Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 7(11): 19664–19695.
DOI:10.3934/math.20221079
Received: 26 February 2022
Revised: 04 August 2022
Accepted: 17 August 2022
Published: 06 September 2022

Research article

Convergence and proposed optimality of adaptive finite element methods for nonlinear optimal control problems

Zuliang Lu[1,2,*], **Fei Cai**[1], **Ruixiang Xu**[1] **and Lu Xing**[1]

[1] Key Laboratory for Nonlinear Science and System Structure, Chongqing Three Gorges University, Chongqing 404000, China

[2] Research Center for Mathematics and Economics, Tianjin University of Finance and Economics, Tianjin 300222, China

* **Correspondence:** Email: zulianglu@sanxiau.edu.cn.

Abstract: This paper investigates the adaptive finite element method for nonlinear optimal control problem, and the research content of reference ([21] H. Leng and Y. Chen, 2017) is extended accordingly. Linear discretisation of the equation of state and the equation of common state is performed using continuous segmentation functions. At the same time, we use the bubble function technique to prove that the posterior error estimates are obtained from the upper and lower bounds. What is more, for the adaptive finite element method, we also consider convergence and quasi-optimality, where we find that the demand $h_0 \ll 1$ on the initial grid is unconstrained for the convergence analysis of the proposed adaptive algorithm for the nonlinear optimal control problem. Simultaneously, some numerical simulation is used to verify our theoretical analysis.

Keywords: nonlinear optimal control problem; adaptive finite element method; convergence; quasi-optimality
Mathematics Subject Classification: 49J20, 65N30

1. Introduction

With the development of computer technology, the finite element method follows closely in the early 1960s. The research on the reliability and validity of finite element analysis promoted the development of the finite element method [3–5, 25, 27]. Applying finite element methods, the emergence of errors has captured the attention of scholars. One of the main sources of errors is the error caused by the discretisation of the model, while researchers dissect various aspects of finite element analysis. In addition, the generation of finite element mesh is also a concern for scholars. In the early conventional FEA, scholars usually used experience, intuition or even guesses to generate meshes to simple judge whether the approximation results are reasonable or not. If it is not reasonable, the grid needs to be redesigned whenever necessary for the efficiency of the analysis and the reliability of the results. Therefore, the emergence of the adaptive finite element method is because the computer, after checking the current conditions, decides whether the solution is accurate enough to meet the determination needs according to the error information obtained during the adjustment process.

To the best of our knowledge, the adaptive finite element method is a numerical method that can automatically adjust the algorithm to improve the process of solving [2]. An appropriate mesh can greatly reduce the errors arising from the discretisation of the finite element approximation process during the replication. According to the current situation, solutions of optimal control problems for nonlinear systems are usually not available. Besides the complexity and diversity of nonlinear equations, it is also very practical to use the adaptive finite element method for solving nonlinear equations.

Adaptive finite element methods have been widely and successfully applied in various linear optimal control problems, for example, Eriksson and Johnson proposed that the adaptive finite element algorithm produces a series of successively refined meshes in which the final mesh satisfies a given error tolerance [11]. Gaevaskaya and Hope et al. proposed an adaptive finite element method for a class of distributed optimal control problems with control constraints is analysed by applying the reliability and discrete local efficiency of the posterior estimator and the quasi-orthogonality property as basic tools [13]. Braess and Carstensen et al. Investigate the residual jump contributions of a standard explicit residual-based a posteriori error estimator for an adaptive finite element method [1]. Hu and Xu et al. performed research on the convergence and optimality of the adaptive nonconforming linear element method for the stokes problem [17]. However, a large number of literature show that the researches on adaptive finite element method for nonlinear optimal control problems have not reached its peak yet.

Recently, for instance, Wriggers and Scherf propose an adaptive finite element technique for nonlinear contact problems [29]. Eriksson and Johnson consider adaptive finite element methods for parabolic problems to a class of nonlinear scalar problems, the authors obtain posteriori error estimates and design corresponding adaptive algorithms [12]. Nowadays, the problem of nonlinear optimal control problems, similar to big data, is the focus of scholars worldwide. Hence it is worthy of investigating the adaptive finite element method for such nonlinear problems.

Many scholars have also studied the prior error estimates of the bilinear optimal control problem. For example, Yang, Demlow, and Dobrowolski et al. investigated the prior error estimates and superconvergence of optimal control problems for bilinear models and give the optimal L^2-norm error estimates and the almost optimal L^∞-norm estimates about the state and co-state variables [7, 8, 31]. Lu investigated a second-order parabolic bilinear optimal control problems and provided a priori error estimates for the finite element solutions of the state equations describing the system [24]. Shen and Yang et al. investigated a quadratic optimal control problem governed by a linear hyperbolic integro differential equation and its finite element approximation, a priori estimates have been carried out using the standard functional analysis techniques, and the existence and regularity of the solution are provided by using these estimates. At the same time, some scholars have analysed the posteriori error estimates of the finite element method for bilinear optimal control problems [15, 26]. Lu, Chen, and Leng et al. discussed the discretisation of Raviart-Thomas mixed finite element for general bilinear optimal control problems, a posteriori error estimates are derived for both the coupled state and the control solutions [18, 23]. Although bilinear optimal control problems are frequently met in applications, they are much more difficult to handle than linear or nonlinear cases. There is little work on nonlinear optimal control problems.

In this paper, we focus on nonlinear optimal control problem with integral control constraints where we deal with the control via adopting piecewise constant discetization while applying continuous piecewise linear discretization for the state and the co-state, respectively. Then a posteriori error estimates is gained. For the convergence and the quasi-optimality, we prove them relying on quasi-orthogonality and discrete local upper bound. Based on the mild assumption to the initial grids, we obtain the proof of convergence and quasi-optimality by means of the solution operator of nonlinear elliptic equations. Finally, some numerical simulations are provided support for our theoretical analysis.

Here are some notations will be used in this paper. Let Ω be a bounded Lipschitz domain in $\mathbb{R}^2$ and $\partial\Omega$ denote the boundary of Ω. We use the standard notation $W^{m,q}(\omega)$ with norm $\|\cdot\|_{m,q,\omega}$ and seminorm $|\cdot|_{m,q,\omega}$ to express the standard Sobolev space for $\omega \subset \Omega$. Moreover, we will omit the subscription if $\omega = \Omega$. For $q = 2$, we denote $W^{m,2}(\Omega)$ by $H^m(\Omega)$ and $\|\cdot\|_m = \|\cdot\|_{m,2}$. Also for $m = 0$ and $q = 2$, we denote $W^{0,2}(\omega) = L^2(\omega)$ and $\|\cdot\|_{0,2,\omega} = \|\cdot\|_{0,\omega}$. Additionally, we observe that $H_0^1(\Omega) = \{v \in H^1(\Omega) : v = 0 \text{ on } \partial\Omega\}$. Let $\mathcal{T}_{h_0}$ be the initial partition of $\bar{\Omega}$ into disjoint triangles. By newest-vertex bisections for $\mathcal{T}_{h_0}$, we can obtain a class $\mathbb{T}$ of conforming partitions. For $\mathcal{T}_h, \tilde{\mathcal{T}}_h \in \mathbb{T}$, we use $\mathcal{T}_h \subset \tilde{\mathcal{T}}_h$ to indicate that $\tilde{\mathcal{T}}_h$ is a refinement of $\mathcal{T}_h$ and $h_T = |T|^{1/2}$, T is the partition diameter. In addition $(\cdot,\cdot)$ denotes the L^2 inner product. Beyond that, let C be a constant which independent of grids size, then we use $A \approx B$ to represent $cA \leq B \leq CA$.

The rest of the paper is organized as follows. In Section 2, we introduce the optimal control problem of our interest and obtain a posteriori error estimates. The relevant algorithms are introduced in Section 3. In Section 4, we use quasi-orthogonality and discrete local upper bounds to prove the convergence of the adaptive finite element method, as well as quasi-optimality in Section 5. We provide an adaptive finite element algorithm and some numerical simulations to verify our theoretical analysis in Section 6. Finally, we summarize the results of this paper and develop a plan for future work.

2. A posteriori error estimates for finite element method

In this paper we mainly enter into meaningful discussions with the following nonlinear optimal control problem governed by nonlinear elliptic equations:

$$\min_{u\in U_{ad}}\left\{\frac{1}{2}\|y-y_d\|_0^2+\frac{\alpha}{2}\|u\|_0^2\right\},$$
$$-\Delta y+\phi(y)=f+u,\quad \text{in } \Omega,$$
$$y=0,\quad \text{on } \partial\Omega,$$

where y is the state variable, u is the control variable, f is a function of the control variable, α is a constant greater than zero, $y_d \in L^2(\Omega)$, $U_{ad} = \{v : v \in L^2(\Omega), \int_\Omega v\, dx \geq 0\}$ is a closed convex subset of $U = L^2(\Omega)$ and $\phi(\cdot) \in W^{2,\infty}(-R, R)$ for any $R > 0$, $\phi'(y) \in L^2(\Omega)$ for any $y \in H^1(\Omega)$, $\phi' \geq 0$. Let $V = H_0^1(\Omega)$, we give the weak formulation to deal with state equation, namely, find $y \in V$ such that

$$a(y,v)+(\phi(y),v)=(f+u,v),\quad \forall\, v\in V,$$

where

$$a(y,v)=\int_\Omega \nabla y\cdot\nabla v\,dx.$$

Then the nonlinear optimal control problem can be restated as follows

$$\min_{u\in U_{ad}}\left\{\frac{1}{2}\|y-y_d\|_0^2+\frac{\alpha}{2}\|u\|_0^2\right\}, \tag{2.1}$$

$$a(y,v)+(\phi(y),v)=(f+u,v),\quad \forall\, v\in V. \tag{2.2}$$

It is well known [14, 20] that the nonlinear optimal control problem has at least one solution (y,u), and that if a pair (y,u) is the solution of the nonlinear optimal control problem, then there is a co-state $p\in V$ such that the triplet (y,p,u) satisfies the following optimality conditions:

$$a(y,v)+(\phi(y),v)=(f+u,v),\quad \forall\, v\in V, \tag{2.3}$$

$$a(q,p)+(\phi'(y)p,q)=(y-y_d,q),\quad \forall\, q\in V, \tag{2.4}$$

$$(\alpha u+p,v-u)\geq 0,\quad \forall\, v\in U_{ad}. \tag{2.5}$$

Due to coercivity of $a(\cdot,\cdot)$, we define a linear operator $S : L^2(\Omega)\to H_0^1(\Omega)$ such that $S(f+u)=y$ and let S^* be the adjoint of S such that $S^*(y-y_d)=p$ where V_h is the continuous piecewise linear finite element space with respect to the partition $\mathcal{T}_h\in\mathbb{T}$. For $\mathcal{T}_h\in\mathbb{T}$, we define U^h as the piecewise constant finite element space with respect to $\mathcal{T}_h$. Let $U_{ad}^h=\{v_h\in U^h : \int_\Omega v_h dx\geq 0\}$. Then we derive the standard finite element discretization for the nonlinear optimal control problem as follows:

$$\min_{u_h\in U_{ad}^h}\left\{\frac{1}{2}\|y_h-y_d\|_0^2+\frac{\alpha}{2}\|u_h\|_0^2\right\}, \tag{2.6}$$

$$a(y_h,v)+(\phi(y_h),v)=(f+u_h,v),\quad \forall\, v\in V_h. \tag{2.7}$$

Similarly the nonlinear optimal control problem (2.6)–(2.7) has a solution (y_h,u_h), and that if a pair $(y_h,u_h)\in V^h\times U^h$ is the solution of (2.6)–(2.7), then there is a co-state $p_h\in V_h$ such that the triplet (y_h,p_h,u_h) satisfies the following optimality conditions:

$$a(y_h,v)+(\phi(y_h),v)=(f+u_h,v),\quad \forall\, v\in V_h, \tag{2.8}$$

$$a(q,p_h)+(\phi'(y_h)p_h,q)=(y_h-y_d,q),\quad \forall\, q\in V_h, \tag{2.9}$$

$$(\alpha u_h+p_h,v_h-u_h)\geq 0,\quad \forall\, v_h\in U_{ad}^h. \tag{2.10}$$

Here we define some error indicators we largely and frequently use in this paper where $\eta(\cdot)$ are error indicators and $osc(\cdot)$ represent the data oscillations. For $\mathcal{T}_h\in\mathbb{T},\ T\in\mathcal{T}_h$, we define

$$\begin{aligned}
&\eta_{1,\mathcal{T}_h}^2(p_h,T)=h_T^2\|\nabla p_h\|_{0,T}^2,\\
&\eta_{2,\mathcal{T}_h}^2(u_h,y_h,T)=h_T^2\|f+u_h-\phi(y_h)\|_{0,T}^2+h_T\|[\nabla y_h]\cdot\mathbf{n}\|_{0,\partial T\backslash\partial\Omega}^2,\\
&\eta_{3,\mathcal{T}_h}^2(y_h,p_h,T)=h_T^2\|y_h-y_d-\phi'(y_h)p_h\|_{0,T}^2+h_T\|[\nabla p_h]\cdot\mathbf{n}\|_{0,\partial T\backslash\partial\Omega}^2,\\
&osc_{\mathcal{T}_h}^2(f,T)=h_T^2\|f-f_T\|_{0,T}^2,\\
&osc_{\mathcal{T}_h}^2(y_h-y_d,T)=h_T^2\|(y_h-y_d)-(y_h-y_d)_T\|_{0,T}^2,
\end{aligned}$$

where $u_h \in U_{ad}^h$, $y_h, p_h \in V_h$, and where f_T is L^2-projection of f onto piecewise constant space on T and $f_T = \frac{\int_T f}{|T|}$. For $\omega \subset \mathcal{T}_h$, we have

$$\eta^2_{1,\mathcal{T}_h}(p_h, \omega) = \sum_{T\in\omega} \eta^2_{1,\mathcal{T}_h}(p_h, T),$$
$$osc^2_{\mathcal{T}_h}(f, \omega) = \sum_{T\in\omega} osc^2_{\mathcal{T}_h}(f, T),$$

by which $\eta^2_{2,\mathcal{T}_h}(u_h, y_h, \omega)$, $\eta^2_{3,\mathcal{T}_h}(u_h, y_h, \omega)$ and $osc^2_{\mathcal{T}_h}(y_h - y_d, \omega)$ can be denoted similarly.

2.1. A posteriori error estimates

Also, a reliable and effective a posteriori error estimates for the nonlinear optimal control problem (2.1)–(2.2) which is presented in next Theorem.

Theorem 2.1. *For $\mathcal{T}_h \in \mathbb{T}$, let (y, p, u) be the exact solution of problem (2.3), (2.4) and (2.5) and (y_h, p_h, u_h) be the solution of problem (2.8), (2.9) and (2.10) with respect to $\mathcal{T}_h$. Then there exist constants c and C such that*

$$\begin{aligned} &\|u - u_h\|_0^2 + \|y - y_h\|_1^2 + \|p - p_h\|_1^2 \\ &\le C\left(\eta^2_{1,\mathcal{T}_h}(p_h, \mathcal{T}_h) + \eta^2_{2,\mathcal{T}_h}(u_h, y_h, \mathcal{T}_h) + \eta^2_{3,\mathcal{T}_h}(y_h, p_h, \mathcal{T}_h)\right), \end{aligned} \tag{2.11}$$

and

$$\begin{aligned} &c\left(\eta^2_{1,\mathcal{T}_h}(p_h, \mathcal{T}_h) + \eta^2_{2,\mathcal{T}_h}(u_h, y_h, \mathcal{T}_h) + \eta^2_{3,\mathcal{T}_h}(y_h, p_h, \mathcal{T}_h)\right) \\ &\le \|u - u_h\|_0^2 + \|y - y_h\|_1^2 + \|p - p_h\|_1^2 + osc^2_{\mathcal{T}_h}(f, \mathcal{T}_h) + osc^2_{\mathcal{T}_h}(y_h - y_d, \mathcal{T}_h). \end{aligned} \tag{2.12}$$

Proof. **Step 1.** We seek functions $y^h, p^h \in V$ satisfying the following auxiliary problems

$$a(y^h, v) + (\phi(y^h), v) = (f + u_h, v), \quad \forall\, v \in V, \tag{2.13}$$
$$a(q, p^h) + (\phi'(y^h)p^h, q) = (y_h - y_d, q), \quad \forall\, q \in V. \tag{2.14}$$

It follows that

$$\begin{aligned} \|u - u_h\|_0^2 &\le C(\alpha u, u - u_h) - C(\alpha u_h, u - u_h) \\ &\le -C(\alpha u_h, u - u_h) \\ &= C(\alpha u^h, u_h - u) + C(\alpha u^h - \alpha u_h, u - u_h). \end{aligned}$$

With the help of the proof of Lemma 3.4 in [14] and the lemma 3.4 in [21], we have

$$\begin{aligned} (\alpha u^h, u_h - u) &= (\alpha u_h + p_h, u_h - u) \\ &\le C\Big(\sum_{T\in\mathcal{T}_h} \|p_h - \pi_h p_h\|^2_{0,T} + \|u - u_h\|_0^2\Big) \\ &\le C(\sigma)\eta^2_{1,\mathcal{T}_h}(p_h, \mathcal{T}_h) + C\sigma\|u - u_h\|_0^2, \end{aligned} \tag{2.15}$$

where $\eta^2_{1,\mathcal{T}_h}(p_h,\mathcal{T}_h)=\sum\limits_{T\in\mathcal{T}_h}\eta^2_{1,\mathcal{T}_h}(p_h,T)$, and

$$\begin{aligned}(\alpha u^h-\alpha u_h,u-u_h)&=(p_h-p^h,u-u_h)\\&\le C(\sigma)\|p^h-p_h\|_0^2+C\sigma\|u-u_h\|_0^2,\end{aligned}\tag{2.16}$$

where π_h is the L^2-projection operator onto piecewise constant space on $\mathcal{T}_h$, $C(\sigma)$ is a universal constant, which depends on σ, σ is an arbitrary positive number, and C is a general universal constant, which can include $C(\sigma)$. Obviously, from (2.15) and (2.16) there holds

$$\|u-u_h\|_0^2\le C(\eta^2_{1,\mathcal{T}_h}(p_h,\mathcal{T}_h)+\|p^h-p_h\|_0^2).$$

Let $e^y=y^h-y_h$, and $e^y_I=\hat{\pi}_h e^y$, where $\hat{\pi}_h$ is the average interpolation operator defined in Lemma 3.2 of [14], then we can obtain

$$\begin{aligned}c\|e^y\|_1^2\le&(\nabla(y^h-y_h),\nabla e^y)+(\phi(y^h)-\phi(y_h),e^y)\\=&(\nabla(y^h-y_h),\nabla(e^y-e^y_I))+(\phi(y^h)-\phi(y_h),e^y-e^y_I)\\=&\sum_{T\in\mathcal{T}_h}\int_T(f+u_h-\phi(y_h))(e^y-e^y_I)dx-\sum_{T\in\mathcal{T}_h}\int_{\partial T}([\nabla y_h]\cdot\mathbf{n})(e^y-e^y_I)dx\\\le&C(\sigma)\sum_{T\in\mathcal{T}_h}h_T^2\int_T(f+u_h-\phi(y_h))^2dx+C(\sigma)\sum_{\partial T\backslash\partial\Omega}h_T\int([\nabla y_h]\cdot\mathbf{n})^2dx+C\sigma\|e^y\|_1^2\\=&C(\sigma)\eta^2_{2,\mathcal{T}_h}(u_h,y_h,\mathcal{T}_h)+C\sigma\|e^y\|_1^2,\end{aligned}$$

where σ is an arbitrary positive number. The definition of η_2 will be given later on. Then let $\sigma=\frac{c}{2C}$, we have

$$\|y^h-y_h\|_1^2\le C\eta^2_{2,\mathcal{T}_h}(u_h,y_h,\mathcal{T}_h).$$

Similarly, let $e^p=p^h-p_h$ and e^p_I be the average interpolation of e^p, then we can get

$$\begin{aligned}c\|e^p\|_1^2\le&(\nabla e^p,\nabla(p^h-p_h))+((\phi'(y^h)(p^h-p_h),e^p)\\=&(\nabla e^p,\nabla(p^h-p_h))+(\phi'(y^h)p^h-\phi'(y_h)p_h,e^p)+((\phi'(y_h)-\phi'(y^h))p_h,e^p)\\=&(\nabla(e^p-e^p_I),\nabla(p^h-p_h))+(\phi'(y^h)p^h-\phi'(y_h)p_h,e^p-e^p_I)+(\nabla e^p_I,\nabla(p^h-p_h))\\&+(\phi'(y^h)p^h-\phi'(y_h)p_h,e^p_I)+((\phi'(y_h)-\phi'(y^h))p_h,e^p)\\=&\sum_{T\in\mathcal{T}_h}\int_T(y_h-y_d-\phi'(y_h)p_h)(e^p-e^p_I)dx\\&-\sum_{\partial T\backslash\partial\Omega}\int_{\partial T}([\nabla p_h]\cdot\mathbf{n})(e^p-e^p_I)-(y^h-y_h,e^p_I)+((\phi'(y_h)-\phi'(y^h))p_h,e^p)dx\\\le&C(\sigma)\sum_{T\in\mathcal{T}_h}h_T^2\int_T(y_h-y_d-\phi'(y_h)p_h)^2dx+C(\sigma)\sum_{\partial T\backslash\partial\Omega}h_T\int_{\partial T}([\nabla p_h]\cdot\mathbf{n})^2dx\\&+C\sigma\sum_{T\in\mathcal{T}_h}h_T^2\int_T|\nabla e^p|^2dx+C\|y^h-y_h\|_0\|e^p\|_0+C\|\phi'(y^h)\end{aligned}$$

$$
\begin{aligned}
&- \phi'(y_h)\|_0\|p_h\|_{0,4}\|e^p\|_{0,4}\\
\leq& C(\sigma)\eta^2_{3,\mathcal{T}_h}(y_h, p_h, \mathcal{T}_h) + C(\sigma)\|y^h - y_h\|_0^2 + C(\sigma)\|p_h\|_1^2\|y^h - y_h\|_0^2 + C\sigma\|p^h - p_h\|_1^2)\\
\leq& C(\sigma)\eta^2_{3,\mathcal{T}_h}(y_h, p_h, \mathcal{T}_h) + C\|y^h - y_h\|_0^2 + C\sigma\|p^h - p_h\|_1^2,
\end{aligned}
$$

in which we apply the embedding theorem $\|v\|_{0,4} \leq C\|v\|_1$ (see [6]) and the property: $\|p_h\|_1 \leq C$, $C(\sigma)\|y^h - y_h\|_0^2 + C(\sigma)\|p_h\|_1^2\|y^h - y_h\|_0^2 \leq C(\sigma)\|y^h - y_h\|_0^2 + C(\sigma)C\|y^h - y_h\|_0^2 \leq C\|y^h - y_h\|_0^2$. The definition of η_3 will be given later on. Absolutely, we can obtain

$$
\|p^h - p_h\|_1^2 \leq C\eta^2_{3,\mathcal{T}_h}(y_h, p_h, \mathcal{T}_h) + C\|y^h - y_h\|_0^2.
$$

Hence we gain

$$
\begin{aligned}
\|p_h - p^h\|_0^2 &\leq C\|p_h - p^h\|_1^2 + C\|y_h - y^h\|_1^2\\
&\leq C(\eta^2_{2,\mathcal{T}_h}(u_h, y_h, \mathcal{T}_h) + \eta^2_{3,\mathcal{T}_h}(y_h, p_h, \mathcal{T}_h)).
\end{aligned}
$$

By the triangular inequality we obtain that

$$
\begin{aligned}
&\|y - y_h\|_1 \leq \|y - y^h\|_1 + \|y^h - y_h\|_1 \leq C(\|u - u_h\|_0 + \|y^h - y_h\|_1),\\
&\|p - p_h\|_1 \leq \|p - p^h\|_1 + \|p^h - p_h\|_1 \leq C(\|y - y_h\|_1 + \|p^h - p_h\|_1).
\end{aligned}
$$

In connection with what we discussed above, we have

$$
\|u - u_h\|_0^2 + \|p - p_h\|_1^2 + \|y - y_h\|_1^2 \leq C(\eta^2_{1,\mathcal{T}_h}(p_h, \mathcal{T}_h) + \eta^2_{2,\mathcal{T}_h}(u_h, y_h, \mathcal{T}_h) + \eta^2_{3,\mathcal{T}_h}(y_h, p_h, \mathcal{T}_h)).
$$

Step 2. Now we are in the position to get the lower bound. Consulting to the proof of Lemma 3.6 in [14] and Lemma 3.4 in [21], we conclude that

$$
h_T^2\|\nabla p_h\|_0^2 \approx \sum_{T\in\mathcal{T}_h} \|p_h - \pi_h p_h\|_{0,T}^2 \leq C(\|u - u_h\|_0^2 + \|p - p_h\|_1^2). \tag{2.17}
$$

Proof. It is easily seen that

$$
\begin{aligned}
\sum_{\mathrm{T}\in\mathcal{T}_h} \|p_h - \pi_h p_h\|_{0,T}^2 &= \sum_{\mathrm{T}\in\mathcal{T}_h} \|p_h - \pi_h p_h\|_{0,T}\|p_h - p + p - \pi_h p + \pi_h p - \pi_h p_h\|_{0,T}\\
&\leq \sum_{\mathrm{T}\in\mathcal{T}_h} \|p_h - \pi_h p_h\|_{0,T}\|p - \pi_h p\| + \frac{1}{3}\sum_{\mathrm{T}\in\mathcal{T}_h} \|p_h - \pi_h p_h\|_{0,T}^2\\
&\quad + C\|p_h - p\|_1^2.
\end{aligned}
$$

Since $u + p = \max(0, \bar{p}) = const$, hence

$$
\pi_h(u + p) = u + p,
$$

such that

$$
\sum_{\mathrm{T}\in\mathcal{T}_h} \|p_h - \pi_h p_h\|_{0,T}\|p - \pi_h p\|_{0,T}
$$

$$= \sum_{T\in\mathcal{T}_h} \|p_h - \pi_h p_h\|_{0,T} \|p + u - \pi_h(p+u) + \pi_h p - u\|_{0,T}$$
$$= \sum_{T\in\mathcal{T}_h} \|p_h - \pi_h p_h\|_{0,T} \|\pi_h(u - u_h) - (u - u_h)\|_{0,T}$$
$$\leq \frac{1}{3} \sum_{T\in\mathcal{T}_h} \|p_h - \pi_h p_h\|_{0,T}^2 + C\|u_h - u\|_1^2.$$

This completes the proof.

According to [28], we empoly the standard bubble function technique to estimate error indicators $\eta_{2,\mathcal{T}_h}(u_h, y_h, \mathcal{T}_h)$ and $\eta_{3,\mathcal{T}_h}(y_h, p_h, \mathcal{T}_h)$. Similar to Chapter 7.2 in [22], there exists polynomials $w_T \in H_0^1(T)$ and $w_{\partial T} \in H_0^1(\partial T \backslash \partial\Omega)$ such that

$$\int_{\partial T} h_T([\nabla p_h]\cdot \mathbf{n})^2 dx = \int_{\partial T} ([\nabla p_h]\cdot \mathbf{n}) w_{\partial T} dx, \tag{2.18}$$
$$\int_T h_T^2((y_h - y_d)_T - \phi'(y_h)p_h)^2 dx = \int_T ((y_h - y_d)_T - \phi'(y_h)p_h) w_T dx, \tag{2.19}$$

and apparently

$$\|w_{\partial T}\|_{1,\partial T\backslash\partial\Omega}^2 \leq C \int_{\partial T} h_T([\nabla p_h]\cdot \mathbf{n})^2 dx, \tag{2.20}$$
$$h_T^{-2}\|w_{\partial T}\|_{0,\partial T\backslash\partial\Omega}^2 \leq C \int_{\partial T} h_T([\nabla p_h]\cdot \mathbf{n})^2 dx, \tag{2.21}$$
$$\|w_T\|_{1,T}^2 \leq C \int_h h_T^2((y_h - y_d)_T - \phi'(y_h)p_h)^2 dx, \tag{2.22}$$
$$h_T^{-2}\|w_T\|_{0,T}^2 \leq C \int_h h_T^2((y_h - y_d)_T - \phi'(y_h)p_h)^2 dx. \tag{2.23}$$

By using the Schwarz inequality, it follows from (2.18), (2.20), and (2.21) that

$$\begin{aligned}
&\int_{\partial T} h_T([\nabla p_h]\cdot \mathbf{n})^2 dx \\
=& \int_{\partial T} ([\nabla p_h]\cdot \mathbf{n}) w_{\partial T} dx \\
=& \int_{\partial T} ([\nabla p_h]\cdot \mathbf{n} - [\nabla p]\cdot \mathbf{n}) w_{\partial T} dx \\
=& \int_{\partial T} \nabla(p_h - p)\nabla w_{\partial T} dx + \operatorname{div}(\Delta(p_h - p)) w_{\partial T} \\
=& \int_{\partial T} \nabla(p_h - p)\nabla w_{\partial T} dx + \int_{\partial T} (y - y_d) dx - \phi'(y)p) w_{\partial T} dx \\
=& \int_{\partial T} \nabla(p_h - p)\nabla w_{\partial T} dx + \int_{\partial T} (y_h - y_d - \phi'(y_h)p_h) w_{\partial T} dx \\
&+ \int_{\partial T} ((y - y_d) - (y_h - y_d)) w_{\partial T} dx + \int_{\partial T} (\phi'(y)p - \phi'(y_h)p_h) w_{\partial T} dx \\
\leq& C(\sigma)\|p_h - p\|_{1,\partial T\backslash\partial\Omega}^2 + C(\sigma)\|y - y_h\|_{0,\partial T\backslash\partial\Omega}^2
\end{aligned}$$

$$
\begin{aligned}
&+\int_{\partial T}\phi'(y)(p-p_h)w_{\partial T}dx+\int_{\partial T}(\phi'(y)-\phi'(y_h))p_hw_{\partial T}dx\\
&+C(\sigma)\int_{\partial T}(y_h-y_d-\phi'(y_h)p_h)^2dx+C\sigma(\|w_{\partial T}\|^2_{1,\partial T\backslash\partial\Omega}+h_T^{-2}\|w_{\partial T}\|^2_{0,\partial T\backslash\partial\Omega})\\
\leq&C(\sigma)\|p_h-p\|^2_{1,\partial T\backslash\partial\Omega}+C(\sigma)\|y-y_h\|^2_{0,\partial T\backslash\partial\Omega}\\
&+C(\sigma)\|\phi'(y)\|^2_{0,\partial T\backslash\partial\Omega}\|(p-p_h)\|^2_{0,\partial T\backslash\partial\Omega}+\int_{\partial T}\tilde{\phi}''(y_h)(y-y_h)p_hw_{\partial T}dx\\
&+C(\sigma)\int_{\partial T}(y_h-y_d-\phi'(y_h)p_h)^2dx+C\sigma(\|w_{\partial T}\|^2_{1,\partial T\backslash\partial\Omega}+h_T^{-2}\|w_{\partial T}\|^2_{0,\partial T\backslash\partial\Omega})\\
\leq&C(\sigma)\|p_h-p\|^2_{1,\partial T\backslash\partial\Omega}+C(\sigma)\|y-y_h\|^2_{0,\partial T\backslash\partial\Omega}\\
&+C(\sigma)\|\tilde{\phi}''(y_h)\|^2_{0,\partial T\backslash\partial\Omega}\|y-y_h\|^2_{0,\partial T\backslash\partial\Omega}\|p_h\|^2_{0,\partial T\backslash\partial\Omega}\\
&+C(\sigma)\int_{\partial T}(y_h-y_d-\phi'(y_h)p_h)^2dx+C\sigma(\|w_{\partial T}\|^2_{1,\partial T\backslash\partial\Omega}+h_T^{-2}\|w_{\partial T}\|^2_{0,\partial T\backslash\partial\Omega})\\
\leq&C(\sigma)\|p_h-p\|^2_{1,\partial T\backslash\partial\Omega}+C(\sigma)\|y-y_h\|^2_{0,\partial T\backslash\partial\Omega}\\
&+C(\sigma)\int_{\partial T}(y_h-y_d-\phi'(y_h)p_h)^2dx+C\sigma\int_{\partial T}h_T([\nabla p_h]\cdot\mathbf{n})^2dx,
\end{aligned}
$$

where σ is an arbitrary positive number and $\phi(\cdot)\in W^{2,\infty}(\Omega)$ has been used. Then let $\sigma=\frac{1}{2C}$ and we have

$$
\begin{aligned}
\int_{\partial T}h_T([\nabla p_h]\cdot\mathbf{n})^2dx\leq&C\|p_h-p\|^2_{1,\partial T\backslash\partial\Omega}+C\|y-y_h\|^2_{0,\partial T\backslash\partial\Omega}\\
&+C\int_{\partial T}(y_h-y_d-\phi'(y_h)p_h)^2dx. \qquad (2.24)
\end{aligned}
$$

Next, it follows from (2.19), (2.22), and (2.23) that

$$
\begin{aligned}
&\int_T h_T^2((y_h-y_d)_T-\phi'(y_h)p_h)^2dx\\
=&\int_T((y_h-y_d)_T-\phi'(y_h)p_h)w_Tdx\\
=&\int_T(y_h-y_d-\phi'(y_h)p_h)w_Tdx+\int_T((y_h-y_d)-(y_h-y_d)_T)w_Tdx\\
\leq&\int_T(y_h-y_d-\phi'(y_h)p_h-(y-y_d)+\phi'(y)p)w_Tdx\\
&+C(\sigma)\int_T h_T^2((y_h-y_d)-(y_h-y_d)_T)^2dx+C\sigma h_T^{-2}\|w_T\|^2_{0,T}\\
=&-\int_T\nabla(p_h-p)\nabla w_Tdx+\int_T((y_h-y_d)-(y-y_d))w_Tdx+\int_T(\phi'(y_h)p_h-\phi'(y)p)w_Tdx\\
&+C(\sigma)\int_T h_T^2((y_h-y_d)-(y_h-y_d)_T)^2dx+C\sigma h_T^{-2}\|w_T\|^2_{0,T}\\
\leq&C(\sigma)\|p_h-p\|^2_{1,T}+C(\sigma)\|(y-y_d)-(y_h-y_d)\|^2_{0,T}\\
&+\int_T\phi'(y_h)(p_h-p)w_Tdx+\int_T(\phi'(y_h)-\phi'(y))pw_Tdx
\end{aligned}
$$

$$
\begin{aligned}
&+ C(\sigma)\int_T h_T^2((y_h-y_d)-(y_h-y_d)_T)^2dx + C\sigma(\|w_T\|_{1,T}^2 + h_T^{-2}\|w_T\|_{0,T}^2)\\
\le & C(\sigma)\|p_h-p\|_{1,T}^2 + C(\sigma)\|y-y_h\|_{0,T}^2 + C(\sigma)\int_T h_T^2((y_h-y_d)-(y_h-y_d)_T)^2dx\\
&+ C(\sigma)\|\phi'(y_h)\|_{0,T}^2\|p_h-p\|_0^2 + C(\sigma)\|\tilde{\phi}''(y)\|_{0,T}^2\|y_h-y\|_{0,T}^2\|p\|_{0,T}^2\\
&+ C\sigma(\|w_T\|_{1,T}^2 + h_T^{-2}\|w_T\|_{0,T}^2)\\
\le & C(\sigma)\|p_h-p\|_{1,T}^2 + C(\sigma)\|y-y_h\|_{0,T}^2 + C(\sigma)\int_T h_T^2((y_h-y_d)-(y_h-y_d)_T)^2dx\\
&+ C\sigma\int_T h_T^2((y_h-y_d)_T-\phi'(y_h)p_h)^2dx,
\end{aligned}
$$

where $\phi(\cdot)\in W^{2,\infty}(\Omega)$ has been used. Absolutely, we can deduce that

$$
\begin{aligned}
&\int_T h_T^2((y_h-y_d)_T-\phi'(y_h)p_h)^2dx\\
\le & C\|p_h-p\|_{1,T}^2 + C\|y-y_h\|_{0,T}^2 + C\int_T h_T^2((y_h-y_d)-(y_h-y_d)_T)^2dx.
\end{aligned}
$$

Then we have

$$
\begin{aligned}
&\int_T h_T^2(y_h-y_d-\phi'(y_h)p_h)^2dx\\
\le & C\int_T h_T^2((y_h-y_d)_T-\phi'(y_h)p_h)^2dx + C\int_T h_T^2((y_h-y_d)-(y_h-y_d)_T)^2dx\\
\le & C\|p_h-p\|_{1,T}^2 + C\|y-y_h\|_{0,T}^2 + C\int_T h_T^2((y_h-y_d)-(y_h-y_d)_T)^2dx. \qquad (2.25)
\end{aligned}
$$

In connection with (2.24) and (2.25) we are easy to gain

$$
\begin{aligned}
\eta_{3,\mathcal{T}_h}^2(y_h,p_h,\mathcal{T}_h) =& \sum_{T\in\mathcal{T}_h}\int_T h_T^2(y_h-y_d-\phi'(y_h)p_h)^2dx\\
&+\sum_{\partial T\backslash\partial\Omega}\int_{\partial T} h_T([\nabla p_h]\cdot\mathbf{n})^2dx\\
\le & C\|p_h-p\|_1^2 + C\|y-y_h\|_0^2 + C\sum_{T\in\mathcal{T}_h}\int_T h_T^2((y_h-y_d)-(y_h-y_d)_T)^2dx\\
\le & C\|p_h-p\|_1^2 + C\|y-y_h\|_1^2 + Cosc_{\mathcal{T}_h}^2(y_h-y_d,\mathcal{T}_h).
\end{aligned}
$$

It can also be deduced that

$$
\begin{aligned}
\eta_{2,\mathcal{T}_h}^2(u_h,y_h,\mathcal{T}_h) =& \sum_{T\in\mathcal{T}_h}\int_T h_T^2(f+u_h-\phi(y_h))^2dx\\
&+\sum_{\partial T\backslash\partial\Omega}\int_{\partial T} h_T([\nabla y_h]\cdot\mathbf{n})^2dx\\
\le & C\|y_h-y\|_1^2 + C\|u-u_h\|_0^2 + C\sum_{T\in\mathcal{T}_h}\int_T h_T^2(f-f_T)^2dx
\end{aligned}
$$

$$\leq C\|y_h - y\|_1^2 + C\|u - u_h\|_0^2 + Cosc^2_{\mathcal{T}_h}(f, \mathcal{T}_h).$$

Above-mentioned results tell the proof of Theorem 2.1 is accomplished. □

3. The algorithms

In this section, we introduce two related algorithms as follows:

Algorithm 3.1. *Adaptive finite element algorithm for nonlinear optimal control problems:*

(0) Given an initial grids $\mathcal{T}_{h_0}$ and construct finite element space $U_{ad}^{h_0}$ and V_{h_0}. Select marking parameter $0 < \theta \leq 1$ and set $k := 0$.

(1) Solve the discrete nonlinear optimal control problem (2.8)–(2.10), then obtain approximate solution $(u_{h_k}, y_{h_k}, p_{h_k})$ with respect to $\mathcal{T}_{h_k}$.

(2) Compute the local error estimator $\eta_{\mathcal{T}_{h_k}}(T)$ for all $T \in \mathcal{T}_{h_k}$.

(3) Select a minimal subset $\mathcal{M}_{h_k}$ of $\mathcal{T}_{h_k}$ such that

$$\eta^2_{\mathcal{T}_{h_k}}(\mathcal{M}_{h_k}) \geq \theta\eta^2_{\mathcal{T}_{h_k}}(\mathcal{T}_{h_k}),$$

where $\eta^2_{\mathcal{T}_{h_k}}(\omega) = \eta^2_{1,\mathcal{T}_h}(p_h, \omega) + \eta^2_{2,\mathcal{T}_h}(u_h, y_h, \omega) + \eta^2_{3,\mathcal{T}_h}(y_h, p_h, \omega)$ for all $\omega \subset \mathcal{T}_{h_k}$.

(4) Refine $\mathcal{M}_{h_k}$ by bisecting $b \geq 1$ times in passing from $\mathcal{T}_{h_k}$ to $\mathcal{T}_{h_{k+1}}$ and generally additional elements are refined in the process in order to ensure that $\mathcal{T}_{h_{k+1}}$ is conforming.

(5) Solve the discrete nonlinear optimal control problem (2.8)–(2.10), then obtain approximate solution $(u_{h_{k+1}}, y_{h_{k+1}}, p_{h_{k+1}})$ with respect to $\mathcal{T}_{h_{k+1}}$.

(6) Set $k = k + 1$ and go to step (2).

Algorithm 3.2. *Given an initial control $u_h^0 \in U_{ad}^h$, then seek (y_h^k, p_h^k, u_h^k) such that*

$$\begin{aligned}
&a(y_h^k, w_h) + (\phi(y_h^k), w_h) = (f + u_h^{k-1}, w_h), \quad \forall\, w_h \in V_h,\\
&a(q_h, p_h^{k-1}) + (\phi(y_h^k)p_h^{k-1}, q_h) = (y_h^k - y_d, q_h), \quad \forall\, q_h \in V_h,\\
&(\alpha u_h^k + p_h^{k-1}, v_h - u_h^k) \geq 0, \quad \forall\, v_h^k \in U_{ad}^h,
\end{aligned}$$

for $k = 1, 2, \cdots$, and apparently

$$u_h^k = \frac{1}{\alpha}\big(-\mathcal{P}_h p_h^k + \max(0, \bar{p}_h^k)\big),$$

where $\mathcal{P}_h$ is the L^2-projection from $L^2(\Omega)$ to U^h and $\bar{p}_h^k = \frac{\int_\Omega p_h^k}{|\Omega|}$.

4. Convergence analysis for adaptive finite element method

In this section we first consider the nonlinear elliptic equations as follows:

$$\begin{cases} -\Delta y + \phi(y) = f, \quad \text{in } \Omega,\\ y = 0, \quad \text{on } \partial\Omega, \end{cases} \tag{4.1}$$

where $f \in L^2(\Omega)$. We introduce the quantity $\mathcal{J}(h)$ in view of the idea in [30] as follows:

$$\mathcal{J}(h) := \sup_{f \in L^2(\Omega),\ \|f\|_{0,\Omega}=1} \inf_{v_h \in v_{\mathcal{T}}} \|S - v_h\|_1,$$

where S is the solution operator for nonlinear elliptic equations. Obviously, $\mathcal{J}(h) \ll 1$ for $h \in (0, h_0)$ if $h_0 \ll 1$. Hereinafter there holds the Lemma about the quantity referring to [10, 30].

Lemma 4.1. *For each $f \in L^2(\Omega)$, there exists a constant C such that*

$$\|Sf - S_h f\|_1 \le C\mathcal{J}(h)\|f\|_0, \tag{4.2}$$

and

$$\|Sf - S_h f\|_0 \le C\mathcal{J}(h)\|Sf - S_h f\|_1, \tag{4.3}$$

where S_h is the discrete solution operator for nonlinear elliptic equations.

Here, Lemma 4.1 is a preparation for Lemma 4.4.

4.1. Local perturbation property

According to [19], the local perturbation property plays an important role for the proof of the convergence. It impels us to combine the sum of the error estimates.

Lemma 4.2. *For $\mathcal{T}_h \in \mathbb{T}$, $T \in \mathcal{T}_h$, let $u_{h_1}, u_{h_2} \in U_{ad}^h$, $y_{h_1}, y_{h_2}, p_{h_1}, p_{h_2} \in V_h$, we have*

$$\eta_{1,\mathcal{T}_h}(p_{h_1}, T) - \eta_{1,\mathcal{T}_h}(p_{h_2}, T) \le Ch_T\|p_{h_1} - p_{h_2}\|_{1,T}, \tag{4.4}$$

$$\eta_{2,\mathcal{T}_h}(u_{h_1}, y_{h_1}, T) - \eta_{2,\mathcal{T}_h}(u_{h_2}, y_{h_2}, T) \le C(h_T\|u_{h_1} - u_{h_2}\|_{0,T} + \|y_{h_1} - y_{h_2}\|_{1,T}), \tag{4.5}$$

$$\eta_{3,\mathcal{T}_h}(y_{h_1}, p_{h_1}, T) - \eta_{3,\mathcal{T}_h}(y_{h_2}, p_{h_2}, T) \le C(h_T\|y_{h_1} - y_{h_2}\|_{0,T} + \|p_{h_1} - p_{h_2}\|_{1,T}), \tag{4.6}$$

$$osc_{\mathcal{T}_h}(y_{h_1} - y_d, T) - osc_{\mathcal{T}_h}(y_{h_2} - y_d, T) \le Ch_T^2\|y_{h_1} - y_{h_2}\|_{1,T}. \tag{4.7}$$

Proof. **Step 1.** According to [16, 21] we have

$$\|v\|_{0,\partial T\backslash\partial\Omega} \le C(h_T^{-1/2}\|v\|_{0,T} + h_T^{1/2}\|v\|_{1,T}). \tag{4.8}$$

By adopting the inverse estimates and (4.8) we obtain

$$\|[\nabla(y_{h_1} - y_{h_2})] \cdot \mathbf{n}\|_{0,\partial T\backslash\partial\Omega} \le Ch_T^{-1/2}\|y_{h_1} - y_{h_2}\|_{1,\omega_T}, \tag{4.9}$$

$$\|[\nabla(p_{h_1} - p_{h_2})] \cdot \mathbf{n}\|_{0,\partial T\backslash\partial\Omega} \le Ch_T^{-1/2}\|p_{h_1} - p_{h_2}\|_{1,\omega_T}, \tag{4.10}$$

where ω_T denotes the patch of elements that share an edge with T. By the definition of $\eta_{1,\mathcal{T}_h}(p_h, T)$ and (4.10) we can deduce that

$$\eta_{1,\mathcal{T}_h}(p_{h_1}, T) \le \eta_{1,\mathcal{T}_h}(p_{h_2}, T) + Ch_T\|[\nabla(p_{h_1} - p_{h_2})] \cdot \mathbf{n}\|_{0,\partial T\backslash\partial\Omega}.$$

Then we have

$$\eta_{1,\mathcal{T}_h}(p_{h_1}, T) - \eta_{1,\mathcal{T}_h}(p_{h_2}, T) \le Ch_T\|[\nabla(p_{h_1} - p_{h_2})] \cdot \mathbf{n}\|_{0,\partial T\backslash\partial\Omega} \le Ch_T\|p_{h_1} - p_{h_2}\|_{1,T},$$

where the proof of (4.4) is finished.
Step 2. By the definition of $\eta_{2,\mathcal{T}_h}(u_h, y_h, T)$ and (4.9), we can deduce that

$$\begin{aligned}
&\eta_{2,\mathcal{T}_h}(u_{h_1}, y_{h_1}, T) - \eta_{2,\mathcal{T}_h}(u_{h_2}, y_{h_2}, T)\\
\le& h_T\|u_{h_1} - u_{h_2}\|_{0,T} + h_T^{1/2}\|[\nabla(y_{h_1} - y_{h_2})]\cdot \mathbf{n}\|_{0,\partial T\setminus\partial\Omega} + h_T\|\phi(y_{h_1}) - \phi(y_{h_2})\|_{0,T}\\
\le& h_T\|u_{h_1} - u_{h_2}\|_{0,T} + C\|y_{h_1} - y_{h_2}\|_{1,\omega_T} + Ch_T\|\tilde{\phi}'(y_{h_1})\|_{0,T}\|y_{h_1} - y_{h_2}\|_{0,T}\\
\le& C(h_T\|u_{h_1} - u_{h_2}\|_{0,T} + \|y_{h_1} - y_{h_2}\|_{1,T}),
\end{aligned}$$

where $\phi(\cdot) \in W^{2,\infty}(\Omega)$ has been used and the proof of (4.5) is finished.
Step 3. By the definition of $\eta_{3,\mathcal{T}_h}(y_h, p_h, T)$ and (4.10) we can derive that

$$\begin{aligned}
&\eta_{3,\mathcal{T}_h}(y_{h_1}, p_{h_1}, T) - \eta_{3,\mathcal{T}_h}(y_{h_2}, p_{h_2}, T)\\
\le& h_T\|y_{h_1} - y_{h_2}\|_{0,T} + h_T^{1/2}\|[\nabla(p_{h_1} - p_{h_2})]\cdot \mathbf{n}\|_{0,\partial T\setminus\partial\Omega} + h_T\|\phi'(y_{h_1})p_{h_1} - \phi'(y_{h_2})p_{h_2}\|_{0,T}\\
\le& h_T\|y_{h_1} - y_{h_2}\|_{0,T} + C\|p_{h_1} - p_{h_2}\|_{1,\omega_T} + h_T\|\phi'(y_{h_1})\|_{0,T}\|p_{h_1} - p_{h_2}\|_{0,T}\\
&+ h_T\|\tilde{\phi}''(y_{h_1})\|_{0,T}\|y_{h_1} - y_{h_2}\|_{0,T}\|p_{h_2}\|_{0,T}\\
\le& h_T\|y_{h_1} - y_{h_2}\|_{0,T} + C\|p_{h_1} - p_{h_2}\|_{1,\omega_T} + Ch_T\|p_{h_1} - p_{h_2}\|_{0,T}\\
\le& C(h_T\|y_{h_1} - y_{h_2}\|_{0,T} + \|p_{h_1} - p_{h_2}\|_{1,T}),
\end{aligned}$$

where $\phi(\cdot) \in W^{2,\infty}(\Omega)$ has been used and the proof of (4.6) is finished.
Step 4. Similarly, we have

$$osc_{\mathcal{T}_h}(y_{h_1} - y_d, T) - osc_{\mathcal{T}_h}(y_{h_2} - y_d, T) \le h_T^2\|y_{h_1} - y_{h_2}\|_{0,T}.$$

In brief, Lemma 4.2 is proved. □

4.2. Error reduction

The authors in [25] demonstrate an error reduction provided the current errors are larger than the desired errors, that is to say, the errors may not be reduced in the process of coarse grids refinement before introducing a node of the refined grids inside each marked element while Dörfler proves a similar result assumption [9].

Lemma 4.3. *Let $\mathcal{T}_h \subset \tilde{\mathcal{T}}_h$ for $\mathcal{T}_h, \tilde{\mathcal{T}}_h \in \mathbb{T}$. $\mathcal{M}_h \subset \mathcal{T}_h$ denotes the set of elements which are marked from $\mathcal{T}_h$ to $\tilde{\mathcal{T}}_h$. Then for $u_h \in U_{ad}^h$, $\tilde{u}_h \in U_{ad}^{\tilde{h}}$, $y_h, p_h \in V_h$, $\tilde{y}_h, \tilde{p}_h \in V_{\tilde{h}}$ and any $\delta, \delta_1 \in (0, 1]$, we have*

$$\begin{aligned}
&\eta^2_{1,\tilde{\mathcal{T}}_h}(\tilde{p}_h, \tilde{\mathcal{T}}_h) - (1+\delta_1)\Big\{\eta^2_{1,\mathcal{T}_h}(p_h, \mathcal{T}_h) - (1 - 2^{-1/2})\eta_{1,\mathcal{T}_h}(p_h, \mathcal{R}_h)\Big\}\\
\le& C\left(1+\delta_1^{-1}\right)h_0^2\|p_h - \tilde{p}_h\|_1^2, \qquad (4.11)
\end{aligned}$$

and

$$\begin{aligned}
&\eta^2_{2,\tilde{\mathcal{T}}_h}(\tilde{u}_h, \tilde{y}_h, \tilde{\mathcal{T}}_h) - (1+\delta)\Big\{\eta^2_{2,\mathcal{T}}(u_h, y_h, \mathcal{T}_h) - \lambda\eta^2_{2,\mathcal{T}_h}(u_h, y_h, \mathcal{M}_h)\Big\}\\
\le& C(1+\delta^{-1})\left(h_0^2\|u_h - \tilde{u}_h\|_0^2 + \|y_h - \tilde{y}_h\|_1^2\right), \qquad (4.12)
\end{aligned}$$

and

$$\eta^2_{3,\tilde{\mathcal{T}}_h}(\tilde{y}_h,\tilde{p}_h,\tilde{\mathcal{T}}_h)-(1+\delta)\Big\{\eta^2_{3,\mathcal{T}_h}(y_h,p_h,\mathcal{T}_h)-\lambda\eta^2_{3,\mathcal{T}_h}(u_h,y_h,\mathcal{M}_h)\Big\}$$
$$\le C(1+\delta^{-1})\Big(h_0^2\|y_h-\tilde{y}_h\|_0^2+\|p_h-\tilde{p}_h\|_1^2\Big), \tag{4.13}$$

and

$$osc^2_{\mathcal{T}_h}(y_h-y_d,\mathcal{T}_h\cap\tilde{\mathcal{T}}_h)-2osc^2_{\tilde{\mathcal{T}}_h}(\tilde{y}_h-y_d,\mathcal{T}_h\cap\tilde{\mathcal{T}}_h)$$
$$\le 2Ch_0^4\|y_h-\tilde{y}_h\|_1^2, \tag{4.14}$$

where $\lambda=1-2^{-\frac{b}{2}},\ h_0=\max_{T\in\mathcal{T}_{h_0}} h_T$ *and* $\mathcal{R}_h$ *denotes the set of elements which are refined from* $\mathcal{T}_h$ *to* $\tilde{\mathcal{T}}_h$*.*

Proof. **Step 1.** Applying the Young's inequality with parameter δ_1 and (4.4) we get

$$\begin{aligned}\eta^2_{1,\tilde{\mathcal{T}}_h}(\tilde{p}_h,\tilde{\mathcal{T}}_h)-\eta^2_{1,\tilde{\mathcal{T}}_h}(p_h,\tilde{\mathcal{T}}_h)\le&Ch_0^2\|p_h-\tilde{p}_h\|_1^2+2\eta_{1,\tilde{\mathcal{T}}_h}(\tilde{p}_h,\tilde{\mathcal{T}}_h)\cdot\eta_{1,\tilde{\mathcal{T}}_h}(p_h,\tilde{\mathcal{T}}_h)\\ \le&C(1+\delta_1^{-1})h_0^2\|p_h-\tilde{p}_h\|_1^2+\delta_1\eta^2_{1,\tilde{\mathcal{T}}_h}(p_h,\tilde{\mathcal{T}}_h).\end{aligned} \tag{4.15}$$

Note that T will be bisected at least one time for the element $T\in\mathcal{R}_h\subset\mathcal{T}_h$, then we have

$$\sum_{T'\in T}\eta_{1,\tilde{\mathcal{T}}_h}(p_h,T')\le 2^{-1/2}\eta_{1,\mathcal{T}_h}(p_h,T).$$

For $T\in\mathcal{T}_h\backslash\mathcal{R}_h$, we gain

$$\eta^2_{1,\tilde{\mathcal{T}}_h}(p_h,T)=\eta^2_{1,\mathcal{T}_h}(p_h,T).$$

In connection with the above estimates we demonstrate that

$$\begin{aligned}&\eta^2_{1,\tilde{\mathcal{T}}_h}(\tilde{p}_h,\tilde{\mathcal{T}}_h)-(1+\delta_1)\Big\{\eta^2_{1,\mathcal{T}_h}(p_h,\mathcal{T}_h)-(1-2^{-1/2})\eta_{1,\mathcal{T}_h}(p_h,\mathcal{R}_h)\Big\}\\ =&\eta^2_{1,\tilde{\mathcal{T}}_h}(p_h,\mathcal{R}_h)+\eta^2_{1,\tilde{\mathcal{T}}_h}(p_h,\mathcal{T}_h\backslash\mathcal{R}_h)-(1+\delta_1)\Big\{\eta^2_{1,\mathcal{T}_h}(p_h,\mathcal{T}_h)-(1-2^{-1/2})\eta_{1,\mathcal{T}_h}(p_h,\mathcal{R}_h)\Big\}\\ \le&2^{-1/2}\eta^2_{1,\tilde{\mathcal{T}}_h}(p_h,\mathcal{R}_h)+\eta^2_{1,\tilde{\mathcal{T}}_h}(p_h,\mathcal{T}_h\backslash\mathcal{R}_h)-(1+\delta_1)\Big\{\eta^2_{1,\mathcal{T}_h}(p_h,\mathcal{T}_h)-(1-2^{-1/2})\eta_{1,\mathcal{T}_h}(p_h,\mathcal{R}_h)\Big\}\\ \le&C\left(1+\delta_1^{-1}\right)h_0^2\|p_h-\tilde{p}_h\|_1^2,\end{aligned}$$

which illustrates (4.11) has been proved.

Step 2. Employing the Young's inequality with parameter δ and (4.5) we obtain

$$\begin{aligned}&\eta^2_{3,\tilde{\mathcal{T}}_h}(\tilde{y}_h,\tilde{p}_h,\tilde{\mathcal{T}}_h)-\eta^2_{3,\tilde{\mathcal{T}}_h}(y_h,p_h,\tilde{\mathcal{T}}_h)\\ \le&C(1+\delta^{-1})h_T^2\Big(\sum_{T\in\mathcal{T}_h}h_T\|\tilde{y}_h-y_h\|^2_{0,T}+\|\tilde{p}_h-p_h\|_1^2\Big)+\delta\eta^2_{3,\tilde{\mathcal{T}}_h}(y_h,p_h,\tilde{\mathcal{T}}_h),\end{aligned}$$

where it is similar to (4.15). Then let $\tilde{\mathcal{T}}_{h_{T'}}=\{T\in\tilde{\mathcal{T}}_h: T\subset T'\}$ where $T'\in\mathcal{M}_h$ is a marked element. For arbitrary $p_h\in V_{\mathcal{T}_h}\subset V_{\tilde{\mathcal{T}}_h}$, we find the jump $[\nabla p_h]=0$ on the interior sides of $\cup\tilde{\mathcal{T}}_{h_{T'}}$. Suppose b is the number of bisections, we can deduce that

$$h_T=|T|^{1/2}\le(2^{-b}|T'|)^{1/2}\le 2^{-\frac{b}{2}}h_{T'},$$

due to refinement by bisection, then we obtain

$$\sum_{T\in\tilde{\mathcal{T}}_{hT'}} \eta^2_{3,\tilde{\mathcal{T}}_h}(y_h, p_h, T) \leq 2^{-\frac{b}{2}}\eta^2_{3,\mathcal{T}_h}(y_h, p_h, T').$$

It is easy to find that

$$\eta^2_{3,\tilde{\mathcal{T}}_h}(y_h, p_h, T) \leq \eta^2_{3,\mathcal{T}_h}(y_h, p_h, T),$$

for any $T \in \mathcal{T}_h \backslash \mathcal{M}_h$. In connection with the above estimates we expound that

$$\begin{aligned}\eta^2_{3,\tilde{\mathcal{T}}_h}(y_h, p_h, \tilde{\mathcal{T}}_h) &= \eta^2_{3,\tilde{\mathcal{T}}_h}(y_h, p_h, \mathcal{M}_h) + \eta^2_{3,\tilde{\mathcal{T}}_h}(y_h, p_h, \mathcal{T}_h\backslash\mathcal{M}_h)\\ &\leq 2^{-\frac{b}{2}}\eta^2_{3,\mathcal{T}_h}(y_h, p_h, \mathcal{M}_h) + \eta^2_{3,\mathcal{T}_h}(y_h, p_h, \mathcal{T}_h\backslash\mathcal{M}_h)\\ &= \eta^2_{3,\mathcal{T}_h}(y_h, p_h, \mathcal{T}_h) - (1 - 2^{-\frac{b}{2}})\eta^2_{3,\mathcal{T}_h}(y_h, p_h, \mathcal{M}_h).\end{aligned}$$

As has been said, we can achieve (4.13) connecting with the above estimates to which the proof of (4.12) is similar.

Step 3. For arbitrary $T \in \mathcal{T}_h \cap \tilde{\mathcal{T}}_h$ via using (4.7) and the Young's inequality we obtain that

$$\begin{aligned}&osc_{\mathcal{T}_h}(y_h - y_d, T) = osc_{\tilde{\mathcal{T}}_h}(y_h - y_d, T),\\ &osc^2_{\mathcal{T}_h}(y_h - y_d, T) - 2osc^2_{\tilde{\mathcal{T}}_h}(\tilde{y}_h - y_d, T) \leq 2Ch_T^4\|\tilde{y}_h - y_h\|^2_{1,T}.\end{aligned}$$

Obviously, (4.14) can be got by summing the above inequality over $T \in \mathcal{T}_h \cap \tilde{\mathcal{T}}_h$. To sum up, Lemma 4.3 is proved. □

4.3. Quasi-orthogonality

As to the proof of the convergence, one of the main obstacle is that there do not have the orthogonality while it is vital to prove the convergence. Thus getting back to the second place we transfer proof of the quasi-orthogonality. The latter is popularly adopted in the adaptive mixed and the nonconforming adaptive finite element methods [19]. Apparently it is true for the following basic relationships with $\mathcal{T}_{h_k}, \mathcal{T}_{h_{k+1}} \in \mathbb{T}$ and $\mathcal{T}_{h_k} \subset \mathcal{T}_{h_{k+1}}$,

$$\|u - u_{h_{k+1}}\|^2_0 = \|u - u_{h_k}\|^2_0 - \|u_{h_k} - u_{h_{k+1}}\|^2_0 - 2(u - u_{h_{k+1}}, u_{h_{k+1}} - u_{h_k}), \tag{4.16}$$

$$\|y - y_{h_{k+1}}\|^2_1 = \|y - y_{h_k}\|^2_1 - \|y_{h_k} - y_{h_{k+1}}\|^2_1 - 2a(y - y_{h_{k+1}}, y_{h_{k+1}} - y_{h_k}), \tag{4.17}$$

$$\|p - p_{h_{k+1}}\|^2_1 = \|p - p_{h_k}\|^2_1 - \|p_{h_k} - p_{h_{k+1}}\|^2_1 - 2a(p - p_{h_{k+1}}, p_{h_{k+1}} - p_{h_k}), \tag{4.18}$$

where (u, y, p) are the solution of (2.3)–(2.5), $(u_{h_k}, y_{h_k}, p_{h_k})$ and $(u_{h_{k+1}}, y_{h_{k+1}}, p_{h_{k+1}})$ are the solution of (2.8)–(2.10) with respect to $\mathcal{T}_{h_k}$ and $\mathcal{T}_{h_{k+1}}$, respectively. Accordingly we have the quasi-orthogonality below.

Lemma 4.4. *For $\mathcal{T}_{h_k}, \mathcal{T}_{h_{k+1}} \in \mathbb{T}$ and $\mathcal{T}_{h_k} \subset \mathcal{T}_{h_{k+1}}$, we have*

$$\begin{aligned}&(1 - \delta)\|u - u_{h_{k+1}}\|^2_0 - \|u - u_{h_k}\|^2_0 + \|u_{h_k} - u_{h_{k+1}}\|^2_0\\ &\leq C\delta^{-1}\left(\eta^2_{1,\mathcal{T}_{h_k}}(p_{h_k}, \mathcal{R}_h) + \mathcal{J}^2(h_0)(\eta^2_{2,\mathcal{T}_{h_k}}(u_{h_k}, y_{h_k}, \mathcal{R}_h) + \eta^2_{3,\mathcal{T}_{h_k}}(y_{h_k}, p_{h_k}, \mathcal{R}_h))\right),\end{aligned} \tag{4.19}$$

and

$$
\begin{aligned}
&(1-\delta)\|y-y_{h_{k+1}}\|_1^2-\|y-y_{h_k}\|_1^2-\|y_{h_k}-y_{h_{k+1}}\|_1^2-\delta\|u-u_{h_{k+1}}\|_0^2\\
\le\,& C\delta^{-1}\left(\eta_{1,\mathcal{T}_{h_k}}^2(p_{h_k},\mathcal{R}_h)+\mathcal{J}^2(h_0)(\eta_{2,\mathcal{T}_{h_k}}^2(u_{h_k},y_{h_k},\mathcal{R}_h)+\eta_{3,\mathcal{T}_{h_k}}^2(y_{h_k},p_{h_k},\mathcal{R}_h))\right),
\end{aligned}
\tag{4.20}
$$

and

$$
\begin{aligned}
&(1-\delta)\|p-p_{h_{k+1}}\|_1^2-\|p-p_{h_k}\|_1^2+\|p_{h_k}-p_{h_{k+1}}\|_1^2-\delta\|y-y_{h_{k+1}}\|_1^2\\
\le\,& C\delta^{-1}\left(\eta_{1,\mathcal{T}_{h_k}}^2(p_{h_k},\mathcal{R}_h)+\mathcal{J}^2(h_0)(\eta_{2,\mathcal{T}_{h_k}}^2(u_{h_k},y_{h_k},\mathcal{R}_h)+\eta_{3,\mathcal{T}_{h_k}}^2(y_{h_k},p_{h_k},\mathcal{R}_h))\right).
\end{aligned}
\tag{4.21}
$$

Proof. **Step 1.** It follows from Lemma 4.3 in [19] that for $U_{ad}^{h_k}\subset U_{ad}^{h_{k+1}}$ we have

$$
\begin{aligned}
\alpha\|u_{h_{k+1}}-u_{h_k}\|_0^2\le&(p_{h_{k+1}}-p_{h_k},u_{h_k}-u_{h_{k+1}})+(\alpha u_{h_k}+p_{h_k},u_{h_k}-u_{h_{k+1}})\\
\le&(S^*_{h_{k+1}}(S_{h_{k+1}}(f+u_{h_k})-y_d)-S^*_{h_k}(S_{h_k}(f+u_{h_k})-y_d),u_{h_k}-u_{h_{k+1}})\\
&+C\eta_{1,\mathcal{T}_{h_k}}(p_{h_k},\mathcal{R}_h)\|u_{h_k}-u_{h_{k+1}}\|_0,
\end{aligned}
\tag{4.22}
$$

where $\mathcal{R}_h$ is the set of elements which are refined from $\mathcal{T}_{h_k}$ to $\mathcal{T}_{h_{k+1}}$.

For the right-hand first item of (4.22) we let $\zeta_h\in H_0^1(\Omega)$ be the solution of the following problem based on Lemma 4.1

$$
a(\zeta_h,q)=(S^*_{h_{k+1}}(S_{h_{k+1}}(f+u_{h_k})-y_d)-S^*_{h_k}(S_{h_k}(f+u_{h_k})-y_d),q),\quad\forall\, q\in H_0^1(\Omega).
$$

Hence we can get

$$
\begin{aligned}
&\|S^*_{h_{k+1}}(S_{h_{k+1}}(f+u_{h_k})-y_d)-S^*_{h_k}(S_{h_k}(f+u_{h_k})-y_d)\|_0^2\\
=&a(\zeta_h,S^*_{h_{k+1}}(S_{h_{k+1}}(f+u_{h_k})-y_d)-S^*_{h_k}(S_{h_k}(f+u_{h_k})-y_d))\\
=&a(\zeta_h-\zeta_{h_k},S^*_{h_{k+1}}(S_{h_{k+1}}(f+u_{h_k})-y_d)-S^*_{h_k}(S_{h_k}(f+u_{h_k})-y_d))\\
&+(S_{h_{k+1}}(f+u_{h_k})-S_{h_k}(f+u_{h_k}),\zeta_{h_k}-\zeta_h)+(S_{h_{k+1}}(f+u_{h_k})-S_{h_k}(f+u_{h_k}),\zeta_h).
\end{aligned}
$$

From the proof of Lemma 3.3 in [19], we infer that

$$
\begin{aligned}
&a(\zeta_h-\zeta_{h_k},S^*_{h_{k+1}}(S_{h_{k+1}}(f+u_{h_k})-y_d)-S^*_{h_k}(S_{h_k}(f+u_{h_k})-y_d))\\
\le&\mathcal{J}(h_0)\|S^*_{h_{k+1}}(S_{h_{k+1}}(f+u_{h_k})-y_d)-S^*_{h_k}(S_{h_k}(f+u_{h_k})-y_d)\|_0\\
&\cdot(\|S_{h_{k+1}}(f+u_{h_k})-S_{h_k}(f+u_{h_k})\|_1+\|S^*_{h_{k+1}}(S_{h_k}(f+u_{h_k})-y_d)-S^*_{h_k}(S_{h_k}(f+u_{h_k})-y_d)\|_1),
\end{aligned}
$$

and

$$
\begin{aligned}
&(S_{h_{k+1}}(f+u_{h_k})-S_{h_k}(f+u_{h_k}),\zeta_{h_k}-\zeta_h)\\
\le&C\mathcal{J}(h_0)\|S_{h_{k+1}}(f+u_{h_k})-S_{h_k}(f+u_{h_k})\|_0\\
&\cdot\|S^*_{h_{k+1}}(S_{h_k}(f+u_{h_k})-y_d)-S^*_{h_k}(S_{h_k}(f+u_{h_k})-y_d)\|_0,
\end{aligned}
$$

and

$$
(S_{h_{k+1}}(f+u_{h_k})-S_{h_k}(f+u_{h_k}),\zeta)
$$

$$\leq \|S_{h_{k+1}}(f+u_{h_k}) - S_{h_k}(f+u_{h_k})\|_0 \cdot \|S^*_{h_{k+1}}(S_{h_k}(f+u_{h_k}) - y_d) - S^*_{h_k}(S_{h_k}(f+u_{h_k}) - y_d)\|_0.$$

Similar to Lemma 4.6 in [4], we derive that

$$\begin{aligned}
&\|S_{h_{k+1}}(f+u_{h_k}) - S_{h_k}(f+u_{h_k})\|_1 \leq C\eta_{2,\mathcal{T}_{h_k}}(u_{h_k}, y_{h_k}, \mathcal{R}_h),\\
&\|S^*_{h_{k+1}}(S_{h_k}(f+u_{h_k}) - y_d) - S^*_{h_k}(S_{h_k}(f+u_{h_k}) - y_d)\|_1 \leq C\eta_{3,\mathcal{T}_{h_k}}(y_{h_k}, p_{h_k}, \mathcal{R}_h).
\end{aligned}$$

In connection with the above estimates we conclude that

$$\begin{aligned}
&\|S^*_{h_{k+1}}(S_{h_k}(f+u_{h_k}) - y_d) - S^*_{h_k}(S_{h_k}(f+u_{h_k}) - y_d)\|_0\\
\leq\ & C\Big(\mathcal{J}(h_0)(\eta_{2,\mathcal{T}_{h_k}}(u_{h_k}, y_{h_k}, \mathcal{R}_h) + \eta_{3,\mathcal{T}_{h_k}}(y_{h_k}, p_{h_k}, \mathcal{R}_h)) + \|S_{h_{k+1}}(f+u_{h_k}) - S_{h_k}(f+u_{h_k})\|_0\Big).
\end{aligned}$$

For the third term at the right end of the above inequality, we let $\varphi_h \in H^1_0(\Omega)$ be the solution of the following problem

$$a(q, \varphi_h) = (S_{h_{k+1}}(f+u_{h_k}) - S_{h_k}(f+u_{h_k}), q), \quad \forall\, q \in H^1_0(\Omega).$$

According to the standard duality theory, we can deduce that

$$\begin{aligned}
&\|S_{h_{k+1}}(f+u_{h_k}) - S_{h_k}(f+u_{h_k})\|_0^2\\
=&a(S_{h_{k+1}}(f+u_{h_k}) - S_{h_k}(f+u_{h_k}), \varphi_h - \varphi_{h_k})\\
\leq& C\mathcal{J}(h_0)\|S_{h_{k+1}}(f+u_{h_k}) - S_{h_k}(f+u_{h_k})\|_1 \cdot \|S_{h_{k+1}}(f+u_{h_k}) - S_{h_k}(f+u_{h_k})\|_0,
\end{aligned}$$

where φ_{h_k} is the standard finite element estimate of φ_h with respect to $V_{\mathcal{T}_{h_k}}$. So we have

$$\begin{aligned}
&\|S^*_{h_{k+1}}(S_{h_k}(f+u_{h_k}) - y_d) - S^*_{h_k}(S_{h_k}(f+u_{h_k}) - y_d)\|_0\\
\leq& C\Big(\mathcal{J}(h_0)(\eta_{2,\mathcal{T}_{h_k}}(u_k, y_k, \mathcal{R}_h) + \eta_{3,\mathcal{T}_{h_k}}(y_{h_k}, p_{h_k}, \mathcal{R}_h))\Big).
\end{aligned}$$

As mentioned above, we can get

$$\begin{aligned}
(p_{h_{k+1}} - p_{h_k}, u_{h_k} - u_{h_{k+1}}) \leq& C\Big(\mathcal{J}(h_0)(\eta_{2,\mathcal{T}_{h_k}}(u_{h_k}, y_{h_k}, \mathcal{R}_h)\\
&+ \eta_{3,\mathcal{T}_{h_k}}(y_{h_k}, p_{h_k}, \mathcal{R}_h))\|u_{h_k} - u_{h_{k+1}}\|_0\Big).
\end{aligned}$$

In combination with (4.22) and above inequality, we deduct that

$$\|u_{h_k} - u_{h_{k+1}}\|_0 \leq C\Big(\eta_{1,\mathcal{T}_{h_k}}(p_{h_k}, \mathcal{R}_h) + \mathcal{J}(h_0)(\eta_{2,\mathcal{T}_{h_k}}(u_{h_k}, y_{h_k}, \mathcal{R}_h) + \eta_{3,\mathcal{T}_{h_k}}(y_{h_k}, p_{h_k}, \mathcal{R}_h))\Big). \tag{4.23}$$

It is easy to derive the desired result (4.19) with the help of (4.16) and (4.23).
Step 2. Our task now is to prove (4.20) and so is (4.21). Obviously we have

$$\begin{aligned}
\|y_{h_{k+1}} - y_{h_k}\|_0 =& \|S_{h_{k+1}}(f+u_{h_{k+1}}) - S_{h_k}(f+u_{h_k})\|_0\\
\leq& \|S_{h_{k+1}}(f+u_{h_{k+1}}) - S_{h_{k+1}}(f+u_{h_k})\|_0 + \|S_{h_{k+1}}(f+u_{h_k}) - S_{h_{h_k}}(f+u_{h_k})\|_0\\
\leq& C(\|u_{h_k} - u_{h_{k+1}}\|_0 + \mathcal{J}(h_0)\eta_{2,\mathcal{T}_{h_k}}(u_{h_k}, y_{h_k}, \mathcal{R}_h)).
\end{aligned} \tag{4.24}$$

By using the Cauchy inequality, we obtain

$$\begin{aligned}
&2a(y-y_{h_{k+1}},y_{h_{k+1}}-y_{h_k})\\
=&2(u-u_{h_{k+1}},y_{h_{k+1}}-y_{h_k})-2(\phi(y)-\phi(y_{h_{k+1}}),y_{h_{k+1}}-y_{h_k})\\
\leq&\delta\|u-u_{h_{k+1}}\|_0^2+\frac{1}{\delta}\|y_{h_{k+1}}-y_{h_k}\|_0^2+(\tilde{\phi}'(y)(y-y_{h_{k+1}}),y_{h_{k+1}}-y_{h_k})\\
\leq&\delta\|u-u_{h_{k+1}}\|_0^2+\frac{1}{\delta}\|y_{h_{k+1}}-y_{h_k}\|_0^2+\|\tilde{\phi}'(y)\|_0\|y-y_{h_{k+1}}\|_0\|y_{h_{k+1}}-y_{h_k}\|_0\\
\leq&\delta\|u-u_{h_{k+1}}\|_0^2+\frac{1}{\delta}\|y_{h_{k+1}}-y_{h_k}\|_0^2+C\left(\delta\|y-y_{h_{k+1}}\|_0^2+\frac{1}{\delta}\|y_{h_{k+1}}-y_{h_k}\|_0^2\right).
\end{aligned} \tag{4.25}$$

It is easy to derive the desired result (4.20) with the assistance of (4.17) and (4.23)–(4.25). □

4.4. Convergence

For $\mathcal{T}_{h_k}\in\mathbb{T}$, we will denote $U_{ad}^{h_k}$, V_{h_k} and the solution $(u_{h_k},y_{h_k},p_{h_k})$ of (2.8)–(2.10) with respect to $\mathcal{T}_{h_k}$ by $U_{ad}^{h_k}$, V_{h_k}, and $(u_{h_k},y_{h_k},p_{h_k})$ and we define some notations before we prove the convergence of Algorithm 3.1 as follows:

$$\begin{aligned}
e_{h_k}^2&=\|u-u_{h_k}\|_0^2+\|y-y_{h_k}\|_1^2+\|p-p_{h_k}\|_1^2,\\
E_{h_k}^2&=\|u_{h_k}-u_{h_{k+1}}\|_0^2+\|y_{h_k}-y_{h_{k+1}}\|_1^2+\|p_{h_k}-p_{h_{k+1}}\|_1^2,\\
\tilde{\eta}_{\mathcal{T}_{h_k}}^2(\omega)&=\eta_{2,\mathcal{T}_{h_k}}^2(u_{h_k},y_{h_k},\omega)+\eta_{3,\mathcal{T}_{h_k}}^2(y_{h_k},p_{h_k},\omega),
\end{aligned}$$

for $\omega\subset\Omega$.

Theorem 4.1. *Let $(\mathcal{T}_{h_k},U_{ad}^{h_k},V_{h_k},u_{h_k},y_{h_k},p_{h_k})$ be the sequence of grids, finite element spaces and discrete solutions produced by the Algorithm 3.1. Then there exist constants $\gamma_1>0$, $\gamma_2>0$ and $\alpha\in(0,1)$, only depending on the shape regularity of initial grids $\mathcal{T}_{h_0}$, b, and the marking parameter $\theta\in(0,1]$, such that*

$$e_{h_{k+1}}^2+\gamma_1\eta_{1,\mathcal{T}_{h_{k+1}}}^2(p_{h_{k+1}},\mathcal{T}_{h_{k+1}})+\gamma_2\tilde{\eta}_{\mathcal{T}_{h_{k+1}}}^2(\mathcal{T}_{h_{k+1}})\leq\alpha\left(e_{h_k}^2+\gamma_1\eta_{1,\mathcal{T}_{h_k}}^2(p_{h_k},\mathcal{T}_{h_k})+\gamma_2\tilde{\eta}_{\mathcal{T}_{h_k}}^2(\mathcal{T}_{h_k})\right), \tag{4.26}$$

provided $h_0\ll1$.

Proof. We get the following results from Theorem 2.1, Lemma 4.3 and Lemma 4.4,

$$e_{h_k}^2\leq C\eta_{\mathcal{T}_{h_k}}^2(\mathcal{T}_{h_k}), \tag{4.27}$$

$$\tilde{\eta}_{\mathcal{T}_{h_{k+1}}}^2(\mathcal{T}_{h_{k+1}})\leq(1+\delta)\Big\{\tilde{\eta}_{\mathcal{T}_{h_k}}^2(\mathcal{T}_{h_k})-\lambda\tilde{\eta}_{\mathcal{T}_{h_k}}^2(\mathcal{M}_{h_k})\Big\}+C\left(1+\frac{1}{\delta}\right)E_{h_k}^2, \tag{4.28}$$

$$\begin{aligned}
\eta_{1,\mathcal{T}_{h_{k+1}}}^2(p_{h_{k+1}},\mathcal{T}_{h_{k+1}})\leq&(1+\delta_1)\Big\{\eta_{1,\mathcal{T}_{h_k}}^2(p_{h_k},\mathcal{T}_{h_k})-(1-2^{-1/2})\eta_{1,\mathcal{T}_{h_k}}^2(p_{h_k},\mathcal{R}_{h_k})\Big\}\\
&+C\left(1+\frac{1}{\delta_1}\right)h_0^2\|p_{h_k}-p_{h_{k+1}}\|_1^2,
\end{aligned} \tag{4.29}$$

$$(1-2\delta)e_{h_{k+1}}^2\leq e_{h_k}^2-E_{h_k}^2+C\frac{1}{\delta}\Big(\eta_{1,\mathcal{T}_{h_k}}^2(p_{h_k},\mathcal{R}_{h_k})+\mathcal{J}(h_0)\tilde{\eta}_{\mathcal{T}_{h_k}}^2(\mathcal{T}_{h_k})\Big), \tag{4.30}$$

where $\mathcal{R}_{h_k}$ is the set of elements which are refined from $\mathcal{T}_{h_k}$ to $\mathcal{T}_{h_{k+1}}$. Applying the upper bound in Theorem 2.1, (4.29) can be simplified into

$$\begin{aligned}\eta^2_{1,\mathcal{T}_{h_{k+1}}}(p_{h_{k+1}},\mathcal{T}_{h_{k+1}}) \le{}& (1+\delta_1)\Big\{\eta^2_{1,\mathcal{T}_{h_k}}(p_{h_k},\mathcal{T}_{h_k}) - (1-2^{-1/2})\eta^2_{1,\mathcal{T}_{h_k}}(p_{h_k},\mathcal{R}_{h_k})\Big\}\\ &+ C\Big(1+\frac{1}{\delta_1}\Big)h_0^2\Big(\eta^2_{1,\mathcal{T}_{h_{k+1}}}(p_{h_{k+1}},\mathcal{T}_{h_{k+1}}) + \tilde{\eta}^2_{\mathcal{T}_{h_{k+1}}}(\mathcal{T}_{h_{k+1}})\\ &+ \eta^2_{1,\mathcal{T}_{h_k}}(p_{h_k},\mathcal{T}_{h_k}) + \tilde{\eta}^2_{\mathcal{T}_{h_k}}(\mathcal{T}_{h_k})\Big). \qquad (4.31)\end{aligned}$$

Multiplying (4.28) and (4.31) with $\tilde{\gamma}_2$ and $\tilde{\gamma}_1$, respectively, and adding the results to (4.30) yields

$$\begin{aligned}&(1-2\delta)e^2_{h_{k+1}} + \tilde{\gamma}_1\eta^2_{1,\mathcal{T}_{h_{k+1}}}(p_{h_{k+1}},\mathcal{T}_{h_{k+1}}) + \tilde{\gamma}_2\tilde{\eta}^2_{\mathcal{T}_{h_{k+1}}}(\mathcal{T}_{h_{k+1}})\\ \le{}& e^2_{h_k} + \tilde{\gamma}_2(1+\delta)\Big\{\tilde{\eta}^2_{\mathcal{T}_{h_k}}(\mathcal{T}_{h_k}) - \lambda\tilde{\eta}^2_{\mathcal{T}_{h_k}}(\mathcal{M}_{h_k})\Big\}\\ &+ \tilde{\gamma}_2 C\Big(1+\frac{1}{\delta}\Big)E^2_{h_k} - E^2_{h_k} - \Big(\tilde{\gamma}_1(1+\delta_1)(1-2^{-1/2}) - C\frac{1}{\delta}\Big)\eta^2_{1,\mathcal{T}_{h_k}}(p_{h_k},\mathcal{R}_{h_k})\\ &+ \tilde{\gamma}_1(1+\delta_1)\eta^2_{1,\mathcal{T}_{h_k}}(p_{h_k},\mathcal{T}_{h_k}) + C\frac{1}{\delta}\mathcal{J}^2(h_0)\tilde{\eta}^2_{\mathcal{T}_{h_k}}(\mathcal{T}_{h_k})\\ &+ \tilde{\gamma}_1 C\Big(1+\frac{1}{\delta_1}\Big)h_0^2\Big(\eta^2_{1,\mathcal{T}_{h_{k+1}}}(p_{h_{k+1}},\mathcal{T}_{h_{k+1}}) + \tilde{\eta}^2_{\mathcal{T}_{h_{k+1}}}(\mathcal{T}_{h_{k+1}}) + \eta^2_{1,\mathcal{T}_{h_k}}(p_{h_k},\mathcal{T}_{h_k}) + \tilde{\eta}^2_{\mathcal{T}_{h_k}}(\mathcal{T}_{h_k})\Big).\end{aligned}$$

If $\tilde{\gamma}_1$ is such that

$$\tilde{\gamma}_1(1+\delta_1)(1-2^{-1/2}) - C\frac{1}{\delta} > 0,$$

and one chooses $\tilde{\gamma}_2$ such that

$$\tilde{\gamma}_2 C\Big(1+\frac{1}{\delta}\Big) = 1, \qquad (4.32)$$

then we have

$$\begin{aligned}&(1-2\delta)e^2_{h_{k+1}} + \tilde{\gamma}_1\Big(1 - C\Big(1+\frac{1}{\delta_1}\Big)h_0^2\Big)\eta^2_{1,\mathcal{T}_{h_{k+1}}}(p_{h_{k+1}},\mathcal{T}_{h_{k+1}})\\ &+ \Big(\tilde{\gamma}_2 - \tilde{\gamma}_1 C\Big(1+\frac{1}{\delta_1}\Big)h_0^2\Big)\tilde{\eta}^2_{\mathcal{T}_{h_{k+1}}}(\mathcal{T}_{h_{k+1}})\\ \le{}& e^2_{h_k} + \tilde{\gamma}_1\Big((1+\delta_1) + C\Big(1+\frac{1}{\delta_1}\Big)h_0^2\Big)\eta^2_{1,\mathcal{T}_{h_k}}(p_{h_k},\mathcal{T}_{h_k})\\ &+ \Big(\tilde{\gamma}_2(1+\delta) + \tilde{\gamma}_1 C\Big(1+\frac{1}{\delta_1}\Big)h_0^2 + C\frac{1}{\delta}\mathcal{J}^2(h_0)\Big)\tilde{\eta}^2_{\mathcal{T}_{h_k}}(\mathcal{T}_{h_k})\\ &- c\Big(\eta^2_{1,\mathcal{T}_{h_k}}(p_{h_k},\mathcal{M}_{h_k}) + \tilde{\eta}^2_{\mathcal{T}_{h_k}}(\mathcal{M}_{h_k})\Big),\end{aligned}$$

where

$$c = \min\Big\{\tilde{\gamma}_2(1+\delta)\lambda, \tilde{\gamma}_1(1+\delta_1)(1-2^{-1/2}) - C\frac{1}{\delta}\Big\}.$$

By using the marking strategy in Algorithm 3.1 and the upper bound in Theorem 2.1 to arrive at

$$\begin{aligned}&(1-2\delta)e^2_{h_{k+1}} + \tilde{\gamma}_1\Big(1 - C\Big(1+\frac{1}{\delta_1}\Big)h_0^2\Big)\eta^2_{1,\mathcal{T}_{h_{k+1}}}(p_{h_{k+1}},\mathcal{T}_{h_{k+1}})\\ &+ \Big(\tilde{\gamma}_2 - \tilde{\gamma}_1 C\Big(1+\frac{1}{\delta_1}\Big)h_0^2\Big)\tilde{\eta}^2_{\mathcal{T}_{h_{k+1}}}(\mathcal{T}_{h_{k+1}})\end{aligned}$$

$$\leq (1 - C\theta\beta)e_{h_k}^2 + \Big(\tilde{\gamma}_1\Big((1 + \delta_1) + C\Big(1 + \frac{1}{\delta_1}\Big)h_0^2\Big) - c\theta(1 - \beta)\Big)\eta_{1,\mathcal{T}_{h_k}}(p_{h_k}, \mathcal{T}_{h_k})$$
$$+ \Big(\tilde{\gamma}_2(1 + \delta) + \tilde{\gamma}_1 C\Big(1 + \frac{1}{\delta_1}\Big)h_0^2 + C\frac{1}{\delta}\mathcal{J}^2(h_0) - c\theta(1 - \beta)\Big)\tilde{\eta}_{\mathcal{T}_{h_k}}(\mathcal{T}_{h_k}),$$

where $\beta \in (0, 1)$. Then we deduce that

$$\begin{aligned} & e_{h_{k+1}}^2 + \gamma_1\eta_{1,\mathcal{T}_{h_{k+1}}}^2(p_{h_{k+1}}, \mathcal{T}_{h_{k+1}}) + \gamma_2\tilde{\eta}_{\mathcal{T}_{h_{k+1}}}^2(\mathcal{T}_{h_{k+1}}) \\ & \leq \alpha_1 e_{h_k}^2 + \alpha_2\gamma_1\eta_{1,\mathcal{T}_{h_k}}^2(p_{h_k}, \mathcal{T}_{h_k}) + \alpha_3\gamma_2\tilde{\eta}_{\mathcal{T}_{h_k}}^2(\mathcal{T}_{h_k}), \end{aligned}$$

where

$$\begin{aligned} \alpha_1 &= \frac{1 - C\theta\beta}{1 - 2\delta}, \\ \gamma_1 &= \frac{\tilde{\gamma}_1\Big(1 - C(1 + \delta_1^{-1})h_0^2\Big)}{1 - 2\delta}, \\ \gamma_2 &= \frac{\tilde{\gamma}_2 - \tilde{\gamma}_1 C(1 + \delta_1^{-1})h_0^2}{1 - 2\delta}, \\ \alpha_2 &= \frac{\tilde{\gamma}_1\Big((1 + \delta_1) + C(1 + \delta_1^{-1})h_0^2\Big) - c\theta(1 - \beta)}{\tilde{\gamma}_1\Big(1 - C(1 + \delta_1^{-1})h_0^2\Big)}, \\ \alpha_3 &= \frac{\tilde{\gamma}_2(1 + \delta) + \tilde{\gamma}_1 C(1 + \delta_1^{-1})h_0^2 + C\delta^{-1}\mathcal{J}^2(h_0) - c\theta(1 - \beta)}{\tilde{\gamma}_2 - \tilde{\gamma}_1 C(1 + \delta_1^{-1})h_0^2}. \end{aligned}$$

As long as $\delta < 1$ and β is small enough, $\alpha_1 \in (0, 1)$ can be guaranteed. To facilitate judgment, we transfer the following adjustments to the above formula:

$$\begin{aligned} \alpha_2 &= \frac{(1 - C(1 + \delta_1^{-1})h_0^2) + \delta_1 + 2C(1 + \delta_1^{-1})h_0^2 - \frac{c\theta(1-\beta)}{\tilde{\gamma}_1}}{1 - C(1 + \delta_1^{-1})h_0^2}, \\ \alpha_3 &= \frac{\tilde{\gamma}_2(1 + \delta) - \tilde{\gamma}_1 C(1 + \delta_1^{-1})h_0^2 + 2\tilde{\gamma}_1 C(1 + \delta_1^{-1})h_0^2 + C\delta^{-1}\mathcal{J}^2(h_0) - c\theta(1 - \beta)}{\tilde{\gamma}_2 - \tilde{\gamma}_1 C(1 + \delta_1^{-1})h_0^2}. \end{aligned}$$

It is absolutely clear that

$$\alpha_2 \in (0, 1),$$

if $h_0 \ll 1$ and δ_1 is sufficiently small. Then consider (4.32) to deduce that

$$\gamma_2 = \frac{\delta^2}{C(1 + \delta)},$$

which can say

$$\alpha_3 \in (0, 1),$$

if $h_0 \ll 1$ and δ is sufficiently small. Therefore if choose $\alpha = \max\{\alpha_1, \alpha_2, \alpha_3\}$, we can derive the expected results.

□

Theorem 4.2. *Let* $(\mathcal{T}_{h_k}, U_{ad}^{h_k}, V_{h_k}, u_{h_k}, y_{h_k}, p_{h_k})$ *be the sequence of grids, finite element spaces and discrete solutions produced by the Algorithm 3.1 and the conditions of Theorem 4.1 keep. Then we have*

$$\|u - u_{h_k}\|_0^2 + \|y - y_{h_k}\|_1^2 + \|p - p_{h_k}\|_1^2 \to 0 \quad \text{as} \quad k \to \infty.$$

Proof. It is obviously true combining Theorem 2.1 and Theorem 4.1. □

5. Quasi-optimality for adaptive finite element algorithm

In this section we consider the quasi-optimality for the adaptive finite element method. Firstly we give the notations interpretation. For $\mathcal{T}_h, \mathcal{T}_{h_1}, \mathcal{T}_{h_2} \in \mathbb{T}$, let $\#\mathcal{T}_h$ be the number of elements in $\mathcal{T}_h$, and $\mathcal{T}_{h_1} \oplus \mathcal{T}_{h_2}$ be the smallest common conforming refinement of $\mathcal{T}_{h_1}$ and $\mathcal{T}_{h_2}$ and satisfies [4, 27]

$$\#(\mathcal{T}_{h_1} \oplus \mathcal{T}_{h_2}) \leq \#\mathcal{T}_{h_1} + \#\mathcal{T}_{h_2} - \#\mathcal{T}_{h_0}. \tag{5.1}$$

According to [19], we need defining a function approximation class

$$\begin{aligned}\mathcal{A}^s :=& \{(u, y, p, y_d, f) \in L^2(\Omega) \times H_0^1(\Omega) \times H_0^1(\Omega) \times L^2(\Omega) \\ &\times L^2(\Omega) : |(u, y, p, y_d, f)|_s < +\infty\},\end{aligned}$$

where

$$\begin{aligned}|(u, y, p, y_d, f)|_s :=& \sup_{N>0} N^s \inf_{\mathcal{T}_h \in \mathbb{T}_N} \inf_{(u_h, y_h, p_h) \in U_{ad}^h \times V_h \times V_h} \{\|u - u_h\|_0^2 \\ &+ \|y - y_h\|_1^2 + \|p - p_h\|_1^2 + osc_{\mathcal{T}_h}^2(f, \mathcal{T}_h)) + osc_{\mathcal{T}_h}^2(y_h - y_d, \mathcal{T}_h)\}^{\frac{1}{2}},\end{aligned}$$

and

$$\mathbb{T}_N := \{\mathcal{T}_h \in \mathbb{T} : \#\mathcal{T}_h - \#\mathcal{T}_{h_0} \leq N_0\}.$$

5.1. Localized upper bound

To illustrate the quasi-optimality of the adaptive finite element method, we need a local upper bound on the distance between nested solutions [4], since the error of this method can only be estimated by using the indicators of refined elements without a buffer layer.

Lemma 5.1. *For* $\mathcal{T}_h, \tilde{\mathcal{T}}_h \in \mathbb{T}$ *and* $\mathcal{T}_h \subset \tilde{\mathcal{T}}_h$*, let* $\mathcal{R}_h$ *be the set of refined elements from* $\mathcal{T}_h$ *to* $\tilde{\mathcal{T}}_h$*. Let* (u_h, y_h, p_h) *and* $(\tilde{u}, \tilde{y}, \tilde{p})$ *be the solutions of (2.8)–(2.10) with respect to* $\mathcal{T}_h$ *and* $\tilde{\mathcal{T}}_h$*, respectively. Then there exists a constant C, depending on the shape regularity of initial grids* $\mathcal{T}_{h_0}$ *and b such that*

$$\|u_h - \tilde{u}\|_0^2 + \|y_h - \tilde{y}\|_1^2 + \|p_h - \tilde{p}\|_1^2 \leq C\eta_{\mathcal{T}_h}^2(\mathcal{R}_h), \tag{5.2}$$

where

$$\eta_{\mathcal{T}_h}^2(\mathcal{R}_h) = \eta_{1,\mathcal{T}_h}^2(p_h, \mathcal{R}_h) + \eta_{2,\mathcal{T}_h}^2(u_h, y_h, \mathcal{R}_h) + \eta_{3,\mathcal{T}_h}^2(y_h, p_h, \mathcal{R}_h).$$

Proof. From (4.23) of Lemma 4.4, we have

$$\|u_h - \tilde{u}\|_0^2 \le \eta^2_{\mathcal{T}_h}(\mathcal{R}_h). \tag{5.3}$$

By Lemma 4.6 in [4], we deduce that

$$\begin{aligned} \|y_h - \tilde{y}\|_1 &= \|S_h(f + u_h) - S_{\tilde{h}}(f + \tilde{u})\|_1 \\ &\le \|S_h(f + u_h) - S_{\tilde{h}}(f + u_h)\|_1 + \|S_{\tilde{h}}(f + u_h) - S_{\tilde{h}}(f + \tilde{u})\|_1 \\ &\le C(\eta_{2,\mathcal{T}_h}(u_h, y_h, \mathcal{R}_h) + \|u_h - \tilde{u}\|_0). \end{aligned} \tag{5.4}$$

The corresponding result is attained by the similar method above

$$\begin{aligned} \|p_h - \tilde{p}\|_1 &= \|S_h^*(S_h(f + u_h) - y_d) - S_{\tilde{h}}^*(S_{\tilde{h}}(f + \tilde{u}) - y_d)\|_1 \\ &\le C(\eta_{3,\mathcal{T}_h}(y_h, p_h, \mathcal{R}_h) + \|y_h - \tilde{y}\|_1). \end{aligned} \tag{5.5}$$

In connection with (5.3), (5.4) and (5.5) we can derive (5.2). □

5.2. Dörfler property

Dörfler introduced a crucial marking and proved strict energy error reduction for the Laplacian provided the initial grids $\mathcal{T}_{h_0}$ satisfying a mild assumption [9]. If the sum of errors satisfy suitable error reductions, the error indicators on the coarse grids must satisfy a Dörfler property on the refined one [19].

Lemma 5.2. *Assume that the marking parameter $\theta \in (0, \theta^*)$, where*

$$\theta^* = \frac{C}{2C(1 + h_0^4) + 1}.$$

For $\mathcal{T}_h, \tilde{\mathcal{T}}_h \in \mathbb{T}$ and $\mathcal{T}_h \subset \tilde{\mathcal{T}}_h$, let (u_h, y_h, p_h) and $(\tilde{u}, \tilde{y}, \tilde{p})$ be the solutions of (2.8)–(2.10) with respect to $\mathcal{T}_h$ and $\tilde{\mathcal{T}}_h$, respectively. If

$$e^2_{\tilde{\mathcal{T}}_h} + osc^2_{\tilde{\mathcal{T}}_h}(\tilde{\mathcal{T}}_h) \le \mu[e^2_{\mathcal{T}_h} + osc^2_{\mathcal{T}_h}(\mathcal{T}_h)], \tag{5.6}$$

is satisfied for $\mu := \frac{1}{2}\left(1 - \frac{\theta}{\theta^*}\right)$. Then, the set $\mathcal{R}_h$ of elements which are refined from $\mathcal{T}_h$ to $\tilde{\mathcal{T}}_h$ satisfies the Dörfler property

$$\eta^2_{\mathcal{T}_h}(\mathcal{R}_h) \ge \theta \eta^2_{\mathcal{T}_h}(\mathcal{T}_h),$$

where

$$\begin{aligned} e^2_{\mathcal{T}_h} &= \|u - u_h\|_0^2 + \|y - y_h\|_1^2 + \|p - p_h\|_1^2, \\ osc^2_{\mathcal{T}_h}(\omega) &= osc^2_{\mathcal{T}_h}(f, \omega) + osc^2_{\mathcal{T}_h}(y_h - y_d, \omega), \end{aligned}$$

for $\omega \subset \mathcal{T}_h$ and $e^2_{\tilde{\mathcal{T}}_h}$, $osc^2_{\tilde{\mathcal{T}}_h}(\tilde{\mathcal{T}}_h)$ similarly defined.

Proof. By the lower bound in Theorem 2.1 and (5.6) to obtain that

$$\begin{aligned} (1 - 2\mu)C\eta^2_{\mathcal{T}_h}(\mathcal{T}_h) &\le (1 - 2\mu)(e^2_{\mathcal{T}_h} + osc^2_{\mathcal{T}_h}(\mathcal{T}_h)) \\ &\le e^2_{\mathcal{T}_h} - 2e^2_{\tilde{\mathcal{T}}_h} + osc^2_{\mathcal{T}_h}(\mathcal{T}_h) - 2osc^2_{\tilde{\mathcal{T}}_h}(\tilde{\mathcal{T}}_h). \end{aligned}$$

It is well-known that there exists the fundamental relationships:

$$\|u-u_h\|_0^2 \le 2\|u-\tilde{u}\|_0^2 + 2\|u_h-\tilde{u}\|_0^2,$$
$$\|y-y_h\|_1^2 \le 2\|y-\tilde{y}\|_1^2 + 2\|y_h-\tilde{y}\|_1^2,$$
$$\|p-p_h\|_1^2 \le 2\|p-\tilde{p}\|_1^2 + 2\|p_h-\tilde{p}\|_1^2.$$

Hence we can get the following result form Lemma 5.1

$$e_{\mathcal{T}_h}^2 - 2e_{\tilde{\mathcal{T}}_h}^2 \le 2C\eta_{\mathcal{T}_h}^2(\mathcal{R}_h). \tag{5.7}$$

For $T \in \mathcal{T}_h \cap \tilde{\mathcal{T}}_h$, we can get the following result from (4.14) of Lemma 4.3

$$osc_{\mathcal{T}_h}^2(y_h-y_d,\mathcal{T}_h\cap\tilde{\mathcal{T}}_h) - 2osc_{\tilde{\mathcal{T}}_h}^2(\tilde{y}-y_d,\mathcal{T}_h\cap\tilde{\mathcal{T}}_h) \le 2C(h_0^4\eta_{\mathcal{T}_h}^2(\mathcal{R}_h)). \tag{5.8}$$

According to Remark 2.1 in [4], we get the following result as the indicator $\eta_{\mathcal{T}_h}(T)$ dominates oscillation $osc_{\mathcal{T}_h}(T)$

$$osc_{\mathcal{T}_h}^2(T) \le \eta_{\mathcal{T}_h}^2(T), \tag{5.9}$$

for all $T \in \mathcal{R}_h$. Then in connection with (5.7), (5.8) and (5.9) one obtains

$$(1-2\mu)C\eta_{\mathcal{T}_h}^2(\mathcal{T}_h) \le (2C(1+h_0^4)+1)\eta_{\mathcal{T}_h}^2(\mathcal{R}_h),$$
$$(1-2\mu)\theta^*\eta_{\mathcal{T}_h}^2(\mathcal{T}_h) \le \eta_{\mathcal{T}_h}^2(\mathcal{R}_h),$$
$$\theta\eta_{\mathcal{T}_h}^2(\mathcal{T}_h) \le \eta_{\mathcal{T}_h}^2(\mathcal{R}_h),$$

where

$$\theta^* = \frac{C}{2C(1+h_0^4)+1}, \quad \text{and} \quad \theta = (1-2\mu)\theta^*.$$

□

Lemma 5.3. *Let* (u, y, p) *and* $(\mathcal{T}_{h_k}, U_{ad}^{h_k}, V_{h_k}, u_{h_k}, y_{h_k}, p_{h_k})$ *be the solution of (2.3)–(2.5) and the sequence of grids, finite element spaces and discrete solutions produced by Algorithm 3.1, respectively. Assume that the marking parameter* θ *satisfies the condition in Lemma 5.2, then the following estimate is valid*

$$\#\mathcal{M}_{h_k} \le C\Big(N^{\frac{1}{2s}}|(u,y,p,y_d,f)|_s^{\frac{1}{s}}\mu^{-\frac{1}{2s}}(e_{h_k}^2 + osc_{\mathcal{T}_{h_k}}^2(\mathcal{T}_{h_k}))^{-\frac{1}{2s}}\Big), \tag{5.10}$$

if $(u, y, p, y_d, f) \in \mathcal{A}^s$.

Proof. Let $\epsilon^2 := \mu N^{-1}(e_{\mathcal{T}_k}^2 + osc_{\mathcal{T}_k}^2(\mathcal{T}_k))$, where N shall be produced in the proof of (5.13) and μ is defined in Lemma 5.2. Because of $(u, y, p, y_d, f) \in \mathcal{A}^s$, there exists a $\mathcal{T}_{h_\epsilon} \in \mathbb{T}$ and a $(u_{h_\epsilon}, y_{h_\epsilon}, p_{h_\epsilon}) \in U_{ad}^{h_\epsilon} \times V_{h_\epsilon} \times V_{h_\epsilon}$ such that

$$\#\mathcal{T}_{h_\epsilon} - \#\mathcal{T}_{h_0} \le |(u,y,p,y_d,f)|_s^{1/s} \in \epsilon^{-1/s}, \tag{5.11}$$

and

$$\|u-u_{h_\epsilon}\|_0^2 + \|y-y_{h_\epsilon}\|_1^2 + \|p-p_{h_\epsilon}\|_1^2 + osc_{\mathcal{T}_{h_\epsilon}}^2(f,\mathcal{T}_{h_\epsilon}) + osc_{\mathcal{T}_{h_\epsilon}}^2(y_{h_\epsilon}-y_d,\mathcal{T}_{h_\epsilon}) \le \epsilon^2. \tag{5.12}$$

Let $(u_{h_*}, y_{h_*}, p_{h_*})$ be the solution of (2.8)–(2.10) with respect to $\mathcal{T}_{h_*}$, where $\mathcal{T}_{h_*} = \mathcal{T}_{h_\epsilon} \oplus \mathcal{T}_{h_k}$ is the smallest common refinement of $\mathcal{T}_{h_\epsilon}$ and $\mathcal{T}_{h_k}$. In the following we will prove the following inequality firstly

$$e_{h_*}^2 + osc^2_{\mathcal{T}_{h_*}}(\mathcal{T}_{h_*}) \le N\Big(e^2_{\mathcal{T}_{h_\epsilon}} + osc^2_{\mathcal{T}_{h_\epsilon}}(\mathcal{T}_{h_\epsilon})\Big), \tag{5.13}$$

where

$$\begin{aligned} e^2_{\mathcal{T}_{h_\epsilon}} &:= \|u - u_{h_\epsilon}\|_0^2 + \|y - y_{h_\epsilon}\|_1^2 + \|p - p_{h_\epsilon}\|_1^2, \\ osc^2_{\mathcal{T}_{h_\epsilon}}(\mathcal{T}_{h_\epsilon}) &:= osc^2_{\mathcal{T}_{h_\epsilon}}(f, \mathcal{T}_{h_\epsilon}) + osc^2_{\mathcal{T}_{h_\epsilon}}(y_{h_\epsilon} - y_d, \mathcal{T}_{h_\epsilon}). \end{aligned}$$

According to the principle of adding one item and subtracting one item, then we have

$$\begin{aligned} \|u - u_{h_\epsilon}\|_0^2 &= \|u - u_{h_*}\|_0^2 + \|u_{h_\epsilon} - u_{h_*}\|_0^2 + 2(u - u_{h_*}, u_{h_*} - u_{h_\epsilon}), \\ \|y - y_{h_\epsilon}\|_1^2 &= \|y - y_{h_*}\|_1^2 + \|y_{h_\epsilon} - y_{h_*}\|_1^2 + 2a(y - y_{h_*}, y_{h_*} - y_{h_\epsilon}), \\ \|p - p_{h_\epsilon}\|_1^2 &= \|p - p_{h_*}\|_1^2 + \|p_{h_\epsilon} - p_{h_*}\|_1^2 + 2a(p - p_{h_*}, p_{h_*} - p_{h_\epsilon}). \end{aligned}$$

With the help of the Young's inequality one obtains

$$\begin{aligned} (u - u_{h_*}, u_{h_*} - u_{h_\epsilon}) &= (u - u_{h_\epsilon}, u_{h_*} - u_{h_\epsilon}) - (u_{h_*} - u_{h_\epsilon}, u_{h_*} - u_{h_\epsilon}) \\ &\le (u - u_{h_\epsilon}, u_{h_*} - u_{h_\epsilon}) \\ &\le \|u - u_{h_\epsilon}\|_0^2 + \frac{1}{4}\|u_{h_*} - u_{h_\epsilon}\|_0^2, \end{aligned}$$

and so are $a(y - y_{h_*}, y_{h_*} - y_{h_\epsilon})$ and $a(p - p_{h_*}, p_{h_*} - p_{h_\epsilon})$. Hence in connection with what we get above we deduce that

$$\begin{aligned} &\|u - u_{h_*}\|_0^2 + \|u_{h_*} - u_{h_\epsilon}\|_0^2 + \|y - y_{h_*}\|_a^2 + \|y_{h_*} - y_{h_\epsilon}\|_1^2 + \|p - p_{h_*}\|_1^2 + \|p_{h_*} - p_{h_\epsilon}\|_1^2 \\ \le\ & 6\Big(\|u - u_{h_\epsilon}\|_0^2 + \|y - y_{h_\epsilon}\|_1^2 + \|p - p_{h_\epsilon}\|_1^2\Big). \end{aligned} \tag{5.14}$$

From Remark 2.1 in [4] and (4.14) in Lemma 4.3 with $\mathcal{T}_h = \tilde{\mathcal{T}}_h = \mathcal{T}_{h_*},\ y = y_{h_*}$ and $\tilde{y} = y_{h_\epsilon}$, we obtain that

$$\begin{aligned} osc^2_{\mathcal{T}_{h_*}}(y_{h_*} - y_d, \mathcal{T}_{h_*}) - 2osc^2_{\mathcal{T}_{h_\epsilon}}(y_{h_\epsilon} - y_d, \mathcal{T}_{h_\epsilon}) &\le osc^2_{\mathcal{T}_{h_*}}(y_{h_*} - y_d, \mathcal{T}_{h_*}) - 2osc^2_{\mathcal{T}_{h_*}}(y_{h_\epsilon} - y_d, \mathcal{T}_{h_*}) \\ &\le 2Nh_0^4\|y_{h_*} - y_{h_\epsilon}\|_1^2. \end{aligned} \tag{5.15}$$

For any $T' \in \mathcal{T}_{h_\epsilon}$, let $\mathcal{T}_{h_{T'}} := \{T \in \mathcal{T}_{h_*} : T \in T'\}$. From the proof of Lemma 4.3 in [19], we derive that

$$\sum_{T \in \mathcal{T}_{h_{T'}}} \|f - f_T\|_{0,T}^2 \le N\|f - f_{T'}\|_{0,T'}^2.$$

And then we can get

$$osc^2_{\mathcal{T}_*}(f, \mathcal{T}_*) \le N(osc^2_{\mathcal{T}_\epsilon}(f, \mathcal{T}_\epsilon)). \tag{5.16}$$

Combining (5.14)–(5.16) to obtain (5.13) and using (5.12) and the definition of ϵ^2, we have

$$e_{h_*}^2 + osc^2_{\mathcal{T}_{h_*}}(\mathcal{T}_{h_*}) \le N\Big(e^2_{\mathcal{T}_{h_\epsilon}} + osc^2_{\mathcal{T}_{h_\epsilon}}(\mathcal{T}_{h_\epsilon})\Big) \le N\epsilon^2 = \mu\Big(e^2_{h_k} + osc^2_{\mathcal{T}_{h_k}}(\mathcal{T}_{h_k})\Big).$$

It is true for the following result from Lemma 5.2

$$\#\mathcal{M}_{h_k} \le \#\mathcal{R}_h \le \#\mathcal{T}_{h_*} - \#\mathcal{T}_{h_k} \le \#\mathcal{T}_{h_\epsilon} - \#\mathcal{T}_{h_0}. \tag{5.17}$$

Combining (5.11), (5.17) and the definition of ϵ^2 to derive the desired result (5.10). □

5.3. Quasi-optimality

The following consequence is the result of previous estimates where the truth is that the number of elements is a dwindle for the errors. Namely, if a given adaptive method is used to approximate the exact solution at a certain convergence rate, the iteratively constructed grids sequence will achieve this rate until a constant factor.

Theorem 5.1. *Let (u, y, p) and $(\mathcal{T}_{h_k}, U_{ad}^{h_k}, V_{h_k}, u_{h_k}, y_{h_k}, p_{h_k})$ be the solution of (2.3)–(2.5) and the sequence of grids, finite element spaces and discrete solutions produced by Algorithm 3.1, respectively. Assume that $\mathcal{T}_{h_0}$ satisfies the condition (b) of Section 4 in [27]. Let $(u, y, p, y_d, f) \in \mathcal{A}_s$, then we have*

$$\#\mathcal{T}_{h_k} - \#\mathcal{T}_{h_0} \leq C|(u, y, p, y_d, f)|_s^{\frac{1}{s}} \Big(e_{h_k}^2 + osc_{\mathcal{T}_{h_k}}^2(\mathcal{T}_{h_k})\Big)^{-\frac{1}{2s}}, \tag{5.18}$$

provided $h_0 \ll 1$.

Proof. It follows from Theorem 2.1 and Lemma 5.3 that

$$e_{h_k}^2 + \gamma_1 \eta_{1,\mathcal{T}_{h_k}}^2(p_{h_k}, \mathcal{T}_{h_k}) + \gamma_2 \tilde{\eta}_{\mathcal{T}_{h_k}}(\mathcal{T}_{h_k}) \approx e_{h_k}^2 + osc_{\mathcal{T}_{h_k}}^2(\mathcal{T}_{h_k}). \tag{5.19}$$

From Lemma 2.3 in [4], we have

$$\#\mathcal{T}_{h_k} - \#\mathcal{T}_{h_0} \leq C \sum_{i=0}^{k-1} \mathcal{M}_{h_i}. \tag{5.20}$$

With the assistance of Lemma 5.3 and (5.20) to gain

$$\#\mathcal{T}_{h_k} - \#\mathcal{T}_{h_0} \leq C \sum_{i=0}^{k-1} \mathcal{M}_{h_i} \leq C\Big(M \sum_{i=0}^{k-1} \Big(e_{h_k}^2 + osc_{\mathcal{T}_{h_k}}^2(\mathcal{T}_{h_k})\Big)^{-\frac{1}{2s}}\Big), \tag{5.21}$$

where

$$M = N^{\frac{1}{2s}} |(u, y, p, y_d, f)|_s^{\frac{1}{s}} \alpha^{-\frac{1}{2s}}.$$

Then it follows from (5.19), (5.21) and Theorem 4.1 that

$$\begin{aligned}\#\mathcal{T}_{h_k} - \#\mathcal{T}_{h_0} &\leq C\Big(M\Big(e_{h_k}^2 + osc_{\mathcal{T}_{h_k}}^2(\mathcal{T}_{h_k})\Big)^{-\frac{1}{2s}} \sum_{i=1}^{k} \alpha^{-\frac{1}{2s}}\Big) \\ &\leq C|(u, y, p, y_d, f)|_s^{\frac{1}{s}} \Big(e_{h_k}^2 + osc_{\mathcal{T}_{h_k}}^2(\mathcal{T}_{h_k})\Big)^{-\frac{1}{2s}},\end{aligned}$$

which tells the proof of Theorem 5.1. □

6. Numerical experiments

In this section, we firstly present an adaptive finite element and then give the adaptive iteration method where the purpose is to provide empirical analysis for our theory.

Example 1. *We consider the nonlinear optimal control problem governed by nonlinear elliptic equations subject to the state equation*

$$-\Delta y + y^3 = f + u, \quad -\Delta p + 3y^2 p = y - y_d,$$

where we choose $\alpha = 1$ *and* $\Omega = [0,1] \times [0,1]$ *and apparently exact solution*

$$u = \frac{1}{\alpha}(\max(0, \bar{p}) - p),$$
$$y = \sin(\pi x_1) + \sin(\pi x_2),$$
$$p = -y.$$

By simple calculation we have $\int_\Omega p dx = -\frac{4}{\pi}$ *which satisfies* $u \in U_{ad}$.

We choose 15 adaptive loops for Example 1, then we plot the profiles of the exact state and the numerical state on adaptively refined grids with $\theta = 0.5$ in Figure 1. It is easy to see that the solution is smooth, but we can find larger gradients in some regions, hence comparing with uniform refinement, adaptive finite element method can provide smaller error. In Figure 2, we provide the triangle refined grids after 6 and 12 adaptive iterations of Algorithm 3.1.

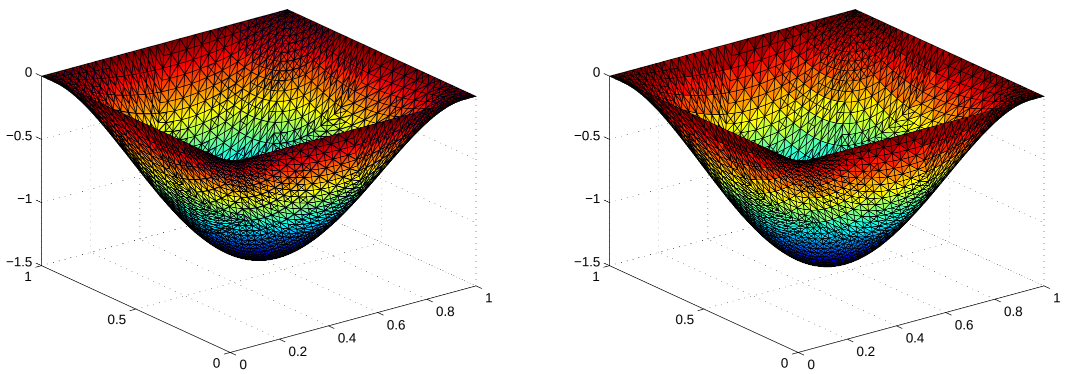

Figure 1. The exact state (left) and the numerical state (right) for Example 1.

Figure 2. The adaptive grids after 6 steps (left) and 12 steps (right) for Example 1.

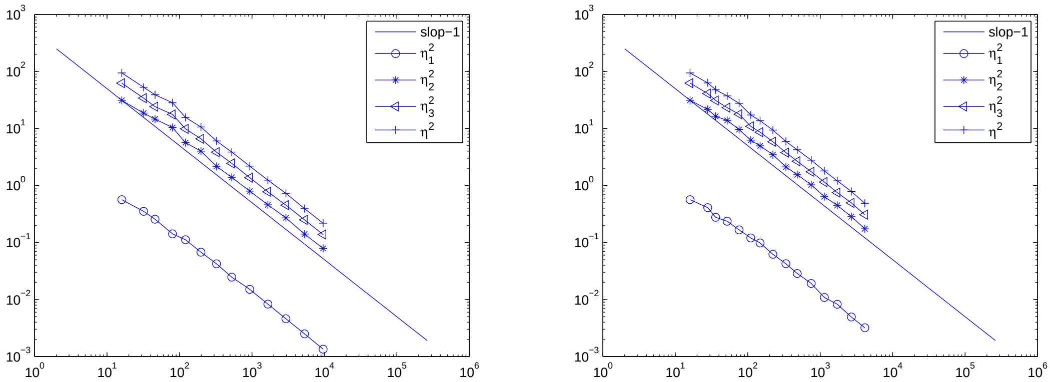

Figure 3. The error estimate compares between adaptively (left) and uniformly (right) refined grids for Example 1.

In Figure 3, we plot the convergence history for the errors where the left is adaptive refinement ($\theta = 0.5$) and the right is uniform refinement ($\theta = 1$). We can find it very intuitively when we provide the optimal convergence rate slope -1 via adopting the linear finite elements, the error reduction can be observed.

Example 2. *We consider the same nonlinear optimal control problem as Example 1 with* $\alpha = 0.1$*,* $\Omega = (-1, 1) \times (0, 1) \cup (-1, 0) \times (-1, 0]$ *and apparently exact solution*

$$p = \begin{cases} -5 \times 10^{10} e^{\frac{1}{m}}, & m < 0, \\ 0, & m \geq 0, \end{cases}$$
$$y = -p,$$
$$m = (x_1 - 0.2)^2 + (x_2 - 0.6)^2 - 0.04,$$

where $u \in U_{ad}$ *can be guaranteed.*

Comparing with Example 1, we provide some plots concerning with 21 adaptive loops for Example 2 with $\theta = 0.5$. In Figure 4, it is more easy to say that the solution is smooth while the large gradients can be found in some regions illustrating that adaptive refinement can obtain smaller errors than the uniformly refinement where we offer the error estimate graphs to explain. In Figure 5, the left plot tells us the error estimates on adaptive refinement ($\theta = 0.5$) and the right shows the error estimates on uniform refinement ($\theta = 1$). With the slope -1 being the optimal convergence rate expected, we see the error reduction from Figure 5. Meanwhile we can also find that the convergence order of the total-error and the error estimate indicators are approaching to straight line slope -1 in which they are roughly parallel where it is showed that the posteriori error estimates we obtained in Section 2 are reliable.

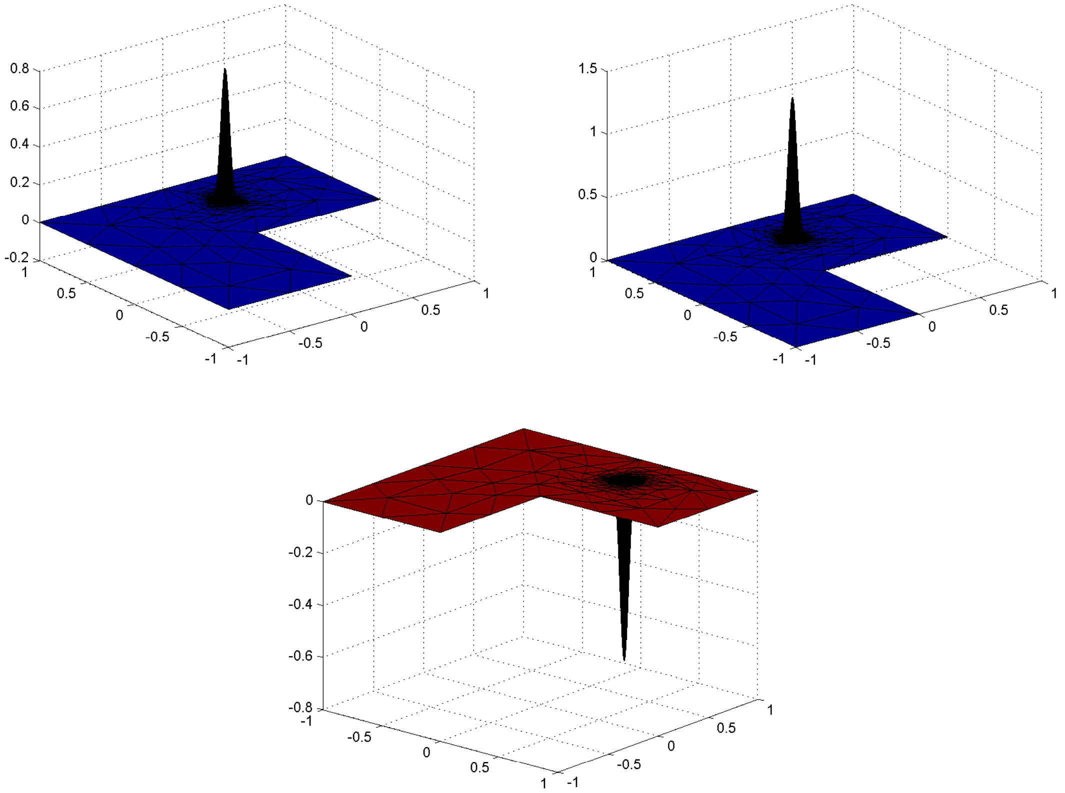

Figure 4. The exact state (upper-middle) and the numerical state (left) and the adjoint state (right) variables for Example 2.

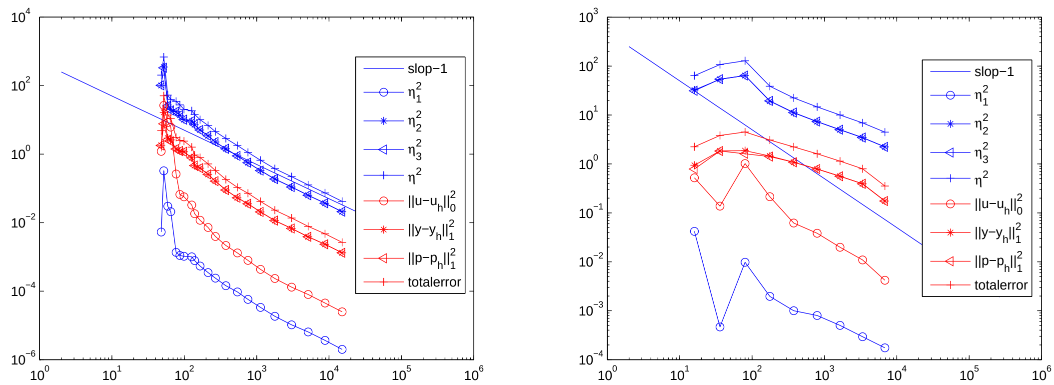

Figure 5. The error estimate compares between adaptively (left) and uniformly (right) refined grids for Example 2.

In Figure 6, we show the adaptive grids after 9 and 19 adaptive iterations for Example 2 of 21 adaptive loops with $\theta = 0.5$. We can find that the grids are concentrated on the regions where the solutions have larger gradients. Only can we note that reduced orders are observed for the uniform refinement because of the singularity of the solutions.

Figure 6. The adaptive grids after 9 steps (left) and 19 steps (right) for Example 2.

7. Conclusions and future expectation

In this paper, we first study the adaptive finite element method for nonlinear optimal control problems and give the corresponding adaptive algorithm. To evaluate the adaptive finite element method, we obtain the a posteriori error estimates for the nonlinear elliptic equations with upper and lower bound convergence and optimality, which are also important indicators for evaluating the algorithms. Therefore, we prove that the sum of the posterior errors of the control, state and covariance variables are convergent, as shown in Theorem 4.2. Based on the local upper bound, we prove the quasi-optimality of the proposed adaptive algorithm, see Section 5. To verify our theoretical analysis, we finally provide some numerical simulations. In previous research papers, the finite element methods for linear optimal control problems were studied. Our innovation is to extend the method of linear optimal control problems to a series of nonlinear optimal control problems.

There are a lot of problems which can not be tackled, such as the $L^2 - L^2$ posteriori error estimates for nonlinear elliptic equations as well as the convergence and quasi-optimality for nonlinear parabolic equations. Furthermore, we note that the analysis in this paper can be generalized to common nonlinear parabolic problems and boundary problems, and we will work on these problems.

Acknowledgments

This work is supported by National Science Foundation of China (11201510), National Social Science Fund of China (19BGL190), Chongqing Research Program of Basic Research and Frontier Technology (cstc2019jcyj-msxmX0280), General project of Chongqing Natural Science Foundation (cstc2021jcyj-msxmX0949), Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJZD-K20200120), Chongqing Key Laboratory of Water Environment Evolution and Pollution Control in Three Gorges Reservoir Area (WEPKL2018YB-04), Science and technology research project of Chongqing Education Commission (KJQN202101212), Research Center for Sustainable Development of Three Gorges Reservoir Area (2019sxxyjd07), Guangdong Basic and Applied Basic Research Foundation of Joint Fund Project (2021A1515111048), and Guangdong Province Characteristic Innovation Project (2021WTSCX120).

Conflict of interest

The authors declare that they have no competing interests.

References

1. D. Braess, C. Carstensen, R. Hoppe, Convergence analysis of a conforming adaptive finite element method for an obstacle problem, *Numer. Math.*, **107** (2007), 455–471. http://doi.org/10.1007/s00211-007-0098-6
2. P. Binev, W. Dahmen, R. Devore, Adaptive finite element methods with convergence rates, *Numer. Math.*, **97** (2004), 219–268. https://doi.org/10.1007/s00211-003-0492-7
3. I. Babuška, W. Rheinboldt, Error estimates for adaptive finite computations, *SIAM J. Numer. Anal.*, **15** (1978), 736–754. https://doi.org/10.1137/0715049
4. J. Cascon, C. Kreuzer, R. Nochetto, K. Siebert, Qusi-optimal convergence rate for an adaptive finite element method, *SIAM J. Numer. Anal.*, **46** (2008), 2524–2550. https://doi.org/10.1137/07069047X
5. Y. Chen, Z. Lu, R. Guo, Error estimates of triangular mixed finite element methods for quasilinear optimal control problems, *Front. Math. China*, **7** (2012), 397–413. https://doi.org/10.1007/s11464-012-0179-4
6. P. Ciarlet, *The finite element method for elliptic problems*, Amsterdam: North-Holland, 1978.
7. A. Demlow, R. Stevenson, Convergence and quasi-optimlity of an adaptive finite element method for controlling L^2 errors, *Numer. Math.*, **117** (2011), 185–218. https://doi.org/10.1007/s00211-010-0349-9
8. M. Dobrowolski, R. Rannacher, Finite element methods for nonlinear elliptic systems of second order, *Math. Nach.*, **94** (1980), 155–172. https://doi.org/10.1002/mana.19800940112
9. W. Dörfler, A convergent adaptive algorithm for Poisson equation, *SIAM J. Numer. Anal.*, **33** (1996), 1106–1124. https://doi.org/10.1137/0733054
10. X. Dai, J. Xu, A. Zhou, Convergence and optimal complexity of adaptive finite element eigenvalue computations, *Numer. Math.*, **110** (2008), 313–355. https://doi.org/10.1007/s00211-008-0169-3
11. K. Eriksson, C. Johnson, An adaptive finite element method for linear elliptic problems, *Math. Comp.*, **50** (1988), 361–383. https://doi.org/10.1090/S0025-5718-1988-0929542-X
12. K. Eriksson, C. Johnson, Adaptive finite element methods for parabolic problems IV: nonlinear problems, *SIAM J. Numer. Anal.*, **32** (1995), 1729–1749. https://doi.org/10.1137/0732078
13. A. Gaevskaya, R. Hoppe, Y. llish, M. Kieweg, Convergence analysis of an adaptive finite element method for distributed control problems with control constrains, In: *Control of coupled partial differential equations*, Basel: Birkhäuser, 2007, 47–68. https://doi.org/10.1007/978-3-7643-7721-2_3
14. L. Ge, W. Liu, D. Yang, L^2 norm equivalent a posteriori error estimate for a constrained optimal control problem, *Int. J. Numer. Anal. Mod.*, **6** (2009), 335–353.

15. L. Ge, W. Liu, D. Yang, Adaptive finite element approximation for a constrained optimal control problem via multi-grids, *J. Sci. Comput.*, **41** (2009), 238. https://doi.org/10.1007/s10915-009-9296-y

16. W. Gong, N. Yan, Adaptive finite element method for elliptic optimal control problems: convergence and optimality, *Numer. Math.*, **135** (2017), 1121–1170. https://doi.org/10.1007/S00211-016-0827-9

17. J. Hu, J. Xu, Convergence and optimality of the adaptive nonconforming linear element method for the stokes problem, *J. Sci. Comput.*, **55** (2013), 125–148. http://doi.org/10.1007/s10915-012-9625-4

18. H. Leng, Y. Chen, Convergence and quasi-optimality of an adaptive finite element method for optimal control problems on L^2-errors, *J. Sci. Comput.*, **73** (2017), 438–458. https://doi.org/10.1007/s10915-017-0425-8

19. H. Leng, Y. Chen, Convergence and quasi-optimality of an adaptive finite element method for optimal control problems with integral control constraint, *Adv. Comput. Math.*, **44** (2018), 367–394. https://doi.org/10.1007/s10444-017-9546-8

20. J. Lions, *Optimal control of systems governed by partial differential equations*, Berlin: Springer, 1971.

21. R. Li, W. Liu, H. Ma, T. Tang, Adaptive finite element approximation for distributed elliptic optimal control problems, *SIAM J. Control Optim.*, **41** (2002), 1321–1349. https://doi.org/10.1137/S0363012901389342

22. W. Liu, N. Yan, *Adaptive finite element methods for optimal control governed by PDEs*, Beijing: Science Press, 2008.

23. Z. Lu, Y. Chen, W. Zheng, A posteriori error estimates of lowest order Raviart-Thomas mixed finite element methods for bilinear optimal control problems, *East Asian J. Appl. Math.*, **2** (2012), 108–125. https://doi.org/10.4208/eajam.080212.260312a

24. Z. Lu, Existence and uniqueness of second order parabolic bilinear optimal control problems, *Lobachevskii J. Math.*, **32** (2011), 320–327. https://doi.org/10.1134/S1995080211040135

25. P. Morin, R. Nochetto, K. Siebert, Data oscillation and convergence of adaptive FEM, *SIAM J. Numer. Anal.*, **38** (2000), 466–488. https://doi.org/10.1137/S0036142999360044

26. W. Shen, D. Yang, W. Liu, Optimal control problem governed by a linear hyperbolic integro-differential equation and its finite element analysis, *Bound. Value Probl.*, **2014** (2014), 173. https://doi.org/10.1186/s13661-014-0173-8

27. R. Stevenson, Optimality of a standard adaptive finite element method, *Found. Comput. Math.*, **7** (2007), 245–269. https://doi.org/10.1007/s10208-005-0183-0

28. R. Verfürth, *A review of a posteriori error estimation and adaptive Grids-Refinement techniques*, Wiley, 1998.

29. P. Wriggers, O. Scherf, An adaptive finite element technique for nonlinear contact problems, In: *Contact mechanics*, Boston, MA: Springer, 1995, 183–194. https://doi.org/10.1007/978-1-4615-1983-6_24

30. J. Xu, A. Zhou, Local and parallel finite element algorithms based on two-grids discretizations, *Adv. Comput. Math.*, **14** (2001), 293–327. https://doi.org/10.1023/A:1012284322811

31. D. Yang, A priori error estimate and superconvergence analysis for an optimal control problem of bilinear type, *J. Comput. Math.*, **26** (2008), 471–487.

AIMS Press