

# CONVERGENCE OF $L^2$ -NORM BASED ADAPTIVE FINITE ELEMENT METHOD FOR ELLIPTIC OPTIMAL CONTROL PROBLEMS

WEI GONG <sup>\*</sup>, NINGNING YAN <sup>◇</sup>, AND ZHAOJIE ZHOU <sup>†</sup>

**Abstract:** This paper aims to study the convergence of adaptive finite element method for control constrained elliptic optimal control problems under  $L^2$ -norm. We prove the contraction property and quasi-optimal complexity for the  $L^2$ -norm errors of both the control, the state and adjoint state variables with  $L^2$ -norm based AFEM, this is in contrast to and improve our previous work [14] where convergence of AFEM based on energy norm had been studied and suboptimal convergence for the control variable was obtained and observed numerically. For the discretization we use variational discretization for the control and piecewise linear and continuous finite elements for the state and adjoint state. Under mild assumptions on the initial mesh and the mesh refinement algorithm to keep the adaptive meshes sufficiently mildly graded we prove the optimal convergence of AFEM for the control problems, numerical results are provided to support our theoretical findings.

**Keywords:** optimal control problem, elliptic equation, control constraints, adaptive finite element method,  $L^2$ -norm error estimate, convergence and optimality

**Subject Classification:** 49J20, 65K10, 65N12, 65N15, 65N30.

## 1. INTRODUCTION

Adaptive finite element method (AFEM for short), aiming to distribute more mesh nodes around the area where singularity of the solution happens to save the computational cost, was firstly proposed by Babuška and Rheinboldt ([1]) and becomes nowadays a popular approach in the community of engineering and scientific computing. The success of AFEM lies in reliable and efficient a posteriori error estimators derived for the underlying problems, which are used as error indicators to detect the location of singularity. We refer to [35] for an overview on various types of error estimators and the applications to different kind of problems.

AFEM was also successfully used for solving optimal control problems (OCPs) governed by partial differential equations, starting from Liu, Yan [23] for residual type a posteriori error estimates and Becker, Kapp, Rannacher [2] for dual-weighted goal-oriented adaptivity. We refer to [15, 16, 21, 24, 25, 26, 27] for the extensions and applications in distributed and boundary control problems, Stokes control problems, time-dependent control problems and so on. Recently, Kohls, Rösch and Siebert derived in [19] an error equivalence property with respect to the  $L^2$ -norm error for the control and energy norm errors for the state and adjoint state, which helps to derive reliable and efficient a posteriori error estimators for optimal control problems with either variational discretization or full control discretization. In [31] the authors considered a posteriori error estimates for elliptic optimal control problems in  $L^2$ -norm by extending the ideas of [19].

The adaptive finite element procedure for solving optimal control problems consists of the following loop

SOLVE  $\rightarrow$  ESTIMATE  $\rightarrow$  MARK  $\rightarrow$  REFINE.

---

*Date:* Today is September 1, 2016.

<sup>\*</sup> NCMIS, LSEC, Institute of Computational Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China. Email: [wgong@lsec.cc.ac.cn](mailto:wgong@lsec.cc.ac.cn).

<sup>◇</sup> LSEC, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China. Email: [ynn@amss.ac.cn](mailto:ynn@amss.ac.cn).

<sup>†</sup> School of Mathematical Sciences, Shandong Normal University, Jinan 250014, China. Email: [zzj534@amss.ac.cn](mailto:zzj534@amss.ac.cn).

The SOLVE step outputs the finite element solutions of the optimal control problems by solving the reduced optimization problems with appropriate optimization algorithms. The ESTIMATE step is based on the a posteriori error estimators which should be reliable and may also be efficient, while the step MARK selects the set of elements for refinement by some strategies, like MAX strategy ([28]) or Dörfler’s marking strategy ([11]), based on the error indicators derived from the contributions of the control, the state and adjoint state approximations. The step REFINE can be done by using iterative or recursive bisection of elements with the minimal refinement condition (see [33, 35]).

Although AFEM has been successfully applied for more than three decades, the convergence analysis is rather recent which started with Dörfler [11] and was further studied in [4, 29, 30, 28, 5], and is still far from mature. Optimal complexity of AFEM, which measures the optimal computational cost of AFEM, is another important issue in AFEM and was firstly addressed by Binev et al. [4] and further studied by Stevenson ([33, 34]). In [5] the authors introduced the notion of quasi error which consists of the sum of the energy error and the scaled estimator, and proved the contraction of the errors and quasi-optimal cardinality of a practical adaptive algorithm which avoids marking for oscillation ([11]) and circumvents the interior node property of mesh refinement ([29, 30]).

Compared with AFEM for boundary value problems, the convergence results of AFEM for solving optimal control problems are scarce and far from satisfactory. Here we give some comments on existing results. Gaevskaya et. al studied in [12] the convergence of AFEM for OCPs with full control discretization. They used the piecewise constant approximation for the control variable and proved an error reduction property for the optimal control, state, adjoint state and the associated co-control variables, some additional requirements on the strict complementarity of the continuous problem and the non-degeneracy property of the discrete control problem were imposed. Moreover, the marking strategy was extended to include the discrete free boundary between the active and inactive control sets. Becker and Mao ([3]) gave a convergence proof for the adaptive algorithm by viewing the control problems as a nonlinear elliptic system of the state and adjoint variables, the adaptive algorithm presented there involved the marking of data oscillation. In [20] the authors proved the plain convergence of the adaptive algorithm, i.e., the sequence of adaptively generated discrete solutions converged to the true solutions of optimal control problems, but no convergence rate and optimality were derived. Recently, Gong and Yan ([13]) gave a rigorous convergence proof for the adaptive finite element algorithm of elliptic optimal control problem in an optimal control framework. The main idea is to show that the energy norm errors of the state and adjoint state variables are equivalent to the boundary value approximations of the state and adjoint state equations up to a higher order term, so that the standard convergence results of AFEM for elliptic boundary value problems can be used. Later on, the ideas were used to prove the convergence of an adaptive multilevel correction finite element method for elliptic optimal control problem. We also mention that in [6] Chen and Leng proved the convergence and quasi-optimality of AFEM for an elliptic control problem with integral type control constraint by using piecewise constant control discretization.

We remark that all the results mentioned above are based on AFEM in energy norm error for both boundary value problems and OCPs. Recall that in the a priori error estimates for optimal control problems ([17]), one can obtain optimal convergence of the control variable by using only optimal  $L^2$ -norm error estimates for the state and the adjoint state variables. This strongly suggests a posteriori error estimates and adaptive algorithm in  $L^2$ -norm. In [14] the authors showed that the convergence of AFEM based on energy norm was suboptimal for the control variable and the numerical experiments confirmed this sub-optimality. This motivates the study in this paper on convergence analysis of AFEM for optimal control problems in  $L^2$ -norm.

Since the Galerkin approximation of second order elliptic equation is only the best approximation in energy norm, it is not straightforward to prove the convergence of AFEM in  $L^2$ -norm. There are already some attempts to prove convergence of AFEM under weaker norms other than the global energy norm. Morin et al. [32] obtained plain convergence of AFEM for controlling weak norms

under quite general assumptions on the marking strategy and norm of interest. Demlow studied in [8] the convergence of AFEM under local energy norm error. Demlow and Steveson proved in [10] the contraction property and optimal convergence rate of AFEM for controlling  $L^2$ -norm with Dörfler's marking strategy under appropriate mesh grading conditions. The results of [10] were then used to prove the quasi-optimality of adaptive finite element methods for controlling local energy errors in [9]. The convergence analysis of AFEM in  $L^2$ -norm presented in [10] is based on the equivalence between the weighted energy norm error and  $L^2$ -norm error under additional assumption on the mesh grading. One also need to impose  $H^2$ -regularity for the underline elliptic equation for the derivation of efficient and reliable a posteriori error estimates. In this paper we follow the approach of [10] to prove the convergence of AFEM for OCPs. We show that the finite element solutions of the optimal control problem are nearly best approximations to the continuous ones from the finite element space in the  $L^2$ -norm up to an oscillation term. Moreover, we show the contraction property and quasi-optimal complexity for the  $L^2$ -norm errors of both the control, the state and adjoint state variables with  $L^2$ -norm based AFEM, this improves the known result of [13] for energy norm based AFEM.

The rest of the paper is organized as follows. In Section 2 we introduce the model elliptic optimal control problem and its finite element approximation, we also derive efficient and reliable a posteriori error estimates in  $L^2$ -norm. The adaptive algorithm based on Dörfler's marking strategy is also presented. In Section 3 we give some connections between the weighted energy norm errors and the  $L^2$ -norm errors, which enable us to give a convergence analysis of the AFEM for optimal control problems, the quasi-optimal cardinality is proved in Section 4. Numerical experiments are carried out in Section 5 to validate our theoretical result.

Let  $\Omega \subset \mathbb{R}^d$  ( $d = 2, 3$ ) be a bounded, convex polygonal or polyhedral domain. We denote by  $W^{m,q}(\Omega)$  the usual Sobolev space of order  $m \geq 0$ ,  $1 \leq q < \infty$  with norm  $\|\cdot\|_{m,q,\Omega}$ . For  $q = 2$  we denote  $W^{m,q}(\Omega)$  by  $H^m(\Omega)$  and  $\|\cdot\|_{m,\Omega} = \|\cdot\|_{m,2,\Omega}$ , which is a Hilbert space. We denote  $\|v\| := \sqrt{(\nabla v, \nabla v)}$  and  $\|v\|_D := (\int_D |\nabla v|^2 dx)^{\frac{1}{2}}$  for the global energy norm and local energy semi-norm over  $\Omega$  and  $D \subset \Omega$ , respectively.

We denote  $C$  a generic positive constant which may stand for different values at its different occurrences but does not depend on mesh size. We use the symbol  $A \lesssim B$  to denote  $A \leq CB$  for some constant  $C$  that is independent of mesh size. If in addition  $B \lesssim A$  holds we use the symbol  $A \simeq B$ .

## 2. ADAPTIVE FINITE ELEMENT METHOD FOR THE OPTIMAL CONTROL PROBLEM

In this section we consider the following elliptic optimal control problem:

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

subject to

$$(2.2) \quad \begin{cases} -\Delta y = f + u & \text{in } \Omega, \\ y = 0 & \text{on } \partial\Omega, \end{cases}$$

where  $\alpha > 0$  is a fixed parameter,  $f$  is a given function,  $y_d \in L^2(\Omega)$  is the desired state and  $U_{ad}$  is the admissible control set with bilateral control constraints:

$$U_{ad} := \left\{ u \in L^2(\Omega), \quad a \leq u \leq b \text{ a.e. in } \Omega \right\},$$

where  $a, b \in \mathbb{R}$  and  $a < b$ .

We introduce the affine linear control-to-state mapping  $S : L^2(\Omega) \rightarrow H_0^1(\Omega)$  such that for each  $f + u \in L^2(\Omega)$  we associate the unique solution of problem (2.2) with  $y = S(f + u)$ , i.e.,

$$(2.3) \quad (\nabla y, \nabla v) = (f + u, v) \quad \forall v \in H_0^1(\Omega).$$

Since  $f$  is fixed, in the following we use  $y = Su$  instead of  $y = S(f + u)$  to emphasize the dependence on  $u$ . Then we are led to a reduced optimization problem  $\min_{u \in U_{ad}} \hat{J}(u) := J(Su, u)$  involving only

the control  $u$ . By standard arguments ([22]) we can prove that the above optimization problem admits a unique solution. Moreover, we can derive the following first order optimality condition:

$$(2.4) \quad (\alpha u + p, v - u) \geq 0 \quad \forall v \in U_{ad},$$

where the adjoint state  $p \in H_0^1(\Omega)$  satisfies

$$(2.5) \quad \begin{cases} -\Delta p = y - y_d & \text{in } \Omega, \\ p = 0 & \text{on } \partial\Omega. \end{cases}$$

It is clear that  $p = S^*(y - y_d)$  where  $S^*$  is the adjoint operator of  $S$  such that

$$(2.6) \quad (\nabla v, \nabla p) = (y - y_d, v) \quad \forall v \in H_0^1(\Omega).$$

Since the elliptic equation (2.2) is linear and self-adjoint we have that  $S = S^*$ . From (2.4) we can derive the pointwise representation of the control  $u$  through the adjoint  $p$ :  $u = P_{[a,b]}(p) := \max\{a, \min\{b, -\frac{p}{\alpha}\}\}$ .

Now we consider the finite element approximation of above optimal control problems. To begin with, let  $\mathcal{T}_0$  be a conforming and quasi-uniform partition of  $\bar{\Omega}$  into disjoint  $d$ -simplices. Each element in  $\mathcal{T}_0$  is assumed to be shape regular in the usual sense (see [7]). We denote the set of all conforming descendants  $\mathcal{T}$  of  $\mathcal{T}_0$  by  $\mathbb{T}$ , which can be generated through uniform or local refinements by newest vertex bisection algorithm. The simplices in any of those partitions are uniformly shape regular depending only on the shape regularity parameters of  $\mathcal{T}_0$  and the dimension  $d$ , we refer to [10, Appendix A] for more details on how to generate such kind of partitions. For each  $\tilde{\mathcal{T}}, \mathcal{T} \in \mathbb{T}$ , we write  $\mathcal{T} \subset \tilde{\mathcal{T}}$  when  $\tilde{\mathcal{T}}$  is a refinement of  $\mathcal{T}$ .

We denote  $h_T = |T|^{\frac{1}{d}}$  for each  $T \in \mathcal{T} \in \mathbb{T}$  with  $|T|$  the volume of  $T$ . Let  $\mathcal{N}_{\mathcal{T}}$  and  $\mathcal{E}_{\mathcal{T}}$  be the sets of vertices and interior edges or faces of  $\mathcal{T}$ . We also denote  $\omega_T$  or  $\tilde{\omega}_T$  the patches of elements sharing a vertex or a facet with  $T$ . Since the Galerkin approximation is not the best approximation in  $L^2$ -norm, we need to impose some restrictions on the mesh for the convergence analysis of AFEM in  $L^2$ -norm. Following the ideas of [10] we define the continuous and piecewise linear mesh function  $h_{\mathcal{T}}$ , such that for any  $z \in \mathcal{N}_{\mathcal{T}}$ ,  $h_{\mathcal{T}}$  is defined as the average of the  $h_{T'}$  over all  $T' \in \mathcal{T}$  where  $z \in T'$ . Then for some constants  $c_{\mathbb{T}}$  and  $C_{\mathbb{T}}$  there holds

$$(2.7) \quad c_{\mathbb{T}} h_T \leq h_{\mathcal{T}}|_T \leq C_{\mathbb{T}} h_T, \quad \forall T \in \mathcal{T}, \mathcal{T} \in \mathbb{T}.$$

In view of the uniform shape regularity of  $\mathbb{T}$  there exists another constant  $\hat{C}_{\mathbb{T}}$  such that

$$\|\nabla h_{\mathcal{T}}\|_{0,\infty,\Omega} \leq \hat{C}_{\mathbb{T}} \quad \forall \mathcal{T} \in \mathbb{T}.$$

Throughout the paper we assume that the partition  $\mathcal{T}$  is sufficiently graded, i.e., for some sufficiently small but fixed constant  $\mu > 0$ , the mesh function  $h_{\mathcal{T}}$  satisfies

$$(2.8) \quad \|\nabla h_{\mathcal{T}}\|_{0,\infty,\Omega} \leq \mu,$$

and in addition, (2.7) holds for some constants  $c_{\mathbb{T}}$  and  $C_{\mathbb{T}}$  that are independent of  $\mu$ . We refer to Appendix A in [10] on how to generate a class of sufficiently mildly graded partitions  $\mathcal{T} \in \mathbb{T}$  for given parameter  $\mu$  such that the mesh function  $h_{\mathcal{T}}$  satisfies (2.7) and (2.8). Given a  $\mu > 0$  we denote the class of such partitions by  $\mathbb{T}_{\mu}$ .

Associated with  $\mathcal{T} \in \mathbb{T}$  we define the continuous and piecewise linear finite element space  $V_{\mathcal{T}} \subset H_0^1(\Omega)$ . Let  $\Pi_{\mathcal{T}} : C(\bar{\Omega}) \rightarrow V_{\mathcal{T}}$  be the standard Lagrange interpolation operator. We define a discrete control-to-state mapping as  $S_{\mathcal{T}} : L^2(\Omega) \rightarrow V_{\mathcal{T}}$  such that  $y_{\mathcal{T}}(u) = S_{\mathcal{T}}(f + u)$  and

$$(2.9) \quad (\nabla y_{\mathcal{T}}(u), \nabla v_{\mathcal{T}}) = (f + u, v_{\mathcal{T}}) \quad \forall v_{\mathcal{T}} \in V_{\mathcal{T}}.$$

Also we denote  $y_{\mathcal{T}}(u) = S_{\mathcal{T}}u$  for simplicity. Then we can formulate a reduced discrete optimization problem  $\min_{u_{\mathcal{T}} \in U_{ad}} \hat{J}(u_{\mathcal{T}}) := J(S_{\mathcal{T}}u_{\mathcal{T}}, u_{\mathcal{T}})$  where we use the variational control discretization proposed by Hinze ([17]). By standard arguments ([22]) we can also prove that the above discrete optimization problem admits a unique solution. Moreover, we can derive the following discrete first order optimality condition:

$$(2.10) \quad (\alpha u_{\mathcal{T}} + p_{\mathcal{T}}, v_{\mathcal{T}} - u_{\mathcal{T}}) \geq 0 \quad \forall v_{\mathcal{T}} \in U_{ad},$$

where the discrete adjoint state  $p_{\mathcal{T}} \in V_{\mathcal{T}}$  satisfies

$$(2.11) \quad (\nabla v_{\mathcal{T}}, \nabla p_{\mathcal{T}}) = (y_{\mathcal{T}} - y_d, v_{\mathcal{T}}) \quad \forall v_{\mathcal{T}} \in V_{\mathcal{T}}$$

with  $y_{\mathcal{T}} = S_{\mathcal{T}} u_{\mathcal{T}}$ . It is clear that  $p_{\mathcal{T}} = S_{\mathcal{T}}^*(y_{\mathcal{T}} - y_d)$  where  $S_{\mathcal{T}}^*$  is the adjoint of  $S_{\mathcal{T}}$ . Similarly, we have  $u_{\mathcal{T}} = P_{[a,b]}(p_{\mathcal{T}}) = \max\{a, \min\{b, -\frac{p_{\mathcal{T}}}{\alpha}\}\}$ , which is not generally a finite element function.

Now we will give some notations for the following purpose. For each element  $T \in \mathcal{T}$  we define the local error indicators  $\eta_{\mathcal{T},y}(u_{\mathcal{T}}, y_{\mathcal{T}}, T)$  contributed to the state equation and  $\eta_{\mathcal{T},p}(y_{\mathcal{T}}, p_{\mathcal{T}}, T)$  contributed to the adjoint state equation by

$$(2.12) \quad \eta_{\mathcal{T},y}^2(u_{\mathcal{T}}, y_{\mathcal{T}}, T) := h_T^4 \|f + u_{\mathcal{T}}\|_{0,T}^2 + \sum_{E \in \mathcal{E}_T, E \subset \partial T} h_E^3 \|[\nabla y_{\mathcal{T}}]_E \cdot n_E\|_{0,E}^2,$$

$$(2.13) \quad \eta_{\mathcal{T},p}^2(y_{\mathcal{T}}, p_{\mathcal{T}}, T) := h_T^4 \|y_{\mathcal{T}} - y_d\|_{0,T}^2 + \sum_{E \in \mathcal{E}_T, E \subset \partial T} h_E^3 \|[\nabla p_{\mathcal{T}}]_E \cdot n_E\|_{0,E}^2,$$

where  $[\nabla y_{\mathcal{T}}]_E \cdot n_E$  denotes the jump of  $\nabla y_{\mathcal{T}}$  across the common side  $E$  of elements  $T^+$  and  $T^-$ ,  $n_E$  denotes the outward normal oriented to  $T^-$ . We also define the local error estimator for the optimal control problem

$$(2.14) \quad \eta_{\mathcal{T}}^2(T) := \eta_{\mathcal{T},y}^2(u_{\mathcal{T}}, y_{\mathcal{T}}, T) + \eta_{\mathcal{T},p}^2(y_{\mathcal{T}}, p_{\mathcal{T}}, T).$$

Then on a subset  $\omega \subset \Omega$ , we define the error estimator  $\eta_{\mathcal{T},y}(u_{\mathcal{T}}, y_{\mathcal{T}}, \omega)$  by

$$(2.15) \quad \eta_{\mathcal{T},y}^2(u_{\mathcal{T}}, y_{\mathcal{T}}, \omega) := \sum_{T \in \mathcal{T}, T \subset \omega} \eta_{\mathcal{T},y}^2(u_{\mathcal{T}}, y_{\mathcal{T}}, T).$$

Thus,  $\eta_{\mathcal{T},y}(u_{\mathcal{T}}, y_{\mathcal{T}}, \mathcal{T})$  constitutes the error estimator for the state equation on  $\Omega$  with respect to the partition  $\mathcal{T}$ . The similar definition applies to the error estimators  $\eta_{\mathcal{T},p}(y_{\mathcal{T}}, p_{\mathcal{T}}, \mathcal{T})$  for the adjoint state equation and  $\eta_{\mathcal{T}}(\mathcal{T})$  for the optimal control problem.

For  $f \in L^2(\Omega)$  we also need to define the data oscillation as (see [29, 30])

$$(2.16) \quad \text{osc}(f, T) := \|h_T^2(f - \bar{f}_T)\|_{0,T}, \quad \text{osc}(f, \mathcal{T}) := \left( \sum_{T \in \mathcal{T}} \text{osc}^2(f, T) \right)^{\frac{1}{2}},$$

where  $\bar{f}_T$  denotes the  $L^2$ -projection of  $f$  onto piecewise constant space on  $T$ . It is easy to see that

$$(2.17) \quad \text{osc}(f_1 + f_2, \mathcal{T}) \leq \text{osc}(f_1, \mathcal{T}) + \text{osc}(f_2, \mathcal{T}), \quad \forall f_1, f_2 \in L^2(\Omega).$$

To derive error estimates in  $L^2$ -norm we need the following assumption:

**Assumption 2.1.** *Assume that  $\Omega$  is convex so that for each  $f + u \in L^2(\Omega)$  problem (2.2) admits a unique solution  $y = Su \in H^2(\Omega) \cap H_0^1(\Omega)$  and*

$$(2.18) \quad \|y\|_{2,\Omega} \leq C_{\text{reg}} \|f + u\|_{0,\Omega}.$$

*The assumption is also valid for the adjoint equation, i.e., for  $S^*$ .*

With above preparations now we are in the position to derive a posteriori error estimates for both the control, the state and adjoint state variables.

**Theorem 2.2.** *Let  $(u, y, p) \in U_{ad} \times H_0^1(\Omega) \times H_0^1(\Omega)$  be the solution of optimal control problem (2.1)-(2.2) and  $(u_{\mathcal{T}}, y_{\mathcal{T}}, p_{\mathcal{T}}) \in U_{ad} \times V_{\mathcal{T}} \times V_{\mathcal{T}}$  be the solution of the discrete problem (2.9)-(2.11). Then we have the a posteriori upper bound*

$$(2.19) \quad \|u - u_{\mathcal{T}}\|_{0,\Omega} + \|y - y_{\mathcal{T}}\|_{0,\Omega} + \|p - p_{\mathcal{T}}\|_{0,\Omega} \leq C_1 \eta_{\mathcal{T}}(\mathcal{T})$$

*and the global lower bound*

$$(2.20) \quad \eta_{\mathcal{T}}(\mathcal{T}) \leq C_2 (\|u - u_{\mathcal{T}}\|_{0,\Omega} + \|y - y_{\mathcal{T}}\|_{0,\Omega} + \|p - p_{\mathcal{T}}\|_{0,\Omega} + \text{osc}_{\mathcal{T}}),$$

*where  $C_1, C_2$  only depend on the shape regularity of  $\mathcal{T}$  and the data oscillation  $\text{osc}_{\mathcal{T}}$  is defined as*

$$\text{osc}_{\mathcal{T}}^2 := \text{osc}^2(f + u_{\mathcal{T}}, \mathcal{T}) + \text{osc}^2(y_{\mathcal{T}} - y_d, \mathcal{T}).$$

*Proof.* Setting  $v = u_{\mathcal{T}} \in U_{ad}$  in (2.4) and  $v_{\mathcal{T}} = u \in U_{ad}$  in (2.10) and noticing that  $p = S^*(Su - y_d)$ ,  $p_{\mathcal{T}} = S_{\mathcal{T}}^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d)$ , we are then led to

$$(2.21) \quad (\alpha u + S^*(Su - y_d), u_{\mathcal{T}} - u) \geq 0,$$

$$(2.22) \quad (\alpha u_{\mathcal{T}} + S_{\mathcal{T}}^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d), u - u_{\mathcal{T}}) \geq 0.$$

Adding the above two inequalities, we conclude from (2.3) and (2.6) that

$$\begin{aligned} & \alpha \|u - u_{\mathcal{T}}\|_{0,\Omega}^2 \leq (S_{\mathcal{T}}^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d) - S^*(Su - y_d), u - u_{\mathcal{T}}) \\ & = (S_{\mathcal{T}}^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d) - S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d), u - u_{\mathcal{T}}) + (S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d) - S^*(Su - y_d), u - u_{\mathcal{T}}) \\ & = (S_{\mathcal{T}}^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d) - S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d), u - u_{\mathcal{T}}) + (S_{\mathcal{T}}u_{\mathcal{T}} - Su, Su - Su_{\mathcal{T}}) \\ & = (S_{\mathcal{T}}^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d) - S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d), u - u_{\mathcal{T}}) + (S_{\mathcal{T}}u_{\mathcal{T}} - Su, Su - S_{\mathcal{T}}u_{\mathcal{T}}) \\ (2.23) \quad & + (S_{\mathcal{T}}u_{\mathcal{T}} - Su, S_{\mathcal{T}}u_{\mathcal{T}} - Su_{\mathcal{T}}). \end{aligned}$$

It follows from Young's inequality that

$$(2.24) \quad \begin{aligned} & \alpha \|u - u_{\mathcal{T}}\|_{0,\Omega}^2 + \|y - y_{\mathcal{T}}\|_{0,\Omega}^2 \\ & \leq C \|Su_{\mathcal{T}} - S_{\mathcal{T}}u_{\mathcal{T}}\|_{0,\Omega}^2 + C \|S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d) - S_{\mathcal{T}}^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d)\|_{0,\Omega}^2, \end{aligned}$$

where we used the fact that  $y = Su$  and  $y_{\mathcal{T}} = S_{\mathcal{T}}u_{\mathcal{T}}$ . Moreover, from (2.3), (2.6) and the triangle inequality we have

$$\begin{aligned} \|p - p_{\mathcal{T}}\|_{0,\Omega} & \leq \|p - S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d)\|_{0,\Omega} + \|S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d) - p_{\mathcal{T}}\|_{0,\Omega} \\ & \leq C \|Su - S_{\mathcal{T}}u_{\mathcal{T}}\|_{0,\Omega} + \|S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d) - p_{\mathcal{T}}\|_{0,\Omega} \\ & \leq C \|u - u_{\mathcal{T}}\|_{0,\Omega} + C \|Su_{\mathcal{T}} - S_{\mathcal{T}}u_{\mathcal{T}}\|_{0,\Omega} + \|S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d) - p_{\mathcal{T}}\|_{0,\Omega}. \end{aligned}$$

Combining the above estimates we are led to

$$(2.25) \quad \begin{aligned} & \alpha \|u - u_{\mathcal{T}}\|_{0,\Omega}^2 + \|y - y_{\mathcal{T}}\|_{0,\Omega}^2 + \|p - p_{\mathcal{T}}\|_{0,\Omega}^2 \\ & \leq C \|Su_{\mathcal{T}} - S_{\mathcal{T}}u_{\mathcal{T}}\|_{0,\Omega}^2 + C \|S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d) - S_{\mathcal{T}}^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d)\|_{0,\Omega}^2. \end{aligned}$$

Note that  $S_{\mathcal{T}}u_{\mathcal{T}}$  and  $S_{\mathcal{T}}^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d)$  are the standard finite element approximations of  $Su_{\mathcal{T}}$  and  $S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d)$  in finite element space  $V_{\mathcal{T}}$ , respectively. Under Assumption 2.1 we can derive from standard a posteriori error estimate for elliptic equation under  $L^2$ -norm that (see [35] for more details)

$$(2.26) \quad \|Su_{\mathcal{T}} - S_{\mathcal{T}}u_{\mathcal{T}}\|_{0,\Omega} \leq C \eta_{\mathcal{T},y}(u_{\mathcal{T}}, y_{\mathcal{T}}, \mathcal{T}),$$

$$(2.27) \quad \|S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d) - S_{\mathcal{T}}^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d)\|_{0,\Omega} \leq C \eta_{\mathcal{T},p}(y_{\mathcal{T}}, p_{\mathcal{T}}, \mathcal{T}).$$

Therefore, combining the above results we are able to derive the upper bound with the constant  $C_1$  independent of the mesh size  $h_{\mathcal{T}}$ .

Now we prove the lower bound. Note that

$$(2.28) \quad \begin{aligned} \|Su_{\mathcal{T}} - S_{\mathcal{T}}u_{\mathcal{T}}\|_{0,\Omega} & \leq \|Su_{\mathcal{T}} - Su\|_{0,\Omega} + \|Su - S_{\mathcal{T}}u_{\mathcal{T}}\|_{0,\Omega} \\ & \leq C \|u - u_{\mathcal{T}}\|_{0,\Omega} + \|y - y_{\mathcal{T}}\|_{0,\Omega}. \end{aligned}$$

Similarly, we can derive that

$$(2.29) \quad \begin{aligned} & \|S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d) - S_{\mathcal{T}}^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d)\|_{0,\Omega} \\ & \leq \|S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d) - S^*(Su - y_d)\|_{0,\Omega} + \|S^*(Su - y_d) - S_{\mathcal{T}}^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d)\|_{0,\Omega} \\ & \leq C \|y - y_{\mathcal{T}}\|_{0,\Omega} + \|p - p_{\mathcal{T}}\|_{0,\Omega}. \end{aligned}$$

Moreover, from standard lower bound error estimate for elliptic equation (see [35] for more details) we have the following global lower bound

$$(2.30) \quad \eta_{\mathcal{T},y}(u_{\mathcal{T}}, y_{\mathcal{T}}, \mathcal{T}) \leq C (\|Su_{\mathcal{T}} - S_{\mathcal{T}}u_{\mathcal{T}}\|_{0,\Omega} + \text{osc}(f + u_{\mathcal{T}}, \mathcal{T})),$$

$$(2.31) \quad \eta_{\mathcal{T},p}(y_{\mathcal{T}}, p_{\mathcal{T}}, \mathcal{T}) \leq C (\|S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d) - S_{\mathcal{T}}^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d)\|_{0,\Omega} + \text{osc}(y_{\mathcal{T}} - y_d, \mathcal{T})).$$

Thus, we can conclude from the above estimates the lower bound with the constant  $C_2$  independent of the mesh size  $h_{\mathcal{T}}$ . This completes the proof.  $\square$

In the following we will present the adaptive algorithm for solving optimal control problems. Note that there are two error estimators  $\eta_{\mathcal{T},y}(u_{\mathcal{T}}, y_{\mathcal{T}}, T)$  and  $\eta_{\mathcal{T},p}(y_{\mathcal{T}}, p_{\mathcal{T}}, T)$  contributed to the state approximation and adjoint state approximation, respectively. We use the sum of the two estimators as our indicators for the marking strategy. The marking algorithm based on Dörfler's strategy for optimal control problems can be described as follows

**Algorithm 2.3.** *Dörfler's marking strategy for OCPs*

- (1) Given a parameter  $0 < \theta < 1$ ;
- (2) Construct a minimal subset  $\mathcal{M} \subset \mathcal{T}$  such that

$$\sum_{T \in \mathcal{M}} \eta_{\mathcal{T}}^2(T) \geq \theta^2 \eta_{\mathcal{T}}^2(\mathcal{T}).$$

- (3) Mark all the elements in  $\mathcal{M}$ .

In the following we will frequently use the abbreviations  $V_k$  for  $V_{\mathcal{T}_k}$ ,  $h_k$  for  $h_{\mathcal{T}_k}$  and  $v_k$  for  $v_{\mathcal{T}_k}$ , and the similar abbreviations for other variables and notations. Now we can present the adaptive finite element algorithm for the optimal control problem as follows.

**Algorithm 2.4.** *Adaptive finite element algorithm for OCPs:*

- (1) Given an initial mesh  $\mathcal{T}_0$  with mesh size  $h_0$ , construct the finite element space  $V_0$ .
- (2) Set  $k = 0$  and solve the optimal control problem (2.9)-(2.11) to obtain  $(u_k, y_k, p_k) \in U_{ad} \times V_k \times V_k$ .
- (3) Compute the local error indicator  $\eta_k(T)$ .
- (4) Construct  $\mathcal{M}_k \subset \mathcal{T}_k$  by the marking Algorithm 2.3.
- (5) Refine  $\mathcal{M}_k$  to get a new conforming mesh  $\mathcal{T}_{k+1}$  by procedure *REFINE* using bisection algorithm.
- (6) Construct the finite element space  $V_{k+1}$  and solve the optimal control problem (2.9)-(2.11) to obtain  $(u_{k+1}, y_{k+1}, p_{k+1}) \in U_{ad} \times V_{k+1} \times V_{k+1}$ .
- (7) Set  $k = k + 1$  and go to Step (3).

In step (5) of Algorithm 2.4 we assume that each marked element in  $\mathcal{M}_k$  is bisected  $r \geq 1$  times to generate a new mesh  $\mathcal{T}_{k+1}$ , and additional elements are refined in the process to ensure that  $\mathcal{T}_{k+1}$  is conforming. We remark that to ensure the mesh grading property (2.8) we have to additionally refine elements other than that of  $\mathcal{M}_k$ . Demlow and Stevenson [10] showed that this can be done by inflating the number of simplices by not more than some fixed multiple which depends on the mesh grading parameter  $\mu$  and can be achieved by the standard newest vertex bisection algorithm with necessary modifications, and the modification does not compromise the quasi-optimality of the resulting algorithm, we refer to Appendix A in [10] for more details.

### 3. CONVERGENCE OF AFEM FOR THE OPTIMAL CONTROL PROBLEM IN $L^2$ -NORM

In this section we intend to prove the contraction property of the  $L^2$ -norm errors of the control, the state and adjoint state with  $L^2$ -norm based AFEM. The proof relies on establishing certain equivalence property between the  $L^2$ -norm error and the weighted energy norm error for the state and adjoint state, and uses the convergence result of AFEM under energy norm.

**3.1. Connections between the weighted energy norm errors and  $L^2$ -norm errors.** In this subsection we collect some results from [10] showing certain relationships between the energy norm and  $L^2$ -norm errors for the finite element approximations of elliptic equations, and their generalizations to optimal control problems.

At first, we show that the  $L^2$ -norm error of finite element approximation of elliptic equation can be bounded by the weighted energy norm as long as the mesh grading is sufficiently mild.

**Lemma 3.1.** ([10, Proposition 3]) *For sufficiently small  $\mu$ , let  $\mathcal{T} \in \mathbb{T}_\mu$ . Under Assumption 2.1 we have that for any  $f, g \in L^2(\Omega)$ ,*

$$(3.1) \quad \|Sf - S_{\mathcal{T}}f\|_{0,\Omega} \lesssim \|h_{\mathcal{T}}(Sf - S_{\mathcal{T}}f)\|,$$

$$(3.2) \quad \|S^*g - S_{\mathcal{T}}^*g\|_{0,\Omega} \lesssim \|h_{\mathcal{T}}(S^*g - S_{\mathcal{T}}^*g)\|.$$

With above result we can establish the connection between the  $L^2$ -norm errors and weighted energy norm errors for the solutions of optimal control problems.

**Lemma 3.2.** *Let  $(u, y, p) \in U_{ad} \times H_0^1(\Omega) \times H_0^1(\Omega)$  be the solution of optimal control problem (2.1)-(2.2) and  $(u_{\mathcal{T}}, y_{\mathcal{T}}, p_{\mathcal{T}}) \in U_{ad} \times V_{\mathcal{T}} \times V_{\mathcal{T}}$  be the solution of the discrete problem (2.9)-(2.11). For sufficiently small  $\mu$ , let  $\mathcal{T} \in \mathbb{T}_\mu$ . Under Assumption 2.1 we have that*

$$(3.3) \quad \|u - u_{\mathcal{T}}\|_{0,\Omega} + \|y - y_{\mathcal{T}}\|_{0,\Omega} + \|p - p_{\mathcal{T}}\|_{0,\Omega} \lesssim \|h_{\mathcal{T}}(y - y_{\mathcal{T}})\| + \|h_{\mathcal{T}}(p - p_{\mathcal{T}})\|$$

provided that  $h_0 \ll 1$ .

*Proof.* From the proof of Theorem 2.2 we can conclude that

$$(3.4) \quad \begin{aligned} & \|u - u_{\mathcal{T}}\|_{0,\Omega} + \|y - y_{\mathcal{T}}\|_{0,\Omega} + \|p - p_{\mathcal{T}}\|_{0,\Omega} \\ & \lesssim \|Su_{\mathcal{T}} - y_{\mathcal{T}}\|_{0,\Omega} + \|S^*(y_{\mathcal{T}} - y_d) - p_{\mathcal{T}}\|_{0,\Omega}. \end{aligned}$$

Recall that  $y_{\mathcal{T}}$  and  $p_{\mathcal{T}}$  are the standard finite element approximations of  $Su_{\mathcal{T}}$  and  $S^*(y_{\mathcal{T}} - y_d)$  in  $V_{\mathcal{T}}$ , respectively. Then it follows from Lemma 3.1 and the triangle inequality that

$$(3.5) \quad \begin{aligned} & \|u - u_{\mathcal{T}}\|_{0,\Omega} + \|y - y_{\mathcal{T}}\|_{0,\Omega} + \|p - p_{\mathcal{T}}\|_{0,\Omega} \\ & \lesssim \|h_{\mathcal{T}}(Su_{\mathcal{T}} - y_{\mathcal{T}})\| + \|h_{\mathcal{T}}(S^*(y_{\mathcal{T}} - y_d) - p_{\mathcal{T}})\| \\ & \lesssim \|h_{\mathcal{T}}(y - y_{\mathcal{T}})\| + \|h_{\mathcal{T}}(p - p_{\mathcal{T}})\| \\ & \quad + \|h_{\mathcal{T}}(y - Su_{\mathcal{T}})\| + \|h_{\mathcal{T}}(p - S^*(y_{\mathcal{T}} - y_d))\|. \end{aligned}$$

After elementary calculation we have

$$(3.6) \quad \begin{aligned} \|h_{\mathcal{T}}(y - Su_{\mathcal{T}})\|^2 &= (\nabla(h_{\mathcal{T}}(y - Su_{\mathcal{T}})), \nabla(h_{\mathcal{T}}(y - Su_{\mathcal{T}}))) \\ &= (\nabla(y - Su_{\mathcal{T}}), \nabla(h_{\mathcal{T}}^2(y - Su_{\mathcal{T}}))) + (\nabla h_{\mathcal{T}}(y - Su_{\mathcal{T}}), \nabla h_{\mathcal{T}}(y - Su_{\mathcal{T}})) \\ &\leq \|h_{\mathcal{T}}\|_{0,\infty,\Omega}^2 \|y - Su_{\mathcal{T}}\|^2 + \|\nabla h_{\mathcal{T}}\|_{0,\infty,\Omega}^2 \|y - Su_{\mathcal{T}}\|_{0,\Omega}^2 \\ &\quad + 2\|h_{\mathcal{T}}\|_{0,\infty,\Omega} \|\nabla h_{\mathcal{T}}\|_{0,\infty,\Omega} \|y - Su_{\mathcal{T}}\| \cdot \|y - Su_{\mathcal{T}}\|_{0,\Omega}. \end{aligned}$$

Note that  $\|h_{\mathcal{T}}\|_{0,\infty,\Omega} \leq h_0$  and  $\|\nabla h_{\mathcal{T}}\|_{0,\infty,\Omega} \leq \mu$ . This combining with the stability of elliptic equation gives

$$(3.7) \quad \|h_{\mathcal{T}}(y - Su_{\mathcal{T}})\| \lesssim (h_0 + \mu) \|u - u_{\mathcal{T}}\|_{0,\Omega}.$$

Similarly, we can prove that

$$(3.8) \quad \begin{aligned} \|h_{\mathcal{T}}(p - S^*(y_{\mathcal{T}} - y_d))\| &\lesssim (h_0 + \mu) \|y - y_{\mathcal{T}}\|_{0,\Omega} \\ &\lesssim (h_0 + \mu) (\|u - u_{\mathcal{T}}\|_{0,\Omega} + \|h_{\mathcal{T}}(y - y_{\mathcal{T}})\|). \end{aligned}$$

Combining the above results we complete the proof if  $h_0 \ll 1$  and  $\mu$  is sufficiently small.  $\square$

By using Proposition 4 in [10] and notice that  $S_{\mathcal{T}}u_{\mathcal{T}}$  and  $S_{\mathcal{T}}^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d)$  are the standard finite element approximations of  $Su_{\mathcal{T}}$  and  $S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d)$  in finite element space  $V_{\mathcal{T}}$ , we have the following a posteriori upper bounds for  $\|h_{\mathcal{T}}(Su_{\mathcal{T}} - S_{\mathcal{T}}u_{\mathcal{T}})\|$  and  $\|h_{\mathcal{T}}(S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d) - S_{\mathcal{T}}^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d))\|$ .

**Lemma 3.3.** *For any  $\mathcal{T} \in \mathbb{T}$ , under Assumption 2.1 we have that*

$$(3.9) \quad \|h_{\mathcal{T}}(Su_{\mathcal{T}} - S_{\mathcal{T}}u_{\mathcal{T}})\| \lesssim \eta_{\mathcal{T},y}(u_{\mathcal{T}}, y_{\mathcal{T}}, \mathcal{T}),$$

$$(3.10) \quad \|h_{\mathcal{T}}(S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d) - S_{\mathcal{T}}^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d))\| \lesssim \eta_{\mathcal{T},p}(y_{\mathcal{T}}, p_{\mathcal{T}}, \mathcal{T}).$$

Now we can derive a posteriori error estimate for the optimal control problem under weighted energy norm.



**Lemma 3.4.** *Let Assumption 2.1 be valid. For any  $\mathcal{T} \in \mathbb{T}$ , there exists a constant  $C_3$  independent of the mesh size of  $\mathcal{T}$  such that*

$$(3.11) \quad \|\|h_{\mathcal{T}}(y - y_{\mathcal{T}})\|\|^2 + \|\|h_{\mathcal{T}}(p - p_{\mathcal{T}})\|\|^2 \leq C_3 \eta_{\mathcal{T}}^2(\mathcal{T}).$$

*Proof.* In view of Lemma 3.3 it suffices to estimate  $\|\|h_{\mathcal{T}}(y - Su_{\mathcal{T}})\|\|$  and  $\|\|h_{\mathcal{T}}(p - S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d))\|\|$ . We use the abbreviation  $e_{\mathcal{T}} = y - Su_{\mathcal{T}}$  and note that

$$(3.12) \quad \begin{aligned} \|\|h_{\mathcal{T}}(y - Su_{\mathcal{T}})\|\|^2 &= (\nabla(h_{\mathcal{T}}e_{\mathcal{T}}), \nabla(h_{\mathcal{T}}e_{\mathcal{T}})) \\ &= (\nabla h_{\mathcal{T}}e_{\mathcal{T}}, \nabla h_{\mathcal{T}}e_{\mathcal{T}}) + (h_{\mathcal{T}}\nabla e_{\mathcal{T}}, h_{\mathcal{T}}\nabla e_{\mathcal{T}}) + 2(\nabla h_{\mathcal{T}}e_{\mathcal{T}}, h_{\mathcal{T}}\nabla e_{\mathcal{T}}) \\ &\leq \|\nabla h_{\mathcal{T}}\|_{0,\infty,\Omega}^2 \|e_{\mathcal{T}}\|_{0,\Omega}^2 + \|h_{\mathcal{T}}\|_{0,\infty,\Omega}^2 \|e_{\mathcal{T}}\|^2 \\ &\quad + 2\|\nabla h_{\mathcal{T}}\|_{0,\infty,\Omega} \|h_{\mathcal{T}}\|_{0,\infty,\Omega} \|e_{\mathcal{T}}\|_{0,\Omega} \|e_{\mathcal{T}}\| \\ &\lesssim \|u - u_{\mathcal{T}}\|_{0,\Omega}^2, \end{aligned}$$

where we used the fact that  $\|h_{\mathcal{T}}\|_{0,\infty,\Omega} \lesssim 1$ ,  $\|\nabla h_{\mathcal{T}}\|_{0,\infty,\Omega} \lesssim 1$  and the stability result for elliptic boundary value problem. Similarly, we can conclude from Lemmas 3.1 and 3.3 that

$$(3.13) \quad \begin{aligned} \|\|h_{\mathcal{T}}(p - S^*(S_{\mathcal{T}}u_{\mathcal{T}} - y_d))\|\|^2 &\lesssim \|y - y_{\mathcal{T}}\|_{0,\Omega}^2 \\ &\lesssim \|y - Su_{\mathcal{T}}\|_{0,\Omega}^2 + \|\|h_{\mathcal{T}}(Su_{\mathcal{T}} - S_{\mathcal{T}}u_{\mathcal{T}})\|\|^2 \\ &\lesssim \|u - u_{\mathcal{T}}\|_{0,\Omega}^2 + \eta_{\mathcal{T},y}^2(u_{\mathcal{T}}, y_{\mathcal{T}}, \mathcal{T}). \end{aligned}$$

Combining Theorem 2.2, (3.12), (3.13), Lemma 3.3 and the triangle inequality we finish the proof.  $\square$

It follows from Theorem 2.2 and Lemma 3.4 that

$$\|\|h_{\mathcal{T}}(y - y_{\mathcal{T}})\|\|^2 + \|\|h_{\mathcal{T}}(p - p_{\mathcal{T}})\|\|^2 \lesssim \|u - u_{\mathcal{T}}\|_{0,\Omega}^2 + \|y - y_{\mathcal{T}}\|_{0,\Omega}^2 + \|p - p_{\mathcal{T}}\|_{0,\Omega}^2 + \text{osc}_{\mathcal{T}}^2.$$

The following stability results for the error estimators are direct consequences of [10, Lemma 1], see also [5, Proposition 3.3].

**Lemma 3.5.** *For any  $\mathcal{T} \in \mathbb{T}$ , let  $u_{\mathcal{T}}, \tilde{u}_{\mathcal{T}} \in U_{ad}$ ,  $y_{\mathcal{T}}, \tilde{y}_{\mathcal{T}}, p_{\mathcal{T}}, \tilde{p}_{\mathcal{T}} \in V_{\mathcal{T}}$ . Under Assumption 2.1 we have that*

$$(3.14) \quad \begin{aligned} |\eta_{\mathcal{T},y}(u_{\mathcal{T}}, y_{\mathcal{T}}, T) - \eta_{\mathcal{T},y}(\tilde{u}_{\mathcal{T}}, \tilde{y}_{\mathcal{T}}, T)| &\lesssim \|h_{\mathcal{T}}(y_{\mathcal{T}} - \tilde{y}_{\mathcal{T}})\|_{\tilde{\omega}_T} + h_T^2 \|u_{\mathcal{T}} - \tilde{u}_{\mathcal{T}}\|_{0,T} \\ &\quad + \|\nabla h_{\mathcal{T}}\|_{0,\infty,\tilde{\omega}_T} \|y_{\mathcal{T}} - \tilde{y}_{\mathcal{T}}\|_{0,\tilde{\omega}_T}, \end{aligned}$$

$$(3.15) \quad \begin{aligned} |\eta_{\mathcal{T},p}(y_{\mathcal{T}}, p_{\mathcal{T}}, T) - \eta_{\mathcal{T},p}(\tilde{y}_{\mathcal{T}}, \tilde{p}_{\mathcal{T}}, T)| &\lesssim \|h_{\mathcal{T}}(p_{\mathcal{T}} - \tilde{p}_{\mathcal{T}})\|_{\tilde{\omega}_T} + h_T^2 \|y_{\mathcal{T}} - \tilde{y}_{\mathcal{T}}\|_{0,T} \\ &\quad + \|\nabla h_{\mathcal{T}}\|_{0,\infty,\tilde{\omega}_T} \|p_{\mathcal{T}} - \tilde{p}_{\mathcal{T}}\|_{0,\tilde{\omega}_T}. \end{aligned}$$

*Proof.* We only prove (3.14), the proof of (3.15) is very similar and we will omit it. Note that

$$\begin{aligned} \eta_{\mathcal{T},y}(u_{\mathcal{T}}, y_{\mathcal{T}}, T) &\leq \eta_{\mathcal{T},y}(\tilde{u}_{\mathcal{T}}, \tilde{y}_{\mathcal{T}}, T) + (h_T^4 \|u_{\mathcal{T}} - \tilde{u}_{\mathcal{T}}\|_{0,T}^2 \\ &\quad + \sum_{E \in \mathcal{E}_T, E \subset \partial T} h_E^3 \|[\nabla(y_{\mathcal{T}} - \tilde{y}_{\mathcal{T}})]_E \cdot n_E\|_{0,E}^2)^{\frac{1}{2}}. \end{aligned}$$

We recall the trace inequality: for any  $T \in \mathcal{T}$  and  $v_T \in V_T$  there holds

$$\|\nabla v_T\|_{0,\partial T} \lesssim h_T^{-\frac{1}{2}} \|\nabla v_T\|_{0,T} + h_T^{\frac{1}{2}} \|\nabla^2 v_T\|_{0,T}.$$

Noting that  $v_T$  is linear polynomial on  $T$ , the second term in above inequality vanishes. Thus, for the edge  $E = T \cap T'$  we have

$$(3.16) \quad \begin{aligned} h_E^{\frac{3}{2}} \|[\nabla(y_{\mathcal{T}} - \tilde{y}_{\mathcal{T}})]_E \cdot n_E\|_{0,E} &\leq h_E^{\frac{3}{2}} (\|\nabla(y_{\mathcal{T}} - \tilde{y}_{\mathcal{T}})_T\|_{0,E} + \|\nabla(y_{\mathcal{T}} - \tilde{y}_{\mathcal{T}})_{T'}\|_{0,E}) \\ &\lesssim h_T (\|\nabla(y_{\mathcal{T}} - \tilde{y}_{\mathcal{T}})\|_{0,T} + \|\nabla(y_{\mathcal{T}} - \tilde{y}_{\mathcal{T}})\|_{0,T'}) \\ &\lesssim \|\|h_{\mathcal{T}}(y_{\mathcal{T}} - \tilde{y}_{\mathcal{T}})\|\|_{T \cup T'} + \|\nabla h_{\mathcal{T}}(y_{\mathcal{T}} - \tilde{y}_{\mathcal{T}})\|_{0,T \cup T'}, \end{aligned}$$

where we used the property (2.7). Summing over the edges of  $T$  we complete the proof.  $\square$

Then we present the following localized upper bounds for the approximations of the optimal control problems.

**Lemma 3.6.** *Let  $(u, y, p) \in U_{ad} \times H_0^1(\Omega) \times H_0^1(\Omega)$  be the solution of optimal control problem (2.1)-(2.2). Given sufficiently small  $\mu$ , let  $\mathcal{T} \in \mathbb{T}_\mu$  and  $\mathcal{T} \subset \tilde{\mathcal{T}} \in \mathbb{T}$ ,  $(u_{\mathcal{T}}, y_{\mathcal{T}}, p_{\mathcal{T}}) \in U_{ad} \times V_{\mathcal{T}} \times V_{\mathcal{T}}$  and  $(u_{\tilde{\mathcal{T}}}, y_{\tilde{\mathcal{T}}}, p_{\tilde{\mathcal{T}}}) \in U_{ad} \times V_{\tilde{\mathcal{T}}} \times V_{\tilde{\mathcal{T}}}$  be the solutions of the discrete problem (2.9)-(2.11) on  $\mathcal{T}$  and  $\tilde{\mathcal{T}}$ , respectively. Then there holds*

$$(3.17) \quad \|u_{\mathcal{T}} - u_{\tilde{\mathcal{T}}}\|_{0,\Omega} + \|y_{\mathcal{T}} - y_{\tilde{\mathcal{T}}}\|_{0,\Omega} + \|p_{\mathcal{T}} - p_{\tilde{\mathcal{T}}}\|_{0,\Omega} \leq C_4 \eta_{\mathcal{T}}(\mathcal{R}_{\mathcal{T} \rightarrow \tilde{\mathcal{T}}}),$$

where  $\mathcal{R}_{\mathcal{T} \rightarrow \tilde{\mathcal{T}}}$  is the subset of elements that are refined from  $\mathcal{T}$  to  $\tilde{\mathcal{T}}$  and  $C_4$  is independent of the mesh size.

*Proof.* Note that  $V_{\mathcal{T}} \subset V_{\tilde{\mathcal{T}}}$ . From (2.10) we have

$$\begin{aligned} (\alpha u_{\mathcal{T}} + S_{\mathcal{T}}^*(y_{\mathcal{T}} - y_d), v_{\mathcal{T}} - u_{\mathcal{T}}) &\geq 0 \quad \forall v_{\mathcal{T}} \in U_{ad}, \\ (\alpha u_{\tilde{\mathcal{T}}} + S_{\tilde{\mathcal{T}}}^*(y_{\tilde{\mathcal{T}}} - y_d), v_{\tilde{\mathcal{T}}} - u_{\tilde{\mathcal{T}}}) &\geq 0 \quad \forall v_{\tilde{\mathcal{T}}} \in U_{ad}. \end{aligned}$$

Setting  $v_{\mathcal{T}} = u_{\tilde{\mathcal{T}}}$  and  $v_{\tilde{\mathcal{T}}} = u_{\mathcal{T}}$  in above inequalities and adding them together, we conclude from (2.9) and (2.11) that

$$\begin{aligned} &\alpha \|u_{\mathcal{T}} - u_{\tilde{\mathcal{T}}}\|_{0,\Omega}^2 \leq (S_{\mathcal{T}}^*(y_{\mathcal{T}} - y_d) - S_{\tilde{\mathcal{T}}}^*(y_{\tilde{\mathcal{T}}} - y_d), u_{\mathcal{T}} - u_{\tilde{\mathcal{T}}}) \\ &= (S_{\mathcal{T}}^*(y_{\mathcal{T}} - y_d) - S_{\tilde{\mathcal{T}}}^*(y_{\mathcal{T}} - y_d), u_{\tilde{\mathcal{T}}} - u_{\mathcal{T}}) + (S_{\tilde{\mathcal{T}}}^*(y_{\mathcal{T}} - y_d) - S_{\tilde{\mathcal{T}}}^*(y_{\tilde{\mathcal{T}}} - y_d), u_{\tilde{\mathcal{T}}} - u_{\mathcal{T}}) \\ &= (S_{\mathcal{T}}^*(y_{\mathcal{T}} - y_d) - S_{\tilde{\mathcal{T}}}^*(y_{\mathcal{T}} - y_d), u_{\tilde{\mathcal{T}}} - u_{\mathcal{T}}) + (S_{\mathcal{T}}u_{\mathcal{T}} - S_{\tilde{\mathcal{T}}}u_{\tilde{\mathcal{T}}}, S_{\tilde{\mathcal{T}}}u_{\tilde{\mathcal{T}}} - S_{\mathcal{T}}u_{\mathcal{T}}) \\ &= (S_{\mathcal{T}}^*(y_{\mathcal{T}} - y_d) - S_{\tilde{\mathcal{T}}}^*(y_{\mathcal{T}} - y_d), u_{\tilde{\mathcal{T}}} - u_{\mathcal{T}}) + (S_{\mathcal{T}}u_{\mathcal{T}} - S_{\tilde{\mathcal{T}}}u_{\tilde{\mathcal{T}}}, S_{\tilde{\mathcal{T}}}u_{\tilde{\mathcal{T}}} - S_{\mathcal{T}}u_{\mathcal{T}}) \\ (3.18) \quad &+ (S_{\mathcal{T}}u_{\mathcal{T}} - S_{\tilde{\mathcal{T}}}u_{\tilde{\mathcal{T}}}, S_{\mathcal{T}}u_{\mathcal{T}} - S_{\tilde{\mathcal{T}}}u_{\tilde{\mathcal{T}}}). \end{aligned}$$

It follows from Young's inequality that

$$(3.19) \quad \begin{aligned} &\alpha \|u_{\mathcal{T}} - u_{\tilde{\mathcal{T}}}\|_{0,\Omega}^2 + \|y_{\mathcal{T}} - y_{\tilde{\mathcal{T}}}\|_{0,\Omega}^2 \\ &\leq C \|S_{\tilde{\mathcal{T}}}u_{\mathcal{T}} - S_{\mathcal{T}}u_{\tilde{\mathcal{T}}}\|_{0,\Omega}^2 + C \|S_{\tilde{\mathcal{T}}}^*(y_{\mathcal{T}} - y_d) - S_{\mathcal{T}}^*(y_{\mathcal{T}} - y_d)\|_{0,\Omega}^2, \end{aligned}$$

where we used the fact that  $y_{\tilde{\mathcal{T}}} = S_{\tilde{\mathcal{T}}}u_{\tilde{\mathcal{T}}}$  and  $y_{\mathcal{T}} = S_{\mathcal{T}}u_{\mathcal{T}}$ . Moreover, from the triangle inequality, the discrete stability of elliptic equation and (3.19) we infer that

$$(3.20) \quad \begin{aligned} \|p_{\mathcal{T}} - p_{\tilde{\mathcal{T}}}\|_{0,\Omega} &\leq \|S_{\tilde{\mathcal{T}}}^*(y_{\mathcal{T}} - y_d) - S_{\tilde{\mathcal{T}}}^*(y_{\tilde{\mathcal{T}}} - y_d)\|_{0,\Omega} + \|S_{\tilde{\mathcal{T}}}^*(y_{\mathcal{T}} - y_d) - S_{\mathcal{T}}^*(y_{\mathcal{T}} - y_d)\|_{0,\Omega} \\ &\leq C \|y_{\mathcal{T}} - y_{\tilde{\mathcal{T}}}\|_{0,\Omega} + \|S_{\tilde{\mathcal{T}}}^*(y_{\mathcal{T}} - y_d) - S_{\mathcal{T}}^*(y_{\mathcal{T}} - y_d)\|_{0,\Omega} \\ &\leq C \|S_{\tilde{\mathcal{T}}}u_{\mathcal{T}} - S_{\mathcal{T}}u_{\tilde{\mathcal{T}}}\|_{0,\Omega} + C \|u_{\mathcal{T}} - u_{\tilde{\mathcal{T}}}\|_{0,\Omega} + C \|S_{\tilde{\mathcal{T}}}^*(y_{\mathcal{T}} - y_d) - S_{\mathcal{T}}^*(y_{\mathcal{T}} - y_d)\|_{0,\Omega} \\ &\leq C \|S_{\tilde{\mathcal{T}}}u_{\mathcal{T}} - S_{\mathcal{T}}u_{\tilde{\mathcal{T}}}\|_{0,\Omega} + C \|S_{\tilde{\mathcal{T}}}^*(y_{\mathcal{T}} - y_d) - S_{\mathcal{T}}^*(y_{\mathcal{T}} - y_d)\|_{0,\Omega}. \end{aligned}$$

Note that  $S_{\mathcal{T}}u_{\mathcal{T}}$  and  $S_{\tilde{\mathcal{T}}}u_{\tilde{\mathcal{T}}}$  are the finite element approximations of  $Su_{\mathcal{T}}$  on  $V_{\mathcal{T}}$  and  $V_{\tilde{\mathcal{T}}}$  associated with partitions  $\mathcal{T}$  and  $\tilde{\mathcal{T}}$ , respectively. Similarly,  $S_{\mathcal{T}}^*(y_{\mathcal{T}} - y_d)$  and  $S_{\tilde{\mathcal{T}}}^*(y_{\mathcal{T}} - y_d)$  are the finite element approximations of  $S^*(y_{\mathcal{T}} - y_d)$  on  $V_{\mathcal{T}}$  and  $V_{\tilde{\mathcal{T}}}$ . Then from Lemma 2 in [10] we conclude that

$$(3.21) \quad \|S_{\mathcal{T}}u_{\mathcal{T}} - S_{\tilde{\mathcal{T}}}u_{\tilde{\mathcal{T}}}\|_{0,\Omega} \leq C \eta_{\mathcal{T},y}(u_{\mathcal{T}}, S_{\mathcal{T}}u_{\mathcal{T}}, \mathcal{R}_{\mathcal{T} \rightarrow \tilde{\mathcal{T}}}),$$

$$(3.22) \quad \|S_{\mathcal{T}}^*(y_{\mathcal{T}} - y_d) - S_{\tilde{\mathcal{T}}}^*(y_{\mathcal{T}} - y_d)\|_{0,\Omega} \leq C \eta_{\mathcal{T},p}(y_{\mathcal{T}} - y_d, S_{\mathcal{T}}^*(y_{\mathcal{T}} - y_d), \mathcal{R}_{\mathcal{T} \rightarrow \tilde{\mathcal{T}}}).$$

Combining (3.19), (3.20), (3.21) and (3.22) we finish the proof.  $\square$

The following lemma presents a quasi-orthogonality result for the solution of elliptic boundary value problem.

**Lemma 3.7.** ([10, Lemma 3]) *For any  $\epsilon > 0$ ,  $\mathcal{T} \in \mathbb{T}$ ,  $\tilde{y}_{\mathcal{T}}, \tilde{p}_{\mathcal{T}} \in V_{\mathcal{T}}$  and  $f, g \in L^2(\Omega)$ , it holds that*

$$(3.23) \quad \begin{aligned} &\|h_{\mathcal{T}}(Sf - S_{\mathcal{T}}f)\|^2 + \|h_{\mathcal{T}}(S_{\mathcal{T}}f - \tilde{y}_{\mathcal{T}})\|^2 - (1 + \epsilon) \|h_{\mathcal{T}}(Sf - \tilde{y}_{\mathcal{T}})\|^2 \\ &\leq \epsilon^{-1} \|\nabla h_{\mathcal{T}}\|_{0,\infty,\Omega}^2 (\|Sf - S_{\mathcal{T}}f\|_{0,\Omega}^2 + \|Sf - \tilde{y}_{\mathcal{T}}\|_{0,\Omega}^2), \end{aligned}$$

$$(3.24) \quad \begin{aligned} &\|h_{\mathcal{T}}(S^*g - S_{\mathcal{T}}^*g)\|^2 + \|h_{\mathcal{T}}(S_{\mathcal{T}}^*g - \tilde{p}_{\mathcal{T}})\|^2 - (1 + \epsilon) \|h_{\mathcal{T}}(S^*g - \tilde{p}_{\mathcal{T}})\|^2 \\ &\leq \epsilon^{-1} \|\nabla h_{\mathcal{T}}\|_{0,\infty,\Omega}^2 (\|S^*g - S_{\mathcal{T}}^*g\|_{0,\Omega}^2 + \|S^*g - \tilde{p}_{\mathcal{T}}\|_{0,\Omega}^2). \end{aligned}$$

We also recall the following inequality which is opposed to Lemma 3.1.

**Lemma 3.8.** ([10, Lemma 4]) *For any  $\mathcal{T} \in \mathbb{T}$  and  $f, g \in L^2(\Omega)$  there holds*

$$(3.25) \quad \| |h_{\mathcal{T}}(Sf - \tilde{y}_{\mathcal{T}})| \| \lesssim \| Sf - \tilde{y}_{\mathcal{T}} \|_{0,\Omega} + \text{osc}(f, \mathcal{T}), \quad \forall \tilde{y}_{\mathcal{T}} \in V_{\mathcal{T}},$$

$$(3.26) \quad \| |h_{\mathcal{T}}(S^*g - \tilde{p}_{\mathcal{T}})| \| \lesssim \| S^*g - \tilde{p}_{\mathcal{T}} \|_{0,\Omega} + \text{osc}(g, \mathcal{T}), \quad \forall \tilde{p}_{\mathcal{T}} \in V_{\mathcal{T}}.$$

As a final preliminary result we show that the  $L^2$  norm errors of the control, the state and adjoint state can be bounded from above by the best approximations of the state and adjoint state variables in finite element space  $V_{\mathcal{T}}$  measured in  $L^2$ -norm, plus data oscillations, if  $\mathcal{T}$  is sufficiently mildly graded. We refer to [10, Corollary 1] for a similar result for elliptic boundary value problem and [7] for the similar Céa's lemma in energy norm.

**Theorem 3.9.** *Let  $(u, y, p) \in U_{ad} \times H_0^1(\Omega) \times H_0^1(\Omega)$  be the solution of optimal control problem (2.1)-(2.2) and  $(u_{\mathcal{T}}, y_{\mathcal{T}}, p_{\mathcal{T}}) \in U_{ad} \times V_{\mathcal{T}} \times V_{\mathcal{T}}$  be the solution of the discrete problem (2.9)-(2.11). Then we have*

$$(3.27) \quad \|u - u_{\mathcal{T}}\|_{0,\Omega} + \|y - y_{\mathcal{T}}\|_{0,\Omega} + \|p - p_{\mathcal{T}}\|_{0,\Omega} \lesssim \inf_{v_{\mathcal{T}} \in V_{\mathcal{T}}} \|y - v_{\mathcal{T}}\|_{0,\Omega} + \text{osc}(f + u, \mathcal{T}) + \inf_{w_{\mathcal{T}} \in V_{\mathcal{T}}} \|p - w_{\mathcal{T}}\|_{0,\Omega} + \text{osc}(y - y_d, \mathcal{T}).$$

*Proof.* From the standard error estimate for elliptic optimal control problem with variational control discretization (see [18, Sec. 3, Thm. 3.4]) we have

$$\begin{aligned} & \|u - u_{\mathcal{T}}\|_{0,\Omega} + \|y - y_{\mathcal{T}}\|_{0,\Omega} + \|p - p_{\mathcal{T}}\|_{0,\Omega} \\ & \lesssim \|y - S_{\mathcal{T}}u\|_{0,\Omega} + \|p - S_{\mathcal{T}}^*(y - y_d)\|_{0,\Omega}. \end{aligned}$$

Recall that  $\|\nabla h_{\mathcal{T}}\|_{0,\infty,\Omega} \leq \mu$ . For any  $v_{\mathcal{T}}, w_{\mathcal{T}} \in V_{\mathcal{T}}$ , it follows from Lemmas 3.1, 3.7 and 3.8 that

$$(3.28) \quad \begin{aligned} & \|u - u_{\mathcal{T}}\|_{0,\Omega} + \|y - y_{\mathcal{T}}\|_{0,\Omega} + \|p - p_{\mathcal{T}}\|_{0,\Omega} \\ & \lesssim \| |h_{\mathcal{T}}(y - S_{\mathcal{T}}u)| \| + \| |h_{\mathcal{T}}(p - S_{\mathcal{T}}^*(y - y_d)) \| \\ & \lesssim \| |h_{\mathcal{T}}(y - v_{\mathcal{T}})| \| + \mu(\|y - S_{\mathcal{T}}u\|_{0,\Omega} + \|y - v_{\mathcal{T}}\|_{0,\Omega}) \\ & \quad + \| |h_{\mathcal{T}}(p - w_{\mathcal{T}})| \| + \mu(\|p - S_{\mathcal{T}}^*(y - y_d)\|_{0,\Omega} + \|p - w_{\mathcal{T}}\|_{0,\Omega}) \\ & \lesssim (1 + \mu)\|y - v_{\mathcal{T}}\|_{0,\Omega} + \text{osc}(f + u, \mathcal{T}) + \mu(\|y - y_{\mathcal{T}}\|_{0,\Omega} + \|y_{\mathcal{T}} - S_{\mathcal{T}}u\|_{0,\Omega}) \\ & \quad + (1 + \mu)\|p - w_{\mathcal{T}}\|_{0,\Omega} + \text{osc}(y - y_d, \mathcal{T}) + \mu(\|p - p_{\mathcal{T}}\|_{0,\Omega} + \|p_{\mathcal{T}} - S_{\mathcal{T}}^*(y - y_d)\|_{0,\Omega}) \\ & \lesssim (1 + \mu)\|y - v_{\mathcal{T}}\|_{0,\Omega} + \text{osc}(f + u, \mathcal{T}) + \mu(\|y - y_{\mathcal{T}}\|_{0,\Omega} + \|u - u_{\mathcal{T}}\|_{0,\Omega}) \\ & \quad + (1 + \mu)\|p - w_{\mathcal{T}}\|_{0,\Omega} + \text{osc}(y - y_d, \mathcal{T}) + \mu(\|p - p_{\mathcal{T}}\|_{0,\Omega} + \|y - y_{\mathcal{T}}\|_{0,\Omega}), \end{aligned}$$

where we used the discrete stability of elliptic equation in the last estimate. By taking  $\mu$  sufficiently small and  $v_{\mathcal{T}}, w_{\mathcal{T}}$  arbitrary we complete the proof.  $\square$

**Remark 3.10.** *Compared to Theorem 3.9 we have an alternative result: there exists a constant  $C_5$  independent of the mesh size such that*

$$(3.29) \quad \begin{aligned} & \|u - u_{\mathcal{T}}\|_{0,\Omega} + \|y - y_{\mathcal{T}}\|_{0,\Omega} + \|p - p_{\mathcal{T}}\|_{0,\Omega} + \text{osc}_{\mathcal{T}} \\ & \leq C_5 \left( \inf_{v_{\mathcal{T}} \in V_{\mathcal{T}}} (\|y - v_{\mathcal{T}}\|_{0,\Omega} + \text{osc}(v_{\mathcal{T}} - y_d, \mathcal{T})) \right. \\ & \quad \left. + \inf_{w_{\mathcal{T}} \in V_{\mathcal{T}}} (\|p - w_{\mathcal{T}}\|_{0,\Omega} + \text{osc}(f + P_{[a,b]}(w_{\mathcal{T}}), \mathcal{T})) \right) \end{aligned}$$

provided that  $h_0 \ll 1$ . In fact, we can conclude from the triangle inequality that

$$\begin{aligned} \text{osc}(f + u, \mathcal{T}) + \text{osc}(y - y_d, \mathcal{T}) & \lesssim \text{osc}(f + P_{[a,b]}(w_{\mathcal{T}}), \mathcal{T}) + \text{osc}(v_{\mathcal{T}} - y_d, \mathcal{T}) \\ & \quad + h_{\mathcal{T}}^2 \|u - P_{[a,b]}(w_{\mathcal{T}})\|_{0,\Omega} + h_{\mathcal{T}}^2 \|y - v_{\mathcal{T}}\|_{0,\Omega}. \end{aligned}$$

From the Lipschitz property of the projection operator  $P_{[a,b]}$  we have

$$\|u - P_{[a,b]}(w_{\mathcal{T}})\|_{0,\Omega} \lesssim \|p - w_{\mathcal{T}}\|_{0,\Omega}.$$

Moreover,

$$\begin{aligned}
\text{osc}(f + u_{\mathcal{T}}, \mathcal{T}) + \text{osc}(y_{\mathcal{T}} - y_d, \mathcal{T}) &\lesssim \text{osc}(f + P_{[a,b]}(w_{\mathcal{T}}), \mathcal{T}) + \text{osc}(v_{\mathcal{T}} - y_d, \mathcal{T}) \\
&\quad + h_{\mathcal{T}}^2 \|u_{\mathcal{T}} - P_{[a,b]}(w_{\mathcal{T}})\|_{0,\Omega} + h_{\mathcal{T}}^2 \|y_{\mathcal{T}} - v_{\mathcal{T}}\|_{0,\Omega} \\
&\lesssim \text{osc}(f + P_{[a,b]}(w_{\mathcal{T}}), \mathcal{T}) + \text{osc}(v_{\mathcal{T}} - y_d, \mathcal{T}) \\
&\quad + h_{\mathcal{T}}^2 (\|p - w_{\mathcal{T}}\|_{0,\Omega} + \|y - v_{\mathcal{T}}\|_{0,\Omega} \\
&\quad + \|p - p_{\mathcal{T}}\|_{0,\Omega} + \|y - y_{\mathcal{T}}\|_{0,\Omega}).
\end{aligned}$$

Combining the above results, the fact that  $h_{\mathcal{T}} \leq h_0$  and using (3.27) we can conclude (3.29) provided that  $h_0 \ll 1$  and  $\mu$  sufficiently small.

**3.2. Convergence analysis of AFEM for OCPs in  $L^2$ .** In this subsection we will prove the convergence of  $L^2$ -norm based AFEM for solving optimal control problems. In the following we assume that  $(u, y, p) \in U_{ad} \times H_0^1(\Omega) \times H_0^1(\Omega)$  is the solution of optimal control problem (2.1)-(2.2) and  $(u_k, y_k, p_k) \in U_{ad} \times V_k \times V_k$  is the solution of the discrete problem (2.9)-(2.11) generated by the adaptive Algorithm 2.4.

At first we prove some quasi-orthogonality properties.

**Lemma 3.11.** *For any  $\epsilon > 0$  there hold*

$$(3.30) \quad \begin{aligned} &\|h_{k+1}(y - y_{k+1})\|^2 + \|h_{k+1}(y_{k+1} - y_k)\|^2 - (1 + \epsilon)\|h_k(y - y_k)\|^2 \\ &\lesssim \epsilon^{-1}(\mu^2 + h_0^2)(\|u - u_{k+1}\|_{0,\Omega}^2 + \|y - y_{k+1}\|_{0,\Omega}^2 + \|y - y_k\|_{0,\Omega}^2), \end{aligned}$$

$$(3.31) \quad \begin{aligned} &\|h_{k+1}(p - p_{k+1})\|^2 + \|h_{k+1}(p_{k+1} - p_k)\|^2 - (1 + \epsilon)\|h_k(p - p_k)\|^2 \\ &\lesssim \epsilon^{-1}(\mu^2 + h_0^2)(\|y - y_{k+1}\|_{0,\Omega}^2 + \|p - p_{k+1}\|_{0,\Omega}^2 + \|p - p_k\|_{0,\Omega}^2). \end{aligned}$$

*Proof.* At first we estimate (3.30). We use the abbreviations  $e_k = y - y_k$  and  $\tilde{e}_k = y_{k+1} - y_k$ , then we have

$$(3.32) \quad \|h_{k+1}e_{k+1}\|^2 = \|h_{k+1}e_k\|^2 - \|h_{k+1}\tilde{e}_k\|^2 - 2(\nabla(h_{k+1}e_{k+1}), \nabla(h_{k+1}\tilde{e}_k)).$$

An elementary calculation gives

$$\begin{aligned}
|(\nabla(h_{k+1}e_{k+1}), \nabla(h_{k+1}\tilde{e}_k))| &= |(\nabla e_{k+1}, \nabla(h_{k+1}^2\tilde{e}_k)) + (|\nabla h_{k+1}|^2 e_{k+1}, \tilde{e}_k) \\
&\quad - (e_k \nabla h_{k+1}, \nabla(h_{k+1}e_{k+1})) + (e_{k+1} \nabla h_{k+1}, \nabla(h_{k+1}e_k))| \\
&\leq \|e_k \nabla h_{k+1}\|_{0,\Omega} \|h_{k+1}e_{k+1}\| + \|e_{k+1} \nabla h_{k+1}\|_{0,\Omega} \|h_{k+1}e_k\| \\
(3.33) \quad &\quad + \|e_{k+1} \nabla h_{k+1}\|_{0,\Omega} \|\tilde{e}_k \nabla h_{k+1}\|_{0,\Omega} + |(\nabla e_{k+1}, \nabla(h_{k+1}^2\tilde{e}_k))|.
\end{aligned}$$

It remains to estimate  $|(\nabla e_{k+1}, \nabla(h_{k+1}^2\tilde{e}_k))|$ . We divide the estimate into two steps. Firstly, from the orthogonality property we have

$$\begin{aligned}
&|(\nabla(Su_{k+1} - y_{k+1}), \nabla(h_{k+1}^2\tilde{e}_k))| \\
&= |(\nabla(Su_{k+1} - y_{k+1}), \nabla(h_{k+1}^2\tilde{e}_k - \Pi_{k+1}(h_{k+1}^2\tilde{e}_k)))| \\
(3.34) \quad &\leq \sum_{T \in \mathcal{T}_{k+1}} \|\nabla(Su_{k+1} - y_{k+1})\|_{0,T} \|\nabla(h_{k+1}^2\tilde{e}_k - \Pi_{k+1}(h_{k+1}^2\tilde{e}_k))\|_{0,T}.
\end{aligned}$$

For each  $T \in \mathcal{T}_{k+1}$  we know that  $h_{k+1}|_T$  and  $\tilde{e}_k$  are linear. So  $|\nabla^2 h_{k+1}^2| \lesssim |\nabla h_{k+1}|^2$ ,  $\nabla^2 h_{k+1} = 0$  and  $\nabla^2 \tilde{e}_k = 0$ . By using the inverse inequality, the standard interpolation error estimate ([7]), the fact that  $\|\nabla h_{k+1}\|_{0,\infty,\Omega} \lesssim 1$  and (2.7) we can derive

$$\begin{aligned}
&\|\nabla(h_{k+1}^2\tilde{e}_k - \Pi_{k+1}(h_{k+1}^2\tilde{e}_k))\|_{0,T} \\
&\lesssim h_T \|\nabla^2(h_{k+1}^2\tilde{e}_k)\|_{0,T} \\
&\lesssim h_T (\|\nabla h_{k+1}^2 \nabla \tilde{e}_k\|_{0,T} + \|\nabla^2 h_{k+1}^2 \tilde{e}_k\|_{0,T}) \\
&\lesssim h_T \|\nabla h_{k+1}\|_{0,\infty,T} (\|h_{k+1}\|_{0,\infty,T} \|\nabla \tilde{e}_k\|_{0,T} + \|\nabla h_{k+1}\|_{0,\infty,T} \|\tilde{e}_k\|_{0,T}) \\
&\lesssim \|\nabla h_{k+1}\|_{0,\infty,T} \|\tilde{e}_k\|_{0,T} (\|h_{k+1}\|_{0,\infty,T} + h_T \|\nabla h_{k+1}\|_{0,\infty,T}) \\
(3.35) \quad &\lesssim \|\nabla h_{k+1}\|_{0,\infty,T} \|h_{k+1}\|_{0,\infty,T} \|\tilde{e}_k\|_{0,T}.
\end{aligned}$$

From (3.34), (3.35) and the stability of elliptic equation we conclude that

$$\begin{aligned}
& |(\nabla(Su_{k+1} - y_{k+1}), \nabla(h_{k+1}^2 \tilde{e}_k))| \\
& \lesssim \sum_{T \in \mathcal{T}_{k+1}} \|\nabla(Su_{k+1} - y_{k+1})\|_{0,T} \|\nabla h_{k+1}\|_{0,\infty,T} \|h_{k+1}\|_{0,\infty,T} \|\tilde{e}_k\|_{0,T} \\
& \lesssim (\|h_{k+1} \nabla e_{k+1}\|_{0,\Omega} + \|h_{k+1} \nabla(y - Su_{k+1})\|_{0,\Omega}) \|\nabla h_{k+1}\|_{0,\infty,\Omega} \|\tilde{e}_k\|_{0,\Omega} \\
(3.36) \quad & \lesssim (\|h_{k+1} e_{k+1}\| + \|\nabla h_{k+1} e_{k+1}\|_{0,\Omega} + \|h_{k+1}\|_{0,\infty,\Omega} \|u - u_{k+1}\|_{0,\Omega}) \|\nabla h_{k+1}\|_{0,\infty,\Omega} \|\tilde{e}_k\|_{0,\Omega}.
\end{aligned}$$

Secondly, from (2.3) we deduce

$$\begin{aligned}
(3.37) \quad & |(\nabla(y - Su_{k+1}), \nabla(h_{k+1}^2 \tilde{e}_k))| = |(u - u_{k+1}, h_{k+1}^2 \tilde{e}_k)| \\
& \leq \|h_{k+1}\|_{0,\infty,\Omega}^2 \|u - u_{k+1}\|_{0,\Omega} \|\tilde{e}_k\|_{0,\Omega}.
\end{aligned}$$

Inserting the above estimates into (3.33) and using Young's inequality we are led to

$$\begin{aligned}
(3.38) \quad & |(\nabla(h_{k+1} e_{k+1}), \nabla(h_{k+1} \tilde{e}_k))| \leq \frac{\delta}{2} (\|h_{k+1} e_{k+1}\|^2 + \|h_{k+1} \tilde{e}_k\|^2) + \|h_{k+1}\|_{0,\infty,\Omega}^2 \|u - u_{k+1}\|_{0,\Omega}^2 \\
& + C(1 + \frac{1}{\delta}) (\|\nabla h_{k+1}\|_{0,\infty,\Omega}^2 + \|h_{k+1}\|_{0,\infty,\Omega}^2) (\|e_{k+1}\|_{0,\Omega}^2 + \|e_k\|_{0,\Omega}^2).
\end{aligned}$$

Combining (3.32) and (3.38) yields

$$\begin{aligned}
(3.39) \quad & (1 - \delta) \|h_{k+1} e_{k+1}\|^2 \leq (1 + \delta) \|h_{k+1} e_k\|^2 - \|h_{k+1} \tilde{e}_k\|^2 + 2 \|h_{k+1}\|_{0,\infty,\Omega}^2 \|u - u_{k+1}\|_{0,\Omega}^2 \\
& + 2C(1 + \frac{1}{\delta}) (\|\nabla h_{k+1}\|_{0,\infty,\Omega}^2 + \|h_{k+1}\|_{0,\infty,\Omega}^2) (\|e_{k+1}\|_{0,\Omega}^2 + \|e_k\|_{0,\Omega}^2).
\end{aligned}$$

Dividing both sides by  $1 - \delta$  and choosing  $\frac{1+\delta}{1-\delta}$  as  $1 + \epsilon$  we arrive at

$$\begin{aligned}
(3.40) \quad & \|h_{k+1} e_{k+1}\|^2 \lesssim (1 + \epsilon) \|h_{k+1} e_k\|^2 - \|h_{k+1} \tilde{e}_k\|^2 + \|h_{k+1}\|_{0,\infty,\Omega}^2 \|u - u_{k+1}\|_{0,\Omega}^2 \\
& + \epsilon^{-1} (\|\nabla h_{k+1}\|_{0,\infty,\Omega}^2 + \|h_{k+1}\|_{0,\infty,\Omega}^2) (\|e_{k+1}\|_{0,\Omega}^2 + \|e_k\|_{0,\Omega}^2).
\end{aligned}$$

Note that  $h_{k+1} \leq h_k$  and  $\|\nabla h_{k+1}\|_{0,\infty,\Omega} \leq \mu$  we have

$$\begin{aligned}
(3.41) \quad & \|h_{k+1} e_k\| \leq \|h_{k+1} \nabla e_k\|_{0,\Omega} + \|\nabla h_{k+1}\|_{0,\infty,\Omega} \|e_k\|_{0,\Omega} \\
& \leq \|h_k \nabla e_k\|_{0,\Omega} + \mu \|e_k\|_{0,\Omega} \\
& \leq \|h_k e_k\| + 2\mu \|e_k\|_{0,\Omega},
\end{aligned}$$

so for any  $\epsilon > 0$  we have

$$(3.42) \quad \|h_{k+1} e_k\|^2 \leq (1 + \epsilon) \|h_k e_k\|^2 + (1 + \frac{1}{\epsilon}) 4\mu^2 \|e_k\|_{0,\Omega}^2,$$

substituting the above result into (3.40) and using  $\|h_{k+1}\|_{0,\infty,\Omega} \leq h_0$  we complete the proof of (3.30) with  $(1 + \epsilon)$  replaced by  $(1 + \epsilon)^2$  which are equivalent. The proof of (3.31) is very similar and we omit it here.  $\square$

We also need the following estimator reduction property, the proof is very similar to [5, Corollary].

**Lemma 3.12.** *For any  $\delta \in (0, 1]$  there hold*

$$\begin{aligned}
(3.43) \quad & \eta_{k+1,y}^2(u_{k+1}, y_{k+1}, \mathcal{T}_{k+1}) - (1 + \delta) \left( \eta_{k,y}^2(u_k, y_k, \mathcal{T}_k) - \lambda \eta_{k,y}^2(u_k, y_k, \mathcal{M}_k) \right) \\
& \leq \delta^{-1} \left( \|h_{k+1}(y_{k+1} - y_k)\|^2 + \mu^2 \|y_{k+1} - y_k\|_{0,\Omega}^2 + h_0^4 \|u_{k+1} - u_k\|_{0,\Omega}^2 \right),
\end{aligned}$$

$$\begin{aligned}
(3.44) \quad & \eta_{k+1,p}^2(y_{k+1}, p_{k+1}, \mathcal{T}_{k+1}) - (1 + \delta) \left( \eta_{k,p}^2(y_k, p_k, \mathcal{T}_k) - \lambda \eta_{k,p}^2(y_k, p_k, \mathcal{M}_k) \right) \\
& \leq \delta^{-1} \left( \|h_{k+1}(p_{k+1} - p_k)\|^2 + \mu^2 \|p_{k+1} - p_k\|_{0,\Omega}^2 + h_0^4 \|y_{k+1} - y_k\|_{0,\Omega}^2 \right),
\end{aligned}$$

where  $\lambda = 1 - 2^{-\frac{3r}{d}}$  and  $r$  is the number of bisections for marked elements in  $\mathcal{M}_k$  during the local mesh refinement.

*Proof.* It follows from (3.14) and Young's inequality that for any  $\delta \in (0, 1]$  and  $T \in \mathcal{T}_{k+1}$  there holds

$$\begin{aligned} & \eta_{k+1,y}^2(u_{k+1}, y_{k+1}, T) - (1 + \delta)\eta_{k+1,y}^2(u_k, y_k, T) \\ \lesssim & \delta^{-1} \left( \|h_{k+1}(y_{k+1} - y_k)\|_{\tilde{\omega}_T}^2 + \|\nabla h_{k+1}\|_{0,\infty,\tilde{\omega}_T}^2 \|y_{k+1} - y_k\|_{0,\tilde{\omega}_T}^2 + h_T^4 \|u_{k+1} - u_k\|_{0,\tilde{\omega}_T}^2 \right). \end{aligned}$$

Summing over  $T \in \mathcal{T}_{k+1}$  and using the fact that the triangulation is shape regular, (2.8) and  $h_{k+1} \leq h_0$ , we have

$$(3.45) \quad \begin{aligned} & \eta_{k+1,y}^2(u_{k+1}, y_{k+1}, \mathcal{T}_{k+1}) - (1 + \delta)\eta_{k+1,y}^2(u_k, y_k, \mathcal{T}_{k+1}) \\ \lesssim & \delta^{-1} \left( \|h_{k+1}(y_{k+1} - y_k)\|^2 + \mu^2 \|y_{k+1} - y_k\|_{0,\Omega}^2 + h_0^4 \|u_{k+1} - u_k\|_{0,\Omega}^2 \right). \end{aligned}$$

For  $T' \in \mathcal{T}_k$  we define  $\mathcal{T}_{T'} = \{T \in \mathcal{T}_{k+1} : T \subset T'\}$ . From the definition of bisection algorithm we know that for a marked element  $T' \in \mathcal{M}_k$  and  $T \in \mathcal{T}_{T'}$  there holds  $h_T \leq 2^{-\frac{r}{d}} h_{T'}$  and  $[\nabla y_k] = 0$  across the edges of  $T$  which lie in the interior of  $T'$ . Therefore,

$$(3.46) \quad \eta_{k+1,y}^2(u_k, y_k, \mathcal{T}_{T'}) \leq 2^{-\frac{3r}{d}} \eta_{k,y}^2(u_k, y_k, T').$$

For  $T' \in \mathcal{T}_k \setminus \mathcal{M}_k$ , we have the monotonicity property  $\eta_{k+1,y}(u_k, y_k, \mathcal{T}_{T'}) \leq \eta_{k,y}(u_k, y_k, T')$  (see, for instance, [5, Remark 2.1]). Summing over  $T \in \mathcal{T}_{k+1}$  we obtain

$$(3.47) \quad \begin{aligned} \eta_{k+1,y}^2(u_k, y_k, \mathcal{T}_{k+1}) & \leq 2^{-\frac{3r}{d}} \eta_{k,y}^2(u_k, y_k, \mathcal{M}_k) + \eta_{k,y}^2(u_k, y_k, \mathcal{T}_k \setminus \mathcal{M}_k) \\ & = \eta_{k,y}^2(u_k, y_k, \mathcal{T}_k) - \lambda \eta_{k,y}^2(u_k, y_k, \mathcal{M}_k). \end{aligned}$$

Combining (3.45) and (3.47) prove (3.43). Similarly, we can prove (3.44).  $\square$

Now we are in the position to prove the contraction property for the weighted energy norm errors of the state and adjoint state.

**Theorem 3.13.** *Let  $(u, y, p) \in U_{ad} \times H_0^1(\Omega) \times H_0^1(\Omega)$  be the solution of optimal control problem (2.1)-(2.2) and  $(u_k, y_k, p_k) \in U_{ad} \times V_k \times V_k$  be the solution of the discrete problem (2.9)-(2.11) generated by the adaptive Algorithm 2.4. Then there exist constant  $\gamma > 0$  and  $\nu \in (0, 1)$  depending on  $c_{\mathbb{T}}, C_{\mathbb{T}}, C_{\text{reg}}$ , the shape regularity of  $\mathcal{T}_0$ , the parameter  $\theta$  in Algorithm 2.3 and the number of times  $r$  that each element in  $\mathcal{T}_k$  is bisected, such that for sufficiently small  $\mu$  it holds*

$$(3.48) \quad \begin{aligned} & \|h_{k+1}(y - y_{k+1})\|^2 + \|h_{k+1}(p - p_{k+1})\|^2 + \gamma \eta_{k+1}^2(\mathcal{T}_{k+1}) \\ \leq & \nu^2 \left( \|h_k(y - y_k)\|^2 + \|h_k(p - p_k)\|^2 + \gamma \eta_k^2(\mathcal{T}_k) \right) \end{aligned}$$

provided that  $h_0 \ll 1$ .

*Proof.* We use the abbreviations

$$\begin{aligned} E_k^2 & := \|h_k(y - y_k)\|^2 + \|h_k(p - p_k)\|^2, \\ \tilde{E}_k^2 & := \|h_{k+1}(y_{k+1} - y_k)\|^2 + \|h_{k+1}(p_{k+1} - p_k)\|^2. \end{aligned}$$

From Lemma 3.4 we have

$$(3.49) \quad E_k^2 \leq C_3 \eta_k^2(\mathcal{T}_k).$$

We can conclude from Lemma 3.11 and Theorem 2.2 that

$$(3.50) \quad E_{k+1}^2 \leq (1 + \epsilon) E_k^2 - \tilde{E}_k^2 + C_6 \epsilon^{-1} (\mu^2 + h_0^2) (\eta_k^2(\mathcal{T}_k) + \eta_{k+1}^2(\mathcal{T}_{k+1})).$$

Moreover, it follows from Lemma 3.12, the triangle inequality and Theorem 2.2 that

$$\begin{aligned}
\eta_{k+1}^2(\mathcal{T}_{k+1}) &\leq (1+\delta)(\eta_k^2(\mathcal{T}_k) - \lambda\eta_k^2(\mathcal{M}_k)) + C_7\delta^{-1}\left(\tilde{E}_k^2 + (\mu^2 + h_0^4)(\eta_k^2(\mathcal{T}_k) + \eta_{k+1}^2(\mathcal{T}_{k+1}))\right) \\
&\leq (1+\delta)(1 - \lambda\theta^2)\eta_k^2(\mathcal{T}_k) + C_7\delta^{-1}\left(\tilde{E}_k^2 + (\mu^2 + h_0^4)(\eta_k^2(\mathcal{T}_k) + \eta_{k+1}^2(\mathcal{T}_{k+1}))\right) \\
&\leq (1+\delta)\left(\left(1 - \frac{1}{2}\lambda\theta^2\right)\eta_k^2(\mathcal{T}_k) - \frac{1}{2C_3}\lambda\theta^2 E_k^2\right) \\
(3.51) \quad &+ C_7\delta^{-1}\left(\tilde{E}_k^2 + (\mu^2 + h_0^4)(\eta_k^2(\mathcal{T}_k) + \eta_{k+1}^2(\mathcal{T}_{k+1}))\right),
\end{aligned}$$

where we used Dörfler's marking strategy in Algorithm 2.3 in the second inequality and (3.49) in the third inequality.

Multiplying (3.51) with  $\tilde{\gamma} = \delta C_7^{-1}$  and adding the result to (3.50) gives

$$\begin{aligned}
E_{k+1}^2 + \tilde{\gamma}\eta_{k+1}^2(\mathcal{T}_{k+1}) &\leq (1+\delta)\left(1 - \frac{1}{2}\lambda\theta^2\right)\tilde{\gamma}\eta_k^2(\mathcal{T}_k) + \left((1+\epsilon) - (1+\delta)\frac{\tilde{\gamma}}{2C_3}\lambda\theta^2\right)E_k^2 \\
(3.52) \quad &+ (\mu^2 + h_0^2)(1 + C_6\epsilon^{-1})(\eta_k^2(\mathcal{T}_k) + \eta_{k+1}^2(\mathcal{T}_{k+1})).
\end{aligned}$$

Fixing a sufficiently small  $\delta$  and a sufficiently small  $\epsilon$  such that

$$(3.53) \quad \tilde{\nu}^2 := \max\left\{\left(1+\delta\right)\left(1 - \frac{1}{2}\lambda\theta^2\right), \left(1+\epsilon\right) - \left(1+\delta\right)\frac{\tilde{\gamma}}{2C_3}\lambda\theta^2\right\} < 1,$$

we conclude from (3.52) that

$$\begin{aligned}
E_{k+1}^2 + \left(\tilde{\gamma} - (\mu^2 + h_0^2)(1 + C_6\epsilon^{-1})\right)\eta_{k+1}^2(\mathcal{T}_{k+1}) \\
(3.54) \quad &\leq \tilde{\nu}^2 E_k^2 + \left(\tilde{\nu}^2\tilde{\gamma} + (\mu^2 + h_0^2)(1 + C_6\epsilon^{-1})\right)\eta_k^2(\mathcal{T}_k).
\end{aligned}$$

Choosing  $\mu$  sufficiently small and  $h_0 \ll 1$  such that

$$(3.55) \quad 0 < \frac{\tilde{\nu}^2\tilde{\gamma} + (\mu^2 + h_0^2)(1 + C_6\epsilon^{-1})}{\tilde{\gamma} - (\mu^2 + h_0^2)(1 + C_6\epsilon^{-1})} \leq \frac{1 + \tilde{\nu}^2}{2} := \nu^2.$$

By choosing  $\gamma := \tilde{\gamma} - (\mu^2 + h_0^2)(1 + C_6\epsilon^{-1})$  we obtain

$$\begin{aligned}
E_{k+1}^2 + \gamma\eta_{k+1}^2(\mathcal{T}_{k+1}) &\leq \tilde{\nu}^2 E_k^2 + \gamma \frac{\tilde{\nu}^2\tilde{\gamma} + (\mu^2 + h_0^2)(1 + C_6\epsilon^{-1})}{\tilde{\gamma} - (\mu^2 + h_0^2)(1 + C_6\epsilon^{-1})}\eta_k^2(\mathcal{T}_k) \\
&\leq \tilde{\nu}^2 E_k^2 + \gamma\nu^2\eta_k^2(\mathcal{T}_k) \\
&\leq \nu^2\left(\frac{2\tilde{\nu}^2}{1 + \tilde{\nu}^2}E_k^2 + \gamma\eta_k^2(\mathcal{T}_k)\right) \\
(3.56) \quad &\leq \nu^2\left(E_k^2 + \gamma\eta_k^2(\mathcal{T}_k)\right)
\end{aligned}$$

in view of (3.53). This completes the proof.  $\square$

We finally show that the  $L^2$ -norm errors for both the control, the state and adjoint state converge to zero with linear rate.

**Theorem 3.14.** *Let  $(u, y, p) \in U_{ad} \times H_0^1(\Omega) \times H_0^1(\Omega)$  be the solution of optimal control problem (2.1)-(2.2) and  $(u_k, y_k, p_k) \in U_{ad} \times V_k \times V_k$  be the solution of the discrete problem (2.9)-(2.11) generated by the adaptive Algorithm 2.4. Let the assumptions in Theorem 3.13 be fulfilled, it holds that for  $k \geq l$*

$$\begin{aligned}
&\|u - u_{k+1}\|_{0,\Omega} + \|y - y_{k+1}\|_{0,\Omega} + \|p - p_{k+1}\|_{0,\Omega} + \text{osc}_{k+1} \\
(3.57) \quad &\lesssim \nu^{k-l}\left(\|u - u_l\|_{0,\Omega} + \|y - y_l\|_{0,\Omega} + \|p - p_l\|_{0,\Omega} + \text{osc}_l\right)
\end{aligned}$$

provided that  $h_0 \ll 1$ .

*Proof.* From the dominance of the indicator over oscillation (see [5, Remark 2.1]) we have

$$\begin{aligned}\text{osc}_{k+1}(f + u_{k+1}, \mathcal{T}_{k+1}) &\leq \eta_{k+1,y}(u_{k+1}, y_{k+1}, \mathcal{T}_{k+1}), \\ \text{osc}_{k+1}(y_{k+1} - y_d, \mathcal{T}_{k+1}) &\leq \eta_{k+1,p}(y_{k+1}, p_{k+1}, \mathcal{T}_{k+1}),\end{aligned}$$

which in turn implies

$$(3.58) \quad \text{osc}_{k+1}^2 \leq \eta_{k+1}^2(\mathcal{T}_{k+1}),$$

this together with Lemma 3.2 yields

$$(3.59) \quad \begin{aligned}& (\|u - u_{k+1}\|_{0,\Omega} + \|y - y_{k+1}\|_{0,\Omega} + \|p - p_{k+1}\|_{0,\Omega} + \text{osc}_{k+1})^2 \\ & \lesssim \|||h_{k+1}(y - y_{k+1})\|||^2 + \|||h_{k+1}(p - p_{k+1})\|||^2 + \gamma\eta_{k+1}^2(\mathcal{T}_{k+1}).\end{aligned}$$

On the other hand, it follows from Lemma 3.4 and Theorem 2.2 that

$$(3.60) \quad \begin{aligned}& \|||h_{k+1}(y - y_{k+1})\|||^2 + \|||h_{k+1}(p - p_{k+1})\|||^2 + \gamma\eta_{k+1}^2(\mathcal{T}_{k+1}) \\ & \lesssim \eta_{k+1}^2(\mathcal{T}_{k+1}) \\ & \lesssim (\|u - u_{k+1}\|_{0,\Omega} + \|y - y_{k+1}\|_{0,\Omega} + \|p - p_{k+1}\|_{0,\Omega} + \text{osc}_{k+1})^2.\end{aligned}$$

Combining (3.59) and (3.60) we arrive at

$$(3.61) \quad \begin{aligned}& (\|u - u_{k+1}\|_{0,\Omega} + \|y - y_{k+1}\|_{0,\Omega} + \|p - p_{k+1}\|_{0,\Omega} + \text{osc}_{k+1})^2 \\ & \simeq \|||h_{k+1}(y - y_{k+1})\|||^2 + \|||h_{k+1}(p - p_{k+1})\|||^2 + \gamma\eta_{k+1}^2(\mathcal{T}_{k+1}),\end{aligned}$$

this together with Theorem 3.13 completes the proof.  $\square$

#### 4. COMPLEXITY OF AFEM FOR THE OPTIMAL CONTROL PROBLEM UNDER $L^2$ -NORM

In this section we prove the quasi-optimal complexity of  $L^2$ -norm based AFEM for solving optimal control problems. To begin with, we follow the idea of [10] (see [5, 33] for the definitions of approximation classes with respect to the energy norm based AFEM) to introduce the approximation class  $\mathcal{A}^s$  for  $s > 0$ :

$$\mathcal{A}^s = \left\{ (y, p) \in H_0^1(\Omega) \times H_0^1(\Omega) : -\Delta y, -\Delta p \in L^2(\Omega), |(y, p)|_{\mathcal{A}^s} < \infty \right\},$$

where

$$(y, p)|_{\mathcal{A}^s} := \sup_{\varepsilon > 0} \varepsilon \inf_{\substack{\mathcal{T} \in \mathbb{T}: \\ v_{\mathcal{T}} \in V_{\mathcal{T}} \\ w_{\mathcal{T}} \in V_{\mathcal{T}}}} \inf_{\substack{(\|y - v_{\mathcal{T}}\|_{0,\Omega} + \text{osc}(v_{\mathcal{T}} - y_d, \mathcal{T})) \\ + \inf_{w_{\mathcal{T}} \in V_{\mathcal{T}}} (\|p - w_{\mathcal{T}}\|_{0,\Omega} + \text{osc}(f + P_{[a,b]}(w_{\mathcal{T}}), \mathcal{T})) \leq \varepsilon}} \left( \#\mathcal{T} - \#\mathcal{T}_0 \right)^s.$$

In our case of  $\Omega$  being convex and  $V_{\mathcal{T}}$  being linear,  $(H^2(\Omega) \cap H_0^1(\Omega))^2 \subset \mathcal{A}^s$  with  $s = 1$  if  $d = 2$  and  $s = \frac{2}{3}$  if  $d = 3$ . However, the class  $\mathcal{A}^s$  is much larger than  $(H^2(\Omega) \cap H_0^1(\Omega))^2$  which makes the  $L^2$ -norm based AFEM attractive, since the rate  $s$  is already realized with uniform refinements if  $y, p \in H^2(\Omega) \cap H_0^1(\Omega)$ . We refer to [10, Sec. 7] for more details.

To prove the optimality of AFEM we shall give the complexity of refinement, we refer to [5, Lemma 2.3] and [34] for related results. The following lemma shows that the difference between the cardinalities of the output and initial partitions can be bounded from above by some multiple of the total number of marked elements.

**Lemma 4.1.** [10, Theorem 4] *Let  $K$  be the total number of calls of bisection algorithm in [10, P. 210], so that  $K$  is no larger than the sum of the cardinalities of all sets of marked simplicies. Then for the output partition  $\mathcal{T}$ , it holds that  $\#\mathcal{T} - \#\mathcal{T}_0 \lesssim K$ .*

To bound the number of marked elements in Dörfler's marking we give the following lemma.

**Lemma 4.2.** *Let  $\theta < \frac{1}{C_2(1+C_4)}$ . Let  $\mu$  be sufficiently small such that Lemma 4.1 is valid. If*

$$(4.1) \quad \begin{aligned}& \|u - u_{\mathcal{T}}\|_{0,\Omega} + \|y - y_{\mathcal{T}}\|_{0,\Omega} + \|p - p_{\mathcal{T}}\|_{0,\Omega} + \text{osc}_{\mathcal{T}} \\ & \leq (1 - \theta C_2(1 + C_4))(\|u - u_k\|_{0,\Omega} + \|y - y_k\|_{0,\Omega} + \|p - p_k\|_{0,\Omega} + \text{osc}_k)\end{aligned}$$



holds for  $\mathcal{T}_k \in \mathbb{T}_\mu$  and  $\mathcal{T}_k \subset \mathcal{T} \in \mathbb{T}$ , then we have

$$(4.2) \quad \eta_k(\mathcal{R}_{\mathcal{T}_k \rightarrow \mathcal{T}}) \geq \theta \eta_k(\mathcal{T}_k).$$

*Proof.* From the triangle inequality we have

$$\begin{aligned} \|u - u_k\|_{0,\Omega} + \|y - y_k\|_{0,\Omega} + \|p - p_k\|_{0,\Omega} &\leq \|u_{\mathcal{T}} - u_k\|_{0,\Omega} + \|y_{\mathcal{T}} - y_k\|_{0,\Omega} + \|p_{\mathcal{T}} - p_k\|_{0,\Omega} \\ &\quad + \|u - u_{\mathcal{T}}\|_{0,\Omega} + \|y - y_{\mathcal{T}}\|_{0,\Omega} + \|p - p_{\mathcal{T}}\|_{0,\Omega}, \\ \text{osc}_k &\leq \text{osc}_k(\mathcal{R}_{\mathcal{T}_k \rightarrow \mathcal{T}}) + \text{osc}_{\mathcal{T}}. \end{aligned}$$

It follows from Theorem 2.2, Lemma 4.1 and the assumption of the lemma that

$$\begin{aligned} \theta(C_4 + 1)\eta_k(\mathcal{T}_k) &\leq \theta C_2(1 + C_4)(\|u - u_k\|_{0,\Omega} + \|y - y_k\|_{0,\Omega} + \|p - p_k\|_{0,\Omega} + \text{osc}_k) \\ &\leq \|u - u_k\|_{0,\Omega} + \|y - y_k\|_{0,\Omega} + \|p - p_k\|_{0,\Omega} + \text{osc}_k \\ &\quad - \|u - u_{\mathcal{T}}\|_{0,\Omega} - \|y - y_{\mathcal{T}}\|_{0,\Omega} - \|p - p_{\mathcal{T}}\|_{0,\Omega} - \text{osc}_{\mathcal{T}} \\ &\leq \|u_{\mathcal{T}} - u_k\|_{0,\Omega} + \|y_{\mathcal{T}} - y_k\|_{0,\Omega} + \|p_{\mathcal{T}} - p_k\|_{0,\Omega} + \text{osc}_k(\mathcal{R}_{\mathcal{T}_k \rightarrow \mathcal{T}}) \\ &\leq (1 + C_4)\eta_k(\mathcal{R}_{\mathcal{T}_k \rightarrow \mathcal{T}}), \end{aligned}$$

this completes the proof.  $\square$

**Lemma 4.3.** *For any  $s > 0$ , let  $(y, p) \in \mathcal{A}^s$ . Assume that  $\mu$  is sufficiently small such that Theorem 3.9 is valid. Then under the assumptions of Lemma 4.2, the number of marked elements  $\mathcal{M}_k \subset \mathcal{T}_k$  defined in Algorithm 2.3 satisfies*

$$(4.3) \quad \#\mathcal{M}_k \lesssim |(y, p)|_{\mathcal{A}^s}^{\frac{1}{s}} (\|u - u_k\|_{0,\Omega} + \|y - y_k\|_{0,\Omega} + \|p - p_k\|_{0,\Omega} + \text{osc}_k)^{-\frac{1}{s}}.$$

*Proof.* Let

$$(4.4) \quad \varepsilon = \frac{1 - \theta C_2(1 + C_4)}{C_5} (\|u - u_k\|_{0,\Omega} + \|y - y_k\|_{0,\Omega} + \|p - p_k\|_{0,\Omega} + \text{osc}_k),$$

where  $C_5$  is defined in Remark 3.10. Let  $\mathcal{T}' \in \mathbb{T}$  and  $y_{\mathcal{T}'}, p_{\mathcal{T}'} \in V_{\mathcal{T}'}$  such that

$$(4.5) \quad \|y - y_{\mathcal{T}'}\|_{0,\Omega} + \text{osc}(y_{\mathcal{T}'} - y_d, \mathcal{T}') + \|p - p_{\mathcal{T}'}\|_{0,\Omega} + \text{osc}(f + P_{[a,b]}(p_{\mathcal{T}'}), \mathcal{T}') \leq \varepsilon.$$

We can conclude from the definition of  $\mathcal{A}^s$  that

$$(4.6) \quad \#\mathcal{T}' - \#\mathcal{T}_0 \lesssim |(y, p)|_{\mathcal{A}^s}^{\frac{1}{s}} \varepsilon^{-\frac{1}{s}}.$$

It can be shown from Appendix A in [10] that  $\mathcal{T}'$  can be refined to a partition  $\mathcal{T}'' \in \mathbb{T}_\mu$  with  $\#\mathcal{T}'' - \#\mathcal{T}_0 \lesssim \#\mathcal{T}' - \#\mathcal{T}_0$  depending on  $\mu$ . Let  $\mathcal{T} := \mathcal{T}'' \oplus \mathcal{T}_k$  be the smallest common refinement of  $\mathcal{T}''$  and  $\mathcal{T}_k$  in  $\mathbb{T}_\mu$ , there holds  $\#\mathcal{T} - \#\mathcal{T}_k \leq \#\mathcal{T}'' - \#\mathcal{T}_0$  (see [33]). In view of  $V_{\mathcal{T}'} \subset V_{\mathcal{T}}$ , the monotonicity of data oscillation ([5, Remark 2.1]) and Remark 3.10 we have

$$\begin{aligned} &\|u - u_{\mathcal{T}}\|_{0,\Omega} + \|y - y_{\mathcal{T}}\|_{0,\Omega} + \|p - p_{\mathcal{T}}\|_{0,\Omega} + \text{osc}_{\mathcal{T}} \\ &\leq C_5 \left( \inf_{v_{\mathcal{T}'} \in V_{\mathcal{T}'}} (\|y - v_{\mathcal{T}'}\|_{0,\Omega} + \text{osc}(v_{\mathcal{T}'} - y_d, \mathcal{T}')) \right. \\ &\quad \left. + \inf_{w_{\mathcal{T}'} \in V_{\mathcal{T}'}} (\|p - w_{\mathcal{T}'}\|_{0,\Omega} + \text{osc}(f + P_{[a,b]}(w_{\mathcal{T}'}), \mathcal{T}')) \right) \\ (4.7) \quad &\leq (1 - \theta C_2(1 + C_4)) (\|u - u_k\|_{0,\Omega} + \|y - y_k\|_{0,\Omega} + \|p - p_k\|_{0,\Omega} + \text{osc}_k), \end{aligned}$$

so we can conclude from Lemma 4.2 that  $\eta_k(\mathcal{R}_{\mathcal{T}_k \rightarrow \mathcal{T}}) \geq \theta \eta_k(\mathcal{T}_k)$ . Note that Algorithm 2.3 selects a minimal set  $\mathcal{M}_k$  of  $\mathcal{T}_k$  with  $\eta_k(\mathcal{M}_k) \geq \theta \eta_k(\mathcal{T}_k)$ . Thus,

$$(4.8) \quad \#\mathcal{M}_k \leq \#\mathcal{R}_{\mathcal{T}_k \rightarrow \mathcal{T}} \leq \#\mathcal{T} - \#\mathcal{T}_k \leq \#\mathcal{T}'' - \#\mathcal{T}_0 \lesssim \#\mathcal{T}' - \#\mathcal{T}_0.$$

Now (4.3) follows from (4.6).  $\square$

Now we are in the position to present our final result on the quasi-optimality of AFEM for solving optimal control problems.

**Theorem 4.4.** *Let  $(u, y, p) \in U_{ad} \times H_0^1(\Omega) \times H_0^1(\Omega)$  be the solution of optimal control problem (2.1)-(2.2) and  $(u_n, y_n, p_n) \in U_{ad} \times V_n \times V_n$  be the solution of the discrete problem (2.9)-(2.11) generated by the adaptive Algorithm 2.4. For some  $s > 0$ , let  $(y, p) \in \mathcal{A}^s$ . Under the assumptions of Lemma 4.2 there holds that*

$$(4.9) \quad \|u - u_n\|_{0,\Omega} + \|y - y_n\|_{0,\Omega} + \|p - p_n\|_{0,\Omega} + \text{osc}_n \lesssim |(y, p)|_{\mathcal{A}^s} (\#\mathcal{T}_n - \#\mathcal{T}_0)^{-s}$$

provided  $h_0 \ll 1$ .

*Proof.* It follows from Lemmas 4.1 and 4.3 that

$$(4.10) \quad \begin{aligned} \#\mathcal{T}_n - \#\mathcal{T}_0 &\lesssim \sum_{k=0}^{n-1} \#\mathcal{M}_k \\ &\lesssim |(y, p)|_{\mathcal{A}^s}^{\frac{1}{s}} \sum_{k=0}^{n-1} (\|u - u_k\|_{0,\Omega} + \|y - y_k\|_{0,\Omega} + \|p - p_k\|_{0,\Omega} + \text{osc}_k)^{-\frac{1}{s}}. \end{aligned}$$

Due to Theorem 3.13 we obtain from  $0 \leq k < n$  that

$$\begin{aligned} &\|u - u_n\|_{0,\Omega} + \|y - y_n\|_{0,\Omega} + \|p - p_n\|_{0,\Omega} + \text{osc}_n \\ &\lesssim \nu^{n-k} \left( \|u - u_k\|_{0,\Omega} + \|y - y_k\|_{0,\Omega} + \|p - p_k\|_{0,\Omega} + \text{osc}_k \right). \end{aligned}$$

Thus,

$$(4.11) \quad \begin{aligned} \#\mathcal{T}_n - \#\mathcal{T}_0 &\lesssim |(y, p)|_{\mathcal{A}^s}^{\frac{1}{s}} (\|u - u_n\|_{0,\Omega} + \|y - y_n\|_{0,\Omega} + \|p - p_n\|_{0,\Omega} + \text{osc}_n)^{-\frac{1}{s}} \sum_{k=0}^{n-1} \nu^{\frac{n-k}{s}} \\ &\lesssim |(y, p)|_{\mathcal{A}^s}^{\frac{1}{s}} (\|u - u_n\|_{0,\Omega} + \|y - y_n\|_{0,\Omega} + \|p - p_n\|_{0,\Omega} + \text{osc}_n)^{-\frac{1}{s}}, \end{aligned}$$

where the last inequality holds due to  $\nu < 1$ . This completes the proof of the theorem.  $\square$

## 5. NUMERICAL EXPERIMENTS

In this section we carry out some numerical tests in two dimensions to support our theoretical results obtained in this paper. As indicated in [10], the additional refinement of elements to ensure the mesh grading property (2.8) seems to be not necessary in practical computations to deliver optimal convergence of  $L^2$ -norm based AFEM. So in current paper we use the practical bisection algorithm without additional refinement, similar phenomenon can be observed for the optimal convergence.

**Example 5.1.** *We consider an example defined on  $\Omega = (0, 1)^2$ . We set  $\alpha = 0.1$  and choose the exact solutions as*

$$\begin{aligned} y &= \arctan(-50x_1 + 100x_2 - 25), \\ p &= 16x_1(1 - x_1)x_2(1 - x_2)(1 + \arctan(200(\frac{1}{16} - (x_1 - 0.5)^2 - (x_2 - 0.5)^2))), \\ u &= \max\{-5, \min\{-1, -\frac{p}{\alpha}\}\}, \end{aligned}$$

*the corresponding  $f$  and  $y_d$  can be obtained after simple calculation. Note that we impose inhomogeneous Dirichlet boundary condition for the state  $y$  and our theoretical analysis applies to this case after some simple adaptations.*

In Figure 1 we plot the profiles of the state and adjoint state variables on adaptively refined mesh with  $\theta = 0.3$  and 20 adaptive loops. Although the solutions are smooth, larger gradients can be observed in certain areas so the adaptive finite element method may deliver much smaller errors compared to the uniform refinement. In Figure 2 and 3 we show the adaptive meshes after 15, 20 and 25 adaptive loops with Dörfler's marking parameter  $\theta = 0.3$ . We can see that the meshes concentrate on the areas where the solutions have larger gradients. Moreover, in Figure 3 we plot the active sets of both the continuous control, the discrete controls by using variational control

discretization and piecewise linear and continuous finite element approximations. We can observe that the active set of variational discretized control is more close to the continuous one compared with full control discretization, this shows the superiority of variational control discretization. In Figure 4 we give the comparisons of convergence history of the  $L^2$ -norm errors of the optimal control, the state and adjoint state and the error estimators on uniformly refined meshes ( $\theta=1$ ) and adaptively refined meshes with  $\theta = 0.3$  and  $\theta = 0.4$ , respectively. Although optimal convergence of second order in  $L^2$ -norm can be observed for both the uniform refinement and adaptive refinement, we have smaller errors for adaptive algorithm which shows the power of AFEM for problems even with  $H^2$ -regularity.

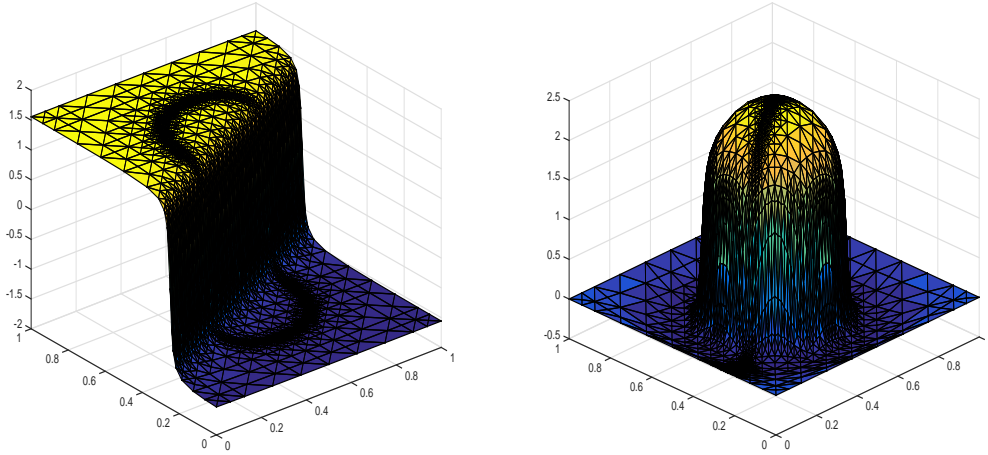


FIGURE 1. The profiles of the numerical state (left) and adjoint state (right) on adaptively refined mesh with  $\theta = 0.3$  and 20 adaptive loops for Example 5.1 generated by Algorithm 2.4.

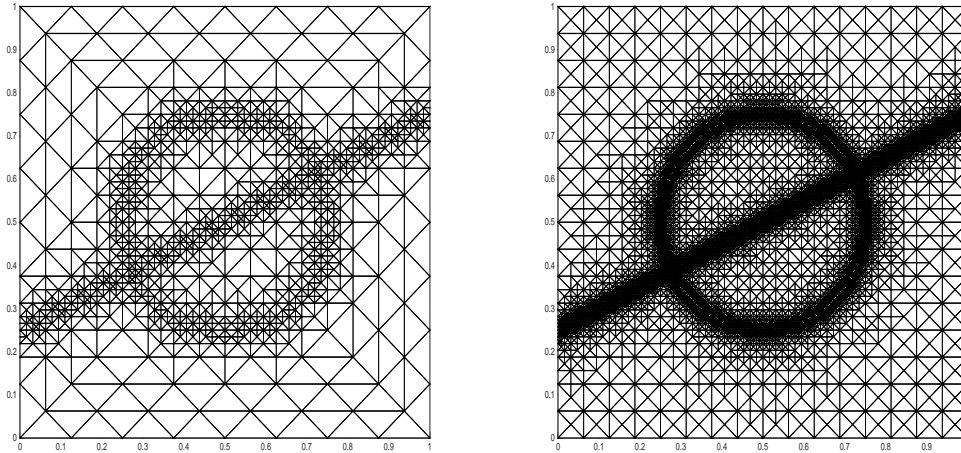


FIGURE 2. The adaptive meshes after 15 steps (left) and 25 steps for Example 5.1 generated by Algorithm 2.4 with  $\theta = 0.3$ .

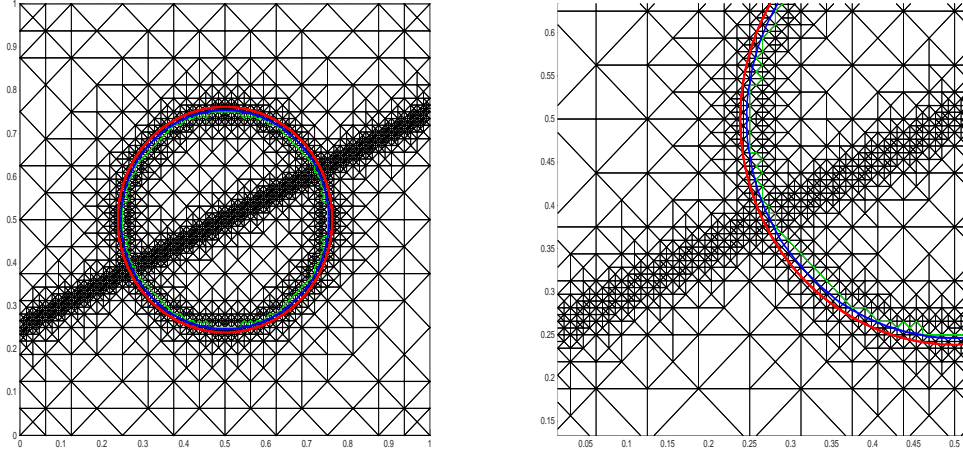


FIGURE 3. The adaptive meshes after 20 steps (left) and zoom in near the left below corner for Example 5.1 generated by Algorithm 2.4 with  $\theta = 0.3$ . The red line depicts the boarder of the active set of the continuous solution, the blue line depicts the boarder of the active set when using variational discretization, and the green line depicts the boarder of the active set obtained by using piecewise linear, continuous controls.

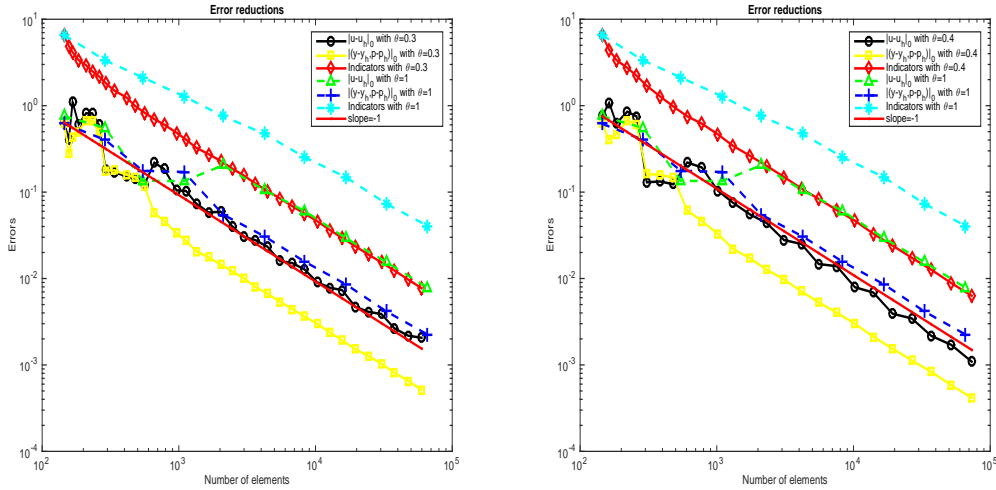


FIGURE 4. The comparisons of convergence history of the optimal control, state and adjoint state and the error estimators on uniformly refined meshes ( $\theta=1$ ) and adaptively refined meshes with  $\theta = 0.3$  (left) and  $\theta = 0.4$  (right), respectively, for Example 5.1 generated by Algorithm 2.4.

**Example 5.2.** In the second example we consider an optimal control problem without explicit solutions. We set  $\Omega = (0, 1)^2$ ,  $\alpha = 10^{-2}$ ,  $a = 10$  and  $b = 15$ . We choose the singular  $f$  and desired state  $y_d$  as

$$f = \sqrt{(x_1 - 1)^2 + (x_2 - 1)^2}^{-1.5}, \quad y_d = \sqrt{x_1^2 + x_2^2}^{-1.9}.$$

Note that  $f$  and  $y_d$  are not in  $L^2(\Omega)$  as we assumed in the paper, the theory we derived above does not apply in this case. However, since the singularity is only located in two points, only some simple modifications need be made in the computations. We intend to use this example to explain that even in convex domain the solution of elliptic equation may have singularity caused by singular data so the adaptive FEM may also find applications.

In Figure 5 we plot the profiles of the state and adjoint state variables on adaptively refined mesh with  $\theta = 0.3$  and 20 adaptive loops. Since  $f$  and  $y_d$  have singularities near the points  $(0, 0)$  and  $(1, 1)$ , respectively, we can observe the corresponding singularities for the state and adjoint state. In Figure 6 and 7 we show the adaptive meshes after 15, 20 and 25 adaptive loops with Dörfler's marking parameter  $\theta = 0.3$ . We can see that the meshes concentrate on the points  $(0, 0)$  and  $(1, 1)$  where the singularities of the solutions are located. Moreover, in Figure 7 we plot the active sets of the discrete controls by using variational control discretization and piecewise linear and continuous finite element approximations. We can observe that the active set of variational discretized control crosses the elements and can give better results, as indicated in Example 5.1. Since we do not have explicit solutions, in Figure 8 we only show the comparisons of convergence history of the error estimators on uniformly refined meshes ( $\theta=1$ ) and adaptively refined meshes with  $\theta = 0.3$  and  $\theta = 0.4$ , respectively. We can observe the optimal second order convergence for the reduction of the error estimators for the adaptive refinement which reflects the optimal convergence of the  $L^2$ -norm of the control, the state and adjoint state because the a posteriori error estimate is reliable and efficient. Note that only reduced orders are observed for the uniform refinement, which is due to the singularity of the solutions caused by singular data.

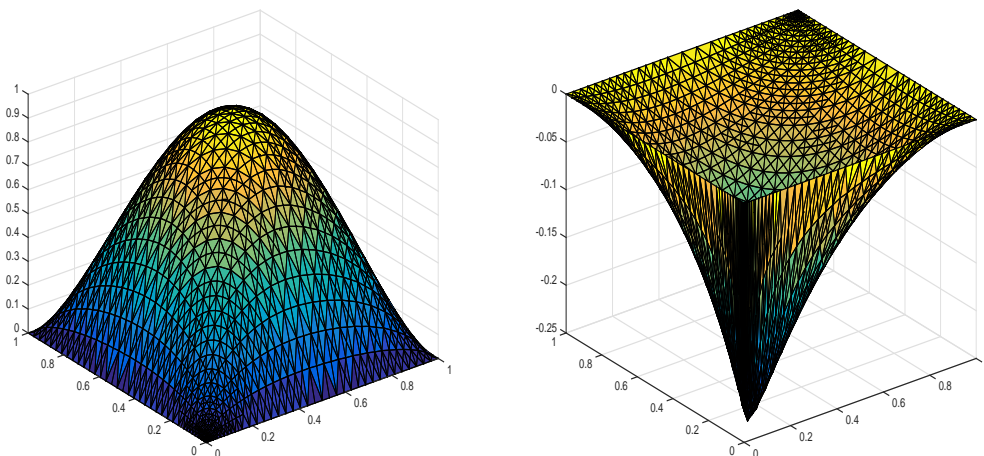


FIGURE 5. The profiles of the numerical state (left) and adjoint state (right) on adaptively refined mesh with  $\theta = 0.3$  and 20 adaptive loops for Example 5.2 generated by Algorithm 2.4.

#### ACKNOWLEDGEMENTS

The first author was supported by the National Basic Research Program of China under grant 2012CB821204 and the National Natural Science Foundation of China under grants 11671391 and 91530204. The second author acknowledged the support of the National Natural Science Foundation of China under grants 11571356 and 91530204. The third author acknowledged the support of the National Natural Science Foundation of China under grant 11301311.

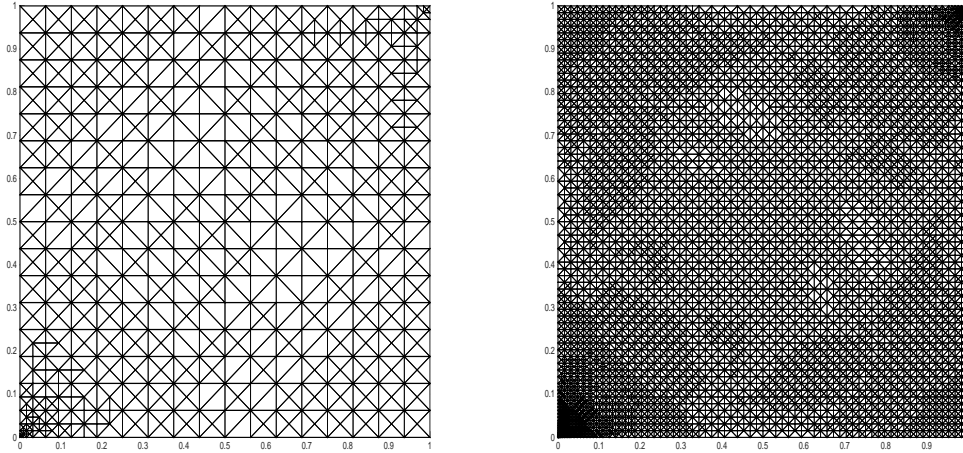


FIGURE 6. The adaptive meshes after 15 steps (left) and 25 steps for Example 5.2 generated by Algorithm 2.4 with  $\theta = 0.3$ .

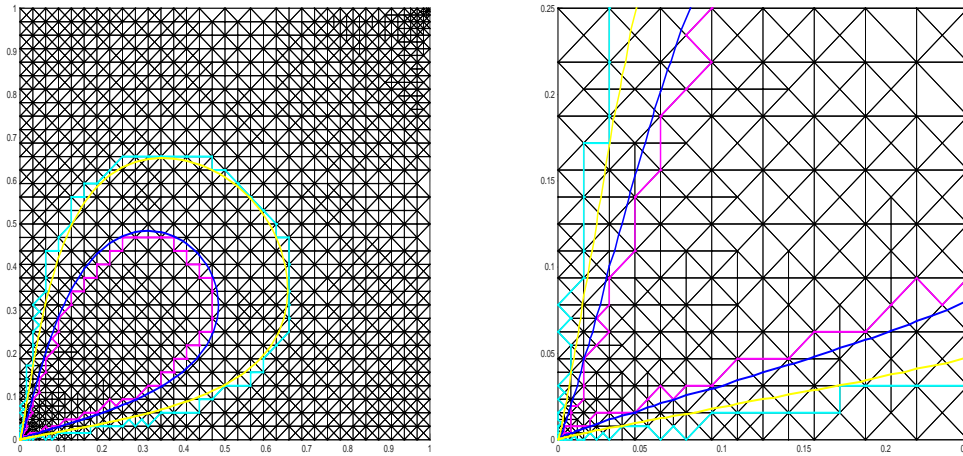


FIGURE 7. The adaptive meshes after 20 steps (left) and zoom in near the origin for Example 5.2 generated by Algorithm 2.4 with  $\theta = 0.3$ . The blue (upper bound) and yellow (lower bound) lines depict the boarder of the active set of the discrete solution when using variational discretization, and the pink (upper bound) and cyan (lower bound) lines depict the boarder of the active set obtained by using piecewise linear, continuous controls.

#### REFERENCES

- [1] I. Babuška and W.C. Rheinboldt, Error estimates for adaptive finite element computations, *SIAM J. Numer. Anal.*, 15 (1978), pp. 736-754.
- [2] 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), pp. 113-132.
- [3] R. Becker and S.P. Mao, Quasi-optimality of an adaptive finite element method for an optimal control problem, *Comput. Methods Appl. Math.*, 11 (2011), pp. 107-128.

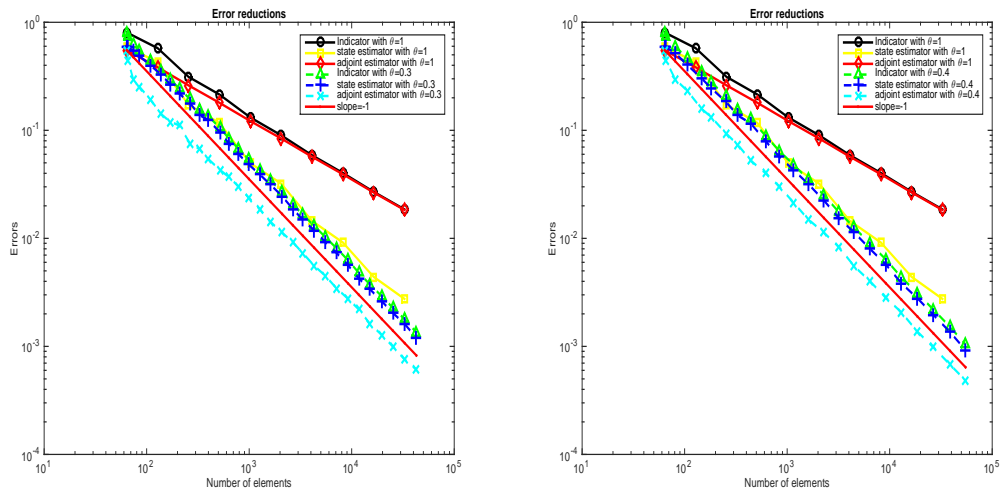


FIGURE 8. The comparisons of convergence history of the optimal control, state and adjoint state and the error estimators on uniformly refined meshes ( $\theta=1$ ) and adaptively refined meshes with  $\theta = 0.3$  (left) and  $\theta = 0.4$  (right), respectively, for Example 5.2 generated by Algorithm 2.4.

- [4] P. Binev, W. Dahmen and R. DeVore, Adaptive finite element methods with convergence rates, *Numer. Math.*, 97 (2004), pp. 219-268.
- [5] 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.
- [6] Y.P. Chen and H.T. Leng, Convergence and quasi-optimality of an adaptive finite element method for a constrained optimal control problem, submitted, 2016.
- [7] P.G. Ciarlet, *The Finite Element Methods for Elliptic Problems*, North-Holland, Amsterdam, 1978.
- [8] A. Demlow, Convergence of an adaptive finite element method for controlling local energy errors, *SIAM J. Numer. Anal.*, 48 (2010), no. 2, pp. 470-497.
- [9] A. Demlow, Quasi-optimality of adaptive finite element methods for controlling local energy errors, *Numer. Math.*, DOI 10.1007/s002111-015-0774-x, 2015.
- [10] A. Demlow and R. Stevenson, Convergence and quasi-optimality of an adaptive finite element method for controlling  $L^2$  errors, *Numer. Math.*, 117 (2011), no. 2, pp. 185-218.
- [11] W. Dörfler, A convergent adaptive algorithm for Poisson's equation, *SIAM J. Numer. Anal.*, 33 (1996), pp. 1106-1124.
- [12] 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.
- [13] W. Gong and N.N. Yan, Adaptive finite element method for elliptic optimal control problems: convergence and optimality, arXiv:1505.01632, *Numer. Math.*, DOI 10.1007/s00211-016-0827-9, 2016.
- [14] W. Gong, H.H. Xie and N.N. Yan, Adaptive multilevel correction method for finite element approximations of elliptic optimal control problems, submitted, 2015, <http://lsec.cc.ac.cn/~wgong/publication.html>.
- [15] M. Hintermüller and R.H.W. Hoppe, Goal-oriented adaptivity in control constrained optimal control of partial differential equations, *SIAM J. Control Optim.*, 47 (2008), no. 4, pp. 1721-1743.
- [16] 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.
- [17] M. Hinze, A variational discretization concept in control constrained optimization: The linear-quadratic case, *Comput. Optim. Appl.*, 30 (2005), pp. 45-63.
- [18] M. Hinze, R. Pinnau, M. Ulbrich and S. Ulbrich, *Optimization with PDE Constraints*, *Math. Model. Theo. Appl.*, 23, Springer, New York, 2009.
- [19] 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), pp. 1832-1861.

- [20] K. Kohls, A. Rösch and K.G. Siebert, Convergence of adaptive finite elements for optimal control problems with control constraints, G. Leugering et al. (eds.), Trends in PDE Constrained Optimization, vol. 165 of Internat. Ser. Numer. Math., 2015, pp. 403-419.
- [21] R. Li, W.B. Liu, H.P. 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.
- [22] J.L. Lions, Optimal Control of Systems Governed by Partial Differential Equations, Springer-Verlag, Berlin, 1971.
- [23] W.B. Liu and N.N. Yan, A posteriori error analysis for convex distributed optimal control problems, Adv. Comp. Math., 15 (2001), no. 1-4, pp. 285-309.
- [24] W.B. Liu and N.N. Yan, A posteriori error estimates for convex boundary control problems, SIAM J. Numer. Anal., 39 (2001), no. 1, pp. 73-99.
- [25] W.B. Liu and N.N. Yan, A posteriori error estimates for optimal problems governed by Stokes equations, SIAM J. Numer. Anal., 40 (2003), pp. 1850-1869.
- [26] W.B. Liu and N.N. Yan, A posteriori error estimates for optimal control problems governed by parabolic equations, Numer. Math., 93 (2003), pp. 497-521.
- [27] W.B. Liu and N.N. Yan, Adaptive Finite Element Methods for Optimal Control Governed by PDEs, Science press, Beijing, 2008.
- [28] K. Mekchay and R.H. Nochetto, Convergence of adaptive finite element methods for general second order linear elliptic PDEs, SIAM J. Numer. Anal., 43 (2005), pp. 1803-1827.
- [29] 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.
- [30] P. Morin, R.H. Nochetto, and K.G. Siebert, Convergence of adaptive finite element methods, SIAM Rev., 44 (2002), pp. 631-658.
- [31] R. Schneider and G. Wachsmuth, A posteriori error estimation for control-constrained, linear-quadratic optimal control problems, SIAM J. Numer. Anal., 54 (2016), no. 2, pp. 1169-1192.
- [32] P. Morin, K. Siebert and A. Veiser, Convergence of finite elements adapted for weaker norms, Applied and Industrial Mathematics in Italy II. Ser. Adv. Math. Appl. Sci., vol. 75, pp. 468-479, World Sci. Publ., Hackensack, NJ, 2007.
- [33] R. Stevenson, Optimality of a standard adaptive finite element method, Found. Comput. Math., 7 (2007), pp. 245-269.
- [34] R. Stevenson, The completion of locally refined simplicial partitions created by bisection, Math. Comput., 77 (2008), pp. 227-241.
- [35] R. Verfürth, A Review of a Posteriori Error Estimates and Adaptive Mesh Refinement Techniques, Wiley-Teubner, New York, 1996.