ruhr.paD
UA Ruhr Zentrum für
partielle Differentialgleichungen

Convergence of adaptive finite elements for optimal control problems with control constraints

K. Kohls, C. Kreuzer, A. Rösch and K.G. Siebert

Preprint 2016-16

CONVERGENCE OF ADAPTIVE FINITE ELEMENTS FOR OPTIMAL CONTROL PROBLEMS WITH CONTROL CONSTRAINS

KRISTINA KOHLS, CHRISTIAN KREUZER, 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. The result covers the variational discretization (Hinze) as well as control discretizations with piecewise discontinuous finite elements. Moreover, the presented theory can be applied to a large class of state equations, to boundary control and boundary observation.

1. Introduction

Convegence and optimality of Adaptive Finite Element Methods (AFEM) is a well studied topic for linear partial differential equations. Without claiming to be exhaustive we refer to [6, 21, 26, 3, 22, 25, 17, 5] as well as the overview article [23] and the references therein.

In contrast, the situation changes, however, when it comes to linear-quadratic optimal control problems with inequality constraints. Resorting to Dörflers marking strategy [6], the first linear convergence result [7] for constrained optimal control problems is based upon some non-degeneracy assumptions on the continuous and the discrete problems and a smallness assumption on the maximal mesh-size of $\mathcal{G}_0$. A smallness assumption on the coarse mesh was also used in the convergence and optimality result of [9]. However these conditions are practically not verifiable since they typically involve asymptotic estimates with unknown constants.

A new approach was proposed by Kohls, Rösch, and Siebert in [15] for optimal control problems with distributed control and the variational discretization by Hinze [11]. We emphasize that the presented theory does not require any smallness assumption or assumptions on the boundary between active and inactive sets and applies to rather general marking strategies. As a drawback, it does not guarantee convergence rates.

In this paper, we will further develop this new approach for a large class of control discretizations, state equations, error estimators, and marking strategies. The approach covers variational discretizations and piecewise discontinuous polynomial controls as well as allows us to deal with boundary controls or/and boundary observation. The restrictions on the linear equation are very general and therefore the results presented in this paper extend the known theory in a significant way.

Date: November 25, 2016.

1991 *Mathematics Subject Classification.* 65N30, 65N12, 49J20.

Key words and phrases. Adaptive finite elements, aposteriori error estimators, convergence analysis, optimal control, control constraints.

The paper is structured as follows. In Section 2 we state the assumptions and the main result. Section 3 is devoted to the framework of a posteriori error estimation. Convergence of the discrete solutions to a solution of an auxiliary problem is shown in Section 4. Convergence of the discrete solutions to the solution of the original optimal control problem and the convergence of the error estimator to zero is contained in Section 5. Finally, in Section 6, we will present some illustrating examples.

2. Statement of the Main Result

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

$$\begin{aligned} &\min_{(u,y)\in\mathbb{U}^{\mathrm{ad}}\times\mathbb{Y}} \mathcal{J}[u,y] = \psi(y) + \frac{\alpha}{2}\|u\|_{\mathbb{U}}^2 \\ &\text{subject to} \qquad y\in\mathbb{Y}: \quad \mathcal{B}[y,\,v] = \langle f+u,\,v\rangle_{\mathbb{Y}\times\mathbb{Y}^*} \qquad \forall v\in\mathbb{Y}, \end{aligned} \tag{2.1}$$

where $f \in \mathbb{Y}^*$ is a given functional. We have a particular interest in bilinear forms $\mathcal{B}$, that arise in the variational formulation of PDEs. Concrete examples of this type of problem are provided in Section 6 below.

For a bounded domain $\Omega \subset \mathbb{R}^d$ with polyhedral boundary, let $(\mathbb{Y}, \langle\cdot,\,\cdot\rangle_{\mathbb{Y}})$ be some Hilbert space of functions with $\mathbb{Y} \subset L^2(\Omega;\mathbb{R}^m)$ for some $m \in \mathbb{N}$. The quantity $\alpha > 0$ is some given cost parameter. We consider a Fréchet differentiable, quadratic and convex functional $\psi : \mathbb{Y} \to \mathbb{R}$ and suppose that ψ' is locally Lipschitz continuous with constant L, i.e., $\|\psi'(y) - \psi'(\bar{y})\|_{\mathbb{Y}^*(\omega)} \le L\|y - \bar{y}\|_{\mathbb{Y}(\omega)}$ for all $y, \bar{y} \in \mathbb{Y}$ and $\omega \subset \Omega$. Hereafter, we assume that the norm $\|\cdot\|_{\mathbb{Y}} = \|\cdot\|_{\mathbb{Y}(\Omega)}$ is sub-additive, i.e., for any measurable subsets $\omega_1, \omega_2 \subset \Omega$ with $|\omega_1 \cap \omega_2| = 0$, we have that

$$\|v\|^2_{\mathbb{Y}(\omega_1)} + \|v\|^2_{\mathbb{Y}(\omega_2)} \le \|v\|^2_{\mathbb{Y}(\omega_1\cup\omega_2)} \tag{2.2a}$$

and

$$\|v\|^2_{\mathbb{Y}(B_\delta\cap\Omega)} \to 0 \quad \text{as } \delta \to 0 \tag{2.2b}$$

where B_δ denotes any ball in $\mathbb{R}^d$ with radius δ. Note that this condition is weaker than absolute continuity of the norm and accommodates the fact that the relevant balls will be inner or outer balls of shape regular elements or patches. We suppose that the bilinear form $\mathcal{B} : \mathbb{Y} \times \mathbb{Y} \to \mathbb{R}$ is continuous and satisfies the inf-sup conditions

$$\inf_{v\in\mathbb{Y}\setminus\{0\}} \sup_{w\in\mathbb{Y}\setminus\{0\}} \frac{\mathcal{B}[v,\,w]}{\|v\|_{\mathbb{Y}}\|w\|_{\mathbb{Y}}} = \inf_{w\in\mathbb{Y}\setminus\{0\}} \sup_{v\in\mathbb{Y}\setminus\{0\}} \frac{\mathcal{B}[v,\,w]}{\|v\|_{\mathbb{Y}}\|w\|_{\mathbb{Y}}} = \beta > 0. \tag{2.3}$$

This is equivalent to the fact that the state equation admits an unique solution which continuously depends on the data $f + u \in \mathbb{Y}^*$. The control space is assumed as $\mathbb{U} = L^2(\Gamma;\mathbb{R}^m)$ for some $m \in \mathbb{N}$, where Γ maybe part of the domain Ω or its boundary such that $\mathbb{Y} \hookrightarrow \mathbb{U} = \mathbb{U}^* \hookrightarrow \mathbb{Y}^*$. The set of admissible controls $\mathbb{U}^{\mathrm{ad}}$ is given by

$$\mathbb{U}^{\mathrm{ad}} = \{u \in \mathbb{U}, u \in \mathcal{C} \text{ a.e. in } \Gamma\} \tag{2.4}$$

with a given closed convex set $\emptyset \ne \mathcal{C} \subset \mathbb{R}^m$. We shall use the notation $\|u\|_{\mathbb{U}(\omega)} := \|u\|_{L^2(\Gamma\cap\overline{\omega})}$ for all measurable sets $\omega \subset \Omega$. Technically there may be some inclusion operators like traces involved, however, they are omitted for the sake of clarity of the presentation.

Turning to the discretization of (2.1), we assume that Ω is meshed exactly by some conforming initial triangulation $\mathcal{G}_0$ and denote by $\mathbb{G}$ the class of all conforming refinements of $\mathcal{G}_0$ that can be constructed using recursive oder iterative refinement by bisection [2, 16, 24]. 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 $q \in \mathbb{N}$, such that we have the following uniform discrete inf-sup conditions

$$\inf_{v \in \mathbb{Y}(\mathcal{G})\setminus\{0\}} \sup_{w \in \mathbb{Y}(\mathcal{G})\setminus\{0\}} \frac{\mathcal{B}[v,\, w]}{\|v\|_{\mathbb{Y}}\|w\|_{\mathbb{Y}}} = \beta(\mathcal{G}) > 0 \tag{2.5a}$$

or

$$\inf_{w \in \mathbb{Y}(\mathcal{G})\setminus\{0\}} \sup_{v \in \mathbb{Y}(\mathcal{G})\setminus\{0\}} \frac{\mathcal{B}[v,\, w]}{\|v\|_{\mathbb{Y}}\|w\|_{\mathbb{Y}}} = \beta(\mathcal{G}) > 0 \tag{2.5b}$$

with $\beta(\mathcal{G}) \geq \gamma > 0$.

In the case of the variational discretization of (2.1) by Hinze [11], we solve the discretized optimal control problem

$$\begin{aligned} &\min_{(U,Y)\in\mathbb{U}^{\mathrm{ad}}\times\mathbb{Y}(\mathcal{G})} \mathcal{J}[U,Y] = \psi(Y) + \frac{\alpha}{2}\|U\|_{\mathbb{U}}^2 \\ &\text{subject to} \qquad Y \in \mathbb{Y}(\mathcal{G}): \quad \mathcal{B}[Y,\, V] = \langle f + U,\, V\rangle \qquad \forall V \in \mathbb{Y}(\mathcal{G}). \end{aligned} \tag{2.6}$$

For discrete controls, we additionally replace the control space $\mathbb{U}$ by a finite element space of discontinuous piecewise polynomials of bounded degree over a conforming, exact and shape-regular triangulation $\mathcal{G}^\Gamma$ of Γ, which is subordinated to some $\mathcal{G} \in \mathbb{G}$ in the sense that $\mathcal{G}^\Gamma$ is either a subset of $\mathcal{G}$ or of its trace grid on the boundary of Ω. Note that the existence of $\mathcal{G}^\Gamma$ for all $\mathcal{G} \in \mathbb{G}$ requires that Γ is the union of elements in $\mathcal{G}_0$ or respectively of some of its boundary sides. With a little abuse of notation, we denote the resulting discretization of $\mathbb{U}$ by $\mathbb{U}(\mathcal{G})$ and assume that it contains the piecewise constant functions over $\mathcal{G}^\Gamma$. Note that this readily implies that the set of discrete admissible controls

$$\mathbb{U}^{\mathrm{ad}}(\mathcal{G}) := \mathbb{U}^{\mathrm{ad}} \cap \mathbb{U}(\mathcal{G}),$$

is nonempty. The disretized optimal control problem reads then as

$$\begin{aligned} &\min_{(U,Y)\in\mathbb{U}^{\mathrm{ad}}(\mathcal{G})\times\mathbb{Y}(\mathcal{G})} \mathcal{J}[U,Y] = \psi(Y) + \frac{\alpha}{2}\|U\|_{\mathbb{U}}^2 \\ &\text{subject to} \qquad Y \in \mathbb{Y}(\mathcal{G}): \quad \mathcal{B}[Y,\, V] = \langle f + U,\, V\rangle \qquad \forall V \in \mathbb{Y}(\mathcal{G}). \end{aligned} \tag{2.7}$$

It is well-known that (2.1) as well as (2.6) respective (2.7) admit unique solution pairs $(\hat{u}, \hat{y})$ and $(\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}})$; compare with [18, 27].

Numerically, the discrete solutions of (2.6) and (2.7) are computed by solving the corresponding first order optimality systems; compare also with [27]. In other words, the control $\hat{u}$ in (2.1) is the orthogonal projection in $\mathbb{U}$ onto the set of admissible controls $\mathbb{U}^{\mathrm{ad}}$ of the adjoint state

$$\hat{p} \in \mathbb{Y}: \qquad \mathcal{B}[v,\, \hat{p}] = \langle \psi'(\hat{y}),\, v\rangle \qquad \forall v \in \mathbb{Y}. \tag{2.8a}$$

In the discrete settings (2.6) and (2.7) we have analogously

$$\hat{P}_{\mathcal{G}} \in \mathbb{Y}(\mathcal{G}): \qquad \mathcal{B}[V,\, \hat{P}_{\mathcal{G}}] = \langle \psi'(\hat{Y}_{\mathcal{G}}),\, V\rangle \qquad \forall V \in \mathbb{Y}(\mathcal{G}) \tag{2.8b}$$

and the discrete control is the orthogonal projection of $\hat{Y}_{\mathcal{G}}$ onto either $\mathbb{U}^{\mathrm{ad}}$ (variational discretization of Hinze) or $\mathbb{U}^{\mathrm{ad}}(\mathcal{G})$ (control discretization). A more detailed presentation is provided in Section 3.

We use the following adaptive algorithm for approximating the exact solution of (2.1). Starting with the initial conforming triangulation $\mathcal{G}_0$ of Ω, we execute the standard adaptive loop

$$\mathsf{SOLVE} \longrightarrow \mathsf{ESTIMATE} \longrightarrow \mathsf{MARK} \longrightarrow \mathsf{REFINE}. \tag{2.9}$$

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

Assumption 2.1 (Properties of modules)**.** For a given grid $\mathcal{G} \in \mathbb{G}$ the four used modules have the following properties.

(1) The output $(\hat{U}_\mathcal{G}, \hat{Y}_\mathcal{G}, \hat{P}_\mathcal{G}) := \mathsf{SOLVE}\big(\mathcal{G}\big) \in \mathbb{U}^{\mathrm{ad}} \times \mathbb{Y}(\mathcal{G}) \times \mathbb{Y}(\mathcal{G})$ is the exact solution of (2.6) or (2.7), respectively .
(2) The output $\{\mathcal{E}_{ocp}(\hat{U}_\mathcal{G}, \hat{Y}_\mathcal{G}, \hat{P}_\mathcal{G}; E)\}_{E\in\mathcal{G}} := \mathsf{ESTIMATE}\big((\hat{U}_\mathcal{G}, \hat{Y}_\mathcal{G}, \hat{P}_\mathcal{G}); \mathcal{G}\big)$ is a *reliable* and *locally efficient* estimator for the error in the norm $\|\cdot\|_{\mathbb{U}\times\mathbb{Y}\times\mathbb{Y}}$. In §3 below we will formulate the detailed requirement for the estimator.
(3) The output $\mathcal{M} = \mathsf{MARK}\big(\{\mathcal{E}_{ocp}(\hat{U}_\mathcal{G}, \hat{Y}_\mathcal{G}, \hat{P}_\mathcal{G}; E)\}_{E\in\mathcal{G}}, \mathcal{G}\big)$ is a subset of elements subject to refinement. We shall allow any marking strategy such that $\mathcal{M}$ contains an element holding an indicator, which is of the size of the maximal one, i. e., there exists $C > 0$ independent of $\{\mathcal{E}_{ocp}((\hat{U}_\mathcal{G}, \hat{Y}_\mathcal{G}, \hat{P}_\mathcal{G}); E)\}_{E\in\mathcal{G}}$ and $\mathcal{G}$, such that

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

All practically relevant marking strategies do have this property; compare with [22, 25].
(4) The output $\mathcal{G}_+ := \mathsf{REFINE}\big(\mathcal{G}, \mathcal{M}\big) \in \mathbb{G}$ 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 paper is the following convergence result.

Theorem 2.2 (Main result)**.** *Let $(\hat{u}, \hat{y}, \hat{p}) \in \mathbb{U}^{\mathrm{ad}} \times \mathbb{Y} \times \mathbb{Y}$ be the exact solution of* (2.1)*. Suppose that $\{\hat{U}_k, \hat{Y}_k, \hat{P}_k\}_{k\ge0} \subset \mathbb{U}^{\mathrm{ad}} \times \mathbb{Y} \times \mathbb{Y}$ is any sequence of discrete solutions generated by the adaptive iteration* (2.9)*, where the modules have the properties stated in Assumption 2.1. Then we have*

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

The proof of this theorem is based upon ideas from the convergence proofs of Morin, Siebert, and Veeser in [22] and Siebert in [25]. It is a two step procedure presented in §4 and §5. In §4 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 §5 then relies on the *steering mechanisms* of (2.9), mainly encoded in properties of $\mathsf{ESTIMATE}$ and MARK, to finally prove $(\hat{u}_\infty, \hat{y}_\infty, \hat{p}_\infty) = (\hat{u}, \hat{y}, \hat{p})$.

3. Aposteriori Error Estimation

In this section we shortly summarize our findings from [13, 14] providing a unified framework for the aposteriori error analysis for control constrained optimal control problems. In what follows we shall use $a \lesssim b$ for $a \le Cb$ with a constant C that may only depend on data of (2.1), the shape regularity of the grids in $\mathbb{G}$, and properties of the discrete spaces such as the polinomial degree, but is independent of the particular triangulation $\mathcal{G} \in \mathbb{G}$. We shall write $a \simeq b$ whenever $a \lesssim b \lesssim a$.

First order optimality systems. The analysis in [13] is based on the characterization of the solutions by the first order optimality systems; compare with (2.8). In order to concretize this concept let $S, S^*\colon \mathbb{Y}^* \to \mathbb{Y}$ be the solution operators of the state and the adjoint equation, i.e., for any $g \in \mathbb{Y}^*$, we have

$$Sg \in \mathbb{Y}: \qquad \mathcal{B}[Sg,\, v] = \langle g,\, v\rangle \qquad \forall\, v \in \mathbb{Y} \tag{3.1}$$

and

$$S^*g \in \mathbb{Y}: \qquad \mathcal{B}[v,\, S^*g] = \langle g,\, v\rangle \qquad \forall\, v \in \mathbb{Y}. \tag{3.2}$$

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

$$\Pi(p) \in \mathbb{U}^{\mathrm{ad}}: \qquad \langle \alpha\Pi(p) + p,\, \Pi(p) - u\rangle \le 0 \qquad \forall u \in \mathbb{U}^{\mathrm{ad}}. \tag{3.3}$$

Note that here we make use of the embedding $\mathbb{Y} \hookrightarrow \mathbb{U} = \mathbb{U}^*$, which e.g. in the case of a boundary control involves a trace operator. Utilizing these operators, we have that $(\hat{u}, \hat{y}) \in \mathbb{U}^{\mathrm{ad}} \times \mathbb{Y}$ is a solution of (2.1) if and only if $(\hat{u}, \hat{y}, \hat{p}) \in \mathbb{U}^{\mathrm{ad}} \times \mathbb{Y} \times \mathbb{Y}$ is the unique solution of the coupled nonlinear system

$$\hat{y} = S(\hat{u} + f), \qquad \hat{p} = S^*(\psi'(\hat{y})), \quad \text{and} \quad \hat{u} = \Pi(\hat{p}); \tag{3.4}$$

compare with [27].

For $\mathcal{G} \in \mathbb{G}$ we next define $S_{\mathcal{G}}, S_{\mathcal{G}}^*\colon \mathbb{Y}(\mathcal{G})^* \to \mathbb{Y}(\mathcal{G})$ to be the discrete solution operators for (3.1) and (3.2), i.e., for any $G \in \mathbb{Y}(\mathcal{G})^*$ we have

$$S_{\mathcal{G}}G \in \mathbb{Y}(\mathcal{G}): \qquad \mathcal{B}[S_{\mathcal{G}}G,\, V] = \langle G,\, V\rangle \qquad \forall\, V \in \mathbb{Y}(\mathcal{G}), \tag{3.5}$$

and

$$S_{\mathcal{G}}^*G \in \mathbb{Y}(\mathcal{G}): \qquad \mathcal{B}[V,\, S_{\mathcal{G}}^*G] = \langle G,\, V\rangle \qquad \forall\, V \in \mathbb{Y}(\mathcal{G}). \tag{3.6}$$

As for the continuous case, we have then that $(\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}) \in \mathbb{U}^{\mathrm{ad}} \times \mathbb{Y}(\mathcal{G})$ solves (2.6) or (2.7) iff $(\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}) \in \mathbb{U}^{\mathrm{ad}} \times \mathbb{Y}(\mathcal{G}) \times \mathbb{Y}(\mathcal{G})$ is the discrete solution of

$$\begin{aligned} &\hat{Y}_{\mathcal{G}} = S_{\mathcal{G}}(\hat{U}_{\mathcal{G}} + f), && && \hat{P}_{\mathcal{G}} = S_{\mathcal{G}}^*(\psi'(\hat{Y}_{\mathcal{G}})), && \text{and} \\ &\hat{U}_{\mathcal{G}} = \Pi(\hat{P}_{\mathcal{G}}) && \text{or} && \hat{U}_{\mathcal{G}} = \Pi_{\mathcal{G}}(\hat{P}_{\mathcal{G}}), && \text{respectively.} \end{aligned} \tag{3.7}$$

Here we used the obvious embedding $f \in \mathbb{Y}^* \hookrightarrow \mathbb{Y}(\mathcal{G})^*$. The former variational discretization of Hinze requires the evaluation of the *continuous* projection operator Π for discrete functions $P \in \mathbb{Y}(\mathcal{G})$. In the latter control discrete case, we have replaced the continuous projection $\Pi : \mathbb{Y} \to \mathbb{U}^{\mathrm{ad}}$ by the discrete projection $\Pi_{\mathcal{G}} : \mathbb{Y}(\mathcal{G}) \to \mathbb{U}^{\mathrm{ad}}(\mathcal{G})$ defined by

$$\langle \alpha\Pi_{\mathcal{G}}(P_{\mathcal{G}}) + P_{\mathcal{G}},\, \Pi_{\mathcal{G}}(p) - U_{\mathcal{G}}\rangle \le 0 \qquad \forall U_{\mathcal{G}} \in \mathbb{U}^{\mathrm{ad}}(\mathcal{G}). \tag{3.8}$$

Moreover, we define the residuals for $Y, P \in \mathbb{Y}$, $u \in \mathbb{U} \hookrightarrow \mathbb{Y}^*$ and $g \in \mathbb{Y}^*$ by

$$\langle \mathcal{R}(Y; u),\, v\rangle := \mathcal{B}[Y,\, v] - \langle u + f,\, v\rangle = \mathcal{B}[Y - S(u + f),\, v], \qquad v \in \mathbb{Y}$$

and

$$\langle \mathcal{R}^*(P; g),\, v\rangle := \mathcal{B}[v,\, Y] - \langle g,\, v\rangle = \mathcal{B}[v,\, Y - S^*g], \qquad v \in \mathbb{Y}.$$

Thanks to the inf-sup stability and continuity of $\mathcal{B}$, we have equivalence of error and residual, i.e.,

$$\beta\|Y - S(u + f)\|_{\mathbb{Y}} \le \|\mathcal{R}(Y; u)\|_{\mathbb{Y}^*} \le \|\mathcal{B}\|\|Y - S(u + f)\|_{\mathbb{Y}}$$

and

$$\beta\|P - S^*g\|_{\mathbb{Y}} \le \|\mathcal{R}^*(P; g)\|_{\mathbb{Y}^*} \le \|\mathcal{B}\|\|P - S^*g\|_{\mathbb{Y}}.$$

Similar arguments yield

$$\|S\|, \|S^*\| \le \frac{\|\mathcal{B}\|}{\beta} \qquad \text{and} \qquad \|S_{\mathcal{G}}\|, \|S^*_{\mathcal{G}}\| \le \frac{\|\mathcal{B}\|}{\beta(\mathcal{G})} \le \frac{\|\mathcal{B}\|}{\gamma}.$$

Basic error equivalence. The main obstacle in the aposteriori error analysis encountered for instance in [19, 10] can be explained as follows. One would like to exploit Galerkin orthogonality in the linear state equation (3.1) and the adjoint equation (3.2). However, we observe that the triplet $(\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}})$ is the Galerkin approximation to the triplet $(\hat{u}, \hat{y}, \hat{p})$ but $\hat{Y}_{\mathcal{G}}$ is not the Galerkin approximation to the solution $\hat{y}$ of the linear problem (3.1) since we have $\hat{y} = S\hat{u}$ but not $\hat{y} = S\hat{U}_{\mathcal{G}}$. The same argument applies to the adjoint states. This observation shows that we cannot directly employ Galerkin orthogonality for single components of (3.4) and the nonlinearity in (3.3) prevents us from making use of Galerkin orthogonality for the system (3.4). The resort to this problem is given by the following result from [13, Theorem 2.2].

Proposition 3.1 (Basic error equivalence)**.** *Let* $(\hat{u}, \hat{p}, \hat{y}) \in \mathbb{W} = \mathbb{U} \times \mathbb{Y} \times \mathbb{Y}$ *be the solution of the optimality system* (3.4)*. Then we have the basic error equivalence*

$$\|(u, y, p) - (\hat{u}, \hat{p}, \hat{y})\|_{\mathbb{W}} \simeq \|(u, y, p) - (\Pi p, S(u + f), S^*(\psi'(y)))\|_{\mathbb{W}}$$

for arbitrary $(u, p, y) \in \mathbb{W}$*.*

For the problem under consideration, the constants hidden in $\simeq$ depend on the inf-sup constant β^{-1}. Employing this error equivalence it is sufficient to construct a reliable and efficient estimator for the right hand side $\|(\hat{U}_{\mathcal{G}}, \hat{Y}_{\mathcal{G}}, \hat{P}_{\mathcal{G}}) - (\bar{u}, \bar{y}, \bar{p})\|_{\mathbb{W}}$ where the functions $\bar{y}$ and $\bar{p}$ are the exact weak solutions to the *linear* problems (3.1) and (3.2) with given source $\hat{U}_{\mathcal{G}} + f$ and $\psi'(\hat{Y}_{\mathcal{G}})$, respectively. They play a similar role as the *elliptic reconstruction* used in the aposteriori error analysis of parabolic problems; compare with [20]. The third term has a different structure. It is zero for the variational discretization and contains a projection error in the case of a control discretization.

Aposteriori error estimation. We realize that $\hat{Y}_{\mathcal{G}}$ is the Galerkin approximation to $\bar{y}$ and $\hat{P}_{\mathcal{G}}$ the one to $\bar{p}$. We therefore can directly employ (existing) estimators for the linear problems (3.1) and (3.2) and their sum then constitutes an estimator for the optimal control problem; compare with [13, Theorem 3.2].

Let us now fix the requirements for estimators of the form

$$\mathcal{E}_y(Y, u; \mathcal{G}) = \left(\sum_{E \in \mathcal{G}} \mathcal{E}_y^2(Y, u; E)\right)^{1/2} \quad \text{and} \quad \mathcal{E}_p(Y, v; \mathcal{G}) = \left(\sum_{E \in \mathcal{G}} \mathcal{E}_p^2(Y, v; E)\right)^{1/2}$$

for the linear problems (3.1) and (3.2). We denote by osc_y and osc_p the typical oscillation terms appearing in a posteriori analysis of PDEs. For any subset $\mathcal{G}' \subset \mathcal{G}$ we set

$$\mathrm{osc}_y(Y, u; \mathcal{G}') = \left(\sum_{E \in \mathcal{G}'} \mathrm{osc}_y^2(Y, u; E)\right)^{1/2}$$

and analogously for osc_p. For $E \in \mathcal{G}$, we let $\mathcal{N}_{\mathcal{G}}(E) := \{E' \in \mathcal{G} | E' \cap E \neq \emptyset\}$ be the set of direct neighbors and $\Omega_{\mathcal{G}}(E) := \bigcup_{E' \in \mathcal{N}_{\mathcal{G}}(E)} E'$ be the corresponding patch and extend this to sub-triangulations $\mathcal{G}' \subset \mathcal{G}$ via $\Omega_{\mathcal{G}}(\mathcal{G}') := \bigcup_{E \in \mathcal{G}'} \Omega_{\mathcal{G}}(E)$.

Remark 3.2. In principle $\Omega_{\mathcal{G}}(E)$ can be replaced by more general neighborhoods $\tilde{\Omega}_{\mathcal{G}}(E)$ of E if needed. In fact, it is only required that the $\tilde{\Omega}_{\mathcal{G}}(E)$ are a connected union of elements including E and that only finitely many of the $\tilde{\Omega}_{\mathcal{G}}(E)$, $E \in \mathcal{G}$, overlap. This directly implies that the number of elements in $\tilde{\Omega}_{\mathcal{G}}(E)$ is uniformly bounded and that the corresponding set of neighbors $\tilde{\mathcal{N}}_{\mathcal{G}}(E) := \{E' \in \mathcal{G} : E \subset \tilde{\Omega}_{\mathcal{G}}(E)\}$ is quasi uniform. In this context, one may think e.g. of the neighbors of a neighborhood of an element $E \in \mathcal{G}$, i. e., $\tilde{\mathcal{N}}_{\mathcal{G}}(E) = \mathcal{N}_{\mathcal{G}}(\Omega_{\mathcal{G}}(E))$.

Assumption 3.3 (Estimators for the linear problems)**.** We suppose that $\mathcal{E}_y$ and $\mathcal{E}_p$ have the following properties:

(1) **Reliability:** The estimators $\mathcal{E}_y$ and $\mathcal{E}_p$ provide an upper bound for the true error, i. e., for any $u \in \mathbb{U}$ and $y \in \mathbb{Y}$, we have

$$\begin{aligned} \|S_{\mathcal{G}}(u+f) - S(u+f)\|_{\mathbb{Y}} &\lesssim \mathcal{E}_y(S_{\mathcal{G}}(u+f), u; \mathcal{G}), \\ \|S^*_{\mathcal{G}}\psi'(y) - S^*\psi'(y)\|_{\mathbb{Y}} &\lesssim \mathcal{E}_p(S^*_{\mathcal{G}}\psi'(y), \psi'(y); \mathcal{G}). \end{aligned}$$

Typically, these bounds are a consequence of the equivalence of the residual and the error, together with proper interpolation estimates.

(2) **Local Efficiency:** The indicators $\mathcal{E}_y$ and $\mathcal{E}_p$ are local lower bounds for the true error up to oscillations, i.e., for any $Y, P \in \mathbb{Y}(\mathcal{G})$ and $u \in \mathbb{U}$ and $y \in \mathbb{Y}$, we have

$$\begin{aligned} \mathcal{E}_y(Y, u; E) &\lesssim \|Y - S(u+f)\|_{\mathbb{Y}(\Omega_{\mathcal{G}}(E))} + \operatorname{osc}_y(Y, u; \mathcal{N}_{\mathcal{G}}(E)), \\ \mathcal{E}_p(P, \psi'(y); E) &\lesssim \|P - S^*\psi'(y)\|_{\mathbb{Y}(\Omega_{\mathcal{G}}(E))} + \operatorname{osc}_p(P, \psi'(y); \mathcal{N}_{\mathcal{G}}(E)). \end{aligned}$$

(3) **Lipschitz continuity of Indicators:** The indicators $\mathcal{E}_y$ and $\mathcal{E}_p$ are Lipschitz continuous with respect to their second arguments, i.e., for $Y, P \in \mathbb{Y}(\mathcal{G})$, $u_1, u_2 \in \mathbb{U}$ and $y_1, y_2 \in \mathbb{Y}$, we have for all $E \in \mathcal{G}$ that

$$\begin{aligned} |\mathcal{E}_y(Y, u_1; E) - \mathcal{E}_y(Y, u_2; E)| &\lesssim \|u_1 - u_2\|_{\mathbb{U}(\Gamma \cap E)}, \\ |\mathcal{E}_p(P, \psi'(y_1); E) - \mathcal{E}_p(P, \psi'(y_2); E)| &\lesssim \|y_1 - y_2\|_{\mathbb{Y}(E)}. \end{aligned}$$

(4) **Regular test-functions:** Testing the residual with more regular functions, we expect additional powers of the mesh-size in the estimate. In particular, we assume that there exists a dense subspace $\mathbb{Y}_s \subset \mathbb{Y}$, $s > 0$, with subadditive norm, such that for all $v \in \mathbb{Y}_s$, and $u \in \mathbb{U}$, $y \in \mathbb{Y}$ we have

$$\begin{aligned} \langle \mathcal{R}(S_{\mathcal{G}}(u+f); u), \, v \rangle &\lesssim \sum_{E \in \mathcal{G}} h_E^s \, \mathcal{E}_y(S_{\mathcal{G}}(u+f), u; E) \|v\|_{\mathbb{Y}_s(\Omega_{\mathcal{G}}(E))}, \\ \langle \mathcal{R}^*(S^*_{\mathcal{G}}\psi'(y); \psi'(y)), \, v \rangle &\lesssim \sum_{E \in \mathcal{G}} h_E^s \, \mathcal{E}_p(S^*_{\mathcal{G}}\psi'(y), \psi'(y); E) \|v\|_{\mathbb{Y}_s(\Omega_{\mathcal{G}}(E))}. \end{aligned}$$

(5) **Oscillation:** The oscillation quantifies the gap between the error and the estimator. We assume that for all $\epsilon > 0$ there exists a continuous and non decreasing $m_\epsilon : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ with $m(0) = 0$, such that for $Y, P \in \mathbb{Y}$, $u \in \mathbb{U}$, and $y \in \mathbb{Y}$, we have that

$$\begin{aligned} \operatorname{osc}_y(Y, u; E) &\lesssim \epsilon + m_\epsilon(|E|)\big(\|Y\|_{\mathbb{Y}(\Omega_k(E))} + \|u\|_{\mathbb{U}(\Omega_k(E))} + \|D\|_{\mathbb{D}(\Omega_k(E))}\big), \\ \operatorname{osc}_p(P, \psi'(y); E) &\lesssim \epsilon + m_\epsilon(|E|)\big(\|P\|_{\mathbb{Y}(\Omega_k(E))} + \|y\|_{\mathbb{Y}(\Omega_k(E))} + \|D\|_{\mathbb{D}(\Omega_k(E))}\big), \end{aligned}$$

where $\mathbb{D}$ denotes another Hilbert space with a norm satisfying (2.2) and $D \in \mathbb{D}$ is given by the data of (2.1).

The estimator for the error of the control function is constructed from the indicators $\mathcal{E}_u(U,p;E) = \|U - \Pi(p)\|_{\mathbb{U}(\Gamma\cap E)} := \|(U-\Pi(p))\chi_E\|_{\mathbb{U}}$. We set

$$\|U - \Pi(p)\|_{\mathbb{U}}^2 = \mathcal{E}_u^2(U,p;\mathcal{G}) = \sum_{E\in\mathcal{G}} \mathcal{E}_u^2(U,p;E) \tag{3.9}$$

and define the estimator of the optimal control problem by

$$\mathcal{E}_{ocp}(\hat U,\hat Y,\hat P;E) := \mathcal{E}_u(\hat U,\hat P;E) + \mathcal{E}_y(\hat Y,\hat U;E) + \mathcal{E}_p(\hat P,\psi'(\hat Y);E), \quad E\in\mathcal{G}.$$

The following result can be found in [13, Theorem 3.2].

Theorem 3.4 (Aposteriori error control). *Let $(\hat u,\hat y,\hat p)$ be the exact solution of (2.1), let $(\hat U_{\mathcal{G}},\hat Y_{\mathcal{G}},\hat P_{\mathcal{G}})$ be the true solution either of (3.7), and suppose Assumption 3.3. Then $\mathcal{E}_{ocp}(\hat U_{\mathcal{G}},\hat Y_{\mathcal{G}},\hat P_{\mathcal{G}};\mathcal{G})$ is an estimator for the optimal control problem which is reliable, i.e.,*

$$\|(\hat U_{\mathcal{G}},\hat Y_{\mathcal{G}},\hat P_{\mathcal{G}}) - (\hat u,\hat y,\hat p)\|_{\mathbb{U}\times\mathbb{Y}\times\mathbb{Y}} \lesssim \mathcal{E}_{ocp}(\hat U_{\mathcal{G}},\hat Y_{\mathcal{G}},\hat P_{\mathcal{G}};\mathcal{G}) \tag{3.10a}$$

and globally efficient, i.e.,

$$\begin{aligned}\mathcal{E}_{ocp}(\hat U_{\mathcal{G}},\hat Y_{\mathcal{G}},\hat P_{\mathcal{G}};\mathcal{G}) \lesssim \|(\hat U_{\mathcal{G}},\hat Y_{\mathcal{G}},\hat P_{\mathcal{G}}) - (\hat u,\hat y,\hat p)\|_{\mathbb{U}\times\mathbb{Y}\times\mathbb{Y}}\\ + \operatorname{osc}_y(\hat Y_{\mathcal{G}},\hat u;\mathcal{G}) + \operatorname{osc}_p(\hat P_{\mathcal{G}},\psi'(\hat y);\mathcal{G}).\end{aligned} \tag{3.10b}$$

4. 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, $\{\mathcal{G}_k, (\hat U_k,\hat Y_k,\hat P_k)\}_{k\ge0}$ is the sequence of grids and discrete solutions generated by (2.9). For the ease of notation we use for $k\ge0$ 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.

Approximation of an admissible control. We start with the limit of the piecewise constant mesh-size function $h_k\colon \Omega\to\mathbb{R}$ of $\mathcal{G}_k$ defined by $h_{k|E} = |E|^{1/d}$, $E\in\mathcal{G}_k$. The behavior of the mesh-size function is directly related to the decomposition

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

The set $\mathcal{G}_k^+$ contains all elements that are not refined after iteration k and we observe that the sequence $\{\mathcal{G}_k^+\}_{k\ge0}$ is nested, i. e., $\mathcal{G}_\ell^+\subset\mathcal{G}_k^+$ for all $k\ge\ell$. The set $\mathcal{G}_k^0$ contains all elements that are refined at least once more after iteration k; in particular, we have for the marked elements that $\mathcal{M}_k\subset\mathcal{G}_k^0$. It is proved in [22, Lemma 4.3] that $h_k\to h_\infty$ uniformly in $L^\infty(\Omega)$.

Moreover, decomposing $\bar\Omega = \Omega_k^+\cup\Omega_k^0 := \Omega(\mathcal{G}_k^+)\cup\Omega(\mathcal{G}_k^0)$, we have the following relation to the behavior of the mesh-size function shown in [25, Corollary 3.3].

Lemma 4.1 (Convergence of the mesh-size functions). *The mesh-size functions h_k converge uniformly to 0 in Ω_k^0 in the following sense*

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

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

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

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

as well as the limiting set of admissible control functions

$$\mathbb{U}^{\mathrm{ad}}_\infty = \mathbb{U}^{\mathrm{ad}} \cap \mathbb{U}_\infty.$$

To handle the variational and the control discretization in the same setting, we set $\mathbb{U}_k \equiv \mathbb{U}$ in the former case. In a first step, we will show that we can approximate an arbitrary element of $\mathbb{U}^{\mathrm{ad}}_\infty$ in an appropriate way.

Lemma 4.2. *Let u be an arbitrary element of $\mathbb{U}^{\mathrm{ad}}_\infty$. Then there exists a sequence of members $u_k \in \mathbb{U}^{\mathrm{ad}} \cap \mathbb{U}_k$ converging to u in $\mathbb{U}$ with the property*

$$u_k = u \text{ on } \Gamma \cap \Omega_k^+, \qquad k \in \mathbb{N}. \tag{4.1}$$

Proof. For the variational discretization the assertion is trivial since we can choose $u_k = u$ for every k.

Let us investigate spaces $\mathbb{U}_k$ of discontinuous functions. Here we set

$$u_k = \begin{cases} u & \text{for } x \in \Gamma \cap \Omega_k^+, \\ \mathcal{P}_k u & \text{else,} \end{cases}$$

where $\mathcal{P}_k$ denotes the L^2-projection onto piecewise constant functions over $\mathcal{G}_k^\Gamma$, i. e., $\mathcal{P}_k|_T u = \frac{1}{|T|}\int_T u \,\mathrm{d}\Gamma$, $T \in \mathcal{G}_k^\Gamma$. Note that, thanks to the convexity of the set $\mathcal{C}$, we have that $\mathcal{P}_k u \in \mathbb{U}^{\mathrm{ad}}$. Thanks to Lemma 4.1, this implies

$$\|u - u_k\|_{\mathbb{U}} = \|u - u_k\|_{\mathbb{U}(\Omega_k^0)} \to 0$$

as $k \to \infty$. □

Going to the limit. Using piecewise polynomials in combination with refinement by bisection leads to nested spaces $\mathbb{Y}_k$, 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 is a Hilbert space. Consequently, the limiting optimal control problem

$$\begin{aligned} &\min_{(u,y)\in \mathbb{U}^{\mathrm{ad}}_\infty \times \mathbb{Y}_\infty} \mathcal{J}[u,y] = \psi(y) + \frac{\alpha}{2}\|u\|^2_{\mathbb{U}} \\ &\text{subject to} \qquad y \in \mathbb{Y}_\infty : \quad \mathcal{B}[y,\, v] = \langle u + f,\, v\rangle \qquad \forall v \in \mathbb{Y}_\infty \end{aligned} \tag{4.2}$$

admits a unique solution $(\hat{u}_\infty, \hat{y}_\infty) \in \mathbb{U}^{\mathrm{ad}}_\infty \times \mathbb{Y}_\infty$. Thanks to the uniform discrete inf-sup stability (2.5), we have that there exists solution operators $S_\infty, S^*_\infty \colon \mathbb{U} \to \mathbb{Y}_\infty$ of the state respectively the adjoint state equations defined by (3.5) with $\mathbb{Y}(\mathcal{G})$ replaced by $\mathbb{Y}_\infty$, and we have $\|S_\infty\|, \|S^*_\infty\| \leq \frac{\|\mathcal{B}\|}{\gamma}$; compare with [22, 25]. The associated first order optimality system then reads as

$$\begin{aligned} &\hat{y}_\infty = S_\infty(\hat{u}_\infty + f), && && \hat{p}_\infty = S^*_\infty(\psi'(\hat{y}_\infty)), && \text{and} \\ &\hat{u}_\infty = \Pi(\hat{p}_\infty) && \text{or} && \hat{u}_\infty = \Pi_\infty(\hat{p}_\infty), && \text{respectively.} \end{aligned} \tag{4.3}$$

The latter control discrete case employs the discrete projection $\Pi_\infty : \mathbb{Y}_\infty \to \mathbb{U}_\infty^{\rm ad}$ defined by

$$\langle \alpha \Pi_\infty(p_\infty) + p_\infty,\, \Pi_\infty(p) - u_\infty \rangle \le 0 \qquad \forall u_\infty \in \mathbb{U}_\infty^{\rm ad}. \tag{4.4}$$

We shall show that (4.2) is in fact the limiting problem of the adaptive iteration (2.9) in that $(\hat{U}_k, \hat{Y}_k, \hat{P}_k) \to (\hat{u}_\infty, \hat{y}_\infty, \hat{p}_\infty)$. An important ingredient for proving this is the following crucial property of the adaptive algorithm shown in [1, Lemma 6.1] and [22, Lemma 4.2].

Proposition 4.3 (Convergence of solution operators). *For any $u, g \in \mathbb{U}$, we have $S_k u \to S_\infty u$ and $S_k^* g \to S_\infty^* g$ in $\mathbb{Y}$ as $k \to \infty$.*

Proposition 4.4 (Boundedness of solution). *The sequence $(\hat{U}_k, \hat{Y}_k, \hat{P}_k)$ is bounded in $\mathbb{W}$.*

Proof. Take the optimal control $\hat{U}_0$ from the coarsest grid. Thanks to Proposition 4.3 we have that the the sequence $\{S_k(\hat{U}_0 + f)\}_{k=0}^\infty$ is bounded in $\mathbb{Y}$. Since $\hat{U}_0$ is feasible for the optimal control problem for all grids, we obtain the boundedness of the optimal objective values $\{\mathcal{J}[\hat{U}_k, \hat{Y}_k]\}_{k=0}^\infty$. The structure of the objective readily implies the boundedness of the sequence $\{\hat{U}_k\}_{k=0}^\infty$ in $\mathbb{U}$. The boundedness of $\{\hat{Y}_k\}_{k=0}^\infty$ and $\{\hat{P}_k\}_{k=0}^\infty$ is a direct implication of Proposition 4.3. □

We next show convergence of the control functions. In this step we have to deal with the nonlinearity of the constrained optimal control problem.

Lemma 4.5 (Convergence of the controls). *The discrete controls $\{\hat{U}_k\}_{k\ge 0}$ converge strongly to $\hat{u}_\infty$, i. e.,*

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

Proof. Due to Lemma 4.2, $\hat{u}_\infty \in \mathbb{U}_\infty^{\rm ad}$ is the limit of a sequence $\{u_k\}_{k\in\mathbb{N}}$ with $u_k \subset \mathbb{U}^{\rm ad} \cap \mathbb{U}_k$ and $u_k = \hat{u}_\infty$ on $\Gamma \cap \Omega_k^+$. Using the optimality of $\hat{u}_\infty$ and $\hat{U}_k$, we find

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

We estimate the three terms on the right-hand side separately. Using Proposition 4.4 and Lemma 4.2, the first term tends to zero.

For the second term, we immediately obtain from $\hat{p}_\infty = S_\infty^*(\psi'(\hat{y}_\infty))$ by the embedding $\mathbb{Y} \hookrightarrow \mathbb{U}$ and Young's inequality, that

$$\begin{aligned}
\langle S_k^*(\psi'(\hat{y}_\infty)) - \hat{p}_\infty,\, \hat{u}_\infty - \hat{U}_k \rangle &= \langle (S_k^* - S_\infty^*)(\psi'(\hat{y}_\infty)),\, \hat{u}_\infty - \hat{U}_k \rangle \\
&\le \frac{\alpha}{2} \|\hat{u}_\infty - \hat{U}_k\|_{\mathbb{U}}^2 + \frac{1}{2\alpha} \|(S_k^* - S_\infty^*)(\psi'(\hat{y}_\infty))\|_{\mathbb{U}}^2 \\
&\le \frac{\alpha}{2} \|\hat{u}_\infty - \hat{U}_k\|_{\mathbb{U}}^2 + c \|(S_k^* - S_\infty^*)(\psi'(\hat{y}_\infty))\|_{\mathbb{Y}}^2.
\end{aligned}$$

We next turn to the third term. Employing the definition of the solution operators S_k and S_k^* in (3.5) and (3.6), we use $\hat{P}_k = S_k^*(\psi'(\hat{Y}_k)) \in \mathbb{Y}_k$ and $\hat{y}_\infty = S_\infty \hat{u}_\infty \in \mathbb{Y}_\infty$ to obtain

$$\begin{aligned}
\langle \hat{P}_k - S_k^*(\psi'(\hat{y}_\infty)),\, \hat{u}_\infty - \hat{U}_k\rangle &= \langle \hat{u}_\infty - \hat{U}_k,\, S_k^*(\psi'(\hat{Y}_k) - \psi'(\hat{y}_\infty))\rangle \\
&= \mathcal{B}[S_k(\hat{u}_\infty - \hat{U}_k),\, S_k^*(\psi'(\hat{Y}_k) - \psi'(\hat{y}_\infty))] \\
&= \langle \psi'(\hat{Y}_k) - \psi'(\hat{y}_\infty),\, S_k(\hat{u}_\infty - \hat{U}_k)\rangle \\
&= \langle \psi'(\hat{Y}_k) - \psi'(\hat{y}_\infty),\, \hat{y}_\infty - \hat{Y}_k\rangle + \langle \psi'(\hat{Y}_k) - \psi'(\hat{y}_\infty),\, (S_k - S_\infty)(\hat{u}_\infty + f)\rangle \\
&\le 0 + \|\psi'(\hat{Y}_k) - \psi'(\hat{y}_\infty)\|_{\mathbb{Y}^*} \|(S_k - S_\infty)(\hat{u}_\infty + f)\|_{\mathbb{Y}} \\
&\lesssim \|\hat{Y}_k - \hat{y}_\infty\|_Y \|(S_k - S_\infty)(\hat{u}_\infty + f)\|_{\mathbb{Y}} \\
&\lesssim \|(S_k - S_\infty)(\hat{u}_\infty + f)\|_{\mathbb{Y}},
\end{aligned}$$

where we used Proposition 4.4 in the last line. Combining above estimates we obtain with Proposition 4.3 that

$$\begin{aligned}
\|\hat{U}_k - \hat{u}_\infty\|_{\mathbb{U}}^2 &\lesssim \|u_k - \hat{u}_\infty\|_{\mathbb{U}} + \|(S_k^* - S_\infty^*)(\psi'(\hat{y}_\infty))\|_{\mathbb{Y}}^2 + \|(S_k - S_\infty)(\hat{u}_\infty + f)\|_{\mathbb{Y}} \\
&\to 0
\end{aligned}$$

as $k \to \infty$. This finishes the proof. □

Convergence $(\hat{U}_k, \hat{Y}_k, \hat{P}_k) \to (\hat{u}_\infty, \hat{y}_\infty, \hat{p}_\infty)$ is now a direct consequence of the linear theory in Proposition 4.3.

Proposition 4.6 (Convergence of discrete solutions)**.** *The Galerkin approximations $\{(\hat{U}_k, \hat{Y}_k, \hat{P}_k)\}_{k\ge 0}$ converge strongly to the solution $(\hat{u}_\infty, \hat{y}_\infty, \hat{p}_\infty)$ of (4.2), i. e.,*

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

Proof. We already know that $\|\hat{U}_k - \hat{u}_\infty\|_{\mathbb{U}} \to 0$ from Lemma 4.5. In combination with Proposition 4.3, this yields for the discrete states that

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

since $\|S_k\| \le C_F$. Writing $\hat{P}_k - \hat{p}_\infty = S_k^*(\psi'(\hat{Y}_k) - \psi'(\hat{y}_\infty)) + (S_k^* - S_\infty)(\psi'(\hat{y}_\infty))$ we finally deduce $\|\hat{P}_k - \hat{p}_\infty\|_{\mathbb{Y}} \to 0$ with the same arguments. □

The convergence of the discrete solutions readily yields an uniform bound on the estimators.

Corollary 4.7 (Uniform estimator bound)**.** *For all $k \ge 0$, we have*

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

Proof. Starting with the global efficiency (3.10b), the assertion follows from Proposition 4.4 together with the properties of the oscillations in Assumption 3.3. □

Corollary 4.8 (Indicators of marked elements)**.** *All indicators of marked elements vanish in the limit, this is,*

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

Proof. For $k \ge 0$ pick up $E_k \in \arg\max\{\mathcal{E}_{ocp}((\hat{U}_k, \hat{Y}_k, \hat{P}_k); E) \mid E \in \mathcal{M}_k\} \neq \emptyset$. We follow [25, Lemma 3.4] and show $\mathcal{E}_{ocp}((\hat{U}_k, \hat{Y}_k, \hat{P}_k); E_k) \to 0$.

We have with the local efficiency of the estimators (see Assumption 3.3) that

$$\begin{aligned}\mathcal{E}_y(\hat{Y}_k, \hat{U}_k; E_k) &\lesssim \|\hat{Y}_k - S(\hat{U}_k + f)\|_{\mathbb{Y}(\Omega_k(E_k))} + \operatorname{osc}_y(\hat{Y}_k, \hat{U}_k; \mathcal{N}_k(E_k)) \\ &\le \|\hat{Y}_k - \hat{y}_\infty\|_{\mathbb{Y}(\Omega_k(E_k))} + \|\hat{y}_\infty\|_{\mathbb{Y}(\Omega_k(E_k))} + \|S\hat{u}_\infty - S\hat{U}_k\|_{\mathbb{Y}(\Omega_k(E_k))} \\ &\quad + \|S(\hat{u}_\infty + f)\|_{\mathbb{Y}(\Omega_k(E_k))} + \operatorname{osc}_y(\hat{Y}_k, \hat{U}_k; \mathcal{N}_k(E_k)) \\ &\to 0\end{aligned}$$

as $k \to \infty$ for the following reasons: By Assumption 2.1(4) all elements in $\mathcal{M}_k$ are refined in $\mathcal{G}_{k+1}$, which implies $E_k \in \mathcal{G}_k^0$. Local quasi-uniformity of $\mathcal{G}_k$ in combination with Lemma 4.1 therefore yields $|\Omega(\mathcal{N}_k(E_k))| \lesssim |E_k| \le \|h_k\|^d_{\infty;\Omega_k^0} \to 0$. Consequently, the terms $\|\hat{y}_\infty\|_{\mathbb{Y}(\Omega_k(E))}$ and $\|S(\hat{u}_\infty + f)\|_{\mathbb{Y}(\Omega_k(E))}$ vanish thanks to assumption (2.2b). Similarly, we conclude for $\epsilon > 0$ from the properties of the oscillation in Assumption 3.3(5), as well as the boundedness of $\|\hat{Y}_k\|_{\mathbb{Y}(\Omega_k(E_k))}$ and $\|\hat{U}_k\|_{\mathbb{Y}(\Omega_k(E_k))}$ (see Proposition 4.4) that

$$\begin{aligned}\operatorname{osc}_y(\hat{Y}_k, \hat{U}_k; \mathcal{N}_k(E)) &\lesssim \epsilon + m_\epsilon(|E|)\big(\|\hat{U}_k\|_{\mathbb{Y}(\Omega_k(E_k))} + \|\hat{Y}_k\|_{\mathbb{Y}(\Omega_k(E_k))} + \|D\|_{\mathbb{D}(\Omega_k(E_k))}\big) \\ &\to \epsilon.\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, this proves $\operatorname{osc}_y(\hat{Y}_k, \hat{U}_k; \mathcal{N}_k(E)) \to 0$.

Finally, Proposition 4.6 implies that also the terms $\|\hat{Y}_k - \hat{y}_\infty\|_{\mathbb{Y}(\Omega_k(E))}$ and $\|S\hat{u}_\infty - S\hat{U}_k\|_{\mathbb{Y}(\Omega_k(E))}$ vanish. The same arguments apply to the indicator contribution of the adjoint equation.

For the control indicator, we have

$$\begin{aligned}\mathcal{E}_u(\hat{U}_k, \hat{P}_k; E_k) &= \|\hat{U}_k - \Pi(\hat{P}_k)\|_{\mathbb{U}(\Gamma\cap E_k)} \\ &\le \|\hat{U}_k - \hat{u}_\infty\|_{\mathbb{U}(\Gamma\cap E_k)} + \|\hat{u}_\infty\|_{\mathbb{U}(\Gamma\cap E_k)} \\ &\quad + \|\Pi(\hat{p}_\infty) - \Pi(\hat{P}_k)\|_{\mathbb{U}(\Gamma\cap E_k)} + \|\Pi(\hat{p}_\infty)\|_{\mathbb{U}(\Gamma\cap E_k)}.\end{aligned}$$

Similar arguments as before can be used to prove that all but the penultimate term on the right-hand side vanish. It thus follows that $\mathcal{E}_u(\hat{U}_k, \hat{P}_k; E_k) \to 0$ as $k \to \infty$, observing that

$$\begin{aligned}\alpha\|\Pi(\hat{p}_\infty) - \Pi(\hat{P}_k)\|^2_{\mathbb{U}} &= \langle \alpha\Pi(\hat{p}_\infty) + \hat{p}_\infty,\, \Pi(\hat{p}_\infty) - \Pi(\hat{P}_k)\rangle \\ &\quad + \langle \alpha\Pi(\hat{P}_k) + \hat{P}_k,\, \Pi(\hat{P}_k) - \Pi(\hat{p}_\infty)\rangle \\ &\quad + \langle \hat{p}_\infty - \hat{P}_k,\, \Pi(\hat{p}_\infty) - \Pi(\hat{P}_k)\rangle \\ &\le \|\hat{p}_\infty - \hat{P}_k\|_{\mathbb{U}}\|\Pi(\hat{p}_\infty) - \Pi(\hat{P}_k)\|_{\mathbb{U}} \to 0\end{aligned}$$

thanks to Proposition (4.6). Concluding, we have proved $\mathcal{E}_{ocp}((\hat{U}_k, \hat{Y}_k, \hat{P}_k); E_k) \to 0$ as $k \to \infty$. □

5. Convergence 2: Making the Right Decisions

In this section we verify the main result, Theorem 2.2, by showing $(\hat{U}_k, \hat{Y}_k, \hat{P}_k) \to (\hat{u}, \hat{y}, \hat{p})$ and $\mathcal{E}_{ocp}(\hat{U}_k, \hat{Y}_k, \hat{P}_k; \mathcal{G}_k) \to 0$. Error convergence requires appropriate decisions in the adaptive iteration, which we have summarized in Assumption 2.1. Estimator convergence is then a consequence of local efficiency as stated in Theorem 3.4.

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

Lemma 5.1 (Convergence of the indicators)**.** *The maximal indicator vanishes in the limit, this is,*

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

Proof. Combining the assumption on marking in Assumption 2.1(3) with the behavior of the indicators on marked elements, which we have analyzed in Corollary 4.8, we find

$$\max\{\mathcal{E}_{ocp}(\hat{U}_k, \hat{Y}_k, \hat{P}_k; E) \mid E \in \mathcal{G}_k\} \le C \max\{\mathcal{E}_{ocp}(\hat{U}_k, \hat{Y}_k, \hat{P}_k; E) \mid E \in \mathcal{M}_k\} \to 0$$

as $k \to \infty$. □

Convergence of the residuals. We next show that the residuals of the state and the adjoint equation of the limiting first order optimality system (4.3) vanish. The proof adapts the techniques from [25, Proposition 3.1] to the situation at hand.

Proposition 5.2 (Convergence of the residual)**.** *For the residuals* $\mathcal{R}$ *of* (3.1) *and* $\mathcal{R}^*$ *of* (3.2)*, we have*

$$\mathcal{R}(\hat{y}_\infty; \hat{u}_\infty) = \mathcal{R}^*(\hat{p}_\infty; \psi'(\hat{y}_\infty)) = 0 \qquad \text{in } \mathbb{Y}^*.$$

In particular, we have $\hat{y}_\infty = S(\hat{u}_\infty + f)$ *and* $\hat{p}_\infty = S^*(\psi'(\hat{y}_\infty))$*.*

Proof. We prove the claim for $\mathcal{R}$. The assertion for $\mathcal{R}^*$ follows along the same lines. Using a density argument it suffices to show $\langle \mathcal{R}(\hat{y}_\infty; \hat{u}_\infty), v\rangle = 0$ for all $v \in \mathbb{Y}_s$ for some $s > 0$; compare Assumption 3.3(4).

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

Let $v \in \mathbb{Y}_s$ with $\|v\|_{\mathbb{Y}_s} = 1$. We next utilize the improved bound in Assumption 3.3(4), decompose $\mathcal{G}_k = \mathcal{G}_\ell^+ \cup (\mathcal{G}_k \setminus \mathcal{G}_\ell^+)$, and recall Corollary 4.7 to bound

$$\begin{aligned}
\langle \mathcal{R}(\hat{Y}_k; \hat{U}_k), v\rangle^2 &\lesssim \sum_{E\in\mathcal{G}_\ell^+} h_E^{2s}\, \mathcal{E}_y^2(\hat{Y}_k, \hat{U}_k; E) + \sum_{E\in\mathcal{G}_k\setminus\mathcal{G}_\ell^+} h_E^{2s}\, \mathcal{E}_y^2(\hat{Y}_k, \hat{U}_k; E) \\
&\lesssim \mathcal{E}_{ocp}^2(\hat{U}_k, \hat{Y}_k, \hat{P}_k; \mathcal{G}_\ell^+) + \|h_\ell\|_{\infty;\Omega_\ell^0}^2 \mathcal{E}_{ocp}^2(\hat{U}_k, \hat{Y}_k, \hat{P}_k; \mathcal{G}_k \setminus \mathcal{G}_\ell^+) \\
&\lesssim \mathcal{E}_{ocp}^2(\hat{U}_k, \hat{Y}_k, \hat{P}_k; \mathcal{G}_\ell^+) + \|h_\ell\|_{\infty;\Omega_\ell^0}^2 \overset{!}{\le} 2\varepsilon
\end{aligned}$$

for any $\varepsilon > 0$. The last inequality can be seen as follows: By Lemma 4.1, we may first choose ℓ large such that $\|h_\ell\|_{\infty;\Omega_\ell^0}^2 \le \varepsilon$. After fixing ℓ, the "pointwise" convergence of the indicators in Lemma 5.1 and $\#\mathcal{G}_\ell^+ < \infty$ implies then $\mathcal{E}_{ocp}^2((\hat{U}_k, \hat{Y}_k, \hat{P}_k); \mathcal{G}_\ell^+) \le \varepsilon$ for sufficiently large $k \ge \ell$. This yields for any fixed $v \in \mathbb{Y}_s$ that

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

observing that $\mathcal{R}$ is continuous with respect to its arguments and recalling the convergence $(\hat{U}_k, \hat{Y}_k) \to (\hat{u}_\infty, \hat{y}_\infty)$ shown in Proposition 4.6. It follows from the density of $\mathbb{Y}_s$ in $\mathbb{Y}$, that $\mathcal{R}(\hat{y}_\infty; \hat{u}_\infty) = 0$ in $\mathbb{Y}^*$. This in turn implies $\hat{y}_\infty = S\hat{u}_\infty$ and finishes the proof. □

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 2.2. Combining Propositions 3.1 and 4.6, we have

$$\begin{aligned}\lim_{k\to\infty} \|(\hat{U}_k, \hat{Y}_k, \hat{P}_k) - (\hat{u}, \hat{p}, \hat{y})\|_{\mathbb{W}} \\ &\simeq \lim_{k\to\infty} \|(\hat{U}_k, \hat{Y}_k, \hat{P}_k) - (\Pi(\hat{P}_k), S(\hat{U}_k + f), S^*(\psi'(\hat{Y}_k)))\|_{\mathbb{W}} \\ &= \|(\hat{u}_\infty, \hat{y}_\infty, \hat{p}_\infty) - (\Pi(\hat{p}_\infty), S(\hat{u}_\infty + f), S^*(\psi'(\hat{y}_\infty))\|_{\mathbb{W}}.\end{aligned}$$

Thanks to Proposition 5.2, in order to prove $\lim_{k\to\infty} \|(\hat{U}_k, \hat{Y}_k, \hat{P}_k) - (\hat{u}, \hat{p}, \hat{y})\|_{\mathbb{W}} = 0$, it suffices to verify that $\hat{u}_\infty = \Pi(\hat{p}_\infty)$. This is trivially satisfied for the variational discretization of Hinze and we can thus concentrate on the control discrete case. Let $k \in \mathbb{N}$, then combining Proposition 4.6 and Lemma 5.1, we conclude for all $E \in \mathcal{G}_k^+$, that

$$\|\hat{u}_\infty - \Pi(\hat{p}_\infty)\|_{\mathbb{U}(E)} \leq \lim_{\ell\to\infty} \mathcal{E}_{ocp}((\hat{U}_\ell, \hat{Y}_\ell, \hat{P}_\ell); E) = 0,$$

i.e., $\hat{u}_\infty = \Pi(\hat{p}_\infty)$ on Ω_k^+. For arbitrary $u \in \mathbb{U}^{\mathrm{ad}}$ let

$$u_k := \begin{cases} \hat{u}_\infty & \text{on } \Omega_k^+ \\ \mathcal{P}_k u & \text{otherwise,} \end{cases}$$

where $\mathcal{P}_k$ denotes the L^2-projection onto the piecewise constant functions over $\mathcal{G}_k^\Gamma$. Obviously, $u_k \in \mathbb{U}_\infty^{\mathrm{ad}}$ and we have that $\|(u - u_k)\chi_{\Omega_k^0}\|_{\mathbb{U}} \to 0$ as $k \to \infty$; compare also with the proof of Lemma 4.2. Therefore, we have

$$\begin{aligned}\langle \hat{p}_\infty + \alpha\hat{u}_\infty,\, u - \hat{u}_\infty\rangle &= \langle \hat{p}_\infty + \alpha\hat{u}_\infty,\, u_k - \hat{u}_\infty\rangle + \langle \hat{p}_\infty + \alpha\hat{u}_\infty,\, u - u_k\rangle \\ &\geq \langle \hat{p}_\infty + \alpha\hat{u}_\infty,\, u - u_k\rangle \\ &= \langle \hat{p}_\infty + \alpha\hat{u}_\infty,\, (u - u_k)\chi_{\Omega_k^0}\rangle + \langle \hat{p}_\infty + \alpha\hat{u}_\infty,\, (u - u_k)\chi_{\Omega_k^+}\rangle \\ &\geq \langle \hat{p}_\infty + \alpha\hat{u}_\infty,\, (u - u_k)\chi_{\Omega_k^0}\rangle \to 0\end{aligned}$$

as $k \to \infty$. Here we have used that $(u - u_k)\chi_{\Omega_k^0} \in \mathbb{U}^{\mathrm{ad}}$ and $\hat{u}_\infty = \Pi(\hat{p}_\infty)$ on Ω_k^+. Since $u \in \mathbb{U}^{\mathrm{ad}}$ was arbitrary, this implies $\hat{u}_\infty = \Pi\hat{p}_\infty$ and therefore $\lim_{k\to\infty} \|(\hat{U}_k, \hat{Y}_k, \hat{P}_k) - (\hat{u}, \hat{p}, \hat{y})\|_{\mathbb{W}} = 0$.

In order to prove convergence of the estimator, we decompose $\mathcal{E}_{ocp}$ for $k \geq \ell$, as in the proof to Proposition 5.2, i.e.

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

We first estimate the second term on the right hand side. The local efficiency in Assumption 3.3(2) in combination with the finite overlap of the patches $\mathcal{N}_k(E)$, (2.2), and the basic error equivalence (Proposition 3.1), allows us to bound

$$
\begin{aligned}
&\mathcal{E}^2_{ocp}(\hat{U}_k, \hat{Y}_k, \hat{P}_k; \mathcal{G}_k \setminus \mathcal{G}^+_\ell) \\
&\qquad \lesssim \|(\hat{U}_k, \hat{Y}_k, \hat{P}_k) - (\Pi(\hat{U}_k), S(\hat{U}_k + f), S^*(\psi'(\hat{Y}_k)))\|^2_{\mathbb{W}(\Omega_{\mathcal{G}}(\mathcal{G}_k \setminus \mathcal{G}^+_\ell))} \\
&\qquad\qquad + \sum_{E \in \mathcal{G}_k \setminus \mathcal{G}^+_\ell} \operatorname{osc}^2_y(\hat{Y}_k, \hat{U}_k; \mathcal{N}_k(E)) + \operatorname{osc}^2_p(\hat{P}_k, \psi'(\hat{Y}_k); \mathcal{N}_k(E)) \\
&\qquad \lesssim \|(\hat{U}_k, \hat{Y}_k, \hat{P}_k) - (\hat{u}, \hat{p}, \hat{y})\|^2_{\mathbb{W}} \\
&\qquad\qquad + \sum_{E \in \mathcal{G}_k \setminus \mathcal{G}^+_\ell} \operatorname{osc}^2_y(\hat{Y}_k, \hat{U}_k; \mathcal{N}_k(E)) + \operatorname{osc}^2_p(\hat{P}_k, \psi'(\hat{Y}_k); \mathcal{N}_k(E)) \\
&\qquad \lesssim \|(\hat{U}_k, \hat{Y}_k, \hat{P}_k) - (\hat{u}, \hat{p}, \hat{y})\|^2_{\mathbb{W}} \\
&\qquad\qquad + \epsilon^2 + m_\epsilon(\|h_\ell\|^d_{\infty;\Omega^0_\ell})^2 \big(\|\hat{Y}_k\|^2_{\mathbb{Y}} + \|\hat{U}_k\|^2_{\mathbb{U}} + \|\hat{P}_k\|_{\mathbb{Y}} + \|D\|^2_{\mathbb{D}}\big),
\end{aligned}
$$

using Assumption 3.3(5) for arbitrary but fixed $\epsilon > 0$, as well as the fact that, thanks to shape regularity, we have $\|h_k\|_{\infty;\Omega_{\mathcal{G}}(\mathcal{G}_k \setminus \mathcal{G}^+_l)} \lesssim \|h_\ell\|_{\infty;\Omega^0_\ell}$. Since $\|\hat{Y}_k\|^2_{\mathbb{Y}} + \|\hat{U}_k\|^2_{\mathbb{U}} + \|\hat{P}_k\|^2_{\mathbb{Y}} \lesssim 1$ (Proposition 4.4), we find

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

By Lemma 4.1 and Assumption 3.3(5), the last term $m_\epsilon(\|h_\ell\|^d_{\infty;\Omega^0_\ell})$ can be made small by choosing ℓ large. After fixing ℓ, we may choose $k \geq \ell$ (as in the proof to Proposition 5.2) such that $\mathcal{E}^2_{ocp}(\hat{U}_k, \hat{Y}_k, \hat{P}_k; \mathcal{G}^+_\ell)$ is small. Moreover, the above established error convergence implies that the term $\|(\hat{U}_k, \hat{Y}_k, \hat{P}_k) - (\hat{u}, \hat{p}, \hat{y})\|^2_{\mathbb{W}}$ is also small possibly after a further increase of k. In summary, we find that

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

for sufficiently large k. Since $\epsilon > 0$ was arbitrary, this yields $\mathcal{E}_{ocp}(\hat{U}_k, \hat{Y}_k, \hat{P}_k; \mathcal{G}_k) \to 0$ as $k \to \infty$ and finishes the proof. □

6. Applications

In this section, we shall demonstrate how the general framework from the previous sections can be used to obtain convergence for specific optimal control problems. To this end, we shall verify Assumption 3.3 for residual based estimaters in the particular cases of an reaction diffusion problem with boundary control and for the Stokes problem with distributed control, and provide a framework, which allows to easily generalise the results to multiple other kinds of estimators. For numerical computations, we refer the reader to [12, 13].

6.1. **A diffusion-reaction problem with boundary control.** We consider the following problem:

$$
-\Delta y + y = f_2 \quad \text{in } \Omega, \qquad \nabla y \cdot \boldsymbol{n} = \begin{cases} u + f_1 & \text{on } \Gamma, \\ 0, & \text{on } \partial\Omega \setminus \Gamma, \end{cases}
$$

where $\Gamma \subset \partial\Omega$ has positive $d-1$ dimensional Hausdorff measure, and $f_1 \in L^2(\Gamma)$ and $f_2 \in L^2(\Omega)$ are given data.

For the state space, we choose $\mathbb{Y} = H^1(\Omega)$ with $\|\cdot\|_{\mathbb{Y}} = \|\cdot\|_{H^1(\Omega)}$ and for the control space we let $\mathbb{U} = L^2(\Gamma)$ and $\|\cdot\|_{\mathbb{U}} = \|\cdot\|_{L^2(\Gamma)}$. Under the additional assumption that $\partial\Omega$ is Lipschitz, the embedding $\mathbb{Y} \hookrightarrow \mathbb{U} \hookrightarrow \mathbb{Y}^*$ is naturally given by the trace operator.

The bilinear form

$$\mathcal{B}[y,\, v] = \int_\Omega \nabla v \cdot \nabla y + vy \,\mathrm{d}x$$

of the weak formulation is continuous and coercive with $\|\mathcal{B}\| = \beta = 1$ and setting $\langle f,\, v\rangle := \langle f_1,\, v\rangle_{L^2(\Omega)} + \langle f_2,\, v\rangle_{L^2(\Gamma)}$, we have $f \in \mathbb{Y}^* = (H^1(\Omega))^*$. Finally, for given $g \in \mathbb{U}$ and desired state $y_d \in L^2(\Omega)$, we define the objective

$$\mathcal{J}[u, y] := \psi(y) + \frac{\alpha}{2}\|u\|_{\mathbb{U}}^2 := \frac{1}{2}\|y - y_d\|_{L^2(\Omega)}^2 + \langle g,\, y\rangle_{L^2(\Gamma)} + \frac{\alpha}{2}\|u\|_{\mathbb{U}}^2.$$

Consequently, the Fréchet derivative of ψ is given by $\psi'(\psi) = \langle y - y_d,\, \cdot\rangle_{L^2(\Omega)} + \langle g,\, \cdot\rangle_{L^2(\Gamma)} \in \mathbb{Y}^*$, which is Lipschitz continuous with constant $L = 1$.

Discretization. We concentrate on the case of discretized controls. Let the initial triangulation $\mathcal{G}_0 \in \mathbb{G}$ of Ω be such that

$$\mathcal{G}_0^\Gamma := \{E \cap \Gamma : E \cap \Gamma \text{ is a } (d-1) \text{ sub-simplex},\ E \in \mathcal{G}_0\}$$

meshes Γ exactly. For $\mathcal{G} \in \mathbb{G}$, we use piecewise polynomials of degrees $\ell_y, \ell_u > 0$ for the state and control discretization, i.e.

$$\mathbb{Y}(\mathcal{G}) := \{V \in C^0(\Omega) : V_{|E} \in \mathbb{P}_{\ell_y}(E), E \in \mathcal{G}\}$$

and

$$\mathbb{U}(\mathcal{G}) := \{V \in L^2(\Gamma) : V_{|T} \in \mathbb{P}_{\ell_u}, T \in \mathcal{G}^\Gamma\}.$$

The control indicators are then given by $\mathcal{E}_u(U, P; E) := \|U - \Pi(P)\|_{L^2(E\cap\Gamma)}$, $E \in \mathcal{G}$; compare with Section 2. In order to obtain a posteriori estimators for the optimal control problem, we have to complement $\mathcal{E}_u$ with estimators for the state and adjoint state.

Residual based estimators. For $\mathcal{G} \in \mathbb{G}$, the standard residual based a posteriori estimator for the state and the adjoint state are given by

$$\begin{aligned}
\mathcal{E}_y(Y, u; E) &= h_E\| - \Delta Y + Y - f_2\|_{L^2(E)} \\
&\qquad + h_E^{1/2}\|[\![\nabla Y]\!]\|_{L^2(\partial E\setminus\Gamma)} + h_E^{1/2}\|[\![\nabla Y]\!] - (U + f_1)\|_{L^2(\partial E\cap\Gamma)}, \\
\mathcal{E}_p(P, \psi'(Y); E) &= h_E\| - \Delta P + P - (Y - y_d)\|_{L^2(E)} \\
&\qquad + h_E^{1/2}\|[\![\nabla P]\!]\|_{L^2(\partial E\setminus\Gamma)} + h_E^{1/2}\|[\![\nabla P]\!] - g\|_{L^2(\partial E\cap\Gamma)}.
\end{aligned}$$

The estimators $\mathcal{E}_y$ and $\mathcal{E}_p$ are reliable and locally efficient with oscillations

$$\begin{aligned}
\operatorname{osc}_y(Y, u; E) &= h_E \inf_{c_2^E \in \mathbb{P}_{2\ell_y - 2}} \|Y - f_2 - c_2^E\|_{L^2(E)} \\
&\qquad + h_E^{1/2} \sum_{S\in\mathcal{G}^\Gamma, S\subset E} \inf_{c_1^S \in \mathbb{P}_{2\ell_y - 1}} \|U + f_1 - c_1^S\|_{L^2(T)}, \\
\operatorname{osc}_p(P, \psi'(Y); E) &= h_E \inf_{c_d^E \in \mathbb{P}_{2\ell_y - 2}} \|P - (Y - y_d) - c_d^E\|_{L^2(E)} \\
&\qquad + h_E^{1/2} \sum_{S\in\mathcal{G}^\Gamma, S\subset E} \inf_{g^S \in \mathbb{P}_{2\ell_y - 1}} \|g - g^S\|_{L^2(T)};
\end{aligned}$$

compare e.g. with [13, 3, 30]. Moreover, $\mathcal{E}_y$ and $\mathcal{E}_p$ are obviously Lipschitz continuous. Concluding, we have that they satisfy Assumptions 3.3(1)–(3). Assumption 3.3(5) follows with $(f_2, f_1), (y_d, g) \in \mathbb{D} = L^2(\Omega) \times L^2(\Gamma)$ and $m_\epsilon(|E|) = |E|^{1/2d}$ independent of $\epsilon > 0$. In order to verify Assumption 3.3(4), we shortly recall the derivation of the upper bound. Since the estimators for the state and the adjoint state have exactly the same structure, we shall restrict our considerations to the former case. Recalling the definition of the residual and using Galerkin orthogonality as well as integration by parts, we obtain with $u \in \mathbb{U}$ and $Y = S_{\mathcal{G}}(u+f) \in \mathbb{Y}(\mathcal{G})$, that

$$\begin{aligned}\langle \mathcal{R}(Y;u),\, v\rangle &= \langle \mathcal{R}(Y;u),\, v - V\rangle \\ &\le \sum_{E\in\mathcal{G}} \|-\Delta Y + Y - f_2\|_{L^2(E)}\|v - V\|_{L^2(E)} \\ &\qquad + \|[\![\nabla Y]\!]\|_{L^2(\partial E\setminus\Gamma)}\|v - V\|_{L^2(\partial E\setminus\Gamma)} \\ &\qquad + \|[\![\nabla Y]\!] - (U + f_1)\|_{L^2(\partial E\cap\Gamma)}\|v - V\|_{L^2(\partial E\cap\Gamma)}\end{aligned}$$

for all $V \in \mathbb{Y}(\mathcal{G})$. Using scaled trace and inverse inequalities and choosing V to be a suitable interpolant of v (see e.g. [4]), yields

$$\langle \mathcal{R}(Y;u),\, v\rangle \lesssim \sum_{E\in\mathcal{G}} h_E^s \mathcal{E}_y(Y, u; E)\|v\|_{H^{1+s}(\Omega_{\mathcal{G}}(E))}$$

for all $v \in H^{1+s}(\Omega)$, $s \in [0, 1]$. Therefore, Assumption 3.3(4) follows with $\mathbb{Y}_s = H^{1+s}(\Omega)$.

6.2. **The Stokes Problem with distributed control.** As an example of a non coercive problem, we shall next consider the following Stokes equations with distributed control:

$$-\Delta \boldsymbol{y} + \nabla q = \boldsymbol{f} + \boldsymbol{u} \quad \text{in } \Omega, \qquad \operatorname{div} \boldsymbol{y} = 0 \quad \text{in } \Omega, \qquad \text{and} \qquad \boldsymbol{y} = 0 \quad \text{on} \partial\Omega,$$

where $\boldsymbol{f} \in L^2(\Omega;\mathbb{R}^d)$ is some given data. The state $y = (\boldsymbol{y}, q)$ is the velocity and the pressure of the fluid and consequently we have for the state space $\mathbb{Y} = H_0^1(\Omega;\mathbb{R}^d) \times L_0^2(\Omega) = H_0^1(\Omega;\mathbb{R}^d) \times \{q \in L^2(\Omega) : \langle q,\, 1\rangle_{L^2(\Omega)} = 0\}$, which is a Hilbert space with norm $\|y\|_{\mathbb{Y}}^2 = \|\nabla \boldsymbol{y}\|_{L^2(\Omega)}^2 + \|q\|_{L^2(\Omega)}^2$ thanks to the Friedrichs inequality $\|\boldsymbol{y}\|_{L^2(\Omega)} \le C_F\|\nabla \boldsymbol{y}\|_{L^2(\Omega)}$. The control space is $\mathbb{U} = L^2(\Omega;\mathbb{R}^d)$ with norm $\|\cdot\|_{\mathbb{U}} = \|\cdot\|_{L^2(\Omega)}$ and the embedding $\mathbb{Y} \hookrightarrow \mathbb{U} \hookrightarrow \mathbb{Y}^*$ is given by $H_0^1(\Omega;\mathbb{R}^d) \hookrightarrow L^2(\Omega;\mathbb{R}^d)$, i. e., $y = (\boldsymbol{y}, q) \in \mathbb{Y}$ implies $\boldsymbol{y} \in L^2(\Omega;\mathbb{R}^d)$ or equivalently $\langle q,\, y\rangle = \langle q,\, \boldsymbol{y}\rangle_{L^2(\Omega)}$ for $q \in L^2(\Omega;\mathbb{R}^d)$. The resulting bilinear form

$$\mathcal{B}[y,\, p] = \mathcal{B}[(\boldsymbol{y}, q),\, (\boldsymbol{p}, r)] := \int_\Omega \nabla \boldsymbol{y} : \nabla \boldsymbol{p} - q \operatorname{div} \boldsymbol{p} + r \operatorname{div} \boldsymbol{y}\, \mathrm{d}x,$$

is continuous and inf-sup stable; compare with [8]. We choose box-constraints for the control i. e.,

$$\mathbb{U}^{\mathrm{ad}} = \{\boldsymbol{u} \in \mathbb{U}\colon \boldsymbol{u} \in \mathcal{C} \text{ a.e. in } \Omega\} \quad \text{with} \quad \mathcal{C} = \{\boldsymbol{v} \in \mathbb{R}^d : \boldsymbol{a} \le \boldsymbol{v} \le \boldsymbol{b}\},$$

where $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$ with $\boldsymbol{a} \le \boldsymbol{b}$ and the inequalities are understood componentwise. This implies

$$\Pi(p)(x) = \max\bigl\{\boldsymbol{a}, \min\{\boldsymbol{b}, \boldsymbol{p}(x)\}\bigr\}, \qquad p = (\boldsymbol{p}, q) \in \mathbb{Y}.$$

We define the objective by

$$\mathcal{J}[\boldsymbol{u}, y] := \frac{1}{2}\|y\|_{\mathbb{Y}}^2 + \frac{\alpha}{2}\|\boldsymbol{u}\|_{\mathbb{U}}^2.$$

In other words $\psi(y) = \frac{1}{2}\|y\|_{\mathbb{Y}}^2$, which is obviously Fréchet differentiable and its derivative ψ' is locally Lipschitz continuous with constant $L = 1$.

Discretization. Since we are dealing with a saddle point problem, we need to resort to an inf-sup stable discretization of $\mathbb{Y}$. For $\mathcal{G} \in \mathbb{G}$, a possible choice is e.g. the common Taylor-Hood element of degree $\ell_y \geq 2$ where $\mathbb{Y}(\mathcal{G}) = \mathbb{V}(\mathcal{G}) \times \mathbb{Q}(\mathcal{G})$ with

$$\begin{aligned}
\mathbb{V}(\mathcal{G}) &= \{\boldsymbol{V} \in C^0(\Omega; \mathbb{R}^d) \colon \boldsymbol{V}_{|E} \in \mathbb{P}_{\ell_y}(E)^d, E \in \mathcal{G}\}, \\
\mathbb{Q}(\mathcal{G}) &= \{Q \in C^0(\Omega) \colon Q_{|E} \in \mathbb{P}_{\ell_y - 1}(E), E \in \mathcal{G} \text{ and } \langle Q,\, 1\rangle_{L^2(\Omega)} = 0\}.
\end{aligned}$$

We choose to discretize the control with discontinuous polynomials of degree $\ell_u \geq 0$, i. e.,

$$\mathbb{U}(\mathcal{G}) = \{\boldsymbol{U} \in L^2(\Omega; \mathbb{R}^d) \colon \boldsymbol{U}_{|E} \in \mathbb{P}_{\ell_u}(E)^d, E \in \mathcal{G}\}.$$

Consequently, we have that

$$\mathbb{U}^{\mathrm{ad}}(\mathcal{G}) = \mathbb{U}(\mathcal{G}) \cap \mathbb{U}^{\mathrm{ad}} = \{\boldsymbol{U} \in \mathbb{U}(\mathcal{G}) \colon \boldsymbol{a} \leq \boldsymbol{U} \leq \boldsymbol{b} \quad \text{a.e. in } \Omega\}.$$

Thanks to the use of box constraints, we have that the restrictions decouple component wise and thus we have e.g. for $\ell_u = 0$ and $x \in E \in \mathcal{G}$, that

$$\Pi_{\mathcal{G}}(P)(x) = \max\left\{\boldsymbol{a}, \min\left\{\boldsymbol{b}, \frac{1}{|E|}\int_E \boldsymbol{P}\,\mathrm{d}x\right\}\right\}, \qquad P = (\boldsymbol{P}, Q) \in \mathbb{Y}(\mathcal{G}),$$

where the maximum and the minimum is taken componentwise.

Residual based estimators. We are now concerned with complementing the control indicator $\mathcal{E}_u(\boldsymbol{U}, P; E) = \|\boldsymbol{U} - \Pi(P)\|_{L^2(E)}$ by residual based estimators for the state and the adjoint state satisfying Assumption 3.3. For $E \in \mathcal{G}$, $\mathcal{G} \in \mathbb{G}$, $Y = (\boldsymbol{Y}, Q), P = (\boldsymbol{P}, R) \in \mathbb{Y}(\mathcal{G})$, and $\boldsymbol{u} \in \mathbb{U}$ let

$$\begin{aligned}
\mathcal{E}_y(Y, \boldsymbol{u}; E) = h_E\| -\Delta\boldsymbol{Y} + \nabla Q - \boldsymbol{f} - \boldsymbol{u}\|_{L^2(E)} \\
+ h_E^{1/2}\|[\![\nabla\boldsymbol{Y}]\!]\|_{L^2(\partial E)} + \|\operatorname{div}\boldsymbol{Y}\|_{L^2(E)}
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{E}_p(P, \psi'(Y); E) = h_E\| -\Delta(\boldsymbol{P} - \boldsymbol{Y}) + \nabla R\|_{L^2(E)} \\
+ h_E^{1/2}\|[\![\nabla\boldsymbol{P} - \nabla\boldsymbol{Y}]\!]\|_{L^2(\partial E)} + \|\operatorname{div}\boldsymbol{Y} - Q\|_{L^2(E)}.
\end{aligned}$$

It is well known [29, 30], that these estimators satisfy Assumptions 3.3(1)–(2) with oscillations

$$\operatorname{osc}_y(Y, \boldsymbol{u}; E) = h_E \inf_{\boldsymbol{c}_E \in \mathbb{P}^d_{\ell_y - 2}} \|\boldsymbol{f} - \boldsymbol{u} - \boldsymbol{c}_E\|_{L^2(E)} \quad \text{and} \quad \operatorname{osc}_p(P, \psi'(Y); E) = 0.$$

Assumptions 3.3(3)–(5) follow then similarly as in Section 6.1 for the diffusion reaction problem; compare also with [13].

6.3. **Other types of estimators.** We summarize, that the convergence of an AFEM for an optimal control problem (2.1) essentially hinges on the properties of estimators for the state and adjoint state equations. Nowadays, for each of countless PDEs, there is a large zoo of estimators available. Of course Assumption 3.3 could be checked for each estimator separately. However, for many of these estimators it is well known that they are locally equivalent to the corresponding residual based ones. We illustrate this principle with the help of the hierarchical estimator in Example 6.1 below. Local equivalence for e.g. estimators based on local problems, on the equilibration of fluxes, or on gradient recovery as well as robust estimators for singularly perturbed reaction diffusion problems can be found in [17]; compare also with [30]. We shall now show convergence of an AFEM' using error estimators $\hat{\mathcal{E}}_y, \hat{\mathcal{E}}_p$, which are locally equivalent to error estimators $\mathcal{E}_y, \mathcal{E}_p$ satisfying Assumption 3.3.

To this end, suppose that the estimators $\hat{\mathcal{E}}_y$ and $\hat{\mathcal{E}}_p$ are organized by some index set $\mathcal{I} = \mathcal{I}(\mathcal{G})$, $\mathcal{G} \in \mathbb{G}$, which can be either the elements $\mathcal{G}$ or the nodes or the sides of $\mathcal{G}$. Additionally, we assume that we have for $\mathcal{G}' \subset \mathcal{G}$ and $\mathcal{I}' \subset \mathcal{I}$ the following local equivalence

$$\mathcal{E}_y\big(\hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; \mathcal{G}'\big) \lesssim \hat{\mathcal{E}}_y\big(\hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; \mathcal{I}(\mathcal{G}')\big) \quad \text{and} \quad \hat{\mathcal{E}}_y\big(\hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; \mathcal{I}'\big) \lesssim \mathcal{E}_y\big(\hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; \mathcal{G}(\mathcal{I}')\big)$$

and corresponding relations for $\mathcal{E}_p$. Here $\mathcal{G}(\mathcal{I}') \subset \mathcal{I}$ and $\mathcal{I}(\mathcal{G}') \subset \mathcal{G}$, such that

$$\mathcal{G}' \subset \mathcal{G}(\mathcal{I}(\mathcal{G}')) \subset \mathcal{N}_\mathcal{G}(\mathcal{G}'), \qquad \#\mathcal{G}(\mathcal{I}') \lesssim \#\mathcal{I}', \qquad \text{and} \qquad \#\mathcal{I}(\mathcal{G}') \lesssim \#\mathcal{G}', \tag{6.1}$$

and only finite many of the $\mathcal{G}(I), I \in \mathcal{I}$ overlap. Similarly as before, we use the convention $\hat{\mathcal{E}}_y(\hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; \mathcal{I}')^2 = \sum_{I \in \mathcal{I}'} \mathcal{E}_y(\hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; I)^2$. Obviously, this implies

$$\hat{\mathcal{E}}_y(\hat{P}_\mathcal{G}, \psi'(\hat{Y}_\mathcal{G}); \mathcal{I}) \eqsim \mathcal{E}_y(\hat{P}_\mathcal{G}, \psi'(\hat{Y}_\mathcal{G}); \mathcal{G})$$

and we have for $E \in \mathcal{G}$ with $\#\mathcal{I}(E) \lesssim 1$, that

$$\mathcal{E}_y\big(\hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; E\big)^2 \lesssim \hat{\mathcal{E}}_y\big(\hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; \mathcal{I}(E)\big)^2 \lesssim \max\big\{\hat{\mathcal{E}}_y\big(\hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; I\big)^2 : I \in \mathcal{I}(E)\big\}.$$

We shall now consider an AFEM' of the form (2.9) with $\mathcal{E}_{ocp}(\hat{U}, \hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; E)$ replaced by

$$\hat{\mathcal{E}}_{ocp}(\hat{U}_\mathcal{G}, \hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; I) := \mathcal{E}_u(\hat{U}_\mathcal{G}, \hat{P}_\mathcal{G}; I) + \hat{\mathcal{E}}_y(\hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; I) + \hat{\mathcal{E}}_p(\hat{P}_\mathcal{G}, \psi'(\hat{Y}_\mathcal{G}); I)$$

and assume that the marking strategy

$$\mathcal{M} = \mathcal{G}(\mathcal{I}_\mathcal{M}) = \mathsf{MARK}(\{\hat{\mathcal{E}}_{ocp}(\hat{U}_\mathcal{G}, \hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; I)\}_{I \in \mathcal{I}}, \mathcal{I}) \subset \mathcal{G}$$

satisfies

(3') $\max\{\hat{\mathcal{E}}_{ocp}(\hat{U}_\mathcal{G}, \hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; I) \mid I \in \mathcal{I}\} \le C \max\{\hat{\mathcal{E}}_{ocp}(\hat{U}_\mathcal{G}, \hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; I) \mid I \in \mathcal{I}_\mathcal{M}\}$;

instead of Assumption 2.1(3). Then we have

$$\begin{aligned} \max\big\{\mathcal{E}_{ocp}(\hat{U}_\mathcal{G}, \hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; E) : E \in \mathcal{G}\big\} &\lesssim \max\big\{\hat{\mathcal{E}}_{ocp}(\hat{U}_\mathcal{G}, \hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; I) : I \in \mathcal{I}\big\} \\ &\lesssim \max\big\{\hat{\mathcal{E}}_{ocp}(\hat{U}_\mathcal{G}, \hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; I) : E \in \mathcal{I}_\mathcal{M}\big\} \\ &\lesssim \max\big\{\mathcal{E}_{ocp}(\hat{U}_\mathcal{G}, \hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; E) : E \in \mathcal{M}\big\}. \end{aligned}$$

In other words, $\mathcal{M}$ satisfies Assumption 2.1(3) for $\mathcal{E}_{ocp}$ and thus Theorem 2.2 implies that the AFEM' converges to the exact solution.

Example 6.1 (Hierarchical estimators). We reconsider the problem of Section 6.1 and restrict ourselves to $\ell_y = 1$. The idea of hierarchical estimators is based upon evaluating the residual on a sufficiently enriched discrete space $\mathbb{Y}(\mathcal{G})' \supsetneq \mathbb{Y}(\mathcal{G})$. Suitable enrichments contain e.g. element and side bubble functions of higher order or on a finer mesh; compare e.g. with [28, 30]. For given $(y, u+f)$ and (p, y), the residuals of the primal and dual problem from Section 6.1 are given by

$$\langle \mathcal{R}(y, u+f),\, v\rangle = \mathcal{B}[y,\, v] - \langle f_1 + u,\, v\rangle_{L^2(\Gamma)} - \langle f_2,\, v\rangle_{L^2(\Omega)}$$

and

$$\langle \mathcal{R}^*(p, \psi'(y)),\, v\rangle = \mathcal{B}[v,\, p] - \langle y - y_d,\, v\rangle_{L^2(\Gamma)} - \langle g,\, v\rangle_{L^2(\Gamma\Omega)}.$$

As index-set for the hierarchical estimators, we use the set $\mathcal{S}$ of all sides of $\mathcal{G}$ and define

$$\mathcal{S}(\mathcal{G}') := \{S \in \mathcal{S} : S \subset E \text{ for some } E \in \mathcal{G}'\}$$

and

$$\mathcal{G}(\mathcal{S}') := \{E \in \mathcal{G} : S \subset E \text{ for some } S \in \mathcal{S}'\}.$$

Obviously this choice satisfies (6.1). For a fixed $S \in \mathcal{S}$, we let z_S be the barycenter of S and consider an enrichment $\mathbb{Y}(\mathcal{G})'$ of $\mathbb{Y}(\mathcal{G})$ that provides for any $S \in \mathcal{S}$ a function $\Phi_S \in \mathbb{Y}(\mathcal{G})' \setminus \mathbb{Y}(\mathcal{G})$ with

$$\Psi_S(z_S) > 0, \qquad \operatorname{supp}(\Phi_S) \subset \omega_S, \quad \text{and} \quad \|\Phi_S\|_{H^1(\Omega)} = 1,$$

where $\omega_S = \Omega(\mathcal{G}(S))$. The side oriented hierarchical indicators on $S \in \mathcal{S}$ are then given by

$$\begin{aligned}\hat{\mathcal{E}}_y(\hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; S) := &\Big|\langle \mathcal{R}(\hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G} + f),\, \Phi_S\rangle\Big| \\ &+ h_S\|\hat{Y}_\mathcal{G} - f_2\|_{L^2(\omega_S)} + h_S^{1/2} \inf_{c_1^S \in \mathbb{R}} \|\hat{U}_\mathcal{G} + f_1 - c_1^S\|_{L^2(S\cap\partial\Omega)}\end{aligned}$$

and

$$\begin{aligned}\hat{\mathcal{E}}_p(\hat{P}_\mathcal{G}, \psi'(\hat{Y}_\mathcal{G}); S) := &\Big|\langle \mathcal{R}^*(\hat{P}_\mathcal{G}, \psi'(\hat{Y}_\mathcal{G})),\, \Phi_S\rangle\Big| \\ &+ h_S\|\hat{P}_\mathcal{G} - (\hat{Y}_\mathcal{G} - y_d)\|_{L^2(\omega_S)} + h_S^{1/2} \inf_{g^S \in \mathbb{R}} \|g - g^S\|_{L^2(S\cap\partial\Omega)}.\end{aligned}$$

Similarly as in [17], it can be shown that the indicators are locally equivalent, i. e., we have

$$\mathcal{E}_y(\hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; E) \lesssim \hat{\mathcal{E}}_y(\hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; \mathcal{S}(E)) \quad \text{and} \quad \hat{\mathcal{E}}_y(\hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; S) \lesssim \mathcal{E}_y(\hat{Y}_\mathcal{G}, \hat{U}_\mathcal{G}; \mathcal{G}(S)),$$

and corresponding relations for the indicators of the adjoint state.

Acknowledgement. The research of Kristina Kohls, Arnd Rösch, and Kunibert G. Siebert was partially supported by the DFG research program 1253 "Optimization with Partial Differential Equations".

References

[1] I. Babuška and M. Vogelius, *Feedback and adaptive finite element solution of one-dimensional boundary value problems*, Numer. Math., 44(1) (1984), pp. 75–102.

[2] E. Bänsch, *Local mesh refinement in 2 and 3 dimensions*, IMPACT Comput. Sci. Eng., 3(3) (1991), pp. 181–191.

[3] J. M. Cascon, C. Kreuzer, R. H. Nochetto, and K. G. Siebert, *Quasi-optimal convergence rate for an adaptive finite element method*, SIAM J. Numer. Anal. 46(5) (2008), pp. 2524–2550.

[4] P. Clément, *Approximation by finite element functions using local regularization*, RAIRO Analyse Numérique, 9(R-2) (1975), pp. 77–84.

[5] L. Diening, C. Kreuzer, and R. Stevenson, *Instance Optimality of the Adaptive Maximum Strategy*, Found. Comput. Math. 16(1) (2016), pp. 33–68.

[6] W. Dörfler, *convergent adaptive algorithm for Poisson's equation*, SIAM J. Numer. Anal. 33(3) (1996), pp. 1106–1124.

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

[8] G. P. Galdi,, *An introduction to the mathematical theory of the Navier-Stokes equations*, Vol. I, Springer Tracts in natural Philosophy, vol. 38, Springer-Verlag, New York, 1994, Linearized steady problems.

[9] W. Gong and N. Yan, *Adaptive finite element method for elliptic optimal control problems: convergence and optimality*, Numerische Mathematik, 2016, DOI 10.1007/s00211-016-0827-9.

[10] M. Hintermüller, R. H. W. Hoppe, Y. Iliash, and M. Kieweg, *An a posteriori error analysis of adaptive finite element methods for distributed elliptic control problems with control constraints*, ESAIM, Control Optim. Calc. Var., 14(3) (2008), pp. 540–560.

[11] M. Hinze, *A variational discretization concept in control constrained optimization: the linear-quadratic case*, Comput. Optim. Appl., 30(1) (2005), pp. 45–61.

[12] K. Kohls, *An Adaptive Finite Element Method for Control-Constrained Optimal Control Problems*, PhD thesis, Fachbereich Mathematik, Unversität Stuttgart, 2013.

[13] K. Kohls, A. Rösch, and K. G. Siebert, *A posteriori error analysis of optimal control problems with control constraints*, SIAM Journal Control and Optimization 52(3): 1832-1861, 2014.

[14] K. Kohls, A. Rösch, and K. G. Siebert, *A posteriori error estimators for control constrained optimal control problems*, in Constrained Optimiziation and Optimal Control for Partial Differential Equations, vol. 160 of Internat. Ser. Numer. Math., Birkhäuser/Springer, Basel, 2012, pp. 431–443.

[15] K. Kohls, A. Rösch, and K. G. Siebert, *Convergence of Adaptive Finite Elements for Optimal Control Problems with Control Constrains*, Trends in PDE Constrained Optimization, vol. 165 of Internat. Ser. Numer. Math., Birkhäuser/Springer, Basel, 2015, pp. 403–419.

[16] I. Kossaczký, *A recursive approach to local mesh refinement in two and three dimensions*, J. Comput. Appl. Math. 55(3) (1994), pp. 275–288.

[17] C. Kreuzer and K. G. Siebert, *Decay rates of adaptive finite elements with Dörfler marking*, Numer. Math., 117(4) (2011), pp. 679–716.

[18] J. L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, vol. 170 of Die Grundlehren der mathematischen Wissenschaften, Springer-Verlag, 1971.

[19] W. Liu and N. Yan, *A posteriori error estimates for distributed convex optimal control problems*, Adv. Comput. Math, 15(1-4) (2001), pp. 285–309.

[20] C. Makridakis and R. H. Nochetto, *Elliptic reconstruction and a posteriori error estimates for parabolic problems*, SIAM J. Numer. Anal., 41 (2003)(4), pp. 1585–1594.

[21] P. Morin, R. H. Nochetto, and K. G. Siebert,*Data oscillation and convergence of adaptive FEM*, SIAM J. Numer. Anal., 38(2) (2000), pp. 466–488.

[22] P. Morin, K. G. Siebert, and A. Veeser, *A basic convergence result for conforming adaptive finite elements*, Math. Models Methods Appl. Sci., 18(5) (2008), pp. 707–737.

[23] R. H. Nochetto, K. G. Siebert, and A. Veeser, *Theory of adaptive finite element methods: an introduction*, in Multiscale, nonlinear and adaptive approximation, Springer, Berlin 2009, pp. 409–542.

[24] A. Schmidt and K. G. Siebert, *Design of adaptive finite element software*, vol. 42 of Lecture Notes in Computational Science and Engineering, Springer-Verlag, Berlin, 2005. The finite element toolbox ALBERTA, With 1 CD-ROM (Unix/Linux).
[25] K. G. Siebert, *A convergence proof for adaptive finite elements without lower bound.*, IMA J. Numer. Anal., 31(3) (2011), pp. 947–970.
[26] R. Stevenson,*Optimality of a standard adaptive finite element method*, Found. Comput. Math., 7(2) (2007), pp. 245–269.
[27] F. Tröltzsch, *Optimal control of partial differential equations*, vol. 112 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2010. Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels.
[28] A. Veeser, *Convergent adaptive finite elements for the nonlinear Laplacian*, Numer. Math., 92(4) (2002), pp. 743–770.
[29] Rüdiger Verfürth, *A posteriori error estimators for the Stokes equations*, Numer. Math. 55(3) (2002), pp. 309–325.
[30] Rüdiger Verfürth, *A posteriori error estimation techniques for finite element methods*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 2013.

Kristina Kohls, Institut für Angewandte Analysis und Numerische Simulation, Fachbereich Mathematik, Universität Stuttgart, Pfaffenwaldring 57, D-70569 Stuttgart, Germany
URL: http://www.ians.uni-stuttgart.de/nmh/

Christian Kreuzer, Fakultät für Mathematik, Ruhr-Universität Bochum, Universitätsstrasse 150 44801 Bochum, Germany
URL: http://www.ruhr-uni-bochum.de/ffm/Lehrstuehle/Kreuzer/index.html
E-mail address: christian.kreuzer@rub.de

Arnd Rösch, Fakultät für Mathematik, Universität Duisburg-Essen, Thea-Leymann-Strasse 9, D-45127 Essen, Germany
URL: http://www.uni-due.de/mathematik/agroesch/
E-mail address: arnd.roesch@uni-due.de

Kunibert G. Siebert, Institut für Angewandte Analysis und Numerische Simulation, Fachbereich Mathematik, Universität Stuttgart, Pfaffenwaldring 57, D-70569 Stuttgart, Germany
URL: http://www.ians.uni-stuttgart.de/nmh/
E-mail address: kg.siebert@ians.uni-stuttgart.de