

Convergence of Adaptive Finite Elements for Optimal Control Problems with Control Constraints

Kristina Kohls, Arnd Rösch, and Kunibert G. Siebert

Abstract We summarize our findings in the analysis of adaptive finite element methods for the efficient discretization of control constrained optimal control problems. We particularly focus on convergence of the adaptive method, i.e., we show that the sequence of adaptively generated discrete solutions converges to the true solution. We restrict the presentation to a simple model problem to highlight the key components of the convergence proof and comment on generalizations of the presented result.

Keywords Adaptive finite elements • A posteriori error estimators • Convergence analysis • Optimal control • Control constraints

Mathematics Subject Classification (2010). 65N30, 65N12, 49J20.

1 Statement of the Main Result

In this summary we analyze adaptive finite element discretizations for control constrained optimal control problems of the form

$$\begin{aligned} \min_{(u,y) \in \mathbb{U}^{\text{ad}} \times \mathbb{Y}} & \frac{1}{2} \|y - y_d\|_{\mathbb{U}}^2 + \frac{\alpha}{2} \|u\|_{\mathbb{U}}^2 \\ \text{subject to} & \quad y \in \mathbb{Y} : \quad \mathcal{B}[y, v] = \langle u, v \rangle \quad v \in \mathbb{Y}. \end{aligned} \tag{1.1}$$

K. Kohls • K.G. Siebert (✉)

Institut für Angewandte Analysis und Numerische Simulation, Fachbereich Mathematik,
Universität Stuttgart, Pfaffenwaldring 57, D-70569 Stuttgart, Germany
e-mail: <http://www.ians.uni-stuttgart.de/nmh/>; kristina.kohls@ians.uni-stuttgart.de;
<http://www.ians.uni-stuttgart.de/nmh/>; kg.siebert@ians.uni-stuttgart.de

A. Rösch

Fakultät für Mathematik, Thea-Leymann-Straße 9, D-45127 Essen, Germany
e-mail: <http://www.uni-due.de/mathematik/agroesch/>; arnd.roesch@uni-due.de

In order to highlight the basic ideas of our convergence analysis we focus on the most simple model problem in the following setting. We let $\Omega \subset \mathbb{R}^d$ be a bounded domain that is meshed exactly by some conforming initial triangulation \mathcal{G}_0 . We consider distributed control in $\mathbb{U} = L_2(\Omega)$ with a non-empty, convex, and closed subset \mathbb{U}^{ad} of admissible controls. We use the $L_2(\Omega)$ scalar product $\langle \cdot, \cdot \rangle$ and write $\|\cdot\|_{\mathbb{U}} = \|\cdot\|_{2;\Omega}$ for its induced norm. The PDE constraint is given by Poisson's problem in the state space $\mathbb{Y} = \dot{H}^1(\Omega)$ equipped with norm $\|\cdot\|_{\mathbb{Y}} = \|\nabla \cdot\|_{2;\Omega}$ and the continuous and coercive bilinear form

$$\mathcal{B}[y, v] = \langle \nabla y, \nabla v \rangle \quad \forall y, v \in \mathbb{Y}.$$

Finally, $y_d \in L_2(\Omega)$ is a desired state and $\alpha > 0$ is some given cost parameter.

Turning to the discretization of (1.1) we denote by \mathbb{G} the class of all conforming refinements of \mathcal{G}_0 that can be constructed using refinement by bisection [13]. For a given grid $\mathcal{G} \in \mathbb{G}$ we let $\mathbb{Y}(\mathcal{G}) \subset \mathbb{Y}$ be a conforming finite element space of piecewise polynomials of fixed degree $q \in \mathbb{N}$. We then consider the variational discretization of (1.1) by Hinze [4], i.e., we solve the discretized optimal control problem

$$\begin{aligned} \min_{(U, Y) \in \mathbb{U}^{\text{ad}} \times \mathbb{Y}(\mathcal{G})} & \frac{1}{2} \|Y - y_d\|_{\mathbb{U}}^2 + \frac{\alpha}{2} \|U\|_{\mathbb{U}}^2 \\ \text{subject to} & \quad Y \in \mathbb{Y}(\mathcal{G}) : \quad \mathcal{B}[Y, V] = \langle U, V \rangle \quad V \in \mathbb{Y}(\mathcal{G}). \end{aligned} \tag{1.2}$$

It is well-known that (1.1) as well as (1.2) admit a unique solution pair (\hat{u}, \hat{y}) , respectively $(\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}})$; compare with [9, 15]. Below we additionally utilize the continuous and discrete adjoint states $\hat{p} \in \mathbb{Y}$, $\hat{P}_{\mathcal{G}} \in \mathbb{Y}(\mathcal{G})$, and consider the solution triplets $(\hat{u}, \hat{y}, \hat{p}) \in \mathbb{U}^{\text{ad}} \times \mathbb{Y} \times \mathbb{Y}$ and $(\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}) \in \mathbb{U}^{\text{ad}} \times \mathbb{Y}(\mathcal{G}) \times \mathbb{Y}(\mathcal{G})$.

We use the following adaptive algorithm for approximating the true solution of (1.1). Starting with the initial conforming triangulation \mathcal{G}_0 of Ω we execute the standard adaptive loop

$$\text{SOLVE} \longrightarrow \text{ESTIMATE} \longrightarrow \text{MARK} \longrightarrow \text{REFINE}. \tag{1.3}$$

In practice, a stopping test is used after ESTIMATE for terminating the iteration; here we shall ignore it for notational convenience.

Assumption 1.1 (Properties of modules). For a given grid $\mathcal{G} \in \mathbb{G}$ the four used modules have the following properties.

1. The output $(\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}) := \text{SOLVE}(\mathcal{G}) \in \mathbb{U}^{\text{ad}} \times \mathbb{Y}(\mathcal{G}) \times \mathbb{Y}(\mathcal{G})$ is the exact solution of (1.2).
2. The output $\{\mathcal{E}_{\mathcal{G}}((\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}); E)\}_{E \in \mathcal{G}} := \text{ESTIMATE}((\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}); \mathcal{G})$ is a *reliable* and *locally efficient* estimator for the error in the norm $\|\cdot\|_{\mathbb{U} \times \mathbb{Y} \times \mathbb{Y}}$. In Sect. 2 below we give an example of such an estimator.

3. The output $\mathcal{M} = \text{MARK}(\{\mathcal{E}_{\mathcal{G}}((\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}); E)\}_{E \in \mathcal{G}}, \mathcal{G})$ is a subset of elements subject to refinement. We shall allow any marking strategy such that \mathcal{M} contains an element holding the maximal indicator, i.e.,

$$\max\{\mathcal{E}_{\mathcal{G}}((\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}); E) \mid E \in \mathcal{G}\} \leq \max\{\mathcal{E}_{\mathcal{G}}((\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}); E) \mid E \in \mathcal{M}\}.$$

All practically relevant marking strategies do have this property.

4. The output $\mathcal{G}_+ := \text{REFINE}(\mathcal{G}, \mathcal{M}) \in \mathbb{G}$ is a conforming refinement of \mathcal{G} such that all elements in \mathcal{M} are bisected at least once, i.e., $\mathcal{G}_+ \cap \mathcal{M} = \emptyset$.

The main contribution of this report is the following convergence result.

Theorem 1.2 (Main result). *Let $(\hat{u}, \hat{y}, \hat{p}) \in \mathbb{U}^{\text{ad}} \times \mathbb{Y} \times \mathbb{Y}$ be the true solution of (1.1). Suppose that $\{\hat{U}_k, \hat{Y}_k, \hat{P}_k\}_{k \geq 0} \subset \mathbb{U}^{\text{ad}} \times \mathbb{Y} \times \mathbb{Y}$ is any sequence of discrete solutions generated by the adaptive iteration (1.3), where the modules have the properties stated in Assumption 1.1. Then we have*

$$\lim_{k \rightarrow \infty} \|(\hat{U}_k, \hat{Y}_k, \hat{P}_k) - (\hat{u}, \hat{y}, \hat{p})\|_{\mathbb{U} \times \mathbb{Y} \times \mathbb{Y}} = 0 \quad \text{and} \quad \lim_{k \rightarrow \infty} \mathcal{E}_{\mathcal{G}_k}(\hat{U}_k, \hat{Y}_k, \hat{P}_k; \mathcal{G}_k) = 0.$$

The proof of this theorem uses results and ideas from the convergence proofs of Morin, Siebert, and Veerer in [12] and Siebert in [14]. It is a two step procedure presented in Sects. 3 and 4. In Sect. 3 we utilize basic *stability properties* of the algorithm to show that the sequence of discrete solutions converges to some triplet $(\hat{u}_{\infty}, \hat{y}_{\infty}, \hat{p}_{\infty})$. The second step in Sect. 4 then relies on the *steering mechanisms* of (1.3), mainly encoded in properties of ESTIMATE and MARK, to finally prove $(\hat{u}_{\infty}, \hat{y}_{\infty}, \hat{p}_{\infty}) = (\hat{u}, \hat{y}, \hat{p})$.

We shortly comment on an existing convergence result for constrained optimal control problems given in [2]. It is based on some non-degeneracy assumptions on the continuous and the discrete problems and a smallness assumption on the maximal mesh-size of \mathcal{G}_0 . Our approach does not require any of these assumptions and it is valid for a larger class of adaptive algorithms. In addition, it can easily be extended in several directions; compare with Sect. 5.

2 Aposteriori Error Estimation

In this section we shortly summarize our findings from [6, 7] providing a unifying framework for the aposteriori error analysis for control constrained optimal control problems. In what follows we shall use $a \lesssim b$ for $a \leq Cb$ with a constant C that may depend on data of (1.1) and the shape regularity of the grids in \mathbb{G} but not on a and b . We shall write $a \simeq b$ whenever $a \lesssim b \lesssim a$.

2.1 First Order Optimality Systems

The analysis in [6] is based on the characterization of the solutions by the first order optimality systems. We let $S, S^*: \mathbb{U} \rightarrow \mathbb{Y}$ be the solution operators of the state and the adjoint equation, i.e., for any $u \in \mathbb{U}$ we have

$$Su \in \mathbb{Y} : \quad \mathcal{B}[Su, v] = \langle u, v \rangle \quad \forall v \in \mathbb{Y} \quad (2.1)$$

and for any $g \in \mathbb{U}$ we have

$$S^*g \in \mathbb{Y} : \quad \mathcal{B}[v, S^*g] = \langle g, v \rangle \quad \forall v \in \mathbb{Y}. \quad (2.2)$$

We denote by $\Pi: \mathbb{U} \rightarrow \mathbb{U}^{\text{ad}}$ the nonlinear projection operator such that $\Pi(p)$ is the best approximation of $-\frac{1}{\alpha}p$ in \mathbb{U}^{ad} , i.e.,

$$\Pi(p) \in \mathbb{U}^{\text{ad}} : \quad \langle \alpha \Pi(p) + p, \Pi(p) - u \rangle \leq 0 \quad \forall u \in \mathbb{U}^{\text{ad}}. \quad (2.3)$$

Utilizing these operators, the continuous solution $(\hat{u}, \hat{y}, \hat{p}) \in \mathbb{U}^{\text{ad}} \times \mathbb{Y} \times \mathbb{Y}$ is the unique solution of the coupled nonlinear system

$$\hat{y} = S\hat{u}, \quad \hat{p} = S^*(\hat{y} - y_d), \quad \hat{u} = \Pi(\hat{p}). \quad (2.4)$$

For $\mathcal{G} \in \mathbb{G}$ we next define $S_{\mathcal{G}}, S_{\mathcal{G}}^*: \mathbb{U} \rightarrow \mathbb{Y}(\mathcal{G})$ to be the discrete solution operators for (2.1) and (2.2), i.e., for any $u \in \mathbb{U}$ we have

$$S_{\mathcal{G}}u \in \mathbb{Y}(\mathcal{G}) : \quad \mathcal{B}[S_{\mathcal{G}}u, V] = \langle u, V \rangle \quad \forall V \in \mathbb{Y}(\mathcal{G}), \quad (2.5)$$

and for any $g \in \mathbb{U}$ we have

$$S_{\mathcal{G}}^*g \in \mathbb{Y}(\mathcal{G}) : \quad \mathcal{B}[V, S_{\mathcal{G}}^*g] = \langle g, V \rangle \quad \forall V \in \mathbb{Y}(\mathcal{G}). \quad (2.6)$$

The discrete solution $(\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}) \in \mathbb{U}^{\text{ad}} \times \mathbb{Y}(\mathcal{G}) \times \mathbb{Y}(\mathcal{G})$ is then uniquely characterized by

$$\hat{Y}_{\mathcal{G}} = S_{\mathcal{G}}\hat{U}_{\mathcal{G}}, \quad \hat{P}_{\mathcal{G}} = S_{\mathcal{G}}^*(\hat{Y}_{\mathcal{G}} - y_d), \quad \hat{U}_{\mathcal{G}} = \Pi(\hat{P}_{\mathcal{G}}). \quad (2.7)$$

Note, that this variational discretization of Hinze requires the evaluation of the *continuous* projection operator Π for discrete functions $P \in \mathbb{Y}(\mathcal{G})$.

We have $\|S\|, \|S^*\|, \|S_{\mathcal{G}}\|, \|S_{\mathcal{G}}^*\| \leq C_F$, employing coercivity of \mathcal{B} with constant 1 in combination with the Friedrichs inequality $\|v\|_{2;\Omega} \leq C_F \|\nabla v\|_{2;\Omega}$ for $v \in \mathring{H}^1(\Omega)$.

2.2 Basic Error Equivalence

The main obstacle in the a posteriori error analysis encountered for instance in [3, 10] can be explained as follows. One would like to exploit Galerkin orthogonality in the linear state equation (2.1) and the adjoint equation (2.2). However, we observe that triplet $(\hat{U}_G, \hat{Y}_G, \hat{P}_G)$ is the Galerkin approximation to the triplet $(\hat{u}, \hat{y}, \hat{p})$ but \hat{Y}_G is not the Galerkin approximation to the solution \hat{u} of the linear problem (2.1) since we have $\hat{y} = S\hat{u}$ but not $\hat{y} = S\hat{U}_G$. The same argument applies to the adjoint states. This observation shows that we cannot directly employ Galerkin orthogonality for single components of (2.4) and the nonlinearity in (2.3) prevents us from making use of Galerkin orthogonality for the system (2.4). The resort to this problem is given by the following result from [6, Theorem 2.2].

Proposition 2.1 (Basic error equivalence). *If we set $\mathbb{W} = \mathbb{U} \times \mathbb{Y} \times \mathbb{Y}$ we have for $\bar{y} = S\hat{U}_G$ and $\bar{p} = S^*(\hat{Y}_G - y_d)$ the basic error equivalence*

$$\|(\hat{U}_G, \hat{Y}_G, \hat{P}_G) - (\hat{u}, \hat{p}, \hat{y})\|_{\mathbb{W}} \simeq \|(\hat{Y}_G, \hat{P}_G) - (\bar{y}, \bar{p})\|_{\mathbb{Y} \times \mathbb{Y}}.$$

For the problem under consideration, the constant hidden in \simeq depends on α^{-1} . For general \mathcal{B} it will in addition depend on the inf-sup constant of \mathcal{B} . Employing this error equivalence it is sufficient to construct a reliable and efficient estimator for the right hand side $\|(\hat{Y}_G, \hat{P}_G) - (\bar{y}, \bar{p})\|_{\mathbb{Y} \times \mathbb{Y}}$. The functions \bar{y} and \bar{p} are solutions to the linear problems (2.1) and (2.2) with given source \hat{U}_G and $\hat{Y}_G - y_d$, respectively. They play a similar role as the *elliptic reconstruction* used in the a posteriori error analysis of parabolic problems; compare with [11].

2.3 A posteriori Error Estimation

We realize that \hat{Y}_G is the Galerkin approximation to \bar{y} and \hat{P}_G the one to \bar{p} . We therefore can directly employ (existing) estimators for the linear problems (2.1) and (2.2) and their sum then constitutes an estimator for the optimal control problem; compare with [6, Theorem 3.2]. For ease of presentation we focus here on the residual estimator. If σ is an interior side we denote by $[[\nabla y]]$ the flux of the normal derivative $\partial_{\bar{n}} y$ across σ . For any subset $\mathcal{G}' \subset \mathcal{G}$ we set $\Omega(\mathcal{G}') := \bigcup_{E \in \mathcal{G}'} E$ and for given $E \in \mathcal{G}$ we denote by $\mathcal{N}_G(E) \subset \mathcal{G}$ the subset consisting of E and its direct neighbors. Finally, we indicate by $\|\cdot\|_{\mathbb{W}(\omega)}$ the natural restriction of $\|\cdot\|_{\mathbb{W}}$ to a subset $\omega \subset \Omega$. We then have the following result.

Theorem 2.2 (A posteriori error control). *For $E \in \mathcal{G}$ we define the indicator*

$$\begin{aligned} \mathcal{E}_G^2((\hat{U}_G, \hat{Y}_G, \hat{P}_G); E) &:= h_E^2 \|\Delta \hat{Y}_G + \hat{U}_G\|_{2;E}^2 + h_E \|[[\nabla \hat{Y}_G]]\|_{2;\partial E \cap \Omega}^2 \\ &\quad + h_E^2 \|\Delta \hat{P}_G + (\hat{Y}_G - y_d)\|_{2;E}^2 + h_E \|[[\nabla \hat{P}_G]]\|_{2;\partial E \cap \Omega}^2. \end{aligned}$$

Then we have the global upper bound

$$\|(\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}) - (\hat{u}, \hat{p}, \hat{y})\|_{\mathbb{W}}^2 \lesssim \mathcal{E}_{\mathcal{G}}^2((\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}); \mathcal{G}) := \sum_{E \in \mathcal{G}} \mathcal{E}_{\mathcal{G}}^2((\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}); E).$$

For any $E \in \mathcal{G}$ we have the local lower bound

$$\begin{aligned} & \mathcal{E}_{\mathcal{G}}^2((\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}); E) \\ & \lesssim \|(\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}) - (\hat{u}, \hat{p}, \hat{y})\|_{\mathbb{W}(\Omega(\mathcal{N}_{\mathcal{G}}(E)))}^2 + \text{osc}_{\mathcal{G}}^2(\hat{U}_{\mathcal{G}}, y_d; \mathcal{N}_{\mathcal{G}}(E)), \end{aligned}$$

where

$$\text{osc}_{\mathcal{G}}^2(\hat{U}_{\mathcal{G}}, y_d; E) := h_E^2 (\|\hat{U}_{\mathcal{G}} - \mathbb{P}_{\mathcal{G}} \hat{U}_{\mathcal{G}}\|_{2; \Omega(\mathcal{N}_{\mathcal{G}}(E))}^2 + \|y_d - \mathbb{P}_{\mathcal{G}} y_d\|_{2; \Omega(\mathcal{N}_{\mathcal{G}}(E))}^2)$$

is the typical oscillation term with the L_2 -projection $\mathbb{P}_{\mathcal{G}}$ onto the set of discontinuous, piecewise polynomials of degree q over \mathcal{G} .

2.4 Bounds for the Residuals

We shortly comment on the derivation of the estimators for the linear problems and thereby recording an important intermediate estimate. For given $u \in \mathbb{U}$ we set $y = Su$ and let $Y = S_{\mathcal{G}}u$ be its Galerkin-approximation in $\mathbb{Y}(\mathcal{G})$. Defining the residual of the state equation (2.1) by

$$\langle \mathcal{R}(S_{\mathcal{G}}u; u), v \rangle = \langle \mathcal{R}(Y; u), v \rangle := \mathcal{B}[Y, v] - \langle u, v \rangle = \mathcal{B}[Y - y, v] \quad \forall v \in \mathbb{Y},$$

we find $\|\mathcal{R}(Y; u)\|_{\mathbb{Y}^*} \simeq \|Y - y\|_{\mathbb{Y}} = \|(S_{\mathcal{G}} - S)u\|_{\mathbb{Y}}$.

Employing Galerkin-orthogonality $\langle \mathcal{R}(Y; u), V \rangle = 0$ for all $V \in \mathbb{Y}(\mathcal{G})$ and using piecewise integration by parts we deduce for any $v \in \mathbb{Y}$ and $V \in \mathbb{V}(\mathcal{G})$ the bound

$$|\langle \mathcal{R}(Y; u), v \rangle| \leq \sum_{E \in \mathcal{G}} \|\Delta Y + u\|_{2; E} \|v - V\|_{2; E} + \frac{1}{2} \|[\![\nabla Y]\!] \|_{2; \partial E \cap \Omega} \|v - V\|_{2; \partial E}.$$

Using for $v \in \mathbb{Y}$ the Scott-Zhang interpolant $V \in \mathbb{V}(\mathcal{G})$ one obtains from interpolation estimates in H^1 by standard arguments the upper bound

$$\|Y - y\|_{\mathbb{Y}} \simeq \|\mathcal{R}(Y; u)\|_{\mathbb{Y}^*} \lesssim \left(\sum_{E \in \mathcal{G}} h_E^2 \|\Delta Y + u\|_{2; E}^2 + h_E \|[\![\nabla Y]\!] \|_{2; \partial E \cap \Omega}^2 \right)^{1/2}.$$

If v is smooth, i.e., $v \in H^2(\Omega) \cap \mathbb{Y}$, we may employ interpolation estimates in H^2 to obtain the *improved bound*

$$|\langle \mathcal{R}(Y; u), v \rangle| \lesssim \left(\sum_{E \in \mathcal{G}} h_E^2 (h_E^2 \|\Delta Y + u\|_{2;E}^2 + h_E \|\llbracket \nabla Y \rrbracket\|_{2;\partial E \cap \Omega}^2) \right)^{1/2} |v|_{H^2(\Omega)}. \tag{2.8}$$

Similar arguments apply to the adjoint problem. For given $g \in \mathbb{U}$ we set $p = S^*g$ and let $P = S_{\mathcal{G}}^*g$ be its Galerkin-approximation in $\mathbb{Y}(\mathcal{G})$. For the residual of (2.2), defined by

$$\langle \mathcal{R}^*(S_{\mathcal{G}}^*g; g), v \rangle = \langle \mathcal{R}^*(P; g), v \rangle := \mathcal{B}[v, P] - \langle g, v \rangle = \mathcal{B}[v, P - p] \quad \forall v \in \mathbb{Y},$$

we have

$$|\langle \mathcal{R}^*(P; g), v \rangle| \lesssim \left(\sum_{E \in \mathcal{G}} h_E^{2s} (h_E^2 \|\Delta P + g\|_{2;E}^2 + h_E \|\llbracket \nabla P \rrbracket\|_{2;\partial E \cap \Omega}^2) \right)^{1/2} |v|_{H^{s+1}(\Omega)} \tag{2.9}$$

for any $v \in H^{s+1}(\Omega) \cap \mathbb{Y}$, $s = 0, 1$. With $s = 0$ we may deduce the upper bound for $\|(S_{\mathcal{G}}^* - S^*)g\|_{\mathbb{Y}} = \|P - p\|_{\mathbb{Y}} \simeq \|\mathcal{R}^*(P; g)\|_{\mathbb{Y}^*}$. The choice $s = 1$ yields the improved estimate for the adjoint problem. Equations (2.8) and (2.9) will become important in Sect. 4 to access *local density* of adaptively generated finite element spaces; compare also with [14, Remark 3.4].

3 Convergence 1: Trusting Stability

In this section we start with the convergence analysis, where we first focus on stability properties of the algorithm that do not depend on the particular decisions taken in MARK. Hereafter, $\{\mathcal{G}_k, (\hat{U}_k, \hat{Y}_k, \hat{P}_k)\}_{k \geq 0}$ is the sequence of grids and discrete solutions generated by (1.3). For ease of notation we use for $k \geq 0$ the short hands $\mathbb{Y}_k = \mathbb{Y}(\mathcal{G}_k)$, $\hat{U}_k = \hat{U}_{\mathcal{G}_k}$, $S_k = S_{\mathcal{G}_k}$ etc.

3.1 A First Limit

Using piecewise polynomials in combination with refinement by bisection leads to nested spaces, i.e., $\mathbb{Y}_k \subset \mathbb{Y}_{k+1}$. This allows us to define the limiting space

$$\mathbb{Y}_{\infty} = \overline{\bigcup_{k \geq 0} \mathbb{Y}_k}^{\|\cdot\|_{\mathbb{Y}}},$$

which is exactly the space that is approximated by the adaptive iteration. It is closed in \mathbb{Y} and therefore a Hilbert space. Consequently, the limiting optimal control problem

$$\begin{aligned} \min_{(u,y) \in \mathbb{U}^{\text{ad}} \times \mathbb{Y}_\infty} & \frac{1}{2} \|y - y_d\|_{\mathbb{U}}^2 + \frac{\alpha}{2} \|u\|_{\mathbb{U}}^2 \\ \text{subject to} & \quad y \in \mathbb{Y}_\infty : \quad \mathcal{B}[y, v] = \langle u, v \rangle \quad v \in \mathbb{Y}_\infty \end{aligned} \quad (3.1)$$

admits a unique solution $(\hat{u}_\infty, \hat{y}_\infty) \in \mathbb{U}^{\text{ad}} \times \mathbb{Y}_\infty$. If $S_\infty, S_\infty^* : \mathbb{U} \rightarrow \mathbb{Y}_\infty$ denote the solution operators of the state respectively the adjoint equation in \mathbb{Y}_∞ the associated first order optimality system reads

$$\hat{y}_\infty = S_\infty \hat{u}_\infty, \quad \hat{p}_\infty = S_\infty^* (\hat{y}_\infty - y_d), \quad \hat{u}_\infty = \Pi(\hat{p}_\infty). \quad (3.2)$$

We next show that in fact (3.1) is the limiting problem of the adaptive iteration (1.3) in that $(\hat{U}_k, \hat{Y}_k, \hat{P}_k) \rightarrow (\hat{u}_\infty, \hat{y}_\infty, \hat{p}_\infty)$. An important ingredient for this proof is the following crucial property of the adaptive algorithm shown in [1, Lemma 6.1] and [12, Lemma 4.2].

Proposition 3.1 (Convergence of solution operators). *For any $u, g \in \mathbb{U}$ we have $S_k u \rightarrow S_\infty u$ and $S_k^* g \rightarrow S_\infty^* g$ in \mathbb{Y} as $k \rightarrow \infty$.*

We next show convergence $\hat{U}_k \rightarrow \hat{u}_\infty$. In this step we have to deal with the nonlinearity of the constrained optimal control problem.

Lemma 3.2 (Convergence of the controls). *The discrete controls $\{\hat{U}_k\}_{k \geq 0}$ converge strongly to \hat{u}_∞ , i.e.,*

$$\lim_{k \rightarrow \infty} \|\hat{U}_k - \hat{u}_\infty\|_{\mathbb{U}} = 0.$$

Proof. Since both $\hat{U}_k = \Pi(\hat{P}_k)$ and $\hat{u}_\infty = \Pi(\hat{p}_\infty)$ are feasible, i.e., $\hat{U}_k, \hat{u}_\infty \in \mathbb{U}^{\text{ad}}$, the definition of Π in (2.3) yields

$$\begin{aligned} \alpha \|\hat{U}_k - \hat{u}_\infty\|_{2;\Omega}^2 &= \langle \alpha \hat{u}_\infty + \hat{p}_\infty, \hat{u}_\infty - \hat{U}_k \rangle + \langle \alpha \hat{U}_k + \hat{P}_k, \hat{U}_k - \hat{u}_\infty \rangle \\ &\quad + \langle \hat{P}_k - \hat{p}_\infty, \hat{u}_\infty - \hat{U}_k \rangle \\ &\leq \langle \hat{P}_k - \hat{p}_\infty, \hat{u}_\infty - \hat{U}_k \rangle \\ &= \langle S_k^* (\hat{y}_\infty - y_d) - \hat{p}_\infty, \hat{u}_\infty - \hat{U}_k \rangle + \langle \hat{P}_k - S_k^* (\hat{y}_\infty - y_d), \hat{u}_\infty - \hat{U}_k \rangle. \end{aligned}$$

We next estimate the last two terms separately. For the first one we immediately obtain from $\hat{p}_\infty = S_\infty(\hat{y}_\infty - y_d)$ by the Cauchy-Schwarz and Young inequalities

$$\begin{aligned} \langle S_k^* (\hat{y}_\infty - y_d) - \hat{p}_\infty, \hat{u}_\infty - \hat{U}_k \rangle &= \langle (S_k^* - S_\infty^*) (\hat{y}_\infty - y_d), \hat{u}_\infty - \hat{U}_k \rangle \\ &\leq \frac{\alpha}{2} \|\hat{u}_\infty - \hat{U}_k\|_{2;\Omega}^2 + \frac{1}{2\alpha} \|(S_k^* - S_\infty^*) (\hat{y}_\infty - y_d)\|_{2;\Omega}^2. \end{aligned}$$

We next turn to the second term. Employing the definition of the solution operators S_k and S_k^* in (2.5) and (2.6) we use $\hat{P}_k = S_k^*(\hat{Y}_k - y_d) \in \mathbb{Y}_k$ and $\hat{y}_\infty = S_\infty \hat{u}_\infty$ to obtain

$$\begin{aligned} \langle \hat{P}_k - S_k^*(\hat{y}_\infty - y_d), \hat{u}_\infty - \hat{U}_k \rangle &= \langle \hat{u}_\infty - \hat{U}_k, S_k^*(\hat{Y}_k - \hat{y}_\infty) \rangle \\ &= \mathcal{B}[S_k(\hat{u}_\infty - \hat{U}_k), S_k^*(\hat{Y}_k - \hat{y}_\infty)] = \langle \hat{Y}_k - \hat{y}_\infty, S_k(\hat{u}_\infty - \hat{U}_k) \rangle \\ &= \langle \hat{Y}_k - \hat{y}_\infty, \hat{y}_\infty - \hat{Y}_k \rangle + \langle \hat{Y}_k - \hat{y}_\infty, (S_k - S_\infty)\hat{u}_\infty \rangle \\ &= -\|\hat{Y}_k - \hat{y}_\infty\|_{2;\Omega}^2 + \frac{1}{2}\|\hat{Y}_k - \hat{y}_\infty\|_{2;\Omega}^2 + \frac{1}{2}\|(S_k - S_\infty)\hat{u}_\infty\|_{2;\Omega}^2 \\ &\leq \frac{1}{2}\|(S_k - S_\infty)\hat{u}_\infty\|_{2;\Omega}^2. \end{aligned}$$

Combining the estimates we have shown

$$\alpha \|\hat{U}_k - \hat{u}_\infty\|_{2;\Omega}^2 \leq \frac{1}{\alpha} \|(S_k^* - S_\infty^*)(\hat{y}_\infty - y_d)\|_{2;\Omega}^2 + \|(S_k - S_\infty)\hat{u}_\infty\|_{2;\Omega}^2 \rightarrow 0$$

as $k \rightarrow \infty$ by Proposition 3.1. This finishes the proof. \square

Convergence $(\hat{U}_k, \hat{Y}_k, \hat{P}_k) \rightarrow (\hat{u}_\infty, \hat{y}_\infty, \hat{p}_\infty)$ is now a direct consequence of the linear theory in Proposition 3.1.

Proposition 3.3 (Convergence of discrete solutions). *The Galerkin approximations $\{(\hat{U}_k, \hat{Y}_k, \hat{P}_k)\}_{k \geq 0}$ converge strongly to the solution $(\hat{u}_\infty, \hat{y}_\infty, \hat{p}_\infty)$ of (3.1), i.e.,*

$$\lim_{k \rightarrow \infty} \|(\hat{U}_k, \hat{Y}_k, \hat{P}_k) - (\hat{u}_\infty, \hat{y}_\infty, \hat{p}_\infty)\|_{\mathbb{U} \times \mathbb{Y} \times \mathbb{Y}} = 0.$$

Proof. We already know $\|\hat{U}_k - \hat{u}_\infty\|_{\mathbb{U}} \rightarrow 0$ from Lemma 3.2. In combination with Proposition 3.1 this yields for the discrete states

$$\begin{aligned} \|\hat{Y}_k - \hat{y}_\infty\|_{\mathbb{Y}} &= \|S_k \hat{U}_k - S_\infty \hat{u}_\infty\|_{\mathbb{Y}} \leq \|S_k(\hat{U}_k - \hat{u}_\infty)\|_{\mathbb{Y}} + \|(S_k - S_\infty)\hat{u}_\infty\|_{\mathbb{Y}} \\ &\leq \|S_k\| \|\hat{U}_k - \hat{u}_\infty\|_{\mathbb{U}} + \|(S_k - S_\infty)\hat{u}_\infty\|_{\mathbb{Y}} \rightarrow 0, \end{aligned}$$

since $\|S_k\| \leq C_F$. Writing $\hat{P}_k - \hat{p}_\infty = S_k^*(\hat{Y}_k - \hat{y}_\infty) + (S_k^* - S_\infty^*)(\hat{y}_\infty - y_d)$ we finally deduce with the same arguments $\|\hat{P}_k - \hat{p}_\infty\|_{\mathbb{Y}} \rightarrow 0$. \square

The convergence of the discrete solutions directly yields a uniform bound on the estimators. The proof follows the ideas in [14, Lemma 3.3] accounting for the situation at hand and using the following important property. Let $\mathcal{G} \in \mathbb{G}$ be given. The finite overlap of the patches $\#\mathcal{N}_{\mathcal{G}}(E) \lesssim 1$ allows us to deduce for any $g \in L_2(\Omega)$ the bound

$$\sum_{E \in \mathcal{G}} \|g\|_{2;\Omega(\mathcal{N}_{\mathcal{G}}(E))}^2 = \sum_{E \in \mathcal{G}} \sum_{E' \in \mathcal{N}_{\mathcal{G}}(E)} \|g\|_{2;E'}^2 \lesssim \sum_{E \in \mathcal{G}} \|g\|_{2;E}^2 = \|g\|_{2;\Omega}^2. \quad (3.3)$$

The constant solely depends on shape-regularity of \mathcal{G} and thus on \mathcal{G}_0 .

Lemma 3.4 (Uniform estimator bound). *For all $k \geq 0$ we have*

$$\mathcal{E}_k((\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}); \mathcal{G}_k) \lesssim 1.$$

Proof. A scaled trace inequality in combination with an inverse estimate yields for the error indicators related to the state equation

$$h_E^2 \|\Delta \hat{Y}_k + \hat{U}_k\|_{2;E}^2 + h_E \|\llbracket \nabla \hat{Y}_k \rrbracket\|_{2;\partial E_k \cap \Omega}^2 \lesssim \|\nabla \hat{Y}_k\|_{2;\Omega(\mathcal{N}_{\mathcal{G}}(E))}^2 + \|\hat{U}_k\|_{2;E}^2.$$

This in turn implies by (3.3)

$$\sum_{E \in \mathcal{G}_k} h_E^2 \|\Delta \hat{Y}_k + \hat{U}_k\|_{2;E}^2 + h_E \|\llbracket \nabla \hat{Y}_k \rrbracket\|_{2;\partial E \cap \Omega}^2 \lesssim \|\nabla \hat{Y}_k\|_{2;\Omega}^2 + \|\hat{U}_k\|_{2;\Omega}^2 \lesssim 1,$$

since $\{\hat{U}_k, \hat{Y}_k\}_{k \geq 0}$ is bounded in $L_2(\Omega) \times \mathring{H}^1(\Omega)$. Similar arguments apply to the estimator contribution related to the adjoint problem. \square

3.2 A Second Limit

We next turn to the limit of the piecewise constant mesh-size function $h_k: \Omega \rightarrow \mathbb{R}$ of \mathcal{G}_k defined by $h_k|_E = |E|^{1/d}$, $E \in \mathcal{G}$. The behavior of the mesh-size function is directly related to the decomposition

$$\mathcal{G}_k^+ := \bigcap_{\ell \geq k} \mathcal{G}_\ell = \{E \in \mathcal{G}_k \mid E \in \mathcal{G}_\ell \ \forall \ell \geq k\}, \quad \text{and} \quad \mathcal{G}_k^0 := \mathcal{G}_k \setminus \mathcal{G}_k^+.$$

The set \mathcal{G}_k^+ contains all elements that are not refined after iteration k and we observe that the sequence $\{\mathcal{G}_k^+\}_{k \geq 0}$ is nested, i.e., $\mathcal{G}_\ell^+ \subset \mathcal{G}_k^+$ for all $k \geq \ell$. The set \mathcal{G}_k^0 contains all elements that are refined at least once more after iteration k ; in particular, $\mathcal{M}_k \subset \mathcal{G}_k^0$. Decomposing $\bar{\Omega} = \Omega_k^+ \cup \Omega_k^0 := \Omega(\mathcal{G}_k^+) \cup \Omega(\mathcal{G}_k^0)$ we have the following connection to the behavior of the mesh-size function shown in [12, Lemma 4.3 and Corollary 4.1].

Lemma 3.5 (Convergence of the mesh-size functions). *The mesh-size functions h_k converge uniformly to 0 in Ω_k^0 in the following sense*

$$\lim_{k \rightarrow \infty} \|h_k \chi_k^0\|_{\infty;\Omega} = \lim_{k \rightarrow \infty} \|h_k\|_{\infty;\Omega_k^0} = 0,$$

where $\chi_k^0 \in L_\infty(\Omega)$ the characteristic function of Ω_k^0 .

Combining convergence of the discrete solutions with the convergence of the mesh-size functions we see that the adaptive algorithm can monitor progress in the following sense.

Lemma 3.6 (Indicators of marked elements). *All indicators of marked elements vanish in the limit, this is,*

$$\lim_{k \rightarrow \infty} \max\{\mathcal{E}_{\mathcal{G}}((\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}); E) \mid E \in \mathcal{M}_k\} = 0.$$

Proof. For $k \geq 0$ pick up $E_k \in \arg \max\{\mathcal{E}_{\mathcal{G}}((\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}); E) \mid E \in \mathcal{M}_k\} \neq \emptyset$. We follow [14, Lemma 3.4] and show $\mathcal{E}_k((\hat{U}_k, \hat{Y}_k, \hat{P}_k); E_k) \rightarrow 0$.

Arguing as in the proof to Lemma 3.4 we find for the indicator contribution of the state equation

$$\begin{aligned} & h_E \|\Delta \hat{Y}_k + \hat{U}_k\|_{2;E_k} + h_E^{1/2} \|[\nabla \hat{Y}_k]\|_{2;\partial E_k \cap \Omega} \lesssim \|\nabla \hat{Y}_k\|_{2;\Omega(\mathcal{N}_k(E_k))} + \|\hat{U}_k\|_{2;E_k} \\ & \leq \|\nabla \hat{y}_{\infty}\|_{2;\Omega(\mathcal{N}_k(E_k))} + \|\hat{u}_{\infty}\|_{2;E_k} + \|\nabla(\hat{Y}_k - \hat{y}_{\infty})\|_{2;\Omega} + \|\hat{U}_k - \hat{u}_{\infty}\|_{2;\Omega} \rightarrow 0 \end{aligned}$$

as $k \rightarrow \infty$ for the following reasons: By Assumption 1.1 (4) all elements in \mathcal{M}_k are refined, which implies $E_k \in \mathcal{G}_k^0$. Local quasi-uniformity of \mathcal{G}_k in combination with Lemma 3.5 therefore yields $|\Omega(\mathcal{N}_k(E_k))| \lesssim |E_k| \leq \|h_k\|_{\infty;\Omega_k^0}^d \rightarrow 0$. Consequently, the first two terms of the right hand side vanish by continuity of $\|\cdot\|_{2;\Omega}$ with respect to the Lebesgue measure. The last two terms converge to 0 by Proposition 3.3. The same arguments apply to the indicator contribution of the adjoint equation, which in summary yields $\mathcal{E}_{\mathcal{G}}((\hat{U}_k, \hat{Y}_k, \hat{P}_k); E_k) \rightarrow 0$ as $k \rightarrow \infty$. \square

4 Convergence 2: Making the Right Decisions

In this section we verify the main result by showing $(\hat{U}_k, \hat{Y}_k, \hat{P}_k) \rightarrow (\hat{u}, \hat{y}, \hat{p})$ and $\mathcal{E}_k(\hat{U}_k, \hat{Y}_k, \hat{P}_k; \mathcal{G}_k) \rightarrow 0$. Error convergence requires appropriate decisions in the adaptive iteration, which we have summarized in Assumption 1.1. Estimator convergence is then a consequence of local efficiency as stated in Theorem 2.2.

4.1 Convergence of the Indicators

We first show that the maximal indicator of all elements vanishes in the limit.

Lemma 4.1 (Convergence of the indicators). *The maximal indicator vanishes in the limit, this is,*

$$\lim_{k \rightarrow \infty} \max\{\mathcal{E}_{\mathcal{G}}((\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}); E) \mid E \in \mathcal{G}_k\} = 0.$$

Proof. Combining the assumption on marking in Assumption 1.1 (3) with the behavior of the indicators on marked elements, which we have analyzed in Lemma 3.6, we find

$$\begin{aligned} \max\{\mathcal{E}_{\mathcal{G}}((\hat{U}_k, \hat{Y}_k, \hat{P}_k); E) \mid E \in \mathcal{G}_k\} \\ \leq \max\{\mathcal{E}_{\mathcal{G}}((\hat{U}_k, \hat{Y}_k, \hat{P}_k); E) \mid E \in \mathcal{M}_k\} \rightarrow 0 \end{aligned}$$

as $k \rightarrow \infty$. \square

4.2 Convergence of the Residuals

We next show that residuals of state and adjoint equation in the limiting first order optimality system (3.2) vanish. The proof adapts the techniques from [14, Proposition 3.1] to the situation at hand.

Proposition 4.2 (Convergence of the residual). *For the residuals \mathcal{R} of (2.1) and \mathcal{R}^* of (2.2) we have*

$$\mathcal{R}(\hat{y}_{\infty}; \hat{u}_{\infty}) = \mathcal{R}^*(\hat{p}_{\infty}; \hat{y}_{\infty} - y_d) = 0 \quad \text{in } \mathbb{Y}^* = H^{-1}(\Omega).$$

Particularly, $\hat{y}_{\infty} = S\hat{u}_{\infty}$ and $\hat{p}_{\infty} = S^*(\hat{y}_{\infty} - y_d)$.

Proof. We prove the claim for \mathcal{R} . The assertion for \mathcal{R}^* follows along the same lines. Using a density argument we only have to show $\langle \mathcal{R}(\hat{y}_{\infty}; \hat{u}_{\infty}), v \rangle = 0$ for all $v \in H^2(\Omega) \cap \hat{H}^1(\Omega)$.

Suppose any pair $k \geq \ell$. Then we have the inclusion $\mathcal{G}_{\ell}^+ \subset \mathcal{G}_k^+ \subset \mathcal{G}_k$ and the sub-triangulation $\mathcal{G}_k \setminus \mathcal{G}_{\ell}^+$ of \mathcal{G}_k covers the sub-domain $\Omega_{\ell}^0 = \Omega(\mathcal{G}_{\ell}^0)$, i.e., we can write $\Omega_{\ell}^0 = \Omega(\mathcal{G}_k \setminus \mathcal{G}_{\ell}^+)$. Moreover, $\|h_k\|_{\infty; \Omega_{\ell}^+} \lesssim 1$ and $\|h_k\|_{\infty; \Omega_{\ell}^0} \leq \|h_{\ell}\|_{\infty; \Omega_{\ell}^0}$.

Pick up any $v \in H^2(\Omega) \cap \hat{H}^1(\Omega)$ with $|v|_{H^2(\Omega)} = 1$. We next utilize the improved bound (2.8) for \mathcal{R} , decompose $\mathcal{G}_k = \mathcal{G}_{\ell}^+ \cup (\mathcal{G}_k \setminus \mathcal{G}_{\ell}^+)$, and recall Lemma 3.4 to bound

$$\begin{aligned} \langle \mathcal{R}(\hat{Y}_k; \hat{U}_k), v \rangle^2 &\lesssim \sum_{E \in \mathcal{G}_{\ell}^+} h_E^2 (h_E^2 \|\Delta \hat{Y}_k + \hat{U}_k\|_{2;E}^2 + h_E \|\llbracket \nabla \hat{Y}_k \rrbracket\|_{2; \partial E \cap \Omega}^2) \\ &\quad + \sum_{E \in \mathcal{G}_k \setminus \mathcal{G}_{\ell}^+} h_E^2 (h_E^2 \|\Delta \hat{Y}_k + \hat{U}_k\|_{2;E}^2 + h_E \|\llbracket \nabla \hat{Y}_k \rrbracket\|_{2; \partial E \cap \Omega}^2) \\ &\lesssim \mathcal{E}_k^2((\hat{U}_k, \hat{Y}_k, \hat{P}_k); \mathcal{G}_{\ell}^+) + \|h_{\ell}\|_{\infty; \Omega_{\ell}^0}^2 \mathcal{E}_k^2((\hat{U}_k, \hat{Y}_k, \hat{P}_k); \mathcal{G}_k \setminus \mathcal{G}_{\ell}^+) \\ &\lesssim \mathcal{E}_k^2((\hat{U}_k, \hat{Y}_k, \hat{P}_k); \mathcal{G}_{\ell}^+) + \|h_{\ell}\|_{\infty; \Omega_{\ell}^0}^2 \stackrel{!}{\leq} 2\varepsilon \end{aligned}$$

for any $\varepsilon > 0$. This can be seen as follows: By Lemma 3.5 we may first choose ℓ large such that $\|h_\ell\|_{\infty; \Omega_\ell^0}^2 \leq \varepsilon$. After fixing ℓ the “point-wise” convergence of the indicators in Lemma 4.1 allows us then to choose a suitable $k \geq \ell$ with $\mathcal{E}_k^2((\hat{U}_k, \hat{Y}_k, \hat{P}_k); \mathcal{G}_\ell^+) \leq \varepsilon$. This yields for any fixed $v \in H^2(\Omega) \cap \hat{H}^1(\Omega)$

$$\langle \mathcal{R}(\hat{y}_\infty; \hat{u}_\infty), v \rangle = \lim_{k \rightarrow \infty} \langle \mathcal{R}(\hat{Y}_k; \hat{U}_k), v \rangle = 0,$$

observing that \mathcal{R} is continuous with respect to its arguments and recalling the convergence $(\hat{U}_k, \hat{Y}_k) \rightarrow (\hat{u}_\infty, \hat{y}_\infty)$ shown in Proposition 3.3. Since v is arbitrary we have shown $\mathcal{R}(\hat{y}_\infty; \hat{u}_\infty) = 0$ in \mathbb{Y}^* . This in turn implies $\hat{y}_\infty = S\hat{u}_\infty$ and finishes the proof. \square

4.3 Convergence of Error and Estimator

We are now in the position to prove the main result, where we again use the abbreviation $\mathbb{W} = \mathbb{U} \times \mathbb{Y} \times \mathbb{Y}$.

Proof of Theorem 1.2. Combining Propositions 2.1, 3.3, and 4.2 we obtain

$$\begin{aligned} \lim_{k \rightarrow \infty} \|(\hat{U}_k, \hat{Y}_k, \hat{P}_k) - (\hat{u}, \hat{p}, \hat{y})\|_{\mathbb{W}} &\simeq \lim_{k \rightarrow \infty} \|(\hat{Y}_k, \hat{P}_k) - (S\hat{U}_k, S^*(\hat{Y}_k - y_d))\|_{\mathbb{Y} \times \mathbb{Y}} \\ &= \|(\hat{y}_\infty, \hat{p}_\infty) - (S\hat{u}_\infty, S^*(\hat{y}_\infty - y_d))\|_{\mathbb{Y} \times \mathbb{Y}} = 0. \end{aligned}$$

This shows convergence of the error.

To show convergence of the estimator we decompose for $k \geq \ell$ as in the proof to Proposition 4.2

$$\mathcal{E}_k^2((\hat{U}_k, \hat{Y}_k, \hat{P}_k); \mathcal{G}_k) = \mathcal{E}_k^2((\hat{U}_k, \hat{Y}_k, \hat{P}_k); \mathcal{G}_\ell^+) + \mathcal{E}_k^2((\hat{U}_k, \hat{Y}_k, \hat{P}_k); \mathcal{G}_k \setminus \mathcal{G}_\ell^+).$$

We first bound the second term on the right hand side. The local lower bound of Theorem 2.2 in combination with the finite overlap of the patches $\mathcal{N}_k(E)$ allows us to bound

$$\begin{aligned} &\mathcal{E}_k^2((\hat{U}_k, \hat{Y}_k, \hat{P}_k); \mathcal{G}_k \setminus \mathcal{G}_\ell^+) \\ &\lesssim \|(\hat{U}_k, \hat{Y}_k, \hat{P}_k) - (\hat{u}, \hat{p}, \hat{y})\|_{\mathbb{W}}^2 + \sum_{E \in \mathcal{G}_k \setminus \mathcal{G}_\ell^+} \text{osc}_k^2(\hat{U}_k, y_d; E) \\ &\lesssim \|(\hat{U}_k, \hat{Y}_k, \hat{P}_k) - (\hat{u}, \hat{p}, \hat{y})\|_{\mathbb{W}}^2 + \|h_\ell\|_{\infty; \Omega_\ell^0}^2 (\|\hat{U}_k\|_{2; \Omega}^2 + \|y_d\|_{2; \Omega}^2), \end{aligned}$$

using (3.3) and the rough estimate

$$\begin{aligned} \text{osc}_k^2(\hat{U}_k, y_d; E) &= h_E^2 (\|\hat{U}_k - \mathbb{P}_{\mathcal{G}_k} \hat{U}_k\|_{2;\Omega(\mathcal{N}_{\mathcal{G}}(E))}^2 + \|y_d - \mathbb{P}_{\mathcal{G}_k} y_d\|_{2;\Omega(\mathcal{N}_{\mathcal{G}}(E))}^2) \\ &\leq \|h_\ell\|_{\infty;\Omega_\ell^0}^2 (\|\hat{U}_k\|_{2;\Omega(\mathcal{N}_k(E))}^2 + \|y_d\|_{2;\Omega(\mathcal{N}_k(E))}^2). \end{aligned}$$

Since $\|\hat{U}_k\|_{2;\Omega}^2 + \|y_d\|_{2;\Omega}^2 \lesssim 1$ we find

$$\begin{aligned} \mathcal{E}_k^2((\hat{U}_k, \hat{Y}_k, \hat{P}_k); \mathcal{G}_k) &\lesssim \mathcal{E}_k^2((\hat{U}_k, \hat{Y}_k, \hat{P}_k); \mathcal{G}_\ell^+) \\ &\quad + \|(\hat{U}_k, \hat{Y}_k, \hat{P}_k) - (\hat{u}, \hat{p}, \hat{y})\|_{\mathbb{W}}^2 + \|h_\ell\|_{\infty;\Omega_\ell^0}^2. \end{aligned}$$

By Lemma 3.5 the last term $\|h_\ell\|_{\infty;\Omega_\ell^0}^2$ can be made small by choosing ℓ large. After fixing ℓ we may choose as in the proof to Proposition 4.2 $k \geq \ell$ such that $\mathcal{E}_k^2((\hat{U}_k, \hat{Y}_k, \hat{P}_k); \mathcal{G}_\ell^+)$ is small. Moreover, the error convergence established above implies that the middle term $\|(\hat{U}_k, \hat{Y}_k, \hat{P}_k) - (\hat{u}, \hat{p}, \hat{y})\|_{\mathbb{W}}^2$ is small too, if we possibly increase k further. In summary, for any $\varepsilon > 0$ we find a k such that

$$\mathcal{E}_k((\hat{U}_k, \hat{Y}_k, \hat{P}_k); \mathcal{G}_k) \leq \varepsilon.$$

This yields $\mathcal{E}_k((\hat{U}_k, \hat{Y}_k, \hat{P}_k); \mathcal{G}_k) \rightarrow 0$ as $k \rightarrow \infty$ and finishes the proof. □

5 Extensions and Outlook

The presented theory has been extended into several directions in the PhD thesis of the first author [5].

5.1 General Linear-Quadratic Optimal Control Problem

The abstract framework can be found in [6, §2.1] and may be summarized as follows. We can allow for continuous, non-coercive bilinear forms $\mathcal{B}: \mathbb{Y} \times \mathbb{Y} \rightarrow \mathbb{R}$ that satisfy an inf-sup condition. This setting includes saddle point problems like the Stokes system and other mixed formulations. More general objectives $\psi(y)$ can replace the simple tracking type functional $\|y - y_d\|_{2;\Omega}^2$. The functional ψ has to be quadratic and strictly convex. Its Fréchet-derivative ψ' has to satisfy a Lipschitz-condition. We may also consider any type of control space such that $\mathbb{Y} \hookrightarrow \mathbb{U} \hookrightarrow \mathbb{Y}^*$ is a Gelfand triple. This then covers more general cases of distributed control as well as Neumann-boundary control.

Admitting a general class of PDE constraints requires appropriate assumptions on the estimators for the linear problems (2.1) and (2.2). Quite weak assumption

are summarized in [14, §2.2.3] comprising other estimators like the hierarchical estimator, an estimator based on local problems on stars, an equilibrated residual estimator, and the ZZ-estimator; compare for instance with [8] for a detailed description of the diverse estimators. We may also weaken the assumption on marking to include marking strategies that adaptively focus on specific estimator contributions, like the indicators for the error in the state or adjoint equation. Such strategies are used in a comparison of adaptive strategies for optimal control problems in [6, §6]. We refer to [14, §2.2.4 and §5] for a sufficient and necessary assumption on marking.

Most of the changes in the presented analysis are then concentrated in the proof to Lemma 3.2. This proof gets inevitably more involved due to the general structure of ψ , where one has to appropriately use convexity of ψ . All other statements can be proven using similar arguments with minor adjustments.

5.2 Discretized Control

Up to now we have concentrated on the variational discretization of Hinze [4]. Here, the precise structure of the set of admissible controls \mathbb{U}^{ad} is not of importance. The actual computation of a discrete solution yet requires the exact computation of $\Pi(P)$ for a discrete function $P \in \mathbb{Y}(\mathcal{G})$. This typically gives restrictions on \mathbb{U}^{ad} , like box-constraints with piecewise constant obstacles.

Very often the control space \mathbb{U} is discretized by a conforming finite element space $\mathbb{U}(\mathcal{G})$. Upon setting $\mathbb{U}^{\text{ad}}(\mathcal{G}) := \mathbb{U}^{\text{ad}} \cap \mathbb{U}(\mathcal{G})$ and assuming that $\mathbb{U}^{\text{ad}}(\mathcal{G})$ is non-empty we can define a discrete projection operator $\Pi_{\mathcal{G}}: \mathbb{U} \rightarrow \mathbb{U}^{\text{ad}}(\mathcal{G})$ for $p \in \mathbb{U}$ by

$$\Pi_{\mathcal{G}}(p) \in \mathbb{U}^{\text{ad}}(\mathcal{G}) : \quad \langle \alpha \Pi_{\mathcal{G}}(p) + p, \Pi_{\mathcal{G}}(p) - U \rangle \leq 0 \quad \forall U \in \mathbb{U}^{\text{ad}}(\mathcal{G}).$$

An efficient computation of $\Pi_{\mathcal{G}}$ benefits from a simple structure of \mathbb{U}^{ad} and a suitable discrete control space $\mathbb{U}(\mathcal{G})$.

We can still consider the general setting of the previous paragraph. However, the analysis of adaptive finite elements for discretized controls gets painstakingly more laborious at several instances that we shortly list.

1. The right hand side in the basic error equivalence in Proposition 2.1 has to be extended by the term $\|\hat{U}_{\mathcal{G}} - \Pi(\hat{U}_{\mathcal{G}})\|_{\mathbb{U}}$ resulting in

$$\|(\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}) - (\hat{u}, \hat{p}, \hat{y})\|_{\mathbb{W}} \simeq \|(\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}) - (\Pi(\hat{P}_{\mathcal{G}}), S(\hat{U}_{\mathcal{G}}), S^* \psi'(\hat{Y}_{\mathcal{G}}))\|_{\mathbb{W}};$$

compare with [6, Theorem 2.2]

2. As a consequence, the element indicators of the estimator in Theorem 2.2 have to be enriched by a term $\|\hat{U}_{\mathcal{G}} - \Pi(\hat{P}_{\mathcal{G}})\|_{\mathbb{U}(E)}^2 = \|\Pi_{\mathcal{G}}(\hat{P}_{\mathcal{G}}) - \Pi(\hat{P}_{\mathcal{G}})\|_{\mathbb{U}(E)}^2$ to grant reliability of the estimator [6, Theorem 3.2]. Frequently this term is estimated further in order to completely avoid the computation of the continuous projection

operator $\Pi(\hat{P}_{\mathcal{G}})$ [3, 10]. This typically results in a non-efficient estimator; compare with [6, Remark 6.1].

3. Nesting of spaces $\mathbb{Y}_k \subset \mathbb{Y}_{k+1}$ is essential to verify with current techniques the point-wise convergence of the discrete solution operators in Proposition 3.1, i.e., $S_k \rightarrow S_\infty$ and $S_k^* \rightarrow S_\infty^*$ as $k \rightarrow \infty$.

Likewise, nesting $\mathbb{U}_k^{\text{ad}} \subset \mathbb{U}_{k+1}^{\text{ad}}$ of the sets of discrete admissible controls is instrumental for proving $\Pi_k(\hat{P}_k) \rightarrow \hat{u}_\infty = \Pi_\infty(\hat{p}_\infty)$ in Lemma 3.2. This nesting poses restrictions on data describing the set of admissible controls \mathbb{U}^{ad} . Typically, such data has to be discrete over \mathcal{G}_0 . In the proof to Lemma 3.2 we additionally have to account for the typical situation $\hat{u}_\infty \notin \mathbb{U}_k^{\text{ad}}$. This increases substantially the complexity of the proof.

4. The finite element spaces $\{\mathbb{Y}_k\}_{k \geq 0}$ are “locally dense” in the subset “ $\Omega_\infty^0 := \lim_{k \rightarrow \infty} \Omega_k^0$ ” of Ω in that $\min_{V \in \mathbb{Y}_k} \|v - V\|_{\mathbb{Y}(\Omega_k^0)} \rightarrow 0$ as $k \rightarrow \infty$; compare with [14, Remark 3.4]. Philosophically speaking, the improved bounds (2.8) and (2.9) for the residuals allow us to access this local density for showing that the residuals $\mathcal{R}(\hat{y}_\infty; \hat{u}_\infty)$, $\mathcal{R}^*(\hat{p}_\infty; \psi'(\hat{y}_\infty)) \in \mathbb{Y}^*$ are not supported in Ω_∞^0 .

The additional contribution for the control error requires to establish the convergence

$$\lim_{k \rightarrow \infty} \|\hat{U}_k - \Pi(\hat{P}_k)\|_{\mathbb{U}(\Omega_k^0)} = \lim_{k \rightarrow \infty} \|\Pi_k(\hat{P}_k) - \Pi(\hat{P}_k)\|_{\mathbb{U}(\Omega_k^0)} \rightarrow 0. \tag{5.1}$$

For $\mathbb{U} = L_2$ and piece-wise constant box-constraints in combination with a discontinuous or a continuous, piecewise linear control discretization one can verify (5.1) employing local density of $\{\mathbb{U}_k\}_{k \geq 0}$ and point-wise properties of Π ; compare with [5, §8.4.2 and §8.4.3]. A characterization of properties of Π and $\mathbb{U}(\mathcal{G})$ that ensure (5.1) is a challenging question and topic of future research.

5. The proof of the estimator convergence in Theorem 1.2 strongly relies on local efficiency of the indicators as stated in Theorem 2.2. For discretized control this requires $\|\hat{U}_{\mathcal{G}} - \Pi(\hat{P}_{\mathcal{G}})\|_{\mathbb{U}(E)}$ to be locally efficient, which can be shown if Π and $\Pi_{\mathcal{G}}$ are locally Lipschitz continuous with uniformly bounded Lipschitz constants. This is typically true in case of distributed control.

In case of Neumann boundary control Lipschitz continuity of Π and $\Pi_{\mathcal{G}}$ involves the trace operator $T : H^1(\Omega) \rightarrow L_2(\partial\Omega)$. We may therefore show *global* efficiency for $\|\hat{U}_{\mathcal{G}} - \Pi(\hat{P}_{\mathcal{G}})\|_{L_2(\partial\Omega)}$ using the trace inequality on Ω . An estimate of $\|\hat{U}_{\mathcal{G}} - \Pi(\hat{P}_{\mathcal{G}})\|_{L_2(\partial E \cap \partial\Omega)}$ needs a local trace inequality on E . The typical scaling arguments yield negative powers of the local mesh-size and thereby ruling out local efficiency. As a consequence, we still can verify the error convergence of Theorem 1.2 but a proof of estimator convergence may require new techniques in that case.

References

1. I. Babuška, M. Vogelius, Feedback and adaptive finite element solution of one-dimensional boundary value problems. *Numer. Math.* **44**, 75–102 (1984)
2. A. Gaevskaya, R.H.W. Hoppe, Y. Iliash, M. Kieweg, Convergence analysis of an adaptive finite element method for distributed control problems with control constraints, in *Control of Coupled Partial Differential Equations*. Volume 155 of International Series of Numerical Mathematics (Birkhäuser, Basel, 2007), pp. 47–68
3. M. Hintermüller, R.H.W. Hoppe, Y. Iliash, M. Kieweg, An a posteriori error analysis of adaptive finite element methods for distributed elliptic control problems with control constraints. *ESAIM Control Optim. Calc. Var.* **14**, 540–560 (2008)
4. M. Hinze, A variational discretization concept in control constrained optimization: the linear-quadratics case. *Comput. Optim. Appl.* **30**, 45–61 (2005)
5. K. Kohls, An adaptive finite element method for control-constrained optimal control problems, PhD thesis, Fachbereich Mathematik, Universität Stuttgart, 2013
6. K. Kohls, A. Rösch, K.G. Siebert, A posteriori error analysis of optimal control problems with control constraints. *SIAM J. Control Optim.* **52**(3), 1832–1861 (2014)
7. K. Kohls, A. Rösch, K.G. Siebert, A posteriori error estimators for control constrained optimal control problems, in *Constrained Optimization and Optimal Control for Partial Differential Equations*, ed. by Leugering et al. International Series of Numerical Mathematics, vol. 160 (Birkhäuser/Springer Basel AG, Basel, 2012), pp. 431–443
8. C. Kreuzer, K.G. Siebert, Decay rates of adaptive finite elements with Dörfler marking. *Numer. Math.* **117**, 679–716 (2011)
9. J.L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*. Volume 170 of Die Grundlehren der Mathematischen Wissenschaften (Springer, Berlin/New York, 1971)
10. W. Liu, N. Yan, A posteriori error estimates for distributed convex optimal control problems. *Adv. Comput. Math.* **15**, 285–309 (2001)
11. C. Makridakis, R.H. Nochetto, Elliptic reconstruction and a posteriori error estimates for parabolic problems. *SIAM J. Numer. Anal.* **41**, 1585–1594 (2003)
12. P. Morin, K.G. Siebert, A. Veiser, A basic convergence result for conforming adaptive finite elements. *Math. Models Methods Appl. Sci.* **18**, 707–737 (2008)
13. A. Schmidt, K.G. Siebert, *Design of Adaptive Finite Element Software*. Volume 42 of Lecture Notes in Computational Science and Engineering (Springer, Berlin, 2005). The finite element toolbox ALBERTA, With 1 CD-ROM (Unix/Linux)
14. K.G. Siebert, A convergence proof for adaptive finite elements without lower bound. *IMA J. Numer. Anal.* **31**, 947–970 (2011)
15. F. Tröltzsch, *Optimal Control of Partial Differential Equations*. Volume 112 of Graduate Studies in Mathematics (American Mathematical Society, Providence, 2010). Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels