Computational Methods in Applied Mathematics
Vol. 11 (2011), No. 2, pp. 107–128

Quasi-optimality of an Adaptive Finite Element Method for an Optimal Control Problem

Roland Becker · Shipeng Mao

Abstract — We prove quasi-optimality of an adaptive finite element algorithm for a model problem of optimal control including control constraints. The quasi-optimility expresses the fact that the decrease of error with respect to the number of mesh cells is optimal up to a constant. The considered algorithm is based on an adaptive marking strategy which compares a standard residual-type a posteriori error estimator with a data approximation term in each step of the algorithm in order to adapt the marking of cells.

2010 Mathematical subject classification: 49J20; 49M15; 49N10; 65K10; 65N12; 65N15; 65N30.

Keywords: Optimal control; adaptive finite elements; convergence of adaptive algorithms; complexity estimates.

Introduction

We consider a model problem of optimal control, the solution of which is approximated by an adaptive finite element method. The analysis of adaptive finite element methods has made important progress in recent years. Based on classical residual-based a posteriori error estimators [2, 11, 21] it has been shown by Dörfler [10] and Morin, Nochetto, and Siebert [18] that an adaptive mesh refinement algorithm converges towards the solution of the Poisson equation. An important further result is the estimation of the dimension of the adaptively constructed discrete spaces by Binev, Dahmen, and DeVore in [5], and Stevenson [20]. The importance of these contributions lays in the fact that they prove quasi-optimality in the following sense: if the solution of the problem can be approximated by the given discretization method on a given family of meshes at a certain rate, the iteratively constructed sequence of meshes will realize this rate up to a constant factor.

In this work, we present an adaptive finite element method for an optimal control problem. Our approach is based on continuous finite elements of fixed degree on locally refined triangular meshes. Following the idea of [4], we use an adaptive marking strategy which either performs the refinement according to a standard residual-type estimator or according

Roland Becker
Laboratoire de Mathématiques Appliquées and EPI Concha INRIA Bordeaux Sud-Ouest, Université de Pau, 64013 Pau Cedex, France
E-mail: `roland.becker@univ-pau.fr`.

Shipeng Mao
LSEC, Institute of Computational Mathematics and EPI Concha INRIA Bordeaux Sud-Ouest, Chinese Academy of Sciences (CAS), Beijing, 100190, PR China
E-mail: `maosp@lsec.cc.ac.cn`.

to a data approximation term. From a computational point of view, the resulting algorithm is simpler than the MNS algorithm [18], which is underlaying the work of [12], where convergence for an optimal control problem similar to the one considered here is shown. Improving upon the known results in the literature, we prove convergence and optimal complexity of the adaptive algorithm without assuming the interior node property of the local refinement algorithm, as is done in [12], for example. The optimality proof, which seems to be new for optimal control problems, is our main contribution, since it gives a rigorous justification of adaptive finite element algorithms in comparison with non-adaptive methods.

The paper is organized as follows: After introduction of the optimal control problem and its discretization on a single mesh in Sections 1 and 2, in Section 3 we define the adaptive algorithm. In Section 4 we establish some lemmata, which are used later on. Section 5 is devoted to the proof of geometrical convergence of the error of the adaptive algorithm under natural assumptions. In Section 6 we prove an asymptotic estimate for the complexity of the sequence of meshes. Finally, we report on some numerical experiments in Section 7.

Throughout the paper we use the following notation. For the norm of the standard Sobolev space $H_0^1(\Omega)$ we write $|u|_1 := \left(\int_\Omega |\nabla u|^2\,dx\right)^{1/2}$. The $L^2(A)$-scalar product and norm are denoted by $\langle\cdot,\cdot\rangle_A$ and $\|\cdot\|_A$, respectively, omitting the subscript in case $A=\Omega$, for either a measurable subset $A\subset\Omega$ or for an edge of a finite element mesh (with obvious modification of the measure).

We work with families of shape regular triangular meshes in the sense of [8]. We denote by h a mesh in a family of admissible meshes $\mathcal{H}$, and by u_h the corresponding finite element solution. The set of cells of mesh h is denoted by $\mathcal{K}_h$, the diameter of $K\in\mathcal{K}_h$ is denoted by ρ_K, and in addition we define $\rho(h) := \max_{K\in\mathcal{K}_h}\rho_K$. As compared to standard notation in finite element literature, h denotes a mesh in a family of meshes $\mathcal{H}$ and *not* a global maximal cell width.

1. The optimal control problem

Let $\Omega\subset\mathbb{R}^2$, be a bounded domain with polygonal boundary $\partial\Omega$. Let $\Omega_B\subset\Omega$ and $\Omega_C\subset\Omega$ be polygonal subdomains. Further let $f\in L^2(\Omega)$ and $\alpha>0$ be given. We consider the following optimization problem:

$$\begin{cases} \displaystyle\inf_{q\in L^2(\Omega_B),u\in H_0^1(\Omega)} \frac{\alpha}{2}\|q\|_{\Omega_B}^2+\frac{1}{2}\|u\|_{\Omega_C}^2 \quad \text{subject to:}\\ -\Delta u = f+q \quad \text{in } \Omega_B, \quad -\Delta u = f \quad \text{in } \Omega\setminus\Omega_B, \quad u=0 \quad \text{on } \partial\Omega,\\ q\geqslant 0 \quad \text{a.e. } \Omega_B. \end{cases} \tag{1.1}$$

This is a linear-quadratic problem. Denoting by $B: L^2(\Omega_B)\to L^2(\Omega)$ the trivial extension operator, we may alternatively write the state equation as

$$-\Delta u = f+Bq \quad \text{in } \Omega, \quad u=0 \quad \text{on } \partial\Omega. \tag{1.2}$$

The state equation can be used to eliminate the state variable, such that we end up with the minimization of a quadratic functional in the control variable q alone. Although conceptionally important, such a formulation hides the main difficulty inherent to optimization problems containing a partial differential equation as constraint: the discretization of the state equation.

We remark that $\alpha > 0$ is necessary for well-posedness. However, in case of finite-dimensional controls (where q is sought in the linear space spanned by given functions $\psi_i \in L^2(\Omega_B), i = 1, \dots m$), $\alpha = 0$ may lead to a well-posed minimization problem. This case will be addressed elsewhere. Here we suppose $\alpha > 0$.

More general linear-quadratic optimal control problems involving non-zero observations $u_d \in L^2(\Omega_C)$, a reference control value $q_d \in L^2(\Omega_B)$, and inhomogenous Dirichlet boundary data can be directly reduced to (1.1). It is well-know that (1.1) admits a unique solution; for this and further results concerning the solution of (1.1) we refer to [16].

Writing $u(q)$ for the unique solution of the state equation for given control, the reduced functional is defined as

$$j(q) := \frac{\alpha}{2}\|q\|^2_{\Omega_B} + \frac{1}{2}\|u(q)\|^2_{\Omega_C}$$

We denote by $L^2_+(\Omega)$ the cone of positive square-integrable functions. Let $Q = L^2_+(\Omega_B)$ be the set of admissible controls.

In terms of the reduced functional, the optimization problem (1.1) simply reads

$$\inf_{q\in Q} j(q). \tag{1.3}$$

We note that j is quadratic since $q \mapsto u(q)$ is affine-linear. Its first- and second-order derivatives are given by

$$j'(q)(p) = \alpha\langle q, p\rangle_{\Omega_B} + \langle u(q), u'(p)\rangle_{\Omega_C},\ j''(p_1, p_2) = \alpha\langle p_1, p_2\rangle_{\Omega_B} + \langle u'(p_1), u'(p_2)\rangle_{\Omega_C}, \tag{1.4}$$

where $u'(p)$ is the solution of (1.2) with control p and $f = 0$.

We observe that $j''(p, p) \geqslant \alpha\|p\|^2_{\Omega_B}$ such that j is strictly convex and the minimization problem (1.3) admits a unique solution which is characterized by the variational inequality

$$j'(q)(p) = \langle\nabla j(q), p\rangle_{\Omega_B} \geqslant 0 \qquad \forall p \in Q. \tag{1.5}$$

Next we define the Lagrange functional by

$$\mathcal{L}(q, u, z) := \frac{\alpha}{2}\|q\|^2_{\Omega_B} + \frac{1}{2}\|u\|^2_{\Omega_C} + \langle f, z\rangle + \langle q, z\rangle_{\Omega_B} - \langle\nabla u, \nabla z\rangle.$$

The first-order necessary conditions, which we also call optimality system, is the variational system

$$\langle\nabla u, \nabla v\rangle - \langle q, v\rangle_{\Omega_B} \;=\; \langle f, v\rangle \quad \forall v \in H^1_0(\Omega), \tag{1.6}$$

$$\langle\nabla v, \nabla z\rangle - \langle u, v\rangle_{\Omega_C} \;=\; 0 \quad \forall v \in H^1_0(\Omega), \tag{1.7}$$

$$\alpha\langle q, p\rangle + \langle z, p\rangle_{\Omega_B} \;\geqslant 0 \qquad \forall p \in Q. \tag{1.8}$$

Let (q, u, z) be the solution of (1.6-1.8). In addition it holds that $\alpha\|q\|^2_{\Omega_B} + \langle z, q\rangle_{\Omega_B} = 0$. Note that equation (1.8) simply translates the inequality $j'(q)(p) = \mathcal{L}_q(q, u, z)(p) \geqslant 0$. We have that $q \geqslant 0$ and $\alpha q + z \geqslant 0$ almost everywhere. The variational inequality also implies that

$$\alpha q + z^- = 0, \tag{1.9}$$

with $x^+ := \max(0, x)$ and $x^- := x - x^+$. We can use (1.9) in order to eliminate the control variable from the system, leading to the nonlinear system of partial differential equations

$$-\Delta u = f - \alpha^{-1}\chi_{\Omega_B} z^-, \quad -\Delta z = \chi_{\Omega_C} u, \tag{1.10}$$

which can be written after the rescaling $\alpha^{-1}z \to z$ as the notationally more convenient system

$$-\Delta u = f - \chi_{\Omega_B} z^-, \quad -\Delta z = \alpha^{-1}\chi_{\Omega_C} u. \tag{1.11}$$

In order to simplify the notation, we consider in the following the reduced system (1.11). See also [15], where the reduced system is used to derive a discrete control space adapted to the adjoint variable via (1.9). Our results carry immediately over to this approach, but the generalization to discretiztaion of (1.6-1.8) is also possible.

In order to simplify the exposition, we introduce some notation. Let $V = H_0^1(\Omega)$ and define $a_0 : V \times V \to \mathbb{R}$ and $a : V \times V \to \mathbb{R}$ by

$$\begin{aligned} a_0((u,z),(v,w)) &:= \langle \nabla u, \nabla v\rangle + \langle \nabla w, \nabla z\rangle + \langle z, v\rangle_{\Omega_B} - \alpha^{-1}\langle u, w\rangle_{\Omega_C}, \\ a((u,z),(v,w)) &:= a_0((u,z),(v,w)) - \langle z^+, v\rangle_{\Omega_B}. \end{aligned} \tag{1.12}$$

Then the variational problem read corresponding to (1.11) reads: Find $(u,z) \in V \times V$ such that for all $(v,w) \in V \times V$

$$a((u,z),(v,w)) := \langle f, v\rangle. \tag{1.13}$$

Let us set for abbreviation

$$|||(u,z)||| = \sqrt{|u|_1^2 + |z|_1^2}. \tag{1.14}$$

The bilinear form a_0 has the following stability property.

Lemma 1.1. *There is a constant $\gamma_{\rm is} > 0$ such that*

$$\sup_{(v,w)\in V\setminus\{0\}\times V\setminus\{0\}} \frac{a_0((u,z),(v,w))}{|||(v,w)|||} \geqslant \gamma_{\rm is} |||(u,z)|||. \tag{1.15}$$

Proof. Testing with $(v,w) = (u,z)$ gives

$$\begin{aligned} a_0((u,z),(v,w)) &= |u|_1^2 + |z|_1^2 + \langle z, u\rangle_{\Omega_B} - \alpha^{-1}\langle u, z\rangle_{\Omega_C} \\ &\geqslant |||(u,z)|||^2 - \left(\|z\|_{\Omega_B}\|u\|_{\Omega_B} + \alpha^{-1}\|z\|_{\Omega_C}\|u\|_{\Omega_C}\right) \\ &\geqslant |||(u,z)|||^2 - \frac{1}{2\varepsilon}\|z\|_{\Omega_B}^2 - \frac{\varepsilon}{2}\|u\|^2 - \frac{\varepsilon}{2}\|z\|^2 - \frac{1}{2\varepsilon\alpha}\|u\|_{\Omega_C}^2. \end{aligned}$$

With the Poincaré inequality

$$\|z\|^2 + \|u\|^2 \leqslant C_\Omega\left(|u|_1 + |z|_1\right)$$

it follows with $\varepsilon = 1/C_\Omega$ that

$$a_0((u,z),(v,w)) \geqslant \frac{1}{2}|||(u,z)|||^2 - \frac{1}{2\varepsilon}\|z\|_{\Omega_B}^2 - \frac{1}{2\alpha\varepsilon}\|u\|_{\Omega_C}^2$$

On the other hand, testing with $(v,w) = (z,-u)$ leads to

$$a_0((u,z),(v,w)) = \|z\|_{\Omega_B}^2 + \alpha^{-1}\|u\|_{\Omega_C}^2.$$

Therefore, choosing $(v,w) = (u,z) + \frac{1}{2\varepsilon}(z,-u)$, we have $|||(v,w)||| \leqslant C\,|||(u,z)|||$ and

$$a_0((u,z),(v,w)) \geqslant \frac{1}{2}|||(u,z)|||^2.$$

The result follows with $\gamma_{\rm is} = 1/(2C)$. □

2. Discretization of the optimal control problem

In this Section, we consider the discretization of the optimal control problem on a fixed triangular mesh $h \in \mathcal{H}$. Let h be a shape-regular partition of Ω into triangles satisfying the standard assumptions [8]. We suppose for simplicity that $\bigcup_{K\in\mathcal{K}_h} K$ covers Ω accurately, and that there exist subsets $\mathcal{K}_h^B \subset \mathcal{K}_h$ and $\mathcal{K}_h^C \subset \mathcal{K}_h$ such that $\bigcup_{K\in\mathcal{K}_h^B} \overline{K} = \overline{\Omega_B}$ and $\bigcup_{K\in\mathcal{K}_h^C} \overline{K} = \overline{\Omega_C}$, respectively. Finally, we denote by N_h the number of cells of mesh $h \in \mathcal{H}$.

The finite-element spaces of $V_h \subset H_0^1(\Omega)$ with $k \geqslant 1$ is defined in standard way

$$V_h \quad := \quad \{v \in H_0^1(\Omega) : \ v|_K \in P^k \text{ for all } K \in \mathcal{K}_h\}.$$

We denote by π_h the L^2-projection on the discrete space of discontinuous piecewise polynomials of order k. For $K \in \mathcal{K}_h$ let $\pi_K u := (\pi_h u)|_K$ such that $\langle u - \pi_K u, w\rangle_K = 0$ for all polynomials w of maximal order k.

Next we introduce the discrete system to be solved : Find $(u_h, z_h) \in V_h \times V_h$ such that

$$a((u_h, z_h), (v_h, w_h)) = \langle f, v_h\rangle \quad \forall (v_h, w_h) \in V_h \times V_h. \tag{2.1}$$

The corresponding discrete control q_h is obtained by $q_h := -z_h^-$, which is not a member of V_h. This approach of indirectly discretizing the control variable has been used in [15] in order to derive a priori error estimates. If we use instead a discretization of the control space by piecewise constants, we would end up with the relation $q_h = -(\pi_h^0 z_h)^-$, where π_h^0 denotes the projection on the piecewise constants, instead.

We solve the coupled system of equations (2.1) by a semi-smooth Newton method [13], based on the propertied of the negative-part-function, combined with a multigrid algorithm for the linear problems arising in each iteration.

Throughout the rest of this paper, we assume that for the solution to the Poisson equation in Ω, the regularity shift $L^2(\Omega) \to H^{1+\tau}(\Omega)$ holds for some τ with $0 < \tau \leqslant 1$. It is well-known that $\tau = 1/2$ for all Lipschitz domains, see for example [19]. This implies that there exists a constant $C > 0$, depending only on the shape regularity of the triangulations and f, such that for $\rho(h)$ denoting the maximal diameter of the elements in mesh h, we have the following relation between the L^2-norm and H^1-norm,

$$\|u - u_h\|^2 + \|z - z_h\|^2 \leqslant C\rho^{2\tau}(h)(|u - u_h|_1^2 + |z - z_h|_1^2). \tag{2.2}$$

We will give a detailed proof of this result in the appendix.

3. Definition of the adaptive algorithm

We define the family of admissible meshes $\mathcal{H}$ in the following recursive way. Starting from an initial mesh h_0, we denote by $\mathcal{R}_{loc}(h, \mathcal{M})$ with $\mathcal{M} \subset \mathcal{K}_h$ the mesh resulting from a local mesh refinement algorithm which subdivides triangles in such a way that at least any edge belonging to a cell of $\mathcal{M}$ is bisected. We make the following assumption concerning the refinement algorithm.

Assumption 3.1. *Let $h_k, k = 0, \dots n$ be a sequence of locally refined triangulations created by the local refinement algorithm, starting from the initial mesh h_0. Let $\mathcal{M}_k \subset \mathcal{K}_{h_k}, k =$*

$0,\dots n-1$ *be the set of marked cells in step* k *and set* $N_k = \#\mathcal{K}_{h_k}$. *Then* $\{h_k\}$ *is uniformly shape regular and we have*

$$N_n \leqslant N_0 + C\sum_{k=0}^{n-1} \#\mathcal{M}_k. \tag{3.1}$$

Assumption 3.1 and especially the complexity estimate (3.1) are known to be true for the newest vertex bisection algorithm [17], see Theorem 2.4 of [5]. It is likely to hold for other local mesh refinement algorithms.

We now define for given $h \in \mathcal{H}$, $\mathcal{M} \subset \mathcal{K}_h$, and $K \in \mathcal{K}_h$ the data approximation term

$$\mu_K := |K|^{1/2}\,\|f - \pi_K f\|_K, \quad \mu_h(\mathcal{M}) := \left(\sum_{K\in\mathcal{M}} \mu_K^2\right)^{1/2} \tag{3.2}$$

and the estimator

$$\begin{aligned}\eta_K^2 &:= \sum_{E\subset\partial K\setminus\partial\Omega} |E|\left(\|[\frac{\partial u_h}{\partial n}]\|_E^2 + \|[\frac{\partial z_h}{\partial n}]\|_E^2\right)\\ &\quad + |K|\left(\|\pi_K f + \Delta u_h - \chi_B z_h^-\|_K^2 + \alpha^{-1}\|\Delta z_h + \alpha^{-1}\chi_C u_h\|_K^2\right), \qquad (3.3)\\ \eta_h^2(\mathcal{M}) &:= \sum_{K\in\mathcal{M}} \eta_K^2.\end{aligned}$$

We set for brevity $\mu_h := \mu_h(\mathcal{K}_h)$ and $\eta_h := \eta_h(\mathcal{K}_h)$.

Remark 3.1. For linear finite elements, $k=1$, the volume term in (3.3) can be removed, see [7]. The quasi-optimality of the resulting AFEM for the Poisson equation has been shown in [3].

The purpose of this article is to analyze the following adaptive finite element algorithm. It generates sequences of meshes $\{h_k\}_k \subset \mathcal{H}$, discrete solutions $\{(u_k, z_k)\}_k$, errors $\{e_k\}_k$, estimators $\{\eta_k\}_k$, data approximation errors $\{\mu_k\}_k$, and sets of marked cells $\{\mathcal{M}_k\}_k$.

Remark 3.2. The refinement is only determined by the data approximation term if it is large compared to the estimator, following the idea of [4].

Remark 3.3. The choice of parameters can be guided by our theoretical results. For our convergence result presented below, the parameters θ, σ, and γ are arbitrary. For our proof of quasi-optimality, we will suppose that the marking parameter θ is small enough. Such an assumption is known from other complexity estimates [5, 20].

4. Estimates for the error, estimator, and data approximation term

For the purpose of the later convergence and complexity proofs, we collect in this section some lemmata, which form the basis of our convergence and complexity analysis.

We first note that the error $e_h := (u-u_h, z-z_h)$ satisfies for all $(v_h, w_h) \in V_h \times V_h$ the Galerkin relation

$$a_0((u-u_h, z-z_h), (v_h, w_h)) = \langle z^+ - z_h^+, v_h\rangle_{\Omega_B} =: S(z, z_h, v_h) \tag{4.1}$$

A similar relation holds for the difference between two successive meshes h and h', $e_{h,h'} := (u_{h'} - u_h, z_{h'} - z_h)$.

The right-hand-side of (4.1) is estimated in the first lemma of this section.

Algorithm 1 $\mathcal{AFEM}$

(0) Choose parameters $0 < \theta, \sigma < 1$, $\gamma > 0$, and an initial mesh h_0, and set $k = 0$.

(1) Solve the discrete optimization problem (2.1) with h replaced by h_k in order to obtain the finite element solutions (u_k, z_k).

(2) Compute the estimator η_k and data approximation term μ_k.

(3) – If $\mu_k^2 \leqslant \gamma\,\eta_k^2$ then find a set $\mathcal{M} \subset \mathcal{K}_{h_k}$ with minimal cardinality such that

$$\eta_k^2(\mathcal{M}) \geqslant \theta\,\eta_k^2. \tag{3.4}$$

– else find a set $\mathcal{M} \subset \mathcal{K}_{h_k}$ with minimal cardinality such that

$$\mu_k^2(\mathcal{M}) \geqslant \sigma\,\mu_k^2. \tag{3.5}$$

(4) Adapt the mesh : $h_{k+1} := \mathcal{R}_{loc}(h_k, \mathcal{M})$.

(5) Set $k := k + 1$ and go to step (1).

Lemma 4.1. *Let S be the term on the right of (4.1). Then we have for all $v \in V$*

$$|S(z, z_h, v)| \leqslant \|z - z_h\|_{\Omega_B} \|v\|_{\Omega_B}. \tag{4.2}$$

Proof. We first use the Cauchy-Schwarz inequality to bound $|S(z, z_h, v)| \leqslant \|z^+ - z_h^+\|_{\Omega_B} \|v\|_{\Omega_B}$. It then remains to show that

$$\|z^+ - z_h^+\|_{\Omega_B} \leqslant \|z - z_h\|_{\Omega_B}. \tag{4.3}$$

Define $\Omega_A := \{x \in \Omega_B : \ z(x) z_h(x) \geqslant 0\}$. Then

$$\begin{aligned} \int_{\Omega_B} |z^+ - z_h^+|^2 &= \int_{\Omega_A} |z^+ - z_h^+|^2 + \int_{\Omega_B \setminus \Omega_A} |z^+ - z_h^+|^2 \\ &\leqslant \int_{\Omega_A} |z - z_h|^2 + \int_{\Omega_B \setminus \Omega_A} |z^+ - z_h^+|^2. \end{aligned}$$

Now, for $x \in \Omega_B \setminus \Omega_A$, one of the terms $z^+(x)$ and $z_h^+(x)$ vanishes. This implies in the case $z^+(x) = 0$, that $|z^+(x) - z_h^+(x)| = |z_h(x)| \leqslant |z_h(x) - z(x)|$. Similarly, in the case $z_h^+(x) = 0$, we have by the same argument that $|z^+(x) - z_h^+(x)| \leqslant |z_h(x) - z(x)|$. It follows that $\|z^+ - z_h^+\|_{\Omega_B \setminus \Omega_A} \leqslant \|z - z_h\|_{\Omega_B \setminus \Omega_A}$, finishing the proof. □

The upper bounds used in the analysis of the adaptive algorithms are given next.

Lemma 4.2. ***(Upper bounds)*** *Under the assumption that $\rho(h_0)$ is sufficiently small, there exists a constant $C_1 > 0$ depending only on the minimum angle of h_0 such that*

$$|||e_h|||^2 \leqslant C_1 \left(\eta_h^2 + \mu_h^2\right). \tag{4.4}$$

In addition, if $\mathcal{M} \subset \mathcal{K}_h$ and $h' = \mathcal{R}_{loc}(h, \mathcal{M})$, there exists a subset $\mathcal{R} \subset \mathcal{K}_h$ with

$$|||e_{h,h'}|||^2 \leqslant C_1 \left(\eta_h^2(\mathcal{R}) + \mu_h^2(\mathcal{R}) + \rho^{2\tau}(h) |||e_h|||^2\right) \tag{4.5}$$

and

$$\#\mathcal{R} \leqslant C_3 \, \#\mathcal{M}. \tag{4.6}$$

Proof. We first prove (4.4).

First, from the stability estimate (1.15), there exists $(v,w) \in V \times V$ with $|||(v,w)||| \leqslant 1$ and

$$\gamma |||(u-u_h, z-z_h)||| \leqslant a_0((u-u_h, z-z_h),(v,w)).$$

With the help of (4.1) and Lemma 4.1, we get with arbitrary $(v_h, w_h) \in V_h \times V_h$

$$\gamma |||(u-u_h, z-z_h)||| \leqslant a_0((u-u_h, z-z_h),(v-v_h, w-w_h)) + \|z-z_h\|_{\Omega_B}\|v_h\|_{\Omega_B}.$$

Since

$$a_0((u,z),(v-v_h,w-w_h)) = a((u,z),(v-v_h,w-w_h)) + \langle z^+, v-v_h\rangle_{\Omega_B}$$

and similarly for $a_0((u_h,z_h),(v-v_h,w-w_h))$, we have

$$a_0((u-u_h,z-z_h),(v-v_h,w-w_h)) = \langle f, v-v_h\rangle - a((u_h,z_h),(v-v_h,w-w_h)) + S(z,z_h,v-v_h).$$

Therefore, we have from (4.7) and Lemma 4.1

$$\begin{aligned}\gamma_{\text{is}} |||(u-u_h, z-z_h)||| \leqslant & \langle f, v-v_h\rangle - a((u_h,z_h),(v-v_h,w-w_h)) \\ & + \|z-z_h\|_{\Omega_B}\left(\|v_h\|_{\Omega_B} + \|v-v_h\|_{\Omega_B}\right).\end{aligned} \tag{4.7}$$

We now chose $v_h = C_h v$ and $w_h = C_h w$ with the Clément interpolation operator $C_h = V \to V_h$. It verifies the interpolation estimate

$$|K|^{-1/2}\|v - C_h v\|_K + |E|^{-1/2}\|v - C_h v\|_E \leqslant C\,|v|_{1,\omega_K}$$

with ω_K denoting the set of neighboring elements of $K \in \mathcal{K}_h$ and $E \subset \partial K$. For the last term in (4.7), we have with (2.2)

$$\|z-z_h\|_{\Omega_B}\left(\|v_h\|_{\Omega_B} + \|v-v_h\|_{\Omega_B}\right) \leqslant C\rho(h)^\tau |z-z_h|_1 \leqslant C\rho(h_0)^\tau |||e_h|||. \tag{4.8}$$

For the other terms on the right-hand side of (4.7) we have

$$\begin{aligned}
&\langle f, v-v_h\rangle - a((u_h,z_h),(v-v_h,w-w_h)) = \langle f, v-C_hv\rangle - \langle \nabla u_h, \nabla(v-C_hv)\rangle \\
&\quad + \langle \nabla(w-C_hw), \nabla z_h\rangle + \langle z_h^-, (v-C_hv)\rangle_{\Omega_B} - \langle \alpha^{-1}u_h, (w-C_hw)\rangle_{\Omega_C} \\
&\leqslant C\Bigg(\sum_{K\in\mathcal{K}_h}\sum_{E\subset\partial K\setminus\partial\Omega} |E|^{1/2}\,\|[\frac{\partial u_h}{\partial n}]\|_{\partial K\setminus\partial\Omega}|v|_{1,\omega_K} \\
&\quad + \sum_{K\in\mathcal{K}_h}|K|^{1/2}\left(\|f-\pi_Kf\|_K|v|_{1,\omega_K} + \|\pi_Kf + \Delta u_h - \chi_B z_h^-\|_K|v|_{1,\omega_K}\right) \\
&\quad + \sum_{K\in\mathcal{K}_h}\sum_{E\subset\partial K\setminus\partial\Omega}|E|^{1/2}\,\|[\frac{\partial z_h}{\partial n}]\|_{\partial K\setminus\partial\Omega}||w|_{1,\omega_K} + \|\Delta z_h + \alpha^{-1}\chi_C u_h\|_K|w|_{1,\omega_K}\Bigg) \\
&\leqslant C\left(\mu_h^2 + \eta_h^2\right)^{1/2}\left(|v|_1^2 + |w|_1^2\right)^{1/2} \leqslant C\left(\mu_h^2+\eta_h^2\right)^{1/2},
\end{aligned}$$

where we have used the shape-regularity of the mesh. This leads with, together with (4.8), to

$$|||e_h|||^2 \leqslant C \left(\eta_h^2 + \mu_h^2 + \rho(h)^{2\tau} |||e_h|||^2\right). \tag{4.9}$$

By the smallness assumption on h_0 we obtain (4.4).

Now we consider (4.5). The arguments which lead to (2.2) are valid, if we replace u, z, v, and w by $u_{h'}$, $z_{h'}$, $v_{h'}$, and $w_{h'}$, respectively. Noting that $C_h v_{h'} - v_{h'}$ and $C_h w_{h'} - w_{h'}$ vanish on all cells, which are not neighbors of refined cells, we can take $\mathcal{R}$ as the set $\mathcal{M}$ augmented by all its neighbors. We then obtain

$$|||e_{h,h'}|||^2 \leqslant C \left(\eta_h^2(\mathcal{R}) + \mu_h^2(\mathcal{R}) + \|z_{h'} - z_h\|^2\right). \tag{4.10}$$

By the triangle inequality, we have $\|z_{h'} - z_h\| \leqslant \|z - z_h\| + \|z - z_{h'}\| \leqslant 2\rho(h)^\tau |||e_h|||$. This yields (4.5). □

The next lemma concerns a lower bound of the error for the optimal control system. The proof is based on standard techniques for lower bounds of elliptic equations, see [21], and is omitted here.

Lemma 4.3. ***(Lower bound)*** *There exists a constant $C_2 > 0$ depending only on the minimum angle of h_0 such that*

$$\eta_h^2 \leqslant C_2 \left(|||e_h|||^2 + \mu_h^2\right). \tag{4.11}$$

The local variant of Lemma 4.3 replacing e_h by $e_{h,h'}$ on the right-hand side of (4.11) only holds for certain type of mesh refinement algorithms. Therefore, we will use instead an estimate for the decrease of the estimator under refinement, which is given next.

Lemma 4.4. ***(Decrease of estimator and data approximation)*** *Let $h' = \mathcal{R}_{loc}(h, \mathcal{M})$. There exist constants $C_4 > 0$ and ξ with $0 < \xi < 1$ depending only on the minimum angle of h_0 such that for any $\delta > 0$*

$$\eta_{h'}^2 \leqslant (1+\delta)\eta_h^2 - \xi(1+\delta)\eta_h^2(\mathcal{M}) + C_4(1 + 1/\delta)\, |||e_{h,h'}|||^2 \tag{4.12}$$

and

$$\mu_{h'}^2 \leqslant \mu_h^2 - \xi\, \mu_h^2(\mathcal{M}). \tag{4.13}$$

Proof. The estimate (4.13) follows from our assumptions on the mesh refinement algorithm and the properties of the local L^2-projection.

The technique developed in [3] leads to

$$\begin{aligned} \sum_{K \in \mathcal{K}_{h'}} \sum_{E \subset \partial K \setminus \partial\Omega} |E|\, \|[\frac{\partial u_{h'}}{\partial n}]\|_E^2 &\leqslant (1+\delta) \sum_{K \in \mathcal{K}_h} \sum_{E \subset \partial K \setminus \partial\Omega} |E|\, \|[\frac{\partial u_h}{\partial n}]\|_E^2 \\ &- \frac{1+\delta}{2} \sum_{K \in \mathcal{M}} \sum_{E \subset \partial K \setminus \partial\Omega} |E|\, \|[\frac{\partial u_h}{\partial n}]\|_E^2 + C_4(1 + 1/\delta)\, |u_{h'} - u_h|_1^2. \end{aligned} \tag{4.14}$$

Indeed, for a new edge produced by the last mesh refinement we have

$$|E|\, \|[\frac{\partial u_{h'}}{\partial n}]\|_E^2 = |E|\, \|[\frac{\partial (u_{h'} - u_h)}{\partial n}]\|_E^2 \leqslant C\, \|\nabla(u_{h'} - u_h)\|_{\omega_E}^2,$$

where ω_E is the patch of cells containing E. Next, for an edge which has not been refined, since for all real numbers x, y and $\delta > 0$ it holds $(x+y)^2 \leqslant (1+\delta)x^2 + (1+1/\delta)y^2$, we have

$$|E|\,\|[\frac{\partial u_{h'}}{\partial n}]\|_E^2 \leqslant (1+\delta)|E|\,\|[\frac{\partial u_h}{\partial n}]\|_E^2 + (1+1/\delta)|E|\,\|[\frac{\partial (u_{h'}-u_h)}{\partial n}]\|_E^2,$$

and the last term can be estimates as before. It remains to consider the case that an edge E is bisected into E_i, $i = 1, 2$. A similar argument ad before combined with the reduction of $|E|$ under bisection leads to

$$\sum_{i=1}^{2} |E_i|\,\|[\frac{\partial u_{h'}}{\partial n}]\|_{E_i}^2 \leqslant (1+\delta - \frac{1+\delta}{2})|E|\,\|[\frac{\partial u_h}{\partial n}]\|_E^2 + C(1+1/\delta)|E|\,\|\nabla(u_{h'}-u_h)\|_{\omega_E}^2.$$

Putting together these estimates gives (4.14). The estimates are identical for the terms corresponding to the jumps of the normal derivatives of z_h.

It remains to bound the volume terms in similar way. We consider the term concerning the residual of the state equation. If $K \notin \mathcal{M}$ we have

$$\begin{aligned}|K|\|\pi_K f + \Delta u_{h'} - \chi_B z_{h'}^-\|_K^2 &\leqslant (1+\delta)|K|\|\pi_K f + \Delta u_h - \chi_B z_h^-\|_K^2 \\ &+ (1+1/\delta)|K|\|\Delta(u_{h'}-u_h)\|_K^2 + (1+1/\delta)|K|\|z_{h'}-z_h\|_K^2.\end{aligned}$$

The first term in the last line is bounded by an inverse estimate.

On the other hand, if $K \in \mathcal{M}$, we have

$$\begin{aligned}\sum_{K'\subset K} |K'|\|\pi_{K'} f + \Delta u_{h'} - \chi_B z_{h'}^-\|_{K'}^2 &\leqslant (1+\delta)(1-\xi)|K|\|\pi_K f + \Delta u_h - \chi_B z_h^-\|_K^2 \\ &+ (1+1/\delta)|K|\|\Delta(u_{h'}-u_h)\|_K^2 + (1+1/\delta)|K|\|z_{h'}-z_h\|_K^2.\end{aligned}$$

Putting these estimates together, we obtain

$$\begin{aligned}\sum_{K'\in\mathcal{K}_{h'}} |K'|\|\pi_{K'} f + \Delta u_{h'} - \chi_B z_{h'}^-\|_{K'}^2 &\leqslant (1+\delta)\sum_{K\in\mathcal{K}_h} |K|\|\pi_K f + \Delta u_h - \chi_B z_h^-\|_K^2 \\ - (1-\xi)(1+\delta)\sum_{K\in\mathcal{K}_h} |K|\|\pi_K f + \Delta u_h - \chi_B z_h^-\|_K^2 &+ (1+1/\delta)C|u_{h'}-u_h|_1^2 + (1+1/\delta)\|z_{h'}-z_h\|^2.\end{aligned}$$

Bounding the last term by Poincaré's inequality yields the result for the volume residuals of the state equation. Finally, we consider the volume terms of the estimator of the adjoint equation. If $K \notin \mathcal{M}$ we have

$$\begin{aligned}|K|\|\Delta z_{h'} + \alpha^{-1}\chi_C u_{h'}\|_K^2 &\leqslant (1+\delta)|K|\|\Delta z_h + \alpha^{-1}\chi_C u_h\|_K^2 \\ &+ (1+1/\delta)|K|\|\Delta(z_{h'}-z_h)\|_K^2 + (1+1/\delta)\alpha^{-1}|K|\|u_{h'}-u_h\|_K^2\end{aligned}$$

and we use again an inverse estimate for the first term on the last line. If $K \in \mathcal{M}$ we have

$$\begin{aligned}\sum_{K'\subset K} |K'|\|\Delta z_{h'} + \alpha^{-1}\chi_C u_{h'}\|_{K'}^2 &\leqslant (1+\delta)(1-\xi)|K|\|\Delta z_h + \alpha^{-1}\chi_C u_h\|_K^2 \\ &+ (1+1/\delta)|K|\|\Delta(z_{h'}-z_h)\|_K^2 + (1+1/\delta)\alpha^{-1}|K|\|u_{h'}-u_h\|_K^2.\end{aligned}$$

Putting these estimates together, we obtain

$$\sum_{K'\in\mathcal{K}_{h'}} |K'|\|\Delta z_{h'} + \alpha^{-1}\chi_C u_{h'}\|_{K'}^2 \leqslant (1+\delta)\sum_{K\in\mathcal{K}_h} |K|\|\Delta z_h + \alpha^{-1}\chi_C u_h\|_K^2$$
$$- (1-\xi)(1+\delta)\sum_{K\in\mathcal{K}_h} |K|\|\Delta z_h + \alpha^{-1}\chi_C u_h\|_K^2$$
$$+ (1+1/\delta)C|z_{h'} - z_h|_1^2 + (1+1/\delta)\alpha^{-1}\|u_{h'} - u_h\|^2.$$

With an application of the Poincaré inequality, we obtain the result for the volume terms of the estimator of the adjoint equation. This concludes the proof. □

The last Lemma deals with the coupling due to control.

Lemma 4.5. ***(coupling)*** *Let κ and ε with $0 < \kappa < 1$ and $0 < \varepsilon \leqslant 1$ be given. If $\rho(h_0)$ is sufficiently small, there holds*

$$(1-\varepsilon)|||e_{h'}|||^2 \leqslant |||e_h|||^2 - (1 - \frac{\kappa^2}{\varepsilon})|||e_{h,h'}|||^2. \tag{4.15}$$

In addition, we have

$$|||e_h|||^2 \leqslant (1+\kappa^2)\left(|||e_{h'}|||^2 + |||e_{h,h'}|||^2\right). \tag{4.16}$$

Proof. We first note that

$$|u - u_{h'}|_1^2 \;=\; |u-u_h|_1^2 - |u_{h'} - u_h|_1^2 + 2\langle\nabla(u - u_{h'}), \nabla(u_{h'} - u_h)\rangle.$$

and similarly

$$|z - z_{h'}|_1^2 \;=\; |z-z_h|_1^2 - |z_{h'} - z_h|_1^2 + 2\langle\nabla(z - z_{h'}), \nabla(z_{h'} - z_h)\rangle.$$

By the Galerkin relation (4.1), (4.2), (2.2), and Young's inequality it follows:

$$\begin{aligned}
&\langle\nabla(u - u_{h'}), \nabla(u_{h'} - u_h)\rangle + \langle\nabla(z - z_{h'}), \nabla(z_{h'} - z_h)\rangle \\
&= \int_{\Omega_B} (z^- - z^-_{h'})(u_{h'} - u_h) + \alpha^{-1}\int_{\Omega_C} (u-u_h)(z_{h'} - z_h) \\
&\leqslant \|z - z_{h'}\|_{\Omega_B}\,\|u_{h'} - u_h\|_{\Omega_B} + \alpha^{-1}\|u - u_{h'}\|_{\Omega_C}\,\|z_{h'} - z_h\|_{\Omega_C} \\
&\leqslant \varepsilon\left(|z - z_{h'}|_1^2 + |u - u_{h'}|_1^2\right) + \frac{C\,\rho(h)^{2\tau}}{4\varepsilon}\left(|u_{h'} - u_h|_1^2 + |z_{h'} - z_h|_1^2\right),
\end{aligned}$$

with $\varepsilon > 0$. Therefore we have

$$|||e_{h'}|||^2 \;\leqslant\; |||e_h|||^2 - |||e_{h,h'}|||^2 + \varepsilon|||e_{h'}|||^2 + \frac{C\,\rho(h)^{2\tau}}{4\varepsilon}|||e_{h,h'}|||^2,$$

which gives (4.15) with $\kappa = \sqrt{\frac{C\,\rho(h_0)^{2\tau}}{4}}$. Finally, (4.16) follows from

$$\begin{aligned}
|||e_h|||^2 \;&\leqslant\; |||e_{h'}|||^2 + |||e_{h,h'}|||^2 \\
&\quad + C\,\rho(h)^{2\tau}\left(|u - u_{h'}|_1^2 + |z - z_{h'}|_1^2\right)\left(|u_{h'} - u_h|_1^2 + |z_{h'} - z_h|_1^2\right) \\
&\leqslant\; |||e_{h'}|||^2 + |||e_{h,h'}|||^2 + \kappa^2|||e_{h'}|||^2 + \kappa^2|||e_{h,h'}|||^2.
\end{aligned}$$

□

5. Convergence proof

We prove convergence of the adaptive algorithm with respect to the following error measure:

$$E_k = \sqrt{|||e_k|||^2 + \beta_1 \eta_{h_k}^2 + \beta_2\, \mu_k^2} \tag{5.1}$$

depending on two constants $\beta_1 > 0$ and $\beta_2 > 0$.

Theorem 5.1. *Let $\{h_k\}_{k\geqslant 0}$ be a sequence of meshes generated by algorithm $\mathcal{AFEM}$ and let $\{u_k, z_k\}_{k\geqslant 0}$ be the corresponding sequence of finite element solutions. There exist constants $\beta_1 > 0$, $\beta_2 > 0$, and $\rho < 1$ such that for all $k = 1, 2, \ldots$*

$$E_{k+1} \leqslant \rho\, E_k. \tag{5.2}$$

provided that $\rho(h_0)$ is sufficiently small.

Proof. Let $e_{k+1,k} := (u_{k+1} - u_k, z_{k+1} - z_k)$. We use (4.15) of Lemma 4.5 and (4.12) of Lemma 4.4 in order to obtain

$$\begin{aligned}(1-\varepsilon)|||e_{k+1}|||^2 + \beta_1 \eta_{k+1}^2 \leqslant |||e_k|||^2 - \left(1 - \frac{\kappa^2}{\varepsilon} - \beta_1 C_4 (1 + 1/\delta)\right) |||e_{k,k+1}|||^2 \\ + \beta_1 (1+\delta) \eta_k^2 - \frac{\beta_1 (1+\delta)}{2} \eta_k^2(\mathcal{M}_k).\end{aligned} \tag{5.3}$$

We take $\varepsilon = 3\kappa^2$ and β_1 such that

$$0 < \beta_1 < \xi\theta/(3C_4). \tag{5.4}$$

This choice of β_1 implies that $\beta_1 C_4 (2 - \xi\theta) < \xi\theta(2/3 - \beta_1 C_4)$ such that $\beta_1 C_4/(2/3 - \beta_1 C_4) < \xi\theta/(2 - \xi\theta)$. This allows us to choose δ such that

$$0 < \frac{\beta_1 C_4}{2/3 - \beta_1 C_4} < \delta < \frac{\xi\theta}{2 - \xi\theta} < 1, \tag{5.5}$$

since $\xi\theta < 1$. The left part of (5.5) implies that $1 + 1/\delta < 2/(3\beta_1 C_4)$. We therefore have

$$1 - \frac{\kappa^2}{\varepsilon} - \beta_1 C_4 (1 + 1/\delta) \geqslant 2/3 - 2/3 = 0.$$

Reporting this inequality into (5.3) gives

$$\begin{aligned}(1-\varepsilon)|||e_{k+1}|||^2 + \beta_1 \eta_{k+1}^2 + \beta_2\, \mu_{k+1}^2 \leqslant |||e_k|||^2 + \beta_1 (1+\delta) \eta_k^2 \\ - \xi\beta_1 (1+\delta) \eta_k^2(\mathcal{M}_k) + \beta_2 \mu_{k+1}^2.\end{aligned} \tag{5.6}$$

We now split the proof into two parts depending on the two cases of the algorithm.

In the first case of the algorithm, we have $\eta_k^2(\mathcal{M}_k) \geqslant \theta\, \eta_k^2$ such that (5.6) becomes (and using $\mu_{k+1} \leqslant \mu_k$)

$$\begin{aligned}(1-\varepsilon)|||e_{k+1}|||^2 + \beta_1 \eta_{k+1}^2 + \beta_2\, \mu_{k+1}^2 &\leqslant |||e_k|||^2 + \beta_1 (1+\delta)(1 - \xi\theta) \eta_k^2 + \beta_2 \mu_k^2 \\ &\leqslant (1-\rho_1)|||e_k|||^2 + (1-\rho_2)\beta_1 \eta_k^2 + (1-\rho_3)\beta_2 \mu_k^2 \\ &\quad + \rho_1 |||e_k|||^2 + (\rho_2 + \delta - \xi\theta(1+\delta))\, \beta_1 \eta_k^2 + \rho_3 \beta_2 \mu_k^2,\end{aligned}$$

with $0 < \rho_i < 1$. This yields convergence, if we can show $\rho_1 > \varepsilon$ and

$$A := \rho_1 |||e_k|||^2 + (\rho_2 + \delta - \xi\theta(1+\delta))\, \beta_1 \eta_k^2 + \rho_3 \beta_2 \mu_k^2 \leqslant 0. \tag{5.7}$$

We set $\rho_2 = (1+\delta)\xi\theta/4$. The second inequality of (5.5) implies that $\delta/(1+\delta) \leqslant \xi\theta/2$ and therefore

$$\rho_2 + \delta - \xi\theta(1+\delta) = \delta - \frac{3}{4}\xi\theta(1+\delta) \leqslant -\frac{1}{4}\xi\theta(1+\delta).$$

This, together with the upper bound (4.4) of Lemma 4.2 and the condition $\mu_k^2 \leqslant \gamma\eta_k^2$, we get

$$\begin{aligned} A &\leqslant \rho_1 |||e_k|||^2 - \frac{\xi}{4}\theta(1+\delta)\beta_1\eta_k^2 + \rho_3\beta_2\mu_k^2 \\ &\leqslant \left(\rho_1 C_1 - \beta_1 \frac{\xi\theta(1+\delta)}{4}\right)\eta_k^2 + (\rho_1 C_1 + \rho_3\beta_2)\mu_k^2 \\ &\leqslant \left(\left(\rho_1 C_1 - \beta_1\frac{\xi\theta(1+\delta)}{4}\right) + \gamma(\rho_1 C_1 + \rho_3\beta_2)\right)\eta_k^2. \end{aligned}$$

Then in order to obtain (5.2), it remains to fulfill the following inequalities:

$$(1+\gamma)C_1\rho_1 - \frac{\beta_1\xi\theta(1+\delta)}{4} + \gamma\beta_2\rho_3 \leqslant 0 \tag{5.8}$$

and

$$0 < \varepsilon < \rho_1 < 1. \tag{5.9}$$

By our assumption on $\rho(h_0)$, let κ^2 be small enough such that

$$0 < \kappa^2 < \frac{\beta_1\theta(1+\delta)}{12C_1(1+\gamma)}. \tag{5.10}$$

This implies that we can choose appropriate ρ_1 such that

$$0 < 3\kappa^2 = \varepsilon < \rho_1 < \frac{\beta_1\theta(1+\delta)}{4C_1(1+\gamma)}, \tag{5.11}$$

which yields the iequality (5.9). Furthermore, the right hand side of (5.11) implies

$$C_1(1+\gamma)\rho_1 < \frac{\beta_1\theta(1+\delta)}{4}, \tag{5.12}$$

and we can therefore choose ρ_3 sufficiently small such that the inequality (5.8) is fulfilled. The fact that β_2 is arbitrary up to now will be used in the second part of the proof. This concludes the convergence proof in the first case.

Now we consider the second case. The decrease of the data approximation term (4.13) yields

$$\mu_{k+1}^2 \leqslant (1-\xi\sigma)\mu_k^2. \tag{5.13}$$

We therefore obtain from (5.6)

$$\begin{aligned} (1-\varepsilon)|||e_{k+1}|||^2 + \beta_1\eta_{k+1}^2 + \beta_2\,\mu_{k+1}^2 &\leqslant |||e_k|||^2 + \beta_1(1+\delta)\eta_k^2 + \beta_2(1-\mu\sigma)\mu_k^2 \\ \leqslant (1-\rho_1)|||e_k|||^2 + \beta_1(1-\rho_2)\eta_k^2 + \beta_2(1-\frac{1}{2}\mu\sigma)\mu_k^2 \\ + \rho_1|||e_k|||^2 + \beta_1(\delta+\rho_2)\eta_k^2 - \frac{1}{2}\beta_2\mu\sigma\mu_k^2. \end{aligned}$$

We thus have convergence if $\rho_1 > \varepsilon$ and

$$A := \rho_1 |||e_k|||^2 + \beta_1(\delta + \rho_2)\eta_k^2 - \frac{1}{2}\beta_2\mu\sigma\mu_k^2 \leqslant 0.$$

Using the global upper bound and $\eta_k^2 \leqslant \gamma^{-1}\mu_k^2$, it turns out that

$$\begin{aligned} A &= \rho_1 |||e_k|||^2 + \beta_1(\delta + \rho_2)\eta_k^2 - \frac{1}{2}\beta_2\mu\sigma\mu_k^2 \\ &\leqslant (\rho_1 C_1 + \beta_1(\delta + \rho_2))\,\eta_k^2 + \left(\rho_1 C_1 - \frac{1}{2}\beta_2\mu\sigma\right)\mu_k^2 \\ &\leqslant \left(\gamma^{-1}(\rho_1 C_1 + \beta_1(\delta + \rho_2)) + \rho_1 C_1 - \frac{1}{2}\beta_2\mu\sigma\right)\mu_k^2. \end{aligned}$$

We choose $\rho_1 > \varepsilon$ and take β_2 large enough such that the following inequality is satisfied:

$$\frac{1}{2}\beta_2\mu\sigma \;\geqslant\; (1 + \gamma^{-1})\rho_1 C_1 + \gamma^{-1}\beta_1(\delta + \rho_2), \tag{5.14}$$

which is possible since β_2 was arbitrary in the first part of the proof. □

6. Quasi-optimality

In order to express the quasi-optimality, we introduce some notation from nonlinear approximation theory, see [5, 9]. Let $\mathcal{H}_N$ be the set of all meshes h which satisfy $N_h \leqslant N$.

Next we define the approximation class

$$\mathcal{W}^s := \Big\{(u, z, f) \in (H_0^1(\Omega), H_0^1(\Omega), L^2(\Omega)) : \|(u, z, f)\|_{\mathcal{W}^s} < +\infty\Big\}. \tag{6.1}$$

with

$$\|(u, z, f)\|_{\mathcal{W}^s} := \sup_{N \geqslant N_0} N^s \inf_{h \in \mathcal{H}_N} \Big(|||e_h|||^2 + \mu_h^2\Big)^{1/2}.$$

We say that an adaptive finite element method realizes optimal convergence rates if, whenever $(u, z, f) \in \mathcal{W}^s$, it produces a sequence of meshes $\{h_k\}$ and corresponding approximations $\{u_k, z_k\}$ such that the error $\{e_k\}$ and data approximation $\{\mu_k\}$ satisfy

$$|||e_k|||^2 + \mu_h^2 \leqslant C\, N_k^{-2s}. \tag{6.2}$$

Theorem 6.1. *Suppose $(f, u, z) \in \mathcal{W}^s$. Let $\{h_k\}_{k \geqslant 0}$ be a sequence of meshes generated by algorithm $\mathcal{AFEM}$ and let $\{V_k\}_{k \geqslant 0}$ and $\{u_k, z_k\}_{k \geqslant 0}$ be the corresponding sequences of finite element spaces and solutions. Let $\varepsilon_k := \sqrt{|||e_k|||^2 + \mu_k^2}$. Assuming the parameters γ and θ to satisfy*

$$0 < \theta < \frac{1}{C_1 C_2(1 + \kappa^2)}, \quad 0 < \gamma \leqslant \frac{1 - C_1 C_2(1 + \kappa^2)\theta}{C_2(1 + (1 + \kappa^2)C_1)}, \tag{6.3}$$

we have the following estimate on the complexity of the algorithm: there exists a constant C such that

$$N_k - N_0 \leqslant C\, \varepsilon_k^{-1/s}, \tag{6.4}$$

provided that $\rho(h_0)$ is sufficiently small.

Proof. From the regularity assumption we have existence of a mesh $h_* \in \mathcal{H}$ with error $e_* = (u - u_{h_*}, z - z_{h_*})$ and number of cells N_* such that for $\lambda > 0$ to be chosen below

$$\varepsilon_{h_*} := \sqrt{|||e_*|||^2 + \mu_{h_*}^2} \leqslant \lambda\, \varepsilon_k := \sqrt{|||e_k|||^2 + \mu_k^2}, \tag{6.5}$$

and

$$N_* \leqslant C \varepsilon_k^{-1/s}. \tag{6.6}$$

Following the proof of Stevenson [20] (proof of Lemma 5.2), we can suppose that h_* is a refinement of h_k, if we replace (6.6) by:

$$N_* - N_k \leqslant C\, \varepsilon_k^{-1/s}. \tag{6.7}$$

Let $\mathcal{M}_* \subset \mathcal{K}_{h_k}$ be the set of edges which have at least to be refined in order to produce h_*. We remember that $\mathcal{M}_k$ denotes the set of marked cells in iteration k.

We will prove below the estimate

$$\#\mathcal{M}_k \leqslant C\, \varepsilon_k^{-1/s}. \tag{6.8}$$

This implies the complexity estimate (6.4) as follows. Let as before $E_l = |||e_l|||^2 + \beta_1 \eta_l^2 + \beta_2\, \mu_l^2$. From Theorem 6.1 we know that for some constant $\rho < 1$

$$E_k \leqslant \rho^{k-l}\, E_l, \quad 0 \leqslant l \leqslant k.$$

We obviously have $\varepsilon_l \leqslant \max(1, \beta_2) E_l$. By the global lower bound (4.11) we also have $E_l \leqslant C\, \varepsilon_l$ with an absolute constant C. This implies

$$\varepsilon_k \leqslant C\, \rho^{k-l}\, \varepsilon_l, \quad 0 \leqslant l \leqslant k. \tag{6.9}$$

The bound (6.9) and Lemma 3.1 imply

$$\begin{aligned} N_{k+1} - N_0 &\leqslant C \sum_{l=0}^{k} \#\mathcal{M}_k \leqslant C \sum_{l=0}^{k} \varepsilon_l^{-1/s} \\ &\leqslant C \left(\sum_{l=0}^{k} \rho_l^{(k-l)/s} \right) \varepsilon_k^{-1/s} \leqslant \frac{C}{1 - \rho^{1/s}}\, \varepsilon_k^{-1/s}. \end{aligned}$$

yielding (6.4).

We now turn the proof of (6.8). As before, we consider the two cases of the algorithm separately. In the first case we have

$$\mu_k^2 \leqslant \gamma \eta_k^2. \tag{6.10}$$

We will prove below that

$$\eta_h^2(\mathcal{M}_*) \geqslant \theta \eta_k^2. \tag{6.11}$$

This implies the estimate (6.8): Since $\mathcal{F}$ is chosen to be the set with minimal cardinality satisfying the bound (6.11), we find that

$$\#\mathcal{M}_k \leqslant \#\mathcal{M}_* \leqslant C(N_* - N_k) \leqslant C\, \varepsilon_k^{-1/s}. \tag{6.12}$$

The proof of (6.11) is obtained as follows. Let $e_{k,*} := ((u_* - u_k), (z_* - z_k))$. We successively use(4.16), (4.5), (6.5) and (4.4) in order to obtain

$$
\begin{aligned}
|||e_k|||^2 &\leqslant (1+\kappa^2)|||e_*|||^2 + (1+\kappa^2)|||e_{k,*}|||^2 \\
&\leqslant \Big((1+\kappa^2) + C_1\rho(h_0)^{2\tau}\Big)|||e_*|||^2 + (1+\kappa^2)C_1\eta_k^2(\mathcal{M}_*) \\
&\quad +(1+\kappa^2)C_1\mu_k^2(\mathcal{M}_*) + C_1\rho(h_0)^{2\tau}\eta_k^2 \\
&\leqslant \Big(\Big((1+\kappa^2) + C_1\rho(h_0)^{2\tau}\Big)\lambda + C_1\rho(h_0)^{2\tau}\Big)\left(\eta_k^2+\mu_k^2\right) + (1+\kappa^2)C_1\eta_k^2(\mathcal{M}_*) \\
&\quad +(1+\kappa^2)C_1\mu_k^2(\mathcal{M}_*) \\
&\leqslant \Big(\Big((1+\kappa^2) + C_1\rho(h_0)^{2\tau}\Big)\lambda + C_1\rho(h_0)^{2\tau}\Big)C_1\eta_k^2 \\
&\quad + \Big(\Big(\Big((1+\kappa^2) + C_1\rho(h_0)^{2\tau}\Big)\lambda + C_1\rho(h_0)^{2\tau}\Big)C_1 + 1 + (1+\kappa^2)C_1\Big)\mu_k^2 \\
&\quad +(1+\kappa^2)C_1\eta_k^2(\mathcal{M}_*).
\end{aligned}
$$

Let $A := \Big(\Big((1+\kappa^2) + C_1\rho(h_0)^{2\tau}\Big)\lambda + C_1\rho(h_0)^{2\tau}\Big)C_1$ and $B := (AC_1 + 1 + (1+\kappa^2)C_1)$. By the upper bond (4.11) and (6.10), in follows that

$$
\begin{aligned}
C_2^{-1}\eta_k^2 &\leqslant |||e_k|||^2 + \mu_k^2 \\
&\leqslant A\eta_k^2 + B\mu_k^2 + (1+\kappa^2)C_1\eta_k^2(\mathcal{M}_*) + \mu_k^2 \\
&\leqslant \left(A(1+\gamma) + (1+(1+\kappa^2)C_1)\gamma\right)\eta_k^2 + (1+\kappa^2)C_1\eta_k^2(\mathcal{M}_*),
\end{aligned}
$$

from which it follows that

$$
\left(\frac{1}{C_2} - \left(A(1+\gamma) + (1+(1+\kappa^2)C_1)\gamma\right)\right)\eta_k^2 \leqslant (1+\kappa^2)C_1\eta_k^2(\mathcal{M}_*).
$$

By the assumption on γ and θ in (6.3), we can choose

$$
\lambda = \frac{\frac{1 - C_1C_2(1+\kappa^2)\theta - (1+(1+\kappa^2)C_1)\gamma}{C_1C_2(1+\gamma)} - C_1\rho(h_0)^{2\tau}}{(1+\kappa^2) + C_1\rho(h_0)^{2\tau}} > 0, \tag{6.13}
$$

provided $\rho(h_0)$ is small enough. We therefore obtain (6.11), completing the proof in the first case.

Now we consider the second case. We thus have

$$
\eta_k^2 \leqslant \gamma^{-1}\mu_k^2. \tag{6.14}
$$

We will prove that

$$
\mu_k^2(\mathcal{M}_*) \geqslant \sigma\mu_k^2. \tag{6.15}
$$

This implies (6.8) as before by the optimality of the choice of $\mathcal{P}$. First we note that by (4.11) and (6.15) we have

$$
|||e_k|||^2 \leqslant C_1\left(\eta_k^2 + \mu_k^2\right) \leqslant C_1(1+\gamma^{-1})\mu_k^2.
$$

This implies together with (6.5) that

$$
\begin{aligned}
\mu_k^2 - \mu_k^2(\mathcal{M}_*) &\leqslant \mu_{h_*}^2 \leqslant \lambda\left(\eta_k^2 + \mu_k^2\right) \\
&\leqslant \lambda\left(1 + C_1(1+\gamma^{-1})\right)\mu_k^2,
\end{aligned}
$$

and therefore with λ small enough, we get

$$\sigma\,\mu_k^2 \leqslant \left(1 - \lambda\left(1 + C_1(1+\gamma^{-1})\right)\right)\mu_k^2 \leqslant \mu_k^2(\mathcal{M}_*).$$

This concludes the proof. □

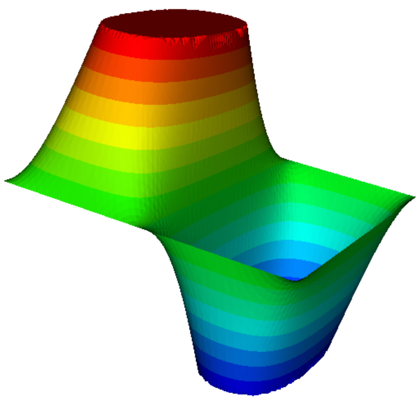
Corollary 6.1. *An argument similar to the one used to prove (4.3) shows that* $\|q - q_h\|_{\Omega_B} = \|z^- - z_h^-\|_{\Omega_B} \leqslant \|z - z_h\|_{\Omega_B} \leqslant C\,|z - z_h|_1$. *Therefore Theorem 5.1 also implies convergence of the control variable and the control could be included in the complexity analysis.*

Let $s > 0$ *and* $(u, z, f) \in \mathcal{W}^s$. *The mesh-independence of the semi-smooth Newton method [14] combined with with multigrid iteration [6, 22], suggests that the algorithm* $\mathcal{AFEM}$ *has optimal work count in the sense that for a given accuracy* $\varepsilon > 0$, *the algorithm provides a discrete solutions* u_h *and* z_h *satisfying* $|||e_h||| \leqslant \varepsilon$ *with a number of operations proportional to* $\varepsilon^{-1/s}$. *The combination of the adaptive algorithm with multigrid requires the introduction of a stopping criterion leading to an additional iteration error. Such an algorithm has been proposed and analyzed for the Poisson problem in [3].*

We finally remark that the regularity assumption $(f, u, z) \in \mathcal{W}^s$ is difficult to verify in practice. However, the a priori error analysis on meshes adapted to corner singularities suggests that $s = 1/2$ if $f \in L^2(\Omega)$ under mild restrictions on the domain, in the considered two-dimensional case, see[1].

7. Numerical experiments

In this Section, we report on two numerical experiments. For the first one the exact solution of the problem is known. We use this example in order to investigate the complexity of the sequence of meshes generated by the adaptive algorithm. The computational domain is $\Omega = (0,1)^2$ and the right-hand side is constructed in such a way that $u(x,y) = z(x,y) = \sin(\pi(x+2y))$. The parameter is α and $Q = \{q \in L^2(\Omega) :\ q_{min} \leqslant q \leqslant q_{max}\}$ with $q_{min} = -50 = -q_{max}$ such that the control has the appearance shown in Figure 7.1.



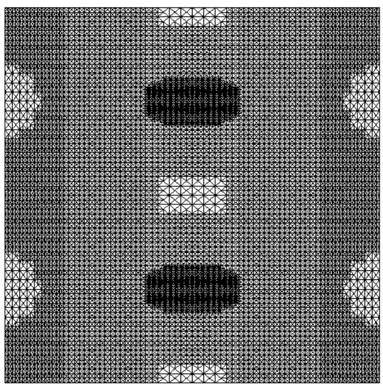


Figure 7.1. Control $-\min(50, \max(-50, -z_h))/\alpha$ and locally refined mesh.

In Table 7.1 the value of the errors and estimators are given on a typical adaptive iteration. The data approximation term has the expected second-order behavior. The error is over-estimated by a factor of 3.5. This is due to the fact that we have for simplicity estimated the constant of the interpolation error by 1.

The computational domain for the second example is the L-shaped domain $\Omega = (-1,1)\times(0,1)\cup(-1,0)\times(-1,0]$. The parameters are $\alpha = 10^{-4}$ and $q_{min} = -10$ ($q_{max} = \infty$), $f = 0$, $u_d = 1$, $\Omega_B = \Omega \cap \{y \geqslant 0\}$, and $\Omega_C := \Omega \cap \{x \leqslant 0\}$. The discrete solutons u_h, z_h, corresponding control, and a typical mesh are shown in Figure 7.2. There is a strong refinement at the re-entrant corner and a long the boundary of $\partial\Omega_B \setminus \partial\Omega$. The first one is however significantly stronger, which is due to the fact that it generates a stronger singularity.

Finally, we make a comparison of the asymptotic behavior of η_h^2 for different refinement parameters θ. Note that $\theta = 1$ leads to uniform refinement, which is known to lead to a loss of convergence rate due to the corner singularity. It can be seen from Figure 7.3 that the adaptive algorithm is able to regain the convergence rate -1. This follows from Theorem 6.1, since the construction of meshes recovering the optimal rate is well known, implying $u \in \mathcal{A}^s$ with $s = 1/2$.

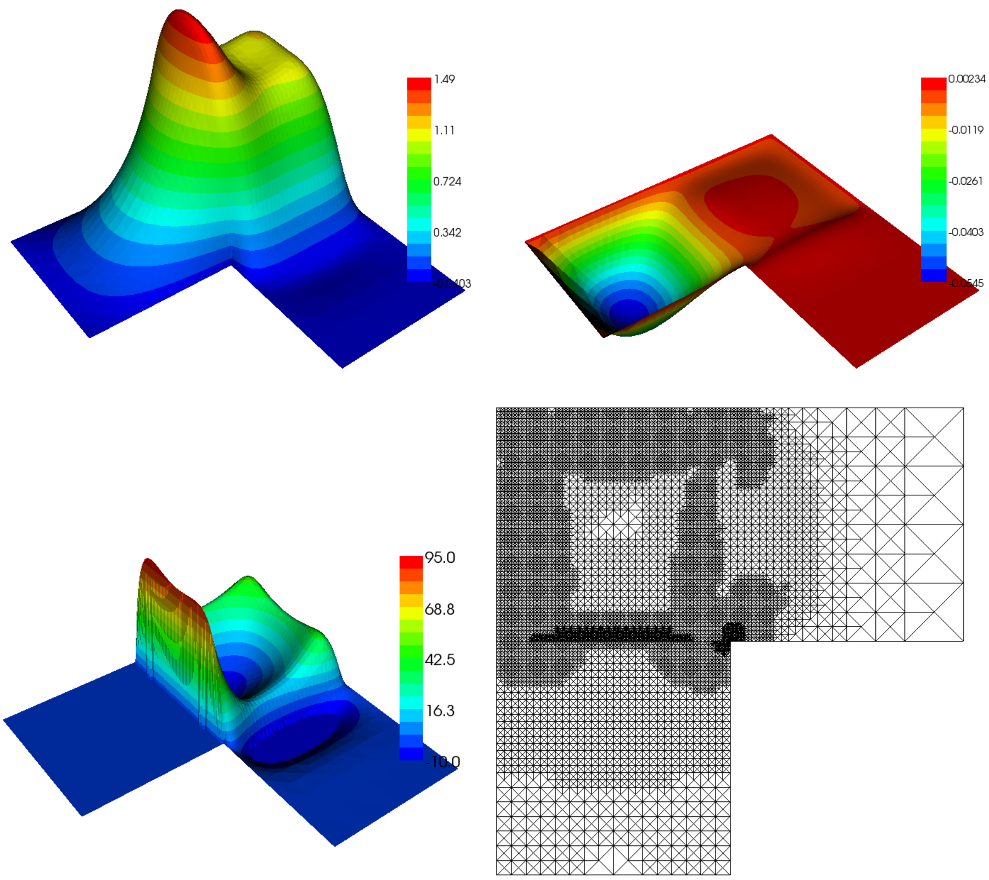


Figure 7.2. Second example: u_h, z_h (scaled by a factor of 10), contol (scaled by a factor 0.01) and locally refined mesh.

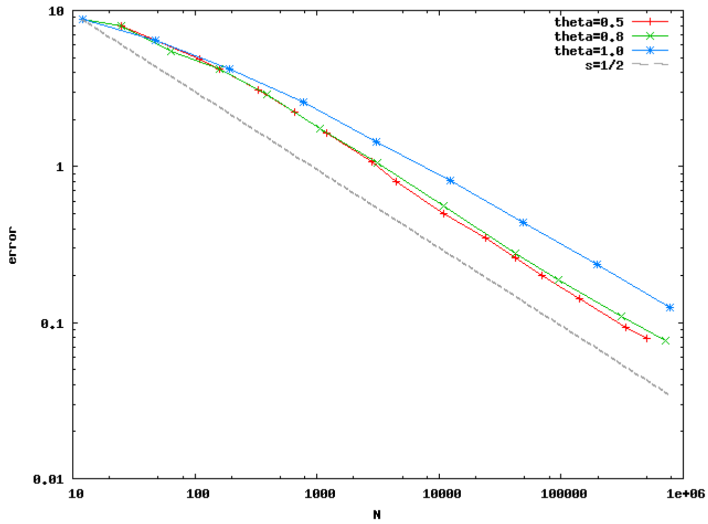


Figure 7.3. Behavior of η_h vs. N_h for $\theta = 0.5, 0.8, 1.0$ compared to first-order decrease $s = 1/2$ (dotted line), log-log-scale.

N_h	e_h	η_h	μ_h	e_h/η_h
100	1.3686	0.9340	4.6864	0.292
364	0.7168	0.2609	2.5108	0.285
1244	0.4042	0.0844	1.4371	0.281
4091	0.2275	0.0453	0.8019	0.283
10752	0.1382	0.0215	0.4846	0.285
24280	0.0871	0.0069	0.3094	0.281
70468	0.0534	0.0026	0.1886	0.283
201632	0.0321	0.0010	0.1132	0.284
457800	0.0204	0.0005	0.0723	0.282

Table 7.1. Adaptive iteration with $\theta = 0.75$.

8. Conclusion

We have proposed a new adaptive algorithm for optimal control based on standard conforming finite elements, using an adaptive marking strategy. For simplicity, we have considered triangular meshes, but the generalization to quadrilaterals and three dimensions seems possible.

We have carried out the proofs of geometric convergence of the error and quasi-optimality in the case that the control variable is eliminated from the system. The generalization to the more commonly studied case of Galerkin discretization of all variables including control (see, for example, [13]).

Appendix

In this appendix we give a proof of the $L^2(\Omega)$-error estimate announced in (2.2).

Lemma 8.1. *Assume that the domain is such that the inverse Laplace-operator allows for the regularity shift $L^2(\Omega) \to H^{1+\tau}(\Omega)$ for $0 < \tau \leqslant 1$. Then we have for h_0 sufficiently small*

$$\|u - u_h\|^2 + \|z - z_h\|^2 \leqslant C\rho^{2\tau}(h)(|u - u_h|_1^2 + |z - z_h|_1^2). \tag{8.1}$$

Proof. Let $\psi_1, \psi_2 \in L^2(\Omega)$ and $\phi \in V$ and $\phi_h \in V_h$ the solutions to the equations

$$\langle \nabla\phi, \nabla v\rangle = \langle \psi_1, v\rangle \; \forall v \in V, \quad \langle \nabla\phi_h, \nabla v_h\rangle = \langle \psi_2, v_h\rangle \; \forall v_h \in V_h. \tag{8.2}$$

Then it follows from the Aubin-Nitsche duality argument and our assumption that

$$\|\phi - \phi_h\| \leqslant C\left(\rho^\tau(h)|\phi - \phi_h|_1 + \|\psi_1 - \psi_2\|\right). \tag{8.3}$$

Therefore we immediately get (with $\alpha q = -z^-$ and $\alpha q_h = -z_h^-$

$$\|u - u_h\| \leqslant C\Bigg(\rho^\tau(h)|u - u_h|_1 + \|q - q_h\|\Bigg)$$

and

$$\begin{aligned} \|z - z_h\| &\leqslant C\Bigg(\rho^\tau(h)|z - z_h|_1 + \|u - u_h\|\Bigg) \\ &\leqslant C\Bigg(\rho^\tau(h)\Bigg(|z - z_h|_1 + |u - u_h|_1\Bigg) + \|q - q_h\|\Bigg). \end{aligned}$$

It therefore remains to show that

$$\|q - q_h\| \leqslant C\rho^\tau(h)\Bigg(|z - z_h|_1 + |u - u_h|_1\Bigg). \tag{8.4}$$

We denote by $u_h(q)$ and $z_h(q)$ the state and adjoint solutions corresponding to the control q, i.e., the solutions to

$$\begin{aligned} \langle \nabla u_h(q), \nabla v_h\rangle &= \langle f, v_h\rangle + \langle q, v_h\rangle_{\Omega_B} \quad \forall v_h \in V_h, \\ \langle \nabla v_h, \nabla z_h(q)\rangle &= \langle u_h(q), v_h\rangle_{\Omega_C} \quad \forall v_h \in V_h. \end{aligned} \tag{8.5}$$

It follows from (8.5) that

$$\begin{aligned} \langle q - q_h, z_h(q) - z_h\rangle_{\Omega_B} &= \langle \nabla(u_h(q) - u_h), \nabla(z_h(q) - z_h)\rangle \\ &= \langle u_h(q) - u_h, u_h(q) - u_h\rangle_{\Omega_C} \geqslant 0. \end{aligned} \tag{8.6}$$

Since for any real numbers a, b it holds $(a^- - b^-)(a^+ - b^+) \geqslant 0$, we have

$$\begin{aligned} \alpha\|q - q_h\|_{\Omega_B}^2 &= \alpha^{-1}\langle z_h^- - z^-, z_h^- - z^-\rangle_{\Omega_B} \\ &= \alpha^{-1}\langle z_h^- - z^-, z_h - z\rangle_{\Omega_B} - \alpha^{-1}\langle z_h^- - z^-, z_h^+ - z^+\rangle_{\Omega_B} \\ &\leqslant \alpha^{-1}\langle z_h^- - z^-, z_h - z\rangle_{\Omega_B} = \langle q - q_h, z_h - z\rangle_{\Omega_B} \\ &= \langle q - q_h, z_h - z_h(q)\rangle_{\Omega_B} + \langle q - q_h, z_h(q) - z\rangle_{\Omega_B} \\ &\leqslant \langle q - q_h, z_h(q) - z\rangle_{\Omega_B}, \end{aligned}$$

such that

$$\alpha\|q - q_h\|_{\Omega_B} \leqslant \|z_h(q) - z\|_{\Omega_B}. \tag{8.7}$$

We next consider the auxiliary problems $w, p \in V$ such that

$$\langle \nabla w, \nabla v\rangle = \langle z_h(q) - z, v\rangle_{\Omega_B} \ \forall v \in V, \quad \langle \nabla v, \nabla p\rangle = \langle w, v\rangle_{\Omega_C} \ \forall v \in V. \tag{8.8}$$

Then it follows with an interpolant w_h of w that

$$\begin{aligned} \|z_h(q) - z\|^2_{\Omega_B} &= \langle \nabla w, \nabla(z_h(q) - z)\rangle \\ &= \langle \nabla(w - w_h), \nabla(z_h(q) - z)\rangle + \langle u_h(q) - u, w_h\rangle_{\Omega_C} \\ &= \langle \nabla(w - w_h), \nabla(z_h(q) - z)\rangle + \langle u_h(q) - u, w_h - w\rangle_{\Omega_C} + \langle u_h(q) - u, w\rangle_{\Omega_C}. \end{aligned}$$

By our assumption we have

$$\langle \nabla(w - w_h), \nabla(z_h(q) - z)\rangle \leqslant C\rho^\tau(h)\|z_h(q) - z\|_{\Omega_B}\, |z_h(q) - z|_1.$$

Interpolation and the Poincaré inequality yield

$$\langle u_h(q) - u, w - w_h\rangle_{\Omega_C} \leqslant C\rho^\tau(h)\|u_h(q) - u\|\|z_h(q) - z\|_{\Omega_B}.$$

In addition we have with an interpolant p_h of p, the regularity shift and the Poincaré inequality

$$\begin{aligned} \langle u_h(q) - u, w\rangle_{\Omega_C} &= \langle \nabla(u_h(q) - u), \nabla p\rangle = \langle \nabla(u_h(q) - u), \nabla(p - p_h)\rangle \\ &\leqslant C\rho^\tau(h)|u_h(q) - u|_1\|w\|_{\Omega_C} \leqslant C\rho^\tau(h)|u_h(q) - u|_1\|z_h(q) - z\|_{\Omega_B}. \end{aligned}$$

Putting these estimates together, we obtain

$$\begin{aligned} \|z_h(q) - z\|_{\Omega_B} &\leqslant C\rho^\tau(h)\Big(|z_h(q) - z|_1 + |u_h(q) - u|_1\Big) \\ &\leqslant C\rho^\tau(h)\Big(|z_h - z|_1 + |u_h - u|_1 + |z_h(q) - z_h|_1 + |u_h(q) - u_h|_1\Big) \\ &\leqslant C\rho^\tau(h)\Big(|z_h - z|_1 + |u_h - u|_1 + \|q - q_h\|_{\Omega_B}\Big) \end{aligned} \tag{8.9}$$

If now h_0 is sufficiently small, we conclude that

$$\alpha\|q - q_h\|_{\Omega_B} \leqslant \|z_h(q) - z\|_{\Omega_B} \leqslant C\rho^\tau(h)\Big(|z_h - z|_1 + |u_h - u|_1\Big),$$

as required. This ends up the proof. □

References

[1] T. Apel, A. Rösch, and G. Winkler, *Optimal control in non-convex domains: a priori discretization error estimates*, Calcolo, 44 (2007), pp. 137–158.

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

[3] R. Becker and S. Mao, *Convergence and quasi-optimal complexity of a simple adaptive finite element method*, M2AN, 43 (2009), pp. 1203–1219.

[4] R. Becker, S. Mao, and Z.-C. Shi, *A convergent adaptive finite element method with optimal complexity*, Electron. Trans. Numer. Anal., 30 (2008), pp. 291–304.

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

[6] J. Bramble and J. Pasciak, *New estimates for multilevel algorithms including the v-cycle*, Math. Comp., 60 (1995), pp. 447–471.

[7] C. Carstensen and R. Verfürth, *Edge residuals dominate a posteriori error estimates for low order finite element methods*, SIAM J. Numer. Anal., 36 (1999), pp. 1571–1587.

[8] P. Ciarlet, *The finite element method for elliptic problems.*, Studies in Mathematics and its Applications. Vol. 4. Amsterdam - New York - Oxford: North-Holland Publishing Company., 1978.

[9] R. DeVore, *Nonlinear approximation.*, in Acta Numerica 1998, A. Iserles, ed., vol. 7, Cambridge University Press, 1998, pp. 51–150.

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

[11] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson, *Introduction to adaptive methods for differential equations*, in Acta Numerica 1995, A. Iserles, ed., Cambridge University Press., 1995, pp. 105–158.

[12] A. Gaevskaya, R. Hoppe, Y. Iliash, and M. Kieweg, *Convergence analysis of an adaptive finite element method for distributed control problems with control constraints*, tech. rep., Houston, 2006.

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

[14] M. Hintermüller and M. Ulbrich, *A mesh-independence result for semismooth Newton methods*, Math. Program., 101 (2004), pp. 151–184.

[15] M. Hinze, *A variational discretization concept in control constrained optimization: The linear-quadratic case*, Comput. Optim. Appl., 30 (2005), pp. 45–61.

[16] J.-L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, vol. 170 of Grundlehren Math. Wiss., Springer-Verlag, Berlin, 1971.

[17] W. F. Mitchell, *A comparison of adaptive refinement techniques for elliptic problems*, ACM Transactions on Mathematical Software (TOMS), 15 (1989), pp. 326–347.

[18] 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.

[19] G. Savare, *Regularity results for elliptic equations in lipschitz domains*, Journal of Functional Analysis, 152 (1998), pp. 176 – 201.

[20] R. Stevenson, *Optimality of a standard adaptive finite element method*, Found. Comput. Math., 7 (2007), pp. 245–269.

[21] R. Verfürth, *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*, Wiley/Teubner, New York-Stuttgart, 1996.

[22] H. Wu and Z. Chen, *Uniform convergence of multigrid v-cycle on adaptively refined finite element meshes for second order elliptic problems*, Sci. in China Ser. A, 49 (2006), pp. 1405–1429.