization. In this example only the lower bound $u \geqslant -0.3$ is active. Figure 3 clearly shows that the active set crosses element edges and is not restricted to finite element edges by our variational discretization for control u, and is much closer to that of the continuous solution compared with full control discretization.

To illustrate the advantage of finite element approximations on adaptive mesh over uniform mesh for solving optimal control problems, we show in the left plot of Figure 4 the error history of the optimal control, state and adjoint state with uniform refinement. We can only observe the reduced orders of convergence which are less than one for the energy norms of the state and adjoint state, and less than two for the L^2-norm of the control. In the right plot of Figure 4 we present the convergence behaviours of the optimal control, state and adjoint state, as well as the error estimators $\eta_{y,h}(y_h,\Omega)$ and $\eta_{p,y}(p_h,\Omega)$ for the state and adjoint state equations with adaptive refinement. In Figure 5 we present the convergence of the error $\|(y-y_h,p-p_h)\|_a$ and error indicator $\eta_h((y_h,p_h),\Omega)$ with $\theta = 0.2$ and $\theta = 0.4$, respectively. It is shown from Figure 5 that the error $\|(y-y_h,p-p_h)\|_a$ is proportional to the a posteriori error estimators, which implies the efficiency of the a posteriori error estimators given in Section 3. Moreover, we can also observe that the convergence order of error $\|(y-y_h,p-p_h)\|_a$ is approximately parallel to the line with slope $-1/2$ which is the optimal convergence rate we can expect by using linear finite elements, this coincides with our theory in Section 4. For the error $\|u-u_h\|_{0,\Omega}$ we can observe the reduction with slope -1, which is better than the results presented in Remark 4.5, and strongly suggests that the convergence rate for the optimal control is not optimal.

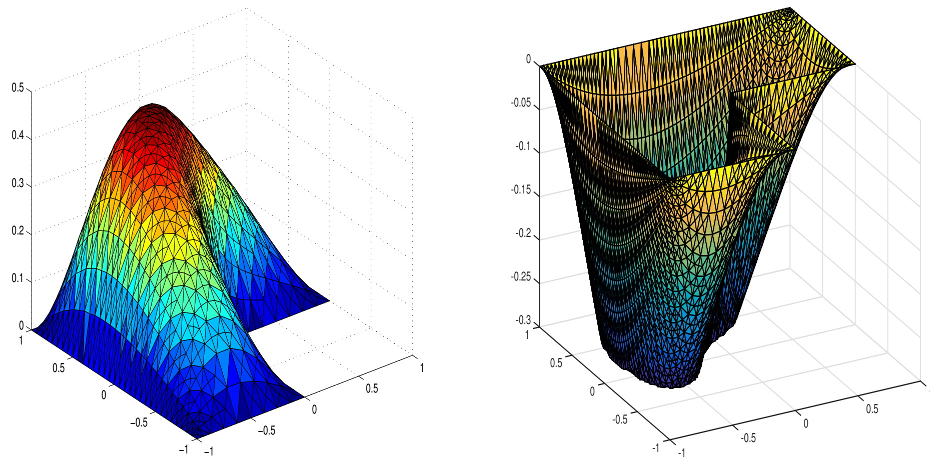

FIGURE 1. The profiles of the discretized optimal state y_h (left) and optimal control u_h (right) for Example 5.1 on adaptively refined mesh.

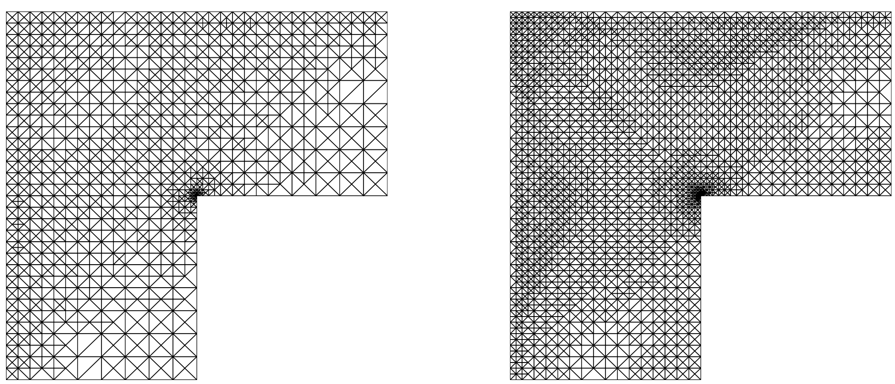

FIGURE 2. The meshes after 10 (left) and 15 (right) adaptive iterations for Example 5.1 generated by Algorithm 3.1 with $\theta = 0.2$.

Acknowledgments

The authors would like to thank two anonymous referees for their careful reviews and many valuable suggestions. The first author was supported by the National Basic Research Program of China under grant 2012CB821204, the National Natural Science Foundation of China under grant 11201464 and 91530204. The second author gratefully acknowledges the support of the National Natural Science Foundation of China (91330202, 11371026, 11001259, 11031006, 2011CB309703), the National Center for Mathematics and Interdisciplinary Science, CAS and the President Foundation of AMSS-CAS. The third author was supported by the National Natural Science Foundation of China under grant 11571356 and 91530204.

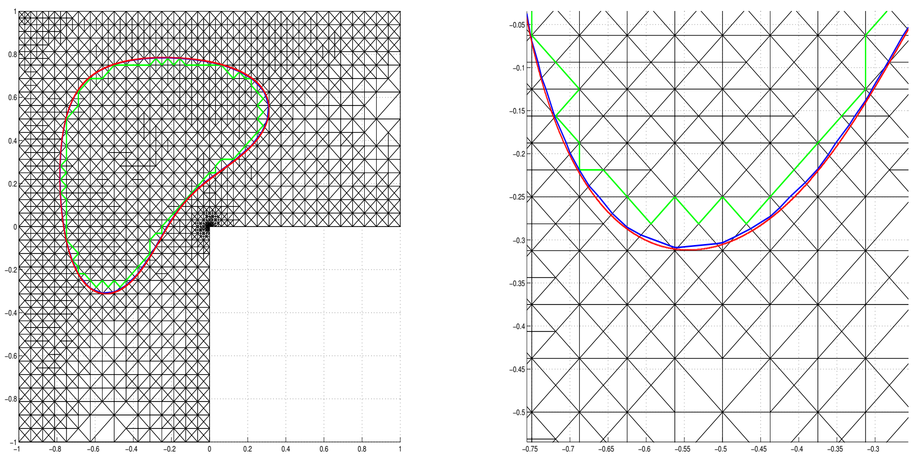

FIGURE 3. Left: The red line depicts the boarder of the active set of the continuous solution, the blue lines depict the boarder of the active set when using variational discretization, and the green line depicts the boarder of the active set obtained by using piecewise linear, continuous controls. Right: Zoom.

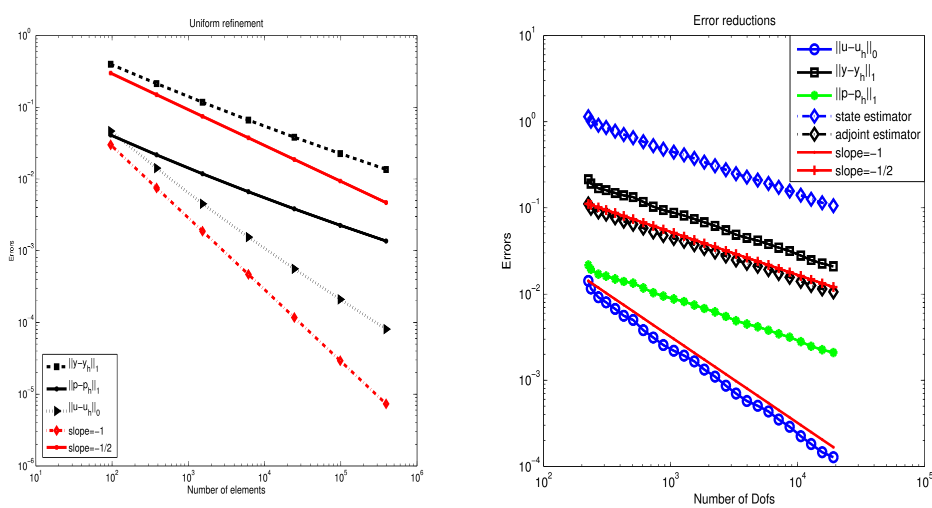

FIGURE 4. The convergence history of the optimal control, state and adjoint state on uniformly refined meshes (left), and the convergence of the errors and estimators on adaptively refined meshes (right) for Example 5.1 generated by Algorithm 3.1.

REFERENCES

[1] T. Apel, A. Rösch, G. Winkler, Optimal control in non-convex domains: a priori discretization error estimates, Calcolo, 44 (2007), no. 3, pp. 137-158.

[2] I. Babuška and W.C. Rheinboldt, Error estimates for adaptive finite element computations, SIAM J. Numer. Anal., 15 (1978), pp. 736-754.

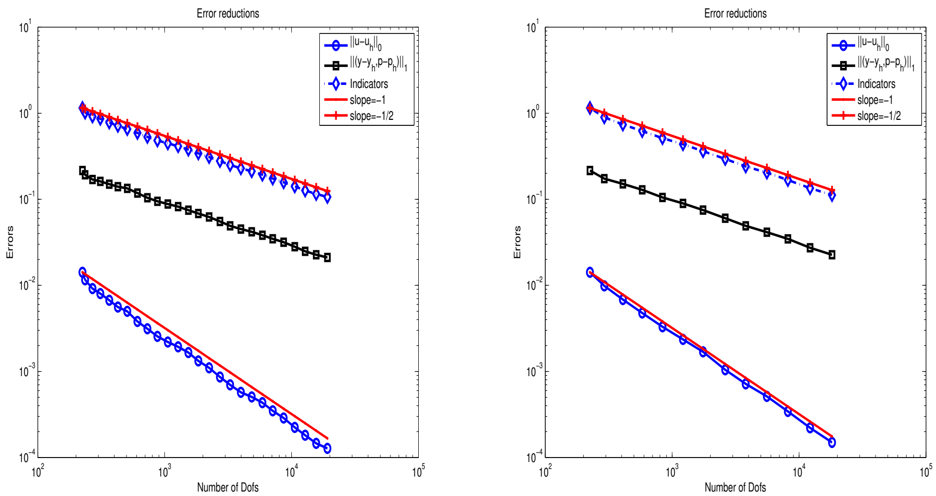

FIGURE 5. The convergence history of the optimal control, the state and adjoint state and error indicator on adaptively refined meshes with $\theta = 0.2$ (left) and $\theta = 0.4$ (right) for Example 5.1 generated by Algorithm 3.1.

[3] R. Becker, H. Kapp and R. Rannacher, Adaptive finite element methods for optimal control of partial differential equations: basic concept, SIAM J. Control Optim., 39 (2000), no. 1, pp. 113-132.

[4] P. Binev, W. Dahmen and R. DeVore, Adaptive finite element methods with convergence rates, Numer. Math., 97 (2004), pp. 219-268.

[5] A. Borzi and V. Schulz, Multigrid methods for PDE optimization, SIAM Review, 51 (2009), pp. 361-395.

[6] J.M. Cascon, C. Kreuzer, R.H. Nochetto and K.G. Siebert, Quasi-optimal convergence rate for an adaptive finite element method, SIAM J. Numer. Anal., 46 (2008), no. 5, pp. 2524-2550.

[7] X.Y. Dai, L.H. He and A.H. Zhou, Convergence and quasi-optimal complexity of adaptive finite element computations for multiple eigenvalues, IMA J. Numer. Anal., 35 (2015), no. 4, pp. 1934-1977.

[8] X.Y. Dai, J.C. Xu and A.H. Zhou, Convergence and optimal complexity of adaptive finite element eigenvalue computations, Numer. Math., 110 (2008), pp. 313-355.

[9] W. Dörfler, A convergent adaptive algorithm for Poisson's equation, SIAM J. Numer. Anal., 33 (1996), pp. 1106-1124.

[10] A. Gaevskaya, R.H.W. Hoppe, Y. Iliash and M. Kieweg, Convergence analysis of an adaptive finite element method for distributed control problems with control constraints, in Control of coupled partial differential equations, vol. 155 of Internat. Ser. Numer. Math., Birkhäuser, Basel, 2007, pp. 47-68.

[11] W. Gong, H. Xie and N. Yan, A multilevel correction method for optimal controls of elliptic equation, SIAM J. Sci. Comput., 37 (2015), no. 5, pp. A2198-A2221.

[12] W. Gong and N. Yan, Adaptive finite element method for elliptic optimal control problems: convergence and optimality, arXiv:1505.01632, Numer. Math., 2016, DOI:10.1007/s00211-016-0827-9.

[13] M. Hintermüller, R.H.W. Hoppe, Y. Iliash and M. Kieweg, An a posteriori error analysis of adaptive finite element methods for distributed elliptic control problems with control constraints, ESAIM Control Optim. Calc. Var., 14 (2008), pp. 540-560.

[14] M. Hintermüller, K. Ito and K. Kunisch, The primal-dual active set strategy as a semismooth Newton method, SIAM J. Optim., 13 (2003), pp. 865-888.

[15] M. Hinze, A variational discretization concept in control constrained optimization: the linear-quadratic case, Comput. Optim. Appl., 30 (2005), pp. 45-63.

[16] M. Hinze, R. Pinnau, M. Ulbrich and S. Ulbrich, Optimization with PDE Constraints, Math. Model. Theo. Appl., 23, Springer, New York, 2009.

[17] K. Kohls, A. Rösch and K.G. Siebert, A posteriori error analysis of optimal control problems with control constraints, SIAM J. Control Optim., 52 (2014), no. 3, pp. 1832-1861.

[18] R.M. Lewis and S.G. Nash, Model problems for the multigrid optimization of systems governed by differential equations, SIAM J. Sci. Comput., 26 (2005), pp. 1811-1837.

[19] R. Li, W. Liu, H. Ma and T. Tang, Adaptive finite element approximation for distributed elliptic optimal control problems, SIAM J. Control Optim., 41 (2002), no. 5, pp. 1321-1349.

[20] Q. Lin and H. Xie, A multilevel correction type of adaptive finite element method for Steklov eigenvalue problems, Proceedings of the International Conference Applications of Mathematics 2012, pp. 134-143.

[21] Q. Lin and H. Xie, A multi-level correction scheme for eigenvalue problems, Math. Comput., 84 (2015), no. 291, pp. 71-88.

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

[23] W. Liu and N. Yan, A posteriori error estimates for optimal boundary control, SIAM J. Numer. Anal., 39 (2001), pp. 73-99.

[24] W. Liu and N. Yan, A posteriori error estimates for distributed convex optimal control problems, Adv. Comput. Math., 15 (2001), pp. 285-309.

[25] W. Liu and N. Yan, A posteriori error estimates for control problems governed by nonlinear elliptic equations, Appl. Numer. Math., 47 (2003), no. 2, pp. 173-187.

[26] W. Liu and N. Yan, Adaptive Finite Element Methods for Optimal Control Governed by PDEs, Science Press, Beijing, 2008.

[27] P. Morin, R.H. Nochetto and K.G. Siebert, Data oscillation and convergence of adaptive FEM, SIAM J. Numer. Anal., 38 (2000), pp. 466-488.

[28] R. Verfürth, A Review of a Posteriori Error Estimates and Adaptive Mesh Refinement Techniques, Wiley-Teubner, New York, 1996.

[29] H. Xie, A type of multilevel method for the Steklov eigenvalue problem, IMA J. Numer. Anal., 34 (2014), no. 2, pp. 592-608.

[30] H. Xie, A multigrid method for eigenvalue problem, J. Comput. Phys., 274 (2014), pp. 550-561.

[31] H. Xie, A multilevel correction type of adaptive finite element method for eigenvalue problems, arXiv: 1201.2308, 2012.

[32] J. Xu, A novel two-grid method for semilinear elliptic equations, SIAM J. Sci. Comput., 15 (1994), pp. 231-237.

[33] A. Zhou, Multi-level adaptive corrections in finite dimensional approximations, J. Comput. Math., 28 (2010), pp. 45-54.