Adaptive finite element methods for sparse PDE–constrained optimization

A. ALLENDES †, F. FUICA ‡

Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile.

AND

E. OTÁROLA§

Departamento de Matemática, Universidad Técnica Federico Santa María, Santiago, Chile.

We propose and analyze reliable and efficient a posteriori error estimators for an optimal control problem that involves a nondifferentiable cost functional, the Poisson problem as state equation and control constraints. To approximate the solution to the state and adjoint equations we consider a piecewise linear finite element method whereas *three different strategies* are used to approximate the control variable: piecewise constant discretization, piecewise linear discretization and the so–called variational discretization approach. For the first two aforementioned solution techniques we devise an error estimator that can be decomposed as the sum of four contributions: two contributions that account for the discretization of the control variable and the associated subgradient, and two contributions related to the discretization of the state and adjoint equations. The error estimator for the variational discretization approach is decomposed only in two contributions that are related to the discretization of the state and adjoint equations. On the basis of the devised a posteriori error estimators, we design simple adaptive strategies that yield optimal rates of convergence for the numerical examples that we perform.

Keywords: PDE–constrained optimization, nondifferentiable objectives, sparse controls, a posteriori error analysis, adaptive finite elements.

1. Introduction.

In this work we shall be interested in the design and analysis of a posteriori error estimators for a nondifferentiable optimal control problem; control constraints are also considered. To make matters precise, we let $\Omega \subset \mathbb{R}^d$, with $d \in \{2,3\}$, be an open and bounded polytopal domain with Lipschitz boundary $\partial \Omega$. Given $f \in L^2(\Omega)$, a desired state $y_\Omega \in L^2(\Omega)$, a regularization parameter $\alpha > 0$, and a sparsity parameter $\beta > 0$, we define the nondifferentiable cost functional

$$J(y,u) := \frac{1}{2} \|y - y_\Omega\|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \|u\|_{L^2(\Omega)}^2 + \beta \|u\|_{L^1(\Omega)}. \tag{1.1}$$

We shall thus be concerned with the nonsmooth optimal control problem: Find

$$\min J(y,u) \tag{1.2}$$

subject to the state equation

$$-\Delta y = u + f \text{ in } \Omega, \qquad y = 0 \text{ on } \partial\Omega, \tag{1.3}$$

and the control constraints

$$u \in \mathbb{U}_{ad}, \quad \mathbb{U}_{ad} := \{v \in L^2(\Omega) : a \leqslant v(x) \leqslant b \text{ a.e. } x \in \Omega\}. \tag{1.4}$$

We immediately comment that, since we are interested in the nondifferentiable scenario, we assume that $a, b \in \mathbb{R}$ are such that $a < 0 < b$. We refer the reader to (Casas *et al.*, 2012c, Remark 2.1) for a discussion.

The design and analysis of solution techniques for problem (1.2)–(1.4) are motivated by the following two observations:

- The cost functional J involves the $L^1(\Omega)$–norm of the control variable. This term, that is a natural measure of the control cost, leads to sparsely supported optimal controls Casas *et al.* (2012c); Stadler (2009); Wachsmuth & Wachsmuth (2011), i.e., *optimal controls that are not zero only in a small region of the considered domain*. This is a desirable feature in applications, for instance, in the optimal placement of discrete actuators Stadler (2009).

- The cost functional J is nondifferentiable. As a consequence, the study of solution techniques for (1.2)–(1.4) present some extra mathematical difficulties compared with the standard case $\alpha > 0$ and $\beta = 0$ that is presented, for instance, in Tröltzsch (2010). Fortunately, these difficulties can be overcame with elements from convex analysis Casas *et al.* (2012c); Wachsmuth & Wachsmuth (2011).

The analysis and finite element discretization of PDE–constrained optimization problems that involve a cost functional containing a $L^1(\Omega)$–control cost term have been considered in a number of works. To the best of our knowledge, the first work that provides an analysis when the state equation is a linear and elliptic PDE is Stadler (2009). In this work, the author utilizes a regularization technique that involves a $L^2(\Omega)$–control cost term, analyzes optimality conditions, and studies the convergence properties of a proposed semismooth Newton method. Later, these results were complemented with rates of convergence with respect to the regularization parameter α in Wachsmuth & Wachsmuth (2011). Subsequently, the authors of Casas *et al.* (2012c) consider a nonlinear version of (1.2)–(1.4) where the state equation is a semilinear elliptic PDE and analyze second order optimality conditions. We refer the reader to the recent work of Casas (2017) for a complete overview of the results available in the literature. Simultaneously with these advances, discretization techniques based on finite element methods and their corresponding *a priori error analyses* have been considered. We refer the reader to Wachsmuth & Wachsmuth (2011), when the state equation is a linear elliptic PDE, to Casas *et al.* (2012b,c) for extensions to the semilinear case, and to Otárola & Salgado (2018) when the state equation (1.3) involves the spectral fractional powers of elliptic operators. We also mention references Casas *et al.* (2017); Casas & Kunisch (2017) for extensions of the aforementioned developments to evolution problems.

As opposed to the available a priori error analysis for finite element approximations of sparse PDE–constrained optimization, the design and analysis of a posteriori error estimators are rather scarce. An a posteriori error estimator is a computable quantity that depends on the discrete solution and data, and provides information about the local quality of the approximate solution. It is an essential ingredient of adaptive finite element methods (AFEMs). The theory for linear second–order elliptic boundary value

problems is well–established Ainsworth & Oden (2000); Nochetto *et al.* (2009); Nochetto & Veeser (2011); Verfürth (2013). In contrast, the theory for constrained optimal control problems is not as developed. The main source of difficulty is its inherent nonlinear feature, which appears as a result of the control constraints. To the best of our knowledge, the first work that provides an advance is Liu & Yan (2001) where the authors propose an estimator and derive a reliability estimate (Liu & Yan, 2001, Theorem 3.1). Subsequently, the analysis was improved in Hintermüller *et al.* (2008) by providing efficiency estimates involving oscillation terms (Hintermüller *et al.*, 2008, Theorems 5.1 and 6.1). An attempt to unify the available results in the literature was later presented in Kohls *et al.* (2014): on the basis of an important error equivalence the analysis is simplified to provide reliable and efficient estimators for the state and adjoint equations. The analysis is based on the energy norm inherited by the state and adjoint equations. Recently, the authors of Schneider & Wachsmuth (2016) provided a general framework that complements the one developed in Kohls *et al.* (2014), and measures the error in a norm that is motivated by the objective. The analysis relies on the convexity of Ω. The common feature in all the previous cited references is that, in contrast to (1.2)–(1.4), $\beta = 0$. For different approaches based on *weighted residual* and *goal–oriented* methods and advances in the semilinear and nonlinear case, the reader is referred to Becker *et al.* (2007); Hintermüller & Hoppe (2008); Meyer *et al.* (2015); Vexler & Wollner (2008).

To the best of our knowledge, the only work that provides an advance concerning a posteriori error analysis for (1.2)–(1.4) is (Wachsmuth & Wachsmuth, 2011). In this work, the authors consider a piecewise constant discretization for the control variable, propose a residual–type a posteriori error estimator. In Theorem 6.2, it is proved that the devised error estimator yields an upper bound for the approximation errors of the state and control variables (the errors committed in the approximation of the associated subgradient and the adjoint variable are not considered). However, no efficiency analysis is provided in Wachsmuth & Wachsmuth (2011). In this work we complement and extend the results presented in (Wachsmuth & Wachsmuth, 2011, Section 6) as follows: We consider *three discretization schemes* for (1.2)–(1.4) that rely on the discretization of the state and adjoint equations with piecewise linear functions. The schemes differ on the type of discretization considered for the control variable: piecewise constant, piecewise linear or variational discretization. For the first two schemes, we design an a posteriori error estimator that accounts for the discretization of the optimal control variable, its associated subgradient, and the state and adjoint variables. The a posteriori error estimator designed for the variational discretization approach only needs to account for the discretization of the state and adjoint variables. We measure the total error in energy–norms and $L^2(\Omega)$–norms and derive, for each scheme, global reliability and local efficiency results in a unified manner. With these estimators at hand, we also design simple adaptive strategies that yield optimal rates of convergence for the numerical examples that we perform.

We organize our exposition as follows. We set notation in Section 2 and briefly recall elements from convex analysis. In Section 3 we present existence and uniqueness results together with first–order necessary and sufficient optimality conditions. In Section 4 we present three finite element discretizations for the optimal control problem (1.2)–(1.4); all of them rely on the discretization of the state and adjoint equations by using piecewise linear functions. To approximate the control variable three strategies are considered: piecewise constant, piecewise linear and variational discretization. The core of our work is Section 5 where, for each discretization presented in Section 4, we design an a posteriori error estimator and derive reliability and local efficiency results. We conclude, in Section 6, with a series of numerical examples that illustrate our theory.

2. Notation and Preliminaries

Let us fix notation and the functional setting in which we will operate. Throughout this work $d \in \{2,3\}$ and $\Omega \subset \mathbb{R}^d$ is an open and bounded polytopal domain with Lipschitz boundary. For a bounded domain $G \subset \mathbb{R}^d$, $L^2(G)$ and $H^1(G)$ denote the standard Lebesgue and Sobolev spaces, respectively, and $H_0^1(G)$ is the subspace of $H^1(G)$ consisting of functions whose trace is zero on ∂G. Let $(\cdot,\cdot)_{L^2(G)}$ and $\|\cdot\|_{L^2(G)}$ denote, respectively, the inner product and norm in $L^2(G)$. The seminorm in $H^1(G)$ is denoted by $|\cdot|_{H^1(G)}$.

If $\mathscr{X}$ and $\mathscr{Y}$ are normed vector spaces, we write $\mathscr{X} \hookrightarrow \mathscr{Y}$ to denote that $\mathscr{X}$ is continuously embedded in $\mathscr{Y}$. We denote by $\mathscr{X}^\star$ the dual of $\mathscr{X}$. The relation $a \lesssim b$ indicates that $a \leqslant Cb$, with a nonessential constant C that might change at each occurrence. Finally, throughout the manuscript we will frequently make use of the following Poincaré inequality

$$\|w\|_{L^2(\Omega)} \leqslant \mathfrak{C}\|\nabla w\|_{L^2(\Omega)} \quad \forall w \in H_0^1(\Omega). \tag{2.1}$$

2.1 *Convex functions and subdifferentials*

In this section we recall some elements from convex analysis that will be essential for the analysis that we will perform.

Let E be a real normed vector space. Let $\eta : E \to \mathbb{R} \cup \{\infty\}$ be convex and proper, and let $v \in E$ with $\eta(v) < \infty$. A subgradient of η at v is a continuous linear functional $v^\star$ on E that satisfies

$$\langle v^\star, w - v \rangle \leqslant \eta(w) - \eta(v) \quad \forall\, w \in E, \tag{2.2}$$

where $\langle \cdot, \cdot \rangle$ denotes the duality pairing between $E^\star$ and E. We immediately remark that a function may admit many subgradients at a point of nondifferentiability. The set of all subgradients of η at v is called subdifferential of η at v and is denoted by $\partial \eta(v)$. By convexity, the subdifferential $\partial \eta(v) \neq \emptyset$ for all points v in the interior of the effective domain of η. Finally, we mention that the subdifferential is monotone, i.e.,

$$\langle v^\star - w^\star, v - w \rangle \geqslant 0 \quad \forall v^\star \in \partial \eta(v),\, \forall w^\star \in \partial \eta(w). \tag{2.3}$$

We refer the reader to Clarke (1990); Schirotzek (2007) for a thorough discussion on convex analysis.

3. Sparse PDE–constrained optimization.

In this section we briefly review the analysis of the nondifferentiable optimal control problem (1.2)–(1.4). We recall existence and uniqueness results together with first–order necessary and sufficient optimality conditions.

For J defined as in (1.1), the nondifferentiable optimal control problem reads:

$$\min_{H_0^1(\Omega) \times \mathbb{U}_{ad}} J(y,u) \tag{3.1}$$

subject to the linear and elliptic PDE

$$(\nabla y, \nabla v)_{L^2(\Omega)} = (u + f, v)_{L^2(\Omega)} \quad \forall v \in H_0^1(\Omega). \tag{3.2}$$

We must immediately notice that the set $\mathbb{U}_{ad}$, defined as in (1.4), is a nonempty, bounded, closed and convex subset of $L^2(\Omega)$.

We define the control-to-state map $\mathscr{L}$ as follows: given $u \in L^2(\Omega)$, $\mathscr{L}$ associates to it a unique state $y \in H_0^1(\Omega)$ that solves (3.2). Since $H_0^1(\Omega) \hookrightarrow L^2(\Omega)$, we may also consider $\mathscr{L}$ acting from $L^2(\Omega)$ into itself. An immediate application of Lax-Milgram Lemma implies that $\mathscr{L}$ is a linear and continuous map. With this operator at hand we define the reduced cost functional

$$j(u) = J(\mathscr{L}u, u) := \frac{1}{2} \|\mathscr{L}u - y_\Omega\|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \|u\|_{L^2(\Omega)}^2 + \beta \|u\|_{L^1(\Omega)},$$

and present the following result; see also (Wachsmuth & Wachsmuth, 2011, Lemma 2.1).

LEMMA 3.1 (well–posedness) The sparse PDE–constrained optimization problem (3.1)–(3.2) has a unique optimal solution $(\bar{y}, \bar{u}) \in H_0^1(\Omega) \times L^2(\Omega)$.

Proof. Since $\mathscr{L}$ is injective and continuous, j is strictly convex and weakly lower semicontinuous. The fact that $\mathbb{U}_{ad}$ is weakly sequentially compact allows us to conclude. $\qquad\square$

In order to obtain optimality conditions for (3.1)–(3.2) we introduce the following ingredients. First, we define the so–called adjoint state p as follows:

$$p \in H_0^1(\Omega): \quad (\nabla w, \nabla p)_{L^2(\Omega)} = (y - y_\Omega, w)_{L^2(\Omega)} \quad \forall w \in H_0^1(\Omega). \tag{3.3}$$

We define the convex and Lipschitz function $\psi : L^1(\Omega) \mapsto \mathbb{R}$ by $\psi(u) := \|u\|_{L^1(\Omega)}$; it corresponds to the nondifferentiable component of the reduced cost functional j. The differential counterpart of the latter is defined by

$$\varphi : L^2(\Omega) \to \mathbb{R}, \quad u \mapsto \varphi(u) := \frac{1}{2} \|\mathscr{L}u - y_\Omega\|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \|u\|_{L^2(\Omega)}^2.$$

Standard arguments reveal that φ is Fréchet differentiable with $\varphi'(u) = p + \alpha u$ (see (Tröltzsch, 2010, Theorem 2.20).

With these ingredients at hand, we present necessary and sufficient optimality conditions for our sparse PDE–constrained optimization problem.

THEOREM 3.1 (optimality conditions) The pair $(\bar{y}, \bar{u}) \in H_0^1(\Omega) \times \mathbb{U}_{ad}$ is optimal for problem (3.1)–(3.2) if and only if $\bar{y} = \mathscr{L}\bar{u}$ and $\bar{u}$ satisfies the variational inequality

$$(\bar{p} + \alpha\bar{u} + \beta\bar{\lambda}, u - \bar{u})_{L^2(\Omega)} \geqslant 0 \quad \forall u \in \mathbb{U}_{ad}, \tag{3.4}$$

where $\bar{p}$ denotes the solution to (3.3) with y replaced by $\bar{y}$ and $\bar{\lambda} \in \partial\psi(\bar{u})$.

Proof. See (Wachsmuth & Wachsmuth, 2011, Lemma 2.2). $\qquad\square$

To present the following result we introduce, for $\mathfrak{a}, \mathfrak{b} \in \mathbb{R}$ the projection formula

$$\Pi_{[\mathfrak{a},\mathfrak{b}]}(v(x)) = \min\{\mathfrak{b}, \max\{\mathfrak{a}, v(x)\}\}. \tag{3.5}$$

COROLLARY 3.1 (projection formula) Let $\bar{y}, \bar{p}, \bar{u}$ and $\bar{\lambda}$ be as in Theorem 3.1. Then, we have that

$$\bar{u}(x) = \Pi_{[a,b]}\left(-\frac{1}{\alpha}\left(\bar{p}(x) + \beta\bar{\lambda}(x)\right)\right), \tag{3.6}$$

and

$$\bar{u}(x) = 0 \quad \Leftrightarrow \quad |\bar{p}(x)| \leqslant \beta, \qquad \bar{\lambda}(x) = \Pi_{[-1,1]}\left(-\frac{1}{\beta}\bar{p}(x)\right). \tag{3.7}$$

Consequently, $\bar{u}, \bar{\lambda} \in H_0^1(\Omega)$, and $\bar{\lambda}$ is uniquely determined.

Proof. See (Casas *et al.*, 2012c, Corollary 3.2). □

We immediately comment that the projection formula (3.7) guarantees the uniqueness of the subgradient $\bar{\lambda}$ (Casas *et al.*, 2012c, Corollary 3.2); this property is not usually obtained in non–differentiable optimization problems.

To summarize, the pair $(\bar{y}, \bar{u})$ is optimal for (3.1)–(3.2) if and only if $(\bar{y}, \bar{p}, \bar{u}) \in H_0^1(\Omega) \times H_0^1(\Omega) \times \mathbb{U}_{ad}$ solves

$$\begin{cases} (\nabla \bar{y}, \nabla v)_{L^2(\Omega)} = (\bar{u} + f, v)_{L^2(\Omega)} & \forall v \in H_0^1(\Omega), \\ (\nabla w, \nabla \bar{p})_{L^2(\Omega)} = (\bar{y} - y_\Omega, w)_{L^2(\Omega)} & \forall w \in H_0^1(\Omega), \\ (\bar{p} + \alpha \bar{u} + \beta \bar{\lambda}, u - \bar{u})_{L^2(\Omega)} \geqslant 0 & \forall u \in \mathbb{U}_{ad}, \end{cases} \tag{3.8}$$

where $\bar{\lambda} \in \partial \psi(\bar{u})$.

4. Finite element discretization.

We present three finite element solution techniques for the nondifferentiable optimal control problem (3.1)–(3.2). All the techniques discretize the state and adjoint equations with piecewise linear functions. However, they differ on the type of discretization technique used for the optimal control variable. In Section 4.1, we consider a piecewise constant discretization, in Section 4.2, an scheme based on piecewise linear functions and, in Section, 4.3 we consider the so–called variational discretization approach.

We begin this section by introducing some finite element notation (Ciarlet, 2002; Ern & Guermond, 2004). Let $\mathscr{T} = \{K\}$ be a conforming partition of $\bar{\Omega}$ into simplices K with size $h_K := \mathrm{diam}(K)$, and set $h_\mathscr{T} := \max_{K \in \mathscr{T}} h_K$. We denote by $\mathbb{T}$ the collection of conforming and shape regular meshes that are refinements of an initial mesh $\mathscr{T}_0$.

Given a mesh $\mathscr{T} \in \mathbb{T}$, we define the finite element space of continuous piecewise polynomials of degree one as

$$\mathbb{V}(\mathscr{T}) = \{v_\mathscr{T} \in C(\bar{\Omega}) : v_\mathscr{T}|_K \in \mathbb{P}_1(K) \forall K \in \mathscr{T}, v_\mathscr{T}|_{\partial \Omega} = 0\}. \tag{4.1}$$

In what follows we will describe the three solution techniques that we will consider for our optimal control problem (3.1)–(3.2).

4.1 *Piecewise constant discretization.*

We define $\mathbb{U}_0(\mathscr{T}) := \{u_\mathscr{T} \in L^2(\Omega) : u_\mathscr{T}|_K \in \mathbb{P}_0(K) \forall K \in \mathscr{T}\}$. The discrete admissible set $\mathbb{U}_{ad,0}(\mathscr{T})$ is thus defined as

$$\mathbb{U}_{ad,0}(\mathscr{T}) := \mathbb{U}_0(\mathscr{T}) \cap \mathbb{U}_{ad}.$$

With these discrete spaces at hand, we propose the following finite element discretization of the optimality system (3.8): Find $(\bar{y}_\mathscr{T}, \bar{p}_\mathscr{T}, \bar{u}_\mathscr{T}) \in \mathbb{V}(\mathscr{T}) \times \mathbb{V}(\mathscr{T}) \times \mathbb{U}_{ad,0}(\mathscr{T})$ such that

$$\begin{cases} (\nabla \bar{y}_\mathscr{T}, \nabla v_\mathscr{T}) = (\bar{u}_\mathscr{T} + f, v_\mathscr{T})_{L^2(\Omega)} & \forall v_\mathscr{T} \in \mathbb{V}(\mathscr{T}), \\ (\nabla w_\mathscr{T}, \nabla \bar{p}_\mathscr{T}) = (\bar{y}_\mathscr{T} - y_\Omega, w_\mathscr{T})_{L^2(\Omega)} & \forall w_\mathscr{T} \in \mathbb{V}(\mathscr{T}), \\ (\bar{p}_\mathscr{T} + \alpha \bar{u}_\mathscr{T} + \beta \bar{\lambda}_\mathscr{T}, u_\mathscr{T} - \bar{u}_\mathscr{T})_{L^2(\Omega)} \geqslant 0 & \forall u_\mathscr{T} \in \mathbb{U}_{ad,0}(\mathscr{T}), \end{cases} \tag{4.2}$$

where $\bar{\lambda}_\mathscr{T} \in \partial \psi(\bar{u}_\mathscr{T})$ and

$$\psi : \mathbb{U}_0(\mathscr{T}) \to \mathbb{R}, \quad u_\mathscr{T} \mapsto \psi(u_\mathscr{T}) = \int_\Omega |u_\mathscr{T}| \mathrm{d}x = \sum_{K \in \mathscr{T}} |u_\mathscr{T}||K|.$$

The next result states discrete projection formulas for $\bar{u}_\mathscr{T}$ and $\bar{\lambda}_\mathscr{T}$.

LEMMA 4.1 (discrete projection formulas in $\mathbb{U}_0(\mathscr{T})$) Let $(\bar{y}_{\mathscr{T}}, \bar{p}_{\mathscr{T}}, \bar{u}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \times \mathbb{V}(\mathscr{T}) \times \mathbb{U}_{ad,0}(\mathscr{T})$ be the solution to (4.2). Then, we have

$$\bar{u}_{\mathscr{T}}|_K = \Pi_{[a,b]}\left(-\frac{1}{\alpha}\left(\frac{1}{|K|}\int_K \bar{p}_{\mathscr{T}}\,\mathrm{d}x + \beta\bar{\lambda}_{\mathscr{T}}|_K\right)\right), \tag{4.3}$$

and

$$\bar{u}_{\mathscr{T}}|_K = 0 \Leftrightarrow \frac{1}{|K|}\left|\int_K \bar{p}_{\mathscr{T}}\,\mathrm{d}x\right| \leqslant \beta, \quad \bar{\lambda}_{\mathscr{T}}|_K = \Pi_{[-1,1]}\left(-\frac{1}{\beta|K|}\int_K \bar{p}_{\mathscr{T}}\,\mathrm{d}x\right). \tag{4.4}$$

Consequently, the discrete subgradient $\bar{\lambda}_{\mathscr{T}}$ is unique.

Proof. We begin by noticing that $\partial\psi(\bar{u}_{\mathscr{T}}) \subset \mathbb{U}_0(\mathscr{T})^*$. Consequently, standard arguments, on the basis of (2.2), allow us to identify $\bar{\lambda}_{\mathscr{T}} \in \partial\psi(\bar{u}_{\mathscr{T}})$ with an element of $\mathbb{U}_0(\mathscr{T})$ that satisfies

$$\bar{\lambda}_{\mathscr{T}} = \sum_{K \in \mathscr{T}} \chi_K \bar{\lambda}_{\mathscr{T}}|_K, \qquad \begin{cases} \bar{\lambda}_{\mathscr{T}}|_K = +1, & \bar{u}_{\mathscr{T}}|_K > 0, \\ \bar{\lambda}_{\mathscr{T}}|_K = -1, & \bar{u}_{\mathscr{T}}|_K < 0, \\ \bar{\lambda}_{\mathscr{T}}|_K \in [-1,1], & \bar{u}_{\mathscr{T}}|_K = 0, \end{cases} \tag{4.5}$$

where χ_K corresponds to the characteristic function of the element $K \in \mathscr{T}$. Thus, since $\bar{u}_{\mathscr{T}} \in \mathbb{U}_{ad,0}(\mathscr{T})$, the variational inequality in the optimality system (4.2) reads

$$\sum_{K \in \mathscr{T}}\left(\int_K \bar{p}_{\mathscr{T}}\,\mathrm{d}x + |K|\left(\alpha\bar{u}_{\mathscr{T}}|_K + \beta\bar{\lambda}_{\mathscr{T}}|_K\right)\right)\left(u_{\mathscr{T}}|_K - \bar{u}_{\mathscr{T}}|_K\right) \geqslant 0,$$

where, for every $K \in \mathscr{T}$, $u_{\mathscr{T}}|_K \in \mathbb{P}_0(K)$ is such that $a \leqslant u_{\mathscr{T}}|_K \leqslant b$. We thus invoke similar arguments to the ones used in the proof of (Tröltzsch, 2010, Lemma 2.26) to obtain the projection formula (4.3).

The proof of (4.4) follows from (3.5), (4.3), and (4.5); see (Casas *et al.*, 2012c, Section 4) for details. This concludes the proof.   □

4.2 *Piecewise linear discretization.*

Let us define $\mathbb{U}_1(\mathscr{T}) := \{u_{\mathscr{T}} \in C(\bar{\Omega}) : u_{\mathscr{T}}|_K \in \mathbb{P}_1(K)\ \forall K \in \mathscr{T}\}$. The discrete admissible set is thus defined as

$$\mathbb{U}_{ad,1}(\mathscr{T}) := \mathbb{U}_1(\mathscr{T}) \cap \mathbb{U}_{ad}.$$

We denote the set of all vertices of the mesh $\mathscr{T}$ by $\mathscr{V}(\mathscr{T})$, and, for $\mathrm{v} \in \mathscr{V}(\mathscr{T})$, we introduce the function $\phi_{\mathrm{v}} \in \mathbb{U}_1(\mathscr{T})$ which is such that $\phi_{\mathrm{v}}(\mathrm{v}') = \delta_{\mathrm{vv}'}$ for all $\mathrm{v}' \in \mathscr{V}(\mathscr{T})$. The set $\{\phi_{\mathrm{v}} : \mathrm{v} \in \mathscr{V}(\mathscr{T})\}$ is the so–called Courant basis of the space $\mathbb{U}_1(\mathscr{T})$ Ern & Guermond (2004); Nochetto *et al.* (2009). We notice that every element $u_{\mathscr{T}} \in \mathbb{U}_1(\mathscr{T})$ can be written as

$$u_{\mathscr{T}} = \sum_{\mathrm{v} \in \mathscr{V}(\mathscr{T})} u_{\mathscr{T}}(\mathrm{v})\phi_{\mathrm{v}}.$$

We follow (Casas *et al.*, 2012b, Section 3) and define, on the space $\mathbb{U}_1(\mathscr{T})$, the discrete inner product $(\cdot,\cdot)_{\mathscr{T}}$ and norm $\|\cdot\|_{\mathscr{T}}$ by

$$(u_{\mathscr{T}}, v_{\mathscr{T}})_{\mathscr{T}} = \sum_{\mathrm{v} \in \mathscr{V}(\mathscr{T})} u_{\mathscr{T}}(\mathrm{v})v_{\mathscr{T}}(\mathrm{v})\int_\Omega \phi_{\mathrm{v}}\,\mathrm{d}x, \quad \|u_{\mathscr{T}}\|_{\mathscr{T}}^2 = (u_{\mathscr{T}}, u_{\mathscr{T}})_{\mathscr{T}}, \tag{4.6}$$

respectively. We also define the discrete nondifferentiable component $\psi_{\mathscr{T}} : \mathbb{U}_1(\mathscr{T}) \to \mathbb{R}$ as $\psi_{\mathscr{T}}(u_{\mathscr{T}}) = \sum_{v \in \mathscr{V}(\mathscr{T})} |u_{\mathscr{T}}(v)| \int_{\Omega} \phi_v \, dx$ and the discrete cost functional

$$J_{\mathscr{T}}(y_{\mathscr{T}}, u_{\mathscr{T}}) = \frac{1}{2}\|y_{\mathscr{T}} - y_{\Omega}\|_{L^2(\Omega)}^2 + \frac{\alpha}{2}\|u_{\mathscr{T}}\|_{\mathscr{T}}^2 + \beta\psi_{\mathscr{T}}(u_{\mathscr{T}}).$$

With this discrete setting at hand, we propose the following finite element discretization of the optimality system (3.8) (Casas *et al.*, 2012b, Theorem 3.3): Find $(\bar{y}_{\mathscr{T}}, \bar{p}_{\mathscr{T}}, \bar{u}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \times \mathbb{V}(\mathscr{T}) \times \mathbb{U}_{ad,1}(\mathscr{T})$ such that

$$\begin{cases} (\nabla\bar{y}_{\mathscr{T}}, \nabla v_{\mathscr{T}}) = (\bar{u}_{\mathscr{T}} + f, v_{\mathscr{T}})_{L^2(\Omega)} & \forall v_{\mathscr{T}} \in \mathbb{V}(\mathscr{T}), \\ (\nabla w_{\mathscr{T}}, \nabla\bar{p}_{\mathscr{T}}) = (\bar{y}_{\mathscr{T}} - y_{\Omega}, w_{\mathscr{T}})_{L^2(\Omega)} & \forall w_{\mathscr{T}} \in \mathbb{V}(\mathscr{T}), \\ (\bar{p}_{\mathscr{T}}, u_{\mathscr{T}} - \bar{u}_{\mathscr{T}})_{L^2(\Omega)} + (\alpha\bar{u}_{\mathscr{T}} + \beta\bar{\lambda}_{\mathscr{T}}, u_{\mathscr{T}} - \bar{u}_{\mathscr{T}})_{\mathscr{T}} \geqslant 0 & \forall u_{\mathscr{T}} \in \mathbb{U}_{ad,1}(\mathscr{T}), \end{cases} \tag{4.7}$$

where $\bar{\lambda}_{\mathscr{T}} \in \partial\psi_{\mathscr{T}}(\bar{u}_{\mathscr{T}})$. To present the following result we introduce the quasi–interpolation operator $\Theta_{\mathscr{T}} : L^1(\Omega) \to \mathbb{U}_1(\mathscr{T})$, that is defined as follows:

$$\Theta_{\mathscr{T}}(w) = \sum_{v \in \mathscr{V}(\mathscr{T})} \theta_v(w)\phi_v, \qquad \theta_v(w) := \frac{\int_{\Omega} w\phi_v \, dx}{\int_{\Omega} \phi_v \, dx}. \tag{4.8}$$

LEMMA 4.2 (discrete projection formulas in $\mathbb{U}_{ad,1}(\mathscr{T})$) Let $(\bar{y}_{\mathscr{T}}, \bar{p}_{\mathscr{T}}, \bar{u}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \times \mathbb{V}(\mathscr{T}) \times \mathbb{U}_{ad,1}(\mathscr{T})$ be the solution to (4.7). Then, for every $v \in \mathscr{V}(\mathscr{T})$, we have

$$\bar{u}_{\mathscr{T}}(v) = \Pi_{[a,b]}\left(-\frac{1}{\alpha}\left(\theta_v(\bar{p}_{\mathscr{T}}) + \beta\bar{\lambda}_{\mathscr{T}}(v)\right)\right), \tag{4.9}$$

and

$$\bar{u}_{\mathscr{T}}(v) = 0 \quad \Leftrightarrow \quad |\theta_v(\bar{p}_{\mathscr{T}})| \leqslant \beta, \qquad \bar{\lambda}_{\mathscr{T}}(v) = \Pi_{[-1,1]}\left(-\frac{1}{\beta}\theta_v(\bar{p}_{\mathscr{T}})\right). \tag{4.10}$$

In particular, the discrete subgradient $\bar{\lambda}_{\mathscr{T}}$ is unique.

Proof. See (Casas *et al.*, 2012b, Lemma 3.4). $\qquad\qquad\square$

REMARK 4.1 (projection formulas) The projection formulas obtained in Lemmas 4.1 and 4.2 are essential ingredients in the numerical resolution of system (4.2) and (4.7), respectively; see **Algorithm 1** in Section 6.

PROPOSITION 4.1 (a priori error estimates) If (4.2) and (4.7) approximate the optimal control problem (3.1)–(3.2) when $\mathbb{U}_{ad}(\mathscr{T}) = \mathbb{U}_{ad,0}(\mathscr{T})$ and $\mathbb{U}_{ad}(\mathscr{T}) = \mathbb{U}_{ad,1}(\mathscr{T})$, respectively and Ω is convex, then, the following a priori error estimates can be derived: For every $h_0 > 0$, there is a constant $C > 0$ such that for all $h_{\mathscr{T}} \leqslant h_0$

$$\|\bar{u} - \bar{u}_{\mathscr{T}}\|_{L^2(\Omega)} \leqslant Ch_{\mathscr{T}}, \tag{4.11}$$

where C is independent of $h_{\mathscr{T}}$.

Proof. For a proof of this result we refer the reader to (Wachsmuth & Wachsmuth, 2011, Proposition 4.5) when $\mathbb{U}_{ad}(\mathscr{T}) = \mathbb{U}_{ad,0}(\mathscr{T})$ and (Casas *et al.*, 2012b, Theorem 3.13) when $\mathbb{U}_{ad}(\mathscr{T}) = \mathbb{U}_{ad,1}(\mathscr{T})$. $\square$

4.3 *Variational discretization*

In what follows we will consider the so–called variational discretization approach introduced by Hinze in (Hinze, 2005). We discretize the state equation with the help of the discrete space (4.1); the admissible set of controls $\mathbb{U}_{ad}$ is not discretized. In spite of this fact, the proposed semidiscrete scheme will induce a discretization of the optimal control and its unique associated subgradient on the basis of projection formulas; see Lemma 4.3 below.

With the aforementioned semidiscrete setting at hand, we propose the following finite element discretization of the optimality system (3.8) (Casas *et al.*, 2012c, Section 5): Find $(\bar{y}_{\mathscr{T}}, \bar{p}_{\mathscr{T}}, \bar{u}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \times \mathbb{V}(\mathscr{T}) \times \mathbb{U}_{ad}$ such that

$$
\begin{cases}
(\nabla \bar{y}_{\mathscr{T}}, \nabla v_{\mathscr{T}}) = (\bar{u}_{\mathscr{T}} + f, v_{\mathscr{T}})_{L^2(\Omega)} & \forall v_{\mathscr{T}} \in \mathbb{V}(\mathscr{T}), \\
(\nabla w_{\mathscr{T}}, \nabla \bar{p}_{\mathscr{T}}) = (\bar{y}_{\mathscr{T}} - y_{\Omega}, w_{\mathscr{T}})_{L^2(\Omega)} & \forall w_{\mathscr{T}} \in \mathbb{V}(\mathscr{T}), \\
(\bar{p}_{\mathscr{T}} + \alpha \bar{u}_{\mathscr{T}} + \beta \bar{\lambda}_{\mathscr{T}}, u_{\mathscr{T}} - \bar{u}_{\mathscr{T}})_{L^2(\Omega)} \geqslant 0 & \forall u_{\mathscr{T}} \in \mathbb{U}_{ad},
\end{cases}
\tag{4.12}
$$

where $\bar{\lambda}_{\mathscr{T}} \in \partial \psi(\bar{u}_{\mathscr{T}})$. We now present projection formulas for the variables $\bar{u}_{\mathscr{T}}$ and $\bar{\lambda}_{\mathscr{T}}$ that make evident how they are implicitly discretized by the semidiscrete scheme (4.12).

LEMMA 4.3 (variational discrete projections) Let $(\bar{y}_{\mathscr{T}}, \bar{p}_{\mathscr{T}}, \bar{u}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \times \mathbb{V}(\mathscr{T}) \times \mathbb{U}_{ad}$ be the solution to (4.12). Then, for all $x \in \Omega$, we have that

$$
\bar{u}_{\mathscr{T}}(x) = \Pi_{[a,b]} \left(-\frac{1}{\alpha} \left(\bar{p}_{\mathscr{T}}(x) + \beta \bar{\lambda}_{\mathscr{T}}(x) \right) \right),
\tag{4.13}
$$

and

$$
\bar{u}_{\mathscr{T}}(x) = 0 \quad \Leftrightarrow \quad |\bar{p}_{\mathscr{T}}(x)| \leqslant \beta, \qquad \bar{\lambda}_{\mathscr{T}}(x) = \Pi_{[-1,1]} \left(-\frac{1}{\beta} \bar{p}_{\mathscr{T}}(x) \right).
\tag{4.14}
$$

In addition, the discrete subgradient $\bar{\lambda}_{\mathscr{T}}$ is unique.

Proof. See (Casas *et al.*, 2012c, Section 5). □

PROPOSITION 4.2 (a priori error estimate) If (4.12) approximate the optimal control problem (3.1)–(3.2) and Ω is convex, then the following a priori error estimate can be derived: For every $h_0 > 0$, there exits $C > 0$ such that for all $h_{\mathscr{T}} \leqslant h_0$,

$$
\|\bar{u} - \bar{u}_{\mathscr{T}}\|_{L^2(\Omega)} \leqslant C h_{\mathscr{T}}^2,
\tag{4.15}
$$

where C is independent of $h_{\mathscr{T}}$.

Proof. See (Wachsmuth & Wachsmuth, 2011, Corollary 4.7) □

5. A posteriori error estimation.

The design and analysis of AFEMs to solve the optimal control problem (3.1)–(3.2) are motivated by the following considerations:

- the a priori error estimates obtained in Casas *et al.* (2012c); Wachsmuth & Wachsmuth (2011) require $\mathscr{T}$ to be quasi–uniform and Ω to be *convex*. In addition, such estimates are valid under the assumption that $h_{\mathscr{T}}$ is sufficiently small. If the condition that Ω is convex is violated, the optimal variables may have singularities and thus exhibit fractional regularity. As a consequence, quasi–uniform refinement of Ω would not result in an efficient solution technique.

- the sparsity term $\psi(u)$ in the cost functional yield an optimal control $\bar{u}$ that is nonzero only in sets of small support in Ω.

It is then natural, to efficiently resolve such a behavior on the optimal control variable and recover optimal rates of convergence when Ω is not convex, to propose AFEMs.

In the next section we will construct three types of a posteriori error estimators; two of them will be based on the following four contributions: two contributions that account for the discretization of the control variable and the associated subgradient, and two contributions related to the discretization of the state and adjoint variables. Instead, the a posteriori error estimator for the variational discretization approach is based only in two contributions: one related to the discretization of the state variable, and another one related to the discretization of the adjoint variable.

5.1 *A posteriori error analysis for the Laplacian*

Since the error estimators that we will propose involve contributions that account for the discretization of the state and adjoint variables, in what follows we summarize some classical a posteriori error estimates for the Laplacian.

Let $g \in L^2(\Omega)$ and consider the following problem: Find $z \in H_0^1(\Omega)$ such that

$$(\nabla z, \nabla v)_{L^2(\Omega)} = (g, v)_{L^2(\Omega)} \quad \forall v \in H_0^1(\Omega). \tag{5.1}$$

We define the Galerkin approximation to (5.1) as the solution to: Find $z_{\mathscr{T}} \in \mathbb{V}(\mathscr{T})$ such that

$$(\nabla z_{\mathscr{T}}, \nabla v_{\mathscr{T}})_{L^2(\Omega)} = (g, v_{\mathscr{T}})_{L^2(\Omega)} \quad \forall v_{\mathscr{T}} \in \mathbb{V}(\mathscr{T}).$$

We define $\mathscr{S}$ as the set of internal $(d-1)$–dimensional interelement boundaries γ of $\mathscr{T}$. For $K \in \mathscr{T}$, let $\mathscr{S}_K$ denote the subset of $\mathscr{S}$ that contains the sides in $\mathscr{S}$ which are sides of K. We also denote by Ω_γ the subset of $\mathscr{T}$ that contains the two elements that have γ as a side. In addition, we define the patch associated with an element $K \in \mathscr{T}$ as

$$\Omega_K := \bigcup_{K' \in \mathscr{T}: \mathscr{S}_K \cap \mathscr{S}_{K'} \neq \emptyset} K'. \tag{5.2}$$

Given a discrete function $z_{\mathscr{T}} \in \mathbb{V}(\mathscr{T})$, we define, for any internal side $\gamma \in \mathscr{S}$, the jump or interelement residual $[\![\nabla z_{\mathscr{T}} \cdot \nu]\!]_\gamma$ by

$$[\![\nabla z_{\mathscr{T}} \cdot \nu]\!]_\gamma = \nu^+ \cdot \nabla z_{\mathscr{T}}|_{K^+} + \nu^- \cdot \nabla z_{\mathscr{T}}|_{K^-},$$

where ν^+, ν^- denote the unit normals to γ pointing outward $K^+, K^- \in \mathscr{T}$, respectively, which are such that $K^+ \neq K^-$ and $\partial K^+ \cap \partial K^- = \gamma$.

With these ingredients at hand, we introduce the following a posteriori error indicators and error estimator

$$\mathscr{E}_{z,K}^2 = h_K^2 \|g\|_{L^2(K)}^2 + h_K \big\|[\![\nabla z_{\mathscr{T}} \cdot \nu]\!]_\gamma\big\|_{L^2(\partial K \setminus \partial \Omega)}^2, \quad \mathscr{E}_z := \left(\sum_{K \in \mathscr{T}} \mathscr{E}_{z,K}^2 \right)^{\frac{1}{2}}, \tag{5.3}$$

respectively. It is well–known that there exists a positive constant $\mathscr{C}_z$ such that the following global reliability result holds:

$$|z - z_{\mathscr{T}}|_{H^1(\Omega)} \leqslant \mathscr{C}_z \mathscr{E}_z. \tag{5.4}$$

We refer the reader to (Ainsworth & Oden, 2000, Section 2.2) and (Verfürth, 2013, Section 1.4) for details.

Let us now define the following a posteriori error indicators and error estimator

$$E_{z,K}^2 = h_K^4 \|g\|_{L^2(K)}^2 + h_K^3 \big\| [\![\nabla z_{\mathscr{T}} \cdot \nu]\!]_\gamma \big\|_{L^2(\partial K \backslash \partial \Omega)}^2, \quad E_z := \left(\sum_{K \in \mathscr{T}} E_{z,K}^2 \right)^{\frac{1}{2}}, \tag{5.5}$$

respectively. If Ω is convex, a duality argument reveals that there exist a positive constant C_z such that

$$\|z - z_{\mathscr{T}}\|_{L^2(\Omega)} \leqslant C_z E_z. \tag{5.6}$$

We refer the reader to (Ainsworth & Oden, 2000, Section 2.4) for details.

REMARK 5.1 (data oscillation) As it is customary in a posteriori error analysis, global reliability properties for residual–type error estimators do not involve oscillation terms Ainsworth & Oden (2000); Nochetto & Veeser (2011); Nochetto *et al.* (2009); Verfürth (2013). Such terms appear when analyzing the asymptotic sharpness of the a posteriori upper bounds (5.4) and (5.6) Ainsworth & Oden (2000); Nochetto & Veeser (2011); Nochetto *et al.* (2009); Verfürth (2013). One may think that the issue of oscillation is specific to standard a posteriori error estimation. However all estimators we are aware of suffer from oscillations of the data that are finer than the mesh–size Nochetto & Veeser (2011); Nochetto *et al.* (2009).

5.2 *Error estimators for sparse PDE–constrained optimization: reliability*

The upper bounds for the errors that we will obtain in our work are constructed using upper bounds on the error between the solution to the discretization (4.2), (4.7) or (4.12) and auxiliary variables that we define in what follows.

Let $(\bar{y}_{\mathscr{T}}, \bar{p}_{\mathscr{T}}, \bar{u}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \times \mathbb{V}(\mathscr{T}) \times \mathbb{U}_{ad}(\mathscr{T})$ be the solution to (4.2), (4.7) or (4.12); $\mathbb{U}_{ad}(\mathscr{T}) = \mathbb{U}_{ad,0}(\mathscr{T})$ for (4.2), $\mathbb{U}_{ad}(\mathscr{T}) = \mathbb{U}_{ad,1}(\mathscr{T})$ for (4.7), and $\mathbb{U}_{ad}(\mathscr{T}) = \mathbb{U}_{ad}$ for (4.12). We define $(\hat{y}, \hat{p}) \in H_0^1(\Omega) \times H_0^1(\Omega)$ as the solution to

$$\begin{cases} (\nabla \hat{y}, \nabla v)_{L^2(\Omega)} &= (\bar{u}_{\mathscr{T}} + f, v)_{L^2(\Omega)} & \forall v \in H_0^1(\Omega), \\ (\nabla w, \nabla \hat{p})_{L^2(\Omega)} &= (\bar{y}_{\mathscr{T}} - y_\Omega, w)_{L^2(\Omega)} & \forall w \in H_0^1(\Omega). \end{cases} \tag{5.7}$$

We notice that $(\bar{y}_{\mathscr{T}}, \bar{p}_{\mathscr{T}})$ can be seen as a finite element approximation of $(\hat{y}, \hat{p})$. This property motivates the following definitions. First, we define

$$\mathscr{E}_{y,K}^2 = h_K^2 \|\bar{u}_{\mathscr{T}} + f\|_{L^2(K)}^2 + h_K \big\| [\![\nabla \bar{y}_{\mathscr{T}} \cdot \nu]\!]_\gamma \big\|_{L^2(\partial K \backslash \partial \Omega)}^2, \quad \mathscr{E}_y^2 := \sum_{K \in \mathscr{T}} \mathscr{E}_{y,K}^2, \tag{5.8}$$

$$\mathscr{E}_{p,K}^2 = h_K^2 \|\bar{y}_{\mathscr{T}} - y_\Omega\|_{L^2(K)}^2 + h_K \big\| [\![\nabla \bar{p}_{\mathscr{T}} \cdot \nu]\!]_\gamma \big\|_{L^2(\partial K \backslash \partial \Omega)}^2, \quad \mathscr{E}_p^2 := \sum_{K \in \mathscr{T}} \mathscr{E}_{p,K}^2. \tag{5.9}$$

In view of the results of the previous section we conclude from (5.4) that there exist constants $\mathscr{C}_y$ and $\mathscr{C}_p$ such that

$$|\hat{y} - \bar{y}_{\mathscr{T}}|_{H^1(\Omega)} \leqslant \mathscr{C}_y \mathscr{E}_y, \qquad |\hat{p} - \bar{p}_{\mathscr{T}}|_{H^1(\Omega)} \leqslant \mathscr{C}_p \mathscr{E}_p. \tag{5.10}$$

Secondly, we define the $L^2(\Omega)$–based a posteriori error indicators and estimators

$$E_{y,K}^2 = h_K^4 \|\bar{u}_{\mathscr{T}} + f\|_{L^2(K)}^2 + h_K^3 \left\|[\![\nabla \bar{y}_{\mathscr{T}} \cdot \nu]\!]_\gamma\right\|_{L^2(\partial K \setminus \partial \Omega)}^2, \quad E_y^2 := \sum_{K \in \mathscr{T}} E_{y,K}^2, \tag{5.11}$$

$$E_{p,K}^2 = h_K^4 \|\bar{y}_{\mathscr{T}} - y_\Omega\|_{L^2(K)}^2 + h_K^3 \left\|[\![\nabla \bar{p}_{\mathscr{T}} \cdot \nu]\!]_\gamma\right\|_{L^2(\partial K \setminus \partial \Omega)}^2, \quad E_p^2 := \sum_{K \in \mathscr{T}} E_{p,K}^2. \tag{5.12}$$

If, in addition, Ω is convex, we thus have from (5.6) that there exist constants C_y and C_p such that

$$\|\hat{y} - \bar{y}_{\mathscr{T}}\|_{L^2(\Omega)} \leqslant C_y E_y, \qquad \|\hat{p} - \bar{p}_{\mathscr{T}}\|_{L^2(\Omega)} \leqslant C_p E_p. \tag{5.13}$$

We now define

$$\tilde{\lambda} := \Pi_{[-1,1]}\left(-\frac{1}{\beta}\bar{p}_{\mathscr{T}}\right), \qquad \tilde{u} := \Pi_{[a,b]}\left(-\frac{1}{\alpha}\left(\bar{p}_{\mathscr{T}} + \beta\tilde{\lambda}\right)\right). \tag{5.14}$$

The following remark is thus necessary.

REMARK 5.2 (properties of $\tilde{u}$ and $\tilde{\lambda}$) We notice two properties which are consequences of definition (5.14). First, $\tilde{\lambda} \in \partial\psi(\tilde{u})$. This will be crucial in the a posteriori error analysis that we will perform in Section 5. Second, if the variational approach is considered, we thus have that

$$\tilde{u} = \bar{u}_{\mathscr{T}}, \qquad \tilde{\lambda} = \bar{\lambda}_{\mathscr{T}}. \tag{5.15}$$

With these ingredients at hand, we define the following a posteriori error indicators and estimators for the optimal control variable and the associated subgradient:

$$E_{u,K}^2 := \|\tilde{u} - \bar{u}_{\mathscr{T}}\|_{L^2(K)}^2, \qquad E_u := \left(\sum_{K \in \mathscr{T}} E_{u,K}^2\right)^{\frac{1}{2}}, \tag{5.16}$$

$$E_{\lambda,K}^2 := \|\tilde{\lambda} - \bar{\lambda}_{\mathscr{T}}\|_{L^2(K)}^2, \qquad E_\lambda := \left(\sum_{K \in \mathscr{T}} E_{\lambda,K}^2\right)^{\frac{1}{2}}. \tag{5.17}$$

Related to new variable $\tilde{u} \in L^2(\Omega)$, we set $(\tilde{y}, \tilde{p}) \in H_0^1(\Omega) \times H_0^1(\Omega)$ to be such that

$$\begin{cases} (\nabla\tilde{y}, \nabla v)_{L^2(\Omega)} &= (\tilde{u} + f, v)_{L^2(\Omega)} \quad \forall v \in H_0^1(\Omega), \\ (\nabla w, \nabla\tilde{p})_{L^2(\Omega)} &= (\tilde{y} - y_\Omega, w)_{L^2(\Omega)} \quad \forall w \in H_0^1(\Omega). \end{cases} \tag{5.18}$$

Finally, we define the errors $e_y = \tilde{y} - \bar{y}_{\mathscr{T}}$, $e_p = \tilde{p} - \bar{p}_{\mathscr{T}}$, $e_\lambda = \tilde{\lambda} - \bar{\lambda}_{\mathscr{T}}$, and $e_u = \tilde{u} - \bar{u}_{\mathscr{T}}$, and, for $e := (e_y, e_p, e_u, e_\lambda)$, the norms

$$\|\!|e|\!\|_\Omega^2 := |e_y|_{H^1(\Omega)}^2 + |e_p|_{H^1(\Omega)}^2 + \|e_u\|_{L^2(\Omega)}^2 + \|e_\lambda\|_{L^2(\Omega)}^2, \tag{5.19}$$

and

$$\|e\|_\Omega^2 := \|e_y\|_{L^2(\Omega)}^2 + \|e_p\|_{L^2(\Omega)}^2 + \|e_u\|_{L^2(\Omega)}^2 + \|e_\lambda\|_{L^2(\Omega)}^2. \tag{5.20}$$

We thus have all the ingredients at hand to develop our a posteriori error analysis.

THEOREM 5.1 (global reliability of E) Let $(\bar{y},\bar{p},\bar{u}) \in H_0^1(\Omega) \times H_0^1(\Omega) \times \mathbb{U}_{ad}$ be the solution to the optimality system (3.8), and $(\bar{y}_{\mathscr{T}},\bar{p}_{\mathscr{T}},\bar{u}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \times \mathbb{V}(\mathscr{T}) \times \mathbb{U}_{ad,1}(\mathscr{T})$ its numerical approximation obtained as the solution to (4.7). If Ω is convex, then

$$\|e\|_{\Omega} \leqslant E, \tag{5.21}$$

where $\|e\|_{\Omega}$ is defined as in (5.20) and

$$E = \left(\sum_{K \in \mathscr{T}} E_K^2 \right)^{\frac{1}{2}}, \qquad E_K^2 = C_{st} E_{y,K}^2 + C_{ad} E_{p,K}^2 + C_{ct} E_{u,K}^2 + C_{sg} E_{\lambda,K}^2. \tag{5.22}$$

The constants C_{st}, C_{ad}, C_{ct} and C_{sg} are independent of the continuous and discrete optimal variables, the size of the elements of the mesh $\mathscr{T}$ and $\#\mathscr{T}$.

Proof. We proceed in five steps.

Step 1. The goal of this step is to control the error $\|\bar{u} - \bar{u}_{\mathscr{T}}\|_{L^2(\Omega)}$. We begin by invoking definitions (5.14) and (5.16) to immediately arrive at the estimate

$$\|\bar{u} - \bar{u}_{\mathscr{T}}\|_{L^2(\Omega)}^2 \leqslant 2\|\bar{u} - \tilde{u}\|_{L^2(\Omega)}^2 + 2E_u^2. \tag{5.23}$$

It thus suffices to control the term $\|\bar{u} - \tilde{u}\|_{L^2(\Omega)}$. To accomplish this task, we first notice that $\tilde{u}$, defined as in (5.14), can be equivalently characterized by

$$(\bar{p}_{\mathscr{T}} + \alpha\tilde{u} + \beta\tilde{\lambda}, u - \tilde{u})_{L^2(\Omega)} \geqslant 0 \quad \forall u \in \mathbb{U}_{ad}.$$

Consequently, by setting $u = \bar{u}$ in the previous variational inequality and $u = \tilde{u}$ in (3.4), we arrive at

$$(\bar{p}_{\mathscr{T}} + \alpha\tilde{u} + \beta\tilde{\lambda}, \bar{u} - \tilde{u})_{L^2(\Omega)} \geqslant 0, \qquad (\bar{p} + \alpha\bar{u} + \beta\bar{\lambda}, \tilde{u} - \bar{u})_{L^2(\Omega)} \geqslant 0.$$

Adding these variational inequalities we thus obtain the following basic estimate

$$\alpha\|\bar{u} - \tilde{u}\|_{L^2(\Omega)}^2 \leqslant (\bar{p} - \bar{p}_{\mathscr{T}}, \tilde{u} - \bar{u})_{L^2(\Omega)} + \beta(\bar{\lambda} - \tilde{\lambda}, \tilde{u} - \bar{u})_{L^2(\Omega)}.$$

Now, since $\bar{\lambda} \in \partial\psi(\bar{u})$ and $\tilde{\lambda} \in \partial\psi(\tilde{u})$, an application of (2.3) yields

$$\beta(\bar{\lambda} - \tilde{\lambda}, \tilde{u} - \bar{u})_{L^2(\Omega)} \leqslant 0.$$

Consequently,

$$\alpha\|\bar{u} - \tilde{u}\|_{L^2(\Omega)}^2 \leqslant (\bar{p} - \bar{p}_{\mathscr{T}}, \tilde{u} - \bar{u})_{L^2(\Omega)}.$$

We now invoke the auxiliary states $\hat{p}$ and $\tilde{p}$, defined as the solution to problems (5.7) and (5.18), respectively, to rewrite the previous expression as follows:

$$\alpha\|\tilde{u} - \bar{u}\|_{L^2(\Omega)}^2 \leqslant (\bar{p} - \tilde{p}, \tilde{u} - \bar{u})_{L^2(\Omega)} + (\tilde{p} - \hat{p}, \tilde{u} - \bar{u})_{L^2(\Omega)} + (\hat{p} - \bar{p}_{\mathscr{T}}, \tilde{u} - \bar{u})_{L^2(\Omega)}.$$

We proceed to bound $(\bar{p} - \tilde{p}, \tilde{u} - \bar{u})_{L^2(\Omega)}$. To accomplish this task, we notice that $\tilde{y} - \bar{y} \in H_0^1(\Omega)$ and $\bar{p} - \tilde{p} \in H_0^1(\Omega)$ solve, for all $v \in H_0^1(\Omega)$ and $w \in H_0^1(\Omega)$,

$$(\nabla(\tilde{y} - \bar{y}), \nabla v)_{L^2(\Omega)} = (\tilde{u} - \bar{u}, v)_{L^2(\Omega)}, \qquad (\nabla w, \nabla(\bar{p} - \tilde{p}))_{L^2(\Omega)} = (\bar{y} - \tilde{y}, w)_{L^2(\Omega)},$$

respectively. Set $v = \bar{p} - \tilde{p}$ and $w = \tilde{y} - \bar{y}$ and conclude that

$$(\bar{p} - \tilde{p}, \tilde{u} - \bar{u})_{L^2(\Omega)} = (\nabla(\tilde{y} - \bar{y}), \nabla(\bar{p} - \tilde{p}))_{L^2(\Omega)} = -\|\tilde{y} - \bar{y}\|_{L^2(\Omega)}^2 \leqslant 0.$$

This result allows us to derive that

$$\alpha \|\tilde{u} - \bar{u}\|_{L^2(\Omega)}^2 \leqslant (\tilde{p} - \hat{p}, \tilde{u} - \bar{u})_{L^2(\Omega)} + (\hat{p} - \bar{p}_{\mathscr{T}}, \tilde{u} - \bar{u})_{L^2(\Omega)},$$

which implies the bounds

$$\|\tilde{u} - \bar{u}\|_{L^2(\Omega)}^2 \leqslant \frac{2}{\alpha^2}\|\tilde{p} - \hat{p}\|_{L^2(\Omega)}^2 + \frac{2}{\alpha^2}\|\hat{p} - \bar{p}_{\mathscr{T}}\|_{L^2(\Omega)}^2$$
$$\leqslant \frac{2}{\alpha^2}\|\tilde{p} - \hat{p}\|_{L^2(\Omega)}^2 + \frac{2}{\alpha^2}C_p^2 E_p^2, \tag{5.24}$$

where, in the last inequality, we have used the a posteriori error estimate (5.13).

To control $\|\tilde{p} - \hat{p}\|_{L^2(\Omega)}$, we notice that $(\nabla w, \nabla(\tilde{p} - \hat{p}))_{L^2(\Omega)} = (\tilde{y} - \bar{y}_{\mathscr{T}}, w)_{L^2(\Omega)}$ for all $w \in H_0^1(\Omega)$. An application of the Poincaré inequality (2.1) thus reveals that

$$\mathfrak{C}^{-2}\|\tilde{p} - \hat{p}\|_{L^2(\Omega)}^2 \leqslant (\nabla(\tilde{p} - \hat{p}), \nabla(\tilde{p} - \hat{p}))_{L^2(\Omega)} \leqslant \|\tilde{p} - \hat{p}\|_{L^2(\Omega)}\|\tilde{y} - \bar{y}_{\mathscr{T}}\|_{L^2(\Omega)},$$

which, in view of the a posteriori estimate (5.13), implies the bound

$$\mathfrak{C}^{-4}\|\tilde{p} - \hat{p}\|_{L^2(\Omega)}^2 \leqslant 2\|\tilde{y} - \hat{y}\|_{L^2(\Omega)}^2 + 2C_y^2 E_y^2. \tag{5.25}$$

Similarly, since $\tilde{y} - \hat{y}$ solves $(\nabla(\tilde{y} - \hat{y}), v)_{L^2(\Omega)} = (\tilde{u} - \bar{u}_{\mathscr{T}}, v)_{L^2(\Omega)}$ for all $v \in H_0^1(\Omega)$, we can conclude that

$$\mathfrak{C}^{-2}\|\tilde{y} - \hat{y}\|_{L^2(\Omega)}^2 \leqslant (\nabla(\tilde{y} - \hat{y}), \nabla(\tilde{y} - \hat{y}))_{L^2(\Omega)} \leqslant \|\tilde{y} - \hat{y}\|_{L^2(\Omega)}\|\tilde{u} - \bar{u}_{\mathscr{T}}\|_{L^2(\Omega)},$$

and thus, invoking definition (5.16), that

$$\|\tilde{y} - \hat{y}\|_{L^2(\Omega)}^2 \leqslant \mathfrak{C}^4 E_u^2.$$

Replacing this estimate into (5.25) we arrive at

$$\|\tilde{p} - \hat{p}\|_{L^2(\Omega)}^2 \leqslant 2\mathfrak{C}^8 E_u^2 + 2\mathfrak{C}^4 C_y^2 E_y^2.$$

On the basis of (5.24), the collection of our previous findings yields the estimate

$$\|\tilde{u} - \bar{u}\|_{L^2(\Omega)}^2 \leqslant \frac{2}{\alpha^2}\left(2\mathfrak{C}^8 E_u^2 + 2\mathfrak{C}^4 C_y^2 E_y^2 + C_p^2 E_p^2\right),$$

which, in view of (5.23), allows us to conclude the a posteriori error estimate

$$\|\bar{u} - \bar{u}_{\mathscr{T}}\|_{L^2(\Omega)}^2 \leqslant \frac{4}{\alpha^2}\left[\left\{2\mathfrak{C}^8 + \frac{\alpha^2}{2}\right\}E_u^2 + 2\mathfrak{C}^4 C_y^2 E_y^2 + C_p^2 E_p^2\right]. \tag{5.26}$$

Step 2. The goal of this step is to bound the error $\|\bar{y} - \bar{y}_{\mathscr{T}}\|_{L^2(\Omega)}$. We begin with

$$\|\bar{y} - \bar{y}_{\mathscr{T}}\|_{L^2(\Omega)}^2 \leqslant 2\|\bar{y} - \hat{y}\|_{L^2(\Omega)}^2 + 2C_y^2 E_y^2, \tag{5.27}$$

which follows from (5.13). Since $\bar{y} - \hat{y}$ solves $(\nabla(\bar{y} - \hat{y}), \nabla v)_{L^2(\Omega)} = (\bar{u} - \bar{u}_{\mathscr{T}}, v)_{L^2(\Omega)}$ for all $v \in H_0^1(\Omega)$, by setting $v = \bar{y} - \hat{y}$ we can conclude that

$$\mathfrak{C}^{-2}\|\bar{y} - \hat{y}\|_{L^2(\Omega)}^2 \leqslant (\nabla(\bar{y} - \hat{y}), \nabla(\bar{y} - \hat{y}))_{L^2(\Omega)} \leqslant \|\bar{y} - \hat{y}\|_{L^2(\Omega)}\|\bar{u} - \bar{u}_{\mathscr{T}}\|_{L^2(\Omega)},$$

which yields the bound $\|\bar{y} - \hat{y}\|_{L^2(\Omega)} \leqslant \mathfrak{C}^2\|\bar{u} - \bar{u}_{\mathscr{T}}\|_{L^2(\Omega)}$. This estimate combined with (5.26) and (5.27) imply that

$$\|\bar{y} \quad - \quad \bar{y}_{\mathscr{T}}\|_{L^2(\Omega)}^2 \quad \leqslant \quad \frac{8}{\alpha^2}\mathfrak{C}^4\left[\left\{2\mathfrak{C}^8 + \frac{\alpha^2}{2}\right\}E_u^2 \quad + \quad C_y^2\left\{2\mathfrak{C}^4 + \frac{\alpha^2}{4\mathfrak{C}^4}\right\}E_y^2 \quad + \quad C_p^2 E_p^2\right]. \quad (5.28)$$

Step 3. We control the term $\|\bar{p} - \bar{p}_{\mathscr{T}}\|_{L^2(\Omega)}$. A simple application of the triangle inequality and the estimate (5.13) reveal that

$$\|\bar{p} - \bar{p}_{\mathscr{T}}\|_{L^2(\Omega)}^2 \leqslant 2\|\bar{p} - \hat{p}\|_{L^2(\Omega)}^2 + 2C_p^2 E_p^2. \quad (5.29)$$

To estimate the term $\|\bar{p} - \hat{p}\|_{L^2(\Omega)}$, we notice that $\bar{p} - \hat{p}$ solves $(\nabla w, \nabla(\bar{p} - \hat{p}))_{L^2(\Omega)} = (\bar{y} - \bar{y}_{\mathscr{T}}, w)_{L^2(\Omega)}$ for all $w \in H_0^1(\Omega)$. Set $w = \bar{p} - \hat{p}$ and conclude that

$$\mathfrak{C}^{-2}\|\bar{p} - \hat{p}\|_{L^2(\Omega)}^2 \leqslant (\nabla(\bar{p} - \hat{p}), \nabla(\bar{p} - \hat{p}))_{L^2(\Omega)} \leqslant \|\bar{p} - \hat{p}\|_{L^2(\Omega)}\|\bar{y} - \bar{y}_{\mathscr{T}}\|_{L^2(\Omega)}.$$

This estimate implies that $\|\bar{p} - \hat{p}\|_{L^2(\Omega)}^2 \leqslant \mathfrak{C}^4\|\bar{y} - \bar{y}_{\mathscr{T}}\|_{L^2(\Omega)}^2$. Therefore, (5.28), (5.29) and the previous estimate allow us to deduce the a posteriori error estimate

$$\|\bar{p} - \bar{p}_{\mathscr{T}}\|_{L^2(\Omega)}^2 \leqslant \frac{16}{\alpha^2}\mathfrak{C}^8\left[\left\{2\mathfrak{C}^8 + \frac{\alpha^2}{2}\right\}E_u^2 + C_y^2\left\{2\mathfrak{C}^4 + \frac{\alpha^2}{4\mathfrak{C}^4}\right\}E_y^2 + C_p^2\left\{1 + \frac{\alpha^2}{8\mathfrak{C}^8}\right\}E_p^2\right]. \quad (5.30)$$

Step 4. The objective of this step is to bound $\|\bar{\lambda} - \bar{\lambda}_{\mathscr{T}}\|_{L^2(\Omega)}$. To accomplish this task we utilize the auxiliary variable $\tilde{\lambda}$, defined as in (5.14), and proceed as follows:

$$\|\bar{\lambda} - \bar{\lambda}_{\mathscr{T}}\|_{L^2(\Omega)}^2 \leqslant 2\|\bar{\lambda} - \tilde{\lambda}\|_{L^2(\Omega)}^2 + 2\|\tilde{\lambda} - \bar{\lambda}_{\mathscr{T}}\|_{L^2(\Omega)}^2$$
$$\leqslant 2\beta^{-2}\|\bar{p} - \bar{p}_{\mathscr{T}}\|_{L^2(\Omega)}^2 + 2E_\lambda^2, \quad (5.31)$$

where, in the last inequality, we have used the projection formula (3.7), the Lipschitz continuity of the operator $\Pi_{[-1,1]}$ and the a posteriori error estimate (5.17). To conclude, we insert the estimate (5.30) into the previous inequality and obtain that

$$\|\bar{\lambda} - \bar{\lambda}_{\mathscr{T}}\|_{L^2(\Omega)}^2 \leqslant \frac{32}{(\alpha\beta)^2}\mathfrak{C}^8\left[\left\{2\mathfrak{C}^8 + \frac{\alpha^2}{2}\right\}E_u^2\right.$$
$$\left. + C_y^2\left\{2\mathfrak{C}^4 + \frac{\alpha^2}{4\mathfrak{C}^4}\right\}E_y^2 + C_p^2\left\{1 + \frac{\alpha^2}{8\mathfrak{C}^8}\right\}E_p^2 + \frac{(\alpha\beta)^2}{16\mathfrak{C}^8}E_\lambda^2\right]. \quad (5.32)$$

Step 5. The desired estimate follows from a collection of the estimates (5.26), (5.28), (5.30) and (5.32). This concludes the proof. □

The previous analysis allows us to derive the following result for a posteriori error estimation based on energy–type norms.

THEOREM 5.2 (global reliability of $\mathscr{E}$) Let $(\bar{y}, \bar{p}, \bar{u}) \in H_0^1(\Omega) \times H_0^1(\Omega) \times \mathbb{U}_{ad}$ be the solution to the optimality system (3.8), and $(\bar{y}_{\mathscr{T}}, \bar{p}_{\mathscr{T}}, \bar{u}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \times \mathbb{V}(\mathscr{T}) \times \mathbb{U}_{ad,0}(\mathscr{T})$ its numerical approximation obtained as the solution to (4.2). Then,

$$\||e\||_{\Omega} \leqslant \mathscr{E}, \tag{5.33}$$

where $\||e\||_{\Omega}$ is defined as in (5.19) and

$$\mathscr{E} = \left(\sum_{K \in \mathscr{T}} \mathscr{E}_K^2 \right)^{\frac{1}{2}}, \quad \mathscr{E}_K^2 := \mathscr{C}_{st}\mathscr{E}_{y,K}^2 + \mathscr{C}_{ad}\mathscr{E}_{p,K}^2 + \mathscr{C}_{ct}E_{u,K}^2 + \mathscr{C}_{sg}E_{\lambda,K}^2. \tag{5.34}$$

The constants $\mathscr{C}_{st}, \mathscr{C}_{ad}, \mathscr{C}_{ct}$ and $\mathscr{C}_{sg}$ are independent of the continuous and discrete optimal variables, the size of the elements of the mesh $\mathscr{T}$ and $\#\mathscr{T}$.

Proof. The arguments elaborated in the Step 1 of the proof of the previous theorem, combined with a series of applications of the Poincaré inequality (2.1), starting from (5.24), allow us to conclude that

$$\|\bar{u} - \bar{u}_{\mathscr{T}}\|_{L^2(\Omega)}^2 \leqslant \frac{4}{\alpha^2}\mathfrak{C}^2 \left[\left\{ 2\mathfrak{C}^6 + \frac{\alpha^2}{2\mathfrak{C}^2} \right\} E_u^2 + 2\mathscr{C}_y^2\mathfrak{C}^4\mathscr{E}_y^2 + \mathscr{C}_p^2\mathscr{E}_p^2 \right]. \tag{5.35}$$

With this bound at hand, we proceed to estimate the term $|\bar{y} - \bar{y}_{\mathscr{T}}|_{H^1(\Omega)}$. A simple application of the triangle inequality and the a posteriori estimate (5.10) yield

$$|\bar{y} - \bar{y}_{\mathscr{T}}|_{H^1(\Omega)}^2 \leqslant 2|\bar{y} - \hat{y}|_{H^1(\Omega)}^2 + 2\mathscr{C}_y^2\mathscr{E}_y^2. \tag{5.36}$$

We invoke the problem that $\bar{y} - \hat{y}$ solves to arrive at $|\bar{y} - \hat{y}|_{H^1(\Omega)}^2 \leqslant \mathfrak{C}^2\|\bar{u} - \bar{u}_{\mathscr{T}}\|_{L^2(\Omega)}^2$. This estimate, together with (5.35) and (5.36), allows us to conclude that

$$|\bar{y} - \bar{y}_{\mathscr{T}}|_{H^1(\Omega)}^2 \leqslant \frac{8}{\alpha^2}\mathfrak{C}^4 \left[\left\{ 2\mathfrak{C}^6 + \frac{\alpha^2}{2\mathfrak{C}^2} \right\} E_u^2 + \mathscr{C}_y^2 \left\{ 2\mathfrak{C}^4 + \frac{\alpha^2}{4\mathfrak{C}^4} \right\} \mathscr{E}_y^2 + \mathscr{C}_p^2\mathscr{E}_p^2 \right]. \tag{5.37}$$

We now control the term $|\bar{p} - \bar{p}_{\mathscr{T}}|_{H^1(\Omega)}$. We begin with the estimate

$$|\bar{p} - \bar{p}_{\mathscr{T}}|_{H^1(\Omega)}^2 \leqslant 2|\bar{p} - \hat{p}|_{H^1(\Omega)}^2 + 2\mathscr{C}_p^2\mathscr{E}_p^2. \tag{5.38}$$

To control $|\bar{p} - \hat{p}|_{H^1(\Omega)}$ we invoke the problem that $\bar{p} - \hat{p}$ solves and obtain that

$$|\bar{p} - \hat{p}|_{H^1(\Omega)}^2 \leqslant \mathfrak{C}^4|\bar{y} - \bar{y}_{\mathscr{T}}|_{H^1(\Omega)}^2.$$

Replacing (5.37) into the previous inequality and invoking (5.38) we derive that

$$|\bar{p} - \bar{p}_{\mathscr{T}}|_{H^1(\Omega)}^2 \leqslant \frac{16}{\alpha^2}\mathfrak{C}^8 \left[\left\{ 2\mathfrak{C}^6 + \frac{\alpha^2}{2\mathfrak{C}^2} \right\} E_u^2 + \mathscr{C}_y^2 \left\{ 2\mathfrak{C}^4 + \frac{\alpha^2}{4\mathfrak{C}^4} \right\} \mathscr{E}_y^2 + \mathscr{C}_p^2 \left\{ 1 + \frac{\alpha^2}{8\mathfrak{C}^8} \right\} \mathscr{E}_p^2 \right]. \tag{5.39}$$

We finally bound $\bar{\lambda} - \bar{\lambda}_{\mathscr{T}}$ in the $L^2(\Omega)$–norm. In view of (5.31) we immediately arrive at

$$\|\bar{\lambda} - \bar{\lambda}_{\mathscr{T}}\|_{L^2(\Omega)}^2 \leqslant 2\mathfrak{C}^2\beta^{-2}|\bar{p} - \bar{p}_{\mathscr{T}}|_{H^1(\Omega)}^2 + 2E_{\lambda}^2.$$

On the basis of (5.39) we can thus conclude that

$$\|\bar{\lambda} - \bar{\lambda}_{\mathscr{T}}\|_{L^2(\Omega)}^2 \leqslant \frac{32}{(\alpha\beta)^2}\mathfrak{C}^{10}\left[\left\{2\mathfrak{C}^6 + \frac{\alpha^2}{2\mathfrak{C}^2}\right\}E_u^2 \right.$$
$$\left. + \mathscr{C}_y^2\left\{2\mathfrak{C}^4 + \frac{\alpha^2}{4\mathfrak{C}^4}\right\}\mathscr{E}_y^2 + \mathscr{C}_p^2\left\{1 + \frac{\alpha^2}{8\mathfrak{C}^8}\right\}\mathscr{E}_p^2 + \frac{(\alpha\beta)^2}{16\mathfrak{C}^{10}}E_\lambda^2\right]. \quad (5.40)$$

The desired results follows upon gathering (5.35), (5.37), (5.39) and (5.40). □

We now provide an a posteriori error estimation result when the variational discretization approach is used to approximate the optimal control problem (3.1)–(3.2).

THEOREM 5.3 (global reliability of $\mathfrak{E}$) Let $(\bar{y}, \bar{p}, \bar{u}) \in H_0^1(\Omega) \times H_0^1(\Omega) \times \mathbb{U}_{ad}$ be the solution to the optimality system (3.8) and $(\bar{y}_{\mathscr{T}}, \bar{p}_{\mathscr{T}}, \bar{u}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \times \mathbb{V}(\mathscr{T}) \times \mathbb{U}_{ad}$ its numerical approximation obtained as the solution to (4.12). If Ω is convex, then

$$\|e\|_\Omega \leqslant \mathfrak{E}, \quad (5.41)$$

where $\|e\|_\Omega$ is defined as in (5.20) and

$$\mathfrak{E} = \left(\sum_{K \in \mathscr{T}} \mathfrak{E}_K^2\right)^{\frac{1}{2}}, \qquad \mathfrak{E}_K^2 = \mathsf{C}_{st}E_{y,K}^2 + \mathsf{C}_{ad}E_{p,K}^2. \quad (5.42)$$

The constants C_{st} and C_{ad} are independent of the continuous and discrete optimal variables, the size of the elements of the mesh $\mathscr{T}$ and $\#\mathscr{T}$.

Proof. The arguments elaborated in the Step 1 of the proof of Theorem 5.1, combined with (5.15), that reads $\tilde{u} = \bar{u}_{\mathscr{T}}$, allow us to deduce the estimate

$$\|\bar{u} - \bar{u}_{\mathscr{T}}\|_{L^2(\Omega)}^2 \leqslant \frac{2}{\alpha^2}\left[\mathfrak{C}^4 C_y^2 E_y^2 + C_p^2 E_p^2\right]. \quad (5.43)$$

To obtain this estimate we have also used that $\hat{y} = \tilde{y}$; compare (5.7) and (5.18).

Now, arguing as in Steps 2 and 3 of the proof of Theorem 5.1, we obtain

$$\|\bar{y} - \bar{y}_{\mathscr{T}}\|_{L^2(\Omega)}^2 \leqslant \frac{4\mathfrak{C}^4}{\alpha^2}\left[C_y^2\left\{\mathfrak{C}^4 + \frac{\alpha^2}{2\mathfrak{C}^4}\right\}E_y^2 + C_p^2 E_p^2\right], \quad (5.44)$$

$$\|\bar{p} - \bar{p}_{\mathscr{T}}\|_{L^2(\Omega)}^2 \leqslant \frac{8\mathfrak{C}^8}{\alpha^2}\left[C_y^2\left\{\mathfrak{C}^4 + \frac{\alpha^2}{2\mathfrak{C}^4}\right\}E_y^2 + C_p^2\left\{1 + \frac{\alpha^2}{8\mathfrak{C}^8}\right\}E_p^2\right]. \quad (5.45)$$

Finally, since, in view of (5.15), we have that $\tilde{\lambda} = \bar{\lambda}_{\mathscr{T}}$, the arguments elaborated in the Step 4 of the proof of Theorem 5.1, allow us to conclude the following a posteriori error estimate:

$$\|\bar{\lambda} - \bar{\lambda}_{\mathscr{T}}\|_{L^2(\Omega)}^2 \leqslant \frac{8\mathfrak{C}^8}{(\alpha\beta)^2}\left[C_y^2\left\{\mathfrak{C}^4 + \frac{\alpha^2}{2\mathfrak{C}^4}\right\}E_y^2 + C_p^2\left\{1 + \frac{\alpha^2}{8\mathfrak{C}^8}\right\}E_p^2\right]. \quad (5.46)$$

The error estimate (5.41) thus follows by collecting the estimates (5.43), (5.44), (5.45), and (5.46). This concludes the proof. □

REMARK 5.3 (Fully computable) Since an upper bound for $\mathfrak{C}$, that appears in (2.1), can be found, for instance, in (Mikhlin, 1986), the a posteriori error estimates (5.21), (5.33) and (5.41) are fully computable under the assumption that the constants C_y, C_p, $\mathscr{C}_y$, and $\mathscr{C}_p$ are known. However, it is well–known that, for the residual–type estimators that we consider in our work, such an assumption does not hold. In the literature, several fully computable a posteriori error estimators, which provide genuine upper bounds for the error, are available. We refer the reader to Ainsworth (2010); Repin & Sauter (2006); Vohralí k (2008) and references therein. If estimators from these references are utilized for the state and adjoint equations, then Theorem 5.2 would provide a fully computable a posteriori error estimate for our optimal control problem.

5.3 *Error estimators for sparse PDE–constrained optimization: efficiency*

In what follows we examine the efficiency properties of the a posteriori error estimators E, $\mathscr{E}$ and $\mathfrak{E}$ which are defined as in (5.22), (5.34), and (5.42), respectively. To accomplish this task, we analyze each of their contributions separately. Before proceeding with such analyses we introduce the following notation: for an edge, triangle or tetrahedron G, let $\mathscr{V}(G)$ be the set of vertices of G. We define, for each element $K \in \mathscr{T}$ and side $\gamma \in \mathscr{S}$, the standard element and edge bubble functions Verfürth (2013, 1989)

$$\beta_K|_K = (d+1)^{(d+1)} \prod_{\mathsf{v} \in \mathscr{V}(K)} \phi_\mathsf{v}, \quad \beta_\gamma|_K = d^d \prod_{\mathsf{v} \in \mathscr{V}(\gamma)} \phi_\mathsf{v}|_K \text{ with } K \subset \Omega_\gamma, \tag{5.47}$$

respectively, where ϕ_v are the barycentric coordinates of K. We recall that Ω_γ corresponds to the patch composed of the two elements of $\mathscr{T}$ sharing γ.

We now present the following error equation associated to the state equation; it follows from the continuous state equation in (3.8) and an application of integration by parts:

$$\sum_{K \in \mathscr{T}} \left(\bar{u}_\mathscr{T} + \Pi_K^\ell(f), v \right)_{L^2(K)} - \sum_{\gamma \in \mathscr{S}} \left([\![\nabla \bar{y}_\mathscr{T} \cdot \mathsf{v}]\!]_\gamma, v \right)_{L^2(\gamma)}$$
$$= (\nabla(\bar{y} - \bar{y}_\mathscr{T}), \nabla v)_{L^2(\Omega)} - (\bar{u} - \bar{u}_\mathscr{T}, v)_{L^2(\Omega)} - \sum_{K \in \mathscr{T}} (f - \Pi_K^\ell(f), v)_{L^2(K)}, \tag{5.48}$$

for all $v \in H_0^1(\Omega)$, where, for $K \in \mathscr{T}$ and $\ell \in \{0, 1\}$, $\Pi_K^\ell(f)$ denotes the $L^2(K)$–orthogonal projection operator onto $\mathbb{P}_\ell(K)$.

On the other hand, similar arguments to the ones that led to (5.48) allow us to conclude the following error equation associated to the adjoint state equation:

$$\sum_{K \in \mathscr{T}} \left(\bar{y}_\mathscr{T} - \Pi_K^\ell(y_\Omega), w \right)_{L^2(K)} - \sum_{\gamma \in \mathscr{S}} \left([\![\nabla \bar{p}_\mathscr{T} \cdot \mathsf{v}]\!]_\gamma, w \right)_{L^2(\gamma)}$$
$$= (\nabla(\bar{p} - \bar{p}_\mathscr{T}), \nabla w)_{L^2(\Omega)} - (\bar{y} - \bar{y}_\mathscr{T}, w)_{L^2(\Omega)} + \sum_{K \in \mathscr{T}} (y_\Omega - \Pi_K^\ell(y_\Omega), w)_{L^2(K)}, \tag{5.49}$$

for all $w \in H_0^1(\Omega)$.

5.3.1 *Efficiency properties of $\mathscr{E}$.* Set $\ell = 0$ in (5.48). We proceed on the basis of standard arguments, as the ones developed in (Ainsworth & Oden, 2000, Section 2.3) and (Verfürth, 2013, Section 1.4), to conclude the following estimates. First, for $K \in \mathscr{T}$, consider $v = (\bar{u}_\mathscr{T} + \Pi_K^0(f))\beta_K$ in (5.48). This yields the estimate

$$h_K^2 \|\bar{u}_\mathscr{T} + \Pi_K^0(f)\|_{L^2(K)}^2 \lesssim |\bar{y} - \bar{y}_\mathscr{T}|_{H^1(K)}^2 + h_K^2 \left(\|\bar{u} - \bar{u}_\mathscr{T}\|_{L^2(K)}^2 + \|f - \Pi_K^0(f)\|_{L^2(K)}^2 \right). \tag{5.50}$$

Second, for $K \in \mathcal{T}$ and $\gamma \in \mathscr{S}_K$, consider $v = [\![\nabla \bar{y}_{\mathcal{T}} \cdot v]\!]_\gamma \beta_\gamma$ in (5.48) and conclude the estimate

$$h_K \|[\![\nabla \bar{y}_{\mathcal{T}} \cdot v]\!]_\gamma\|^2_{L^2(\gamma)} \lesssim \sum_{K' \in \Omega_\gamma} \left(|\bar{y} - \bar{y}_{\mathcal{T}}|^2_{H^1(K')} + h_K^2 \left(\|\bar{u} - \bar{u}_{\mathcal{T}}\|^2_{L^2(K')} + \|f - \Pi^0_{K'}(f)\|^2_{L^2(K')} \right) \right). \quad (5.51)$$

We are now in position to derive the following local efficiency result. To accomplish this task, for $\ell \in \{0,1\}$, $g \in L^2(\Omega)$ and $\mathcal{M} \subset \mathcal{T}$, we define

$$\operatorname{osc}_{\mathcal{T},\ell}(g; \mathcal{M}) := \left(\sum_{K \in \mathcal{M}} h_K^{2(\ell+1)} \|g - \Pi^\ell_K(g)\|^2_{L^2(K)} \right)^{\frac{1}{2}}. \quad (5.52)$$

LEMMA 5.1 (local efficiency of $\mathscr{E}_{y,K}$ and $\mathscr{E}_{p,K}$) Let $(\bar{y}, \bar{p}, \bar{u}) \in H^1_0(\Omega) \times H^1_0(\Omega) \times \mathbb{U}_{ad}$ be the solution to the optimality system (3.8) and $(\bar{y}_{\mathcal{T}}, \bar{p}_{\mathcal{T}}, \bar{u}_{\mathcal{T}}) \in \mathbb{V}(\mathcal{T}) \times \mathbb{V}(\mathcal{T}) \times \mathbb{U}_{ad,0}(\mathcal{T})$ its numerical approximation obtained as the solution to (4.2). Then, for $K \in \mathcal{T}$, the local error indicators $\mathscr{E}_{y,K}$ and $\mathscr{E}_{p,K}$, defined as in (5.8) and (5.9), respectively, satisfy that

$$\mathscr{E}_{y,K} \lesssim |\bar{y} - \bar{y}_{\mathcal{T}}|_{H^1(\Omega_K)} + h_K \|\bar{u} - \bar{u}_{\mathcal{T}}\|_{L^2(\Omega_K)} + \operatorname{osc}_{\mathcal{T},0}(f; \Omega_K), \quad (5.53)$$

and

$$\mathscr{E}_{p,K} \lesssim |\bar{p} - \bar{p}_{\mathcal{T}}|_{H^1(\Omega_K)} + h_K \|\bar{y} - \bar{y}_{\mathcal{T}}\|_{L^2(\Omega_K)} + \operatorname{osc}_{\mathcal{T},0}(y_\Omega; \Omega_K), \quad (5.54)$$

where Ω_K is defined as in (5.2) and the hidden constants are independent of the optimal variables, their approximations, the size of the elements in the mesh $\mathcal{T}$ and $\#\mathcal{T}$.

Proof. Let $K \in \mathcal{T}$. We first control the term $h_K^2 \|\bar{u}_{\mathcal{T}} + f\|^2_{L^2(K)}$ in (5.8). A simple application of the triangle inequality yields

$$\|\bar{u}_{\mathcal{T}} + f\|_{L^2(K)} \leq \|\bar{u}_{\mathcal{T}} + \Pi^0_K(f)\|_{L^2(K)} + \|f - \Pi^0_K(f)\|_{L^2(K)}.$$

We thus apply the estimate (5.50) and conclude that

$$h_K^2 \|\bar{u}_{\mathcal{T}} + f\|^2_{L^2(K)} \lesssim |\bar{y} - \bar{y}_{\mathcal{T}}|^2_{H^1(K)} + h_K^2 \|\bar{u} - \bar{u}_{\mathcal{T}}\|^2_{L^2(K)} + \operatorname{osc}^2_{\mathcal{T},0}(f; K). \quad (5.55)$$

Let $K \in \mathcal{T}$ and $\gamma \in \mathscr{S}$. The control of the term $h_K \|[\![\nabla \bar{y}_{\mathcal{T}} \cdot v]\!]_\gamma\|^2_{L^2(\gamma)}$, in (5.8), follows from (5.51) in combination with (5.55). This bound and (5.51) yield (5.53).

The proof of (5.54) follows similar arguments but based on the error equation (5.49). This concludes the proof. $\qquad \square$

The next result gives the global efficiency property of the estimator $\mathscr{E}$.

THEOREM 5.4 (global efficiency of $\mathscr{E}$) Let $(\bar{y}, \bar{p}, \bar{u}) \in H^1_0(\Omega) \times H^1_0(\Omega) \times \mathbb{U}_{ad}$ be the solution to the optimality system (3.8) and $(\bar{y}_{\mathcal{T}}, \bar{p}_{\mathcal{T}}, \bar{u}_{\mathcal{T}}) \in \mathbb{V}(\mathcal{T}) \times \mathbb{V}(\mathcal{T}) \times \mathbb{U}_{ad,0}(\mathcal{T})$ its numerical approximation obtained as the solution to (4.2). Then, the error estimator $\mathscr{E}$, defined as in (5.34), satisfies that

$$\mathscr{E} \lesssim \|e\|_\Omega + \operatorname{osc}_{\mathcal{T},0}(f; \mathcal{T}) + \operatorname{osc}_{\mathcal{T},0}(y_\Omega; \mathcal{T}), \quad (5.56)$$

where $e = (e_y, e_p, e_u, e_\lambda)$, $\|\cdot\|_\Omega$ is defined as in (5.19), and the hidden constant is independent of the optimal variables, their approximations, the size of the elements in the mesh $\mathcal{T}$ and $\#\mathcal{T}$.

Proof. In view of the definition of $\mathscr{E}_y$, given by (5.8), the estimate (5.53) immediately yields

$$\mathscr{E}_y \lesssim |\bar{y} - \bar{y}_{\mathscr{T}}|_{H^1(\Omega)} + \|\bar{u} - \bar{u}_{\mathscr{T}}\|_{L^2(\Omega)} + \mathrm{osc}_{\mathscr{T},0}(f;\mathscr{T}), \tag{5.57}$$

where we have used that Ω is bounded and the finite overlapping property of stars: each element K is contained in at most $d+2$ patches $\Omega_{K'}$.

On the basis of (5.9), similar arguments reveal that

$$\mathscr{E}_p \lesssim |\bar{p} - \bar{p}_{\mathscr{T}}|_{H^1(\Omega)} + \|\bar{y} - \bar{y}_{\mathscr{T}}\|_{L^2(\Omega)} + \mathrm{osc}_{\mathscr{T},0}(y_\Omega;\mathscr{T}). \tag{5.58}$$

We now study the efficiency of the estimator E_λ, which is defined as in (5.17). A trivial application of a triangle inequality yields

$$E_\lambda = \|\tilde{\lambda} - \bar{\lambda}_{\mathscr{T}}\|_{L^2(\Omega)} \leqslant \|\tilde{\lambda} - \bar{\lambda}\|_{L^2(\Omega)} + \|\bar{\lambda} - \bar{\lambda}_{\mathscr{T}}\|_{L^2(\Omega)}. \tag{5.59}$$

It thus suffices to bound $\|\tilde{\lambda} - \bar{\lambda}\|_{L^2(\Omega)}$. To accomplish this task, we invoke the projection formula (3.7), definition (5.14), and the Lipschitz continuity of $\Pi_{[-1,1]}$ to conclude that

$$\|\tilde{\lambda} - \bar{\lambda}\|_{L^2(\Omega)} \leqslant \beta^{-1}\|\bar{p} - \bar{p}_{\mathscr{T}}\|_{L^2(\Omega)}. \tag{5.60}$$

The control of the estimator E_u, which is defined as in (5.16), follows similar arguments. In fact, an application of a triangle inequality yields

$$E_u = \|\tilde{u} - \bar{u}_{\mathscr{T}}\|_{L^2(\Omega)} \leqslant \|\tilde{u} - \bar{u}\|_{L^2(\Omega)} + \|\bar{u} - \bar{u}_{\mathscr{T}}\|_{L^2(\Omega)}. \tag{5.61}$$

Since $\tilde{u}$ is defined as in (5.14) and $\Pi_{[a,b]}$ is Lipschitz continuous, the projection formula (3.6) and the estimate (5.60) imply that

$$\begin{aligned}
\|\tilde{u} - \bar{u}\|_{L^2(\Omega)} &\leqslant \|\alpha^{-1}(\bar{p} + \beta\bar{\lambda}) - \alpha^{-1}(\bar{p}_{\mathscr{T}} + \beta\tilde{\lambda})\|_{L^2(\Omega)} \\
&\leqslant \alpha^{-1}\|\bar{p} - \bar{p}_{\mathscr{T}}\|_{L^2(\Omega)} + \alpha^{-1}\beta\|\tilde{\lambda} - \bar{\lambda}\|_{L^2(\Omega)} \leqslant 2\alpha^{-1}\|\bar{p} - \bar{p}_{\mathscr{T}}\|_{L^2(\Omega)}.
\end{aligned} \tag{5.62}$$

The desired estimate (5.56) is thus a consequence of the estimates (5.57)–(5.62) combined with the Poincaré inequality (2.1). This concludes the proof. □

5.3.2 *Efficiency properties of E.* We begin by invoking the error equation (5.48), associated to the state equation, with $\ell = 1$ and $v \in H_0^1(\Omega)$ such that $v|_K \in C^2(K)$ for all $K \in \mathscr{T}$. Integration by parts allows us to conclude that

$$\begin{aligned}
\sum_{K \in \mathscr{T}} \left(\bar{u}_{\mathscr{T}} + \Pi_K^1(f), v\right)_{L^2(K)} &- \sum_{\gamma \in \mathscr{S}} \left[(\llbracket \nabla \bar{y}_{\mathscr{T}} \cdot v \rrbracket_\gamma, v)_{L^2(\gamma)} + (\bar{y} - \bar{y}_{\mathscr{T}}, \llbracket \nabla v \cdot v \rrbracket_\gamma)_{L^2(\gamma)} \right] \\
&= -\sum_{K \in \mathscr{T}} \left((\bar{y} - \bar{y}_{\mathscr{T}}, \Delta v)_{L^2(K)} + (\bar{u} - \bar{u}_{\mathscr{T}}, v)_{L^2(K)} + (f - \Pi_K^1(f), v)_{L^2(K)} \right).
\end{aligned} \tag{5.63}$$

Let $K \in \mathscr{T}$. Consider $v = \delta_K := \left(\bar{u}_{\mathscr{T}} + \Pi_K^1(f)\right)\beta_K^2$ in (5.63) and use that $\delta_K|_\gamma \equiv 0$ and $\nabla\delta_K|_\gamma \equiv 0$ for all $\gamma \in \mathscr{S}$ to conclude that

$$\|(\bar{u}_{\mathscr{T}} + \Pi_K^1(f))\beta_K\|_{L^2(K)}^2 = -(\bar{y} - \bar{y}_{\mathscr{T}}, \Delta\delta_K)_{L^2(K)} - (\bar{u} - \bar{u}_{\mathscr{T}}, \delta_K)_{L^2(K)} - (f - \Pi_K^1(f), \delta_K)_{L^2(K)}. \tag{5.64}$$

On the basis of the fact that $\Delta(\bar{u}_{\mathscr{T}} + \Pi_K^1(f)) = 0$, basic computations reveal that

$$\Delta\delta_K = 4\nabla(\bar{u}_{\mathscr{T}} + \Pi_K^1(f)) \cdot \nabla\beta_K\beta_K + 2(\bar{u}_{\mathscr{T}} + \Pi_K^1(f))(\beta_K\Delta\beta_K + \nabla\beta_K \cdot \nabla\beta_K).$$

Consequently, using properties of the standard element bubble function β_K and an inverse inequality we arrive at

$$\|\Delta\delta_K\|_{L^2(K)} \lesssim h_K^{-1}\|\nabla(\bar{u}_{\mathscr{T}} + \Pi_K^1(f))\|_{L^2(K)} + h_K^{-2}\|\bar{u}_{\mathscr{T}} + \Pi_K^1(f)\|_{L^2(K)}$$
$$\lesssim h_K^{-2}\|\bar{u}_{\mathscr{T}} + \Pi_K^1(f)\|_{L^2(K)}.$$

With the previous estimate at hand and using, again, properties of the element bubble function β_K we can conclude that

$$h_K^4\|\bar{u}_{\mathscr{T}} + \Pi_K^1(f)\|_{L^2(K)}^2 \lesssim h_K^4\|(\bar{u}_{\mathscr{T}} + \Pi_K^1(f))\beta_K\|_{L^2(K)}^2$$
$$\lesssim \|\bar{y} - \bar{y}_{\mathscr{T}}\|_{L^2(K)}^2 + h_K^4\left(\|\bar{u} - \bar{u}_{\mathscr{T}}\|_{L^2(K)}^2 + \|f - \Pi_K^1(f)\|_{L^2(K)}^2\right). \quad (5.65)$$

Let $K \in \mathscr{T}$ and $\gamma \in \mathscr{S}_K$. We recall that the patch composed of the two elements of $\mathscr{T}$ sharing γ is denoted by $\Omega_\gamma = K \cup K'$ with $K' \in \mathscr{T}$ and introduce the following edge bubble function

$$\zeta_\gamma|_{\Omega_\gamma} = d^{4d}\left(\prod_{\mathsf{v}\in\mathscr{V}(\gamma)} \phi_{\mathsf{v}}|_K\phi_{\mathsf{v}}|_{K'}\right)^2, \quad (5.66)$$

where, for any $\mathsf{v} \in \mathscr{V}(\gamma)$, $\phi_{\mathsf{v}}|_K$ and $\phi_{\mathsf{v}}|_{K'}$ are understood now as functions over Ω_γ. We notice that $\zeta_\gamma \in \mathbb{P}_{4d}(\Omega_\gamma)$, $\zeta_\gamma \in C^2(\Omega_\gamma)$, and $\zeta_\gamma = 0$ on $\partial\Omega_\gamma$. In addition, we have that

$$\nabla\zeta_\gamma = 0 \text{ on } \partial\Omega_\gamma, \quad [\![\nabla\zeta_\gamma \cdot \mathsf{v}]\!]_\gamma = 0 \text{ on } \gamma. \quad (5.67)$$

We thus consider $v = \delta_\gamma := [\![\nabla\bar{y}_{\mathscr{T}} \cdot \mathsf{v}]\!]_\gamma\zeta_\gamma$ in (5.63) and invoke (5.67) to obtain that

$$\sum_{K'\in\Omega_\gamma}\left(\bar{u}_{\mathscr{T}} + \Pi_{K'}^1(f), \delta_\gamma\right)_{L^2(K')} - \left([\![\nabla\bar{y}_{\mathscr{T}} \cdot \mathsf{v}]\!]_\gamma, \delta_\gamma\right)_{L^2(\gamma)}$$
$$= -\sum_{K'\in\Omega_\gamma}\left((\bar{y} - \bar{y}_{\mathscr{T}}, \Delta\delta_\gamma)_{L^2(K')} + (\bar{u} - \bar{u}_{\mathscr{T}}, \delta_\gamma)_{L^2(K')} + (f - \Pi_{K'}^1(f), \delta_\gamma)_{L^2(K')}\right).$$

Now, notice that, since $[\![\nabla\bar{y}_{\mathscr{T}} \cdot \mathsf{v}]\!]_\gamma \in \mathbb{R}$, we obtain, for $K' \in \Omega_\gamma$, that

$$\|\delta_\gamma\|_{L^2(K')}^2 \lesssim |K'|[\![\nabla\bar{y}_{\mathscr{T}} \cdot \mathsf{v}]\!]_\gamma^2 \lesssim \frac{|K'|}{|\gamma|}\|[\![\nabla\bar{y}_{\mathscr{T}} \cdot \mathsf{v}]\!]_\gamma\|_{L^2(\gamma)}^2 \lesssim h_K\|[\![\nabla\bar{y}_{\mathscr{T}} \cdot \mathsf{v}]\!]_\gamma\|_{L^2(\gamma)}^2,$$

where, to derive the last inequality we have used the shape regularity property of the family $\{\mathscr{T}\}$. Similarly, we have that

$$\|\Delta\delta_\gamma\|_{L^2(K)}^2 \lesssim h_K^{-3}\|[\![\nabla\bar{y}_{\mathscr{T}} \cdot \mathsf{v}]\!]_\gamma\|_{L^2(\gamma)}^2.$$

With these estimates at hand, we thus use standard arguments, the shape regularity property of the family $\{\mathscr{T}\}$, and the estimate (5.65) to arrive at

$$h_K^3\|[\![\nabla\bar{y}_{\mathscr{T}} \cdot \mathsf{v}]\!]_\gamma\|_{L^2(\gamma)}^2 \lesssim \sum_{K'\in\Omega_\gamma}\left(\|\bar{y} - \bar{y}_{\mathscr{T}}\|_{L^2(K')}^2 + h_K^4\|\bar{u} - \bar{u}_{\mathscr{T}}\|_{L^2(K')}^2\right) + \text{osc}_{\mathscr{T},1}^2(f;\Omega_\gamma). \quad (5.68)$$

On other hand, similar arguments to the ones that led to (5.63) allow us to arrive at the following error equation associated to the adjoint state equation:

$$\sum_{K \in \mathscr{T}} \left(\bar{y}_{\mathscr{T}} - \Pi_K^1(y_\Omega), w \right)_{L^2(K)} - \sum_{\gamma \in \mathscr{S}} \left[\left([\![\nabla \bar{p}_{\mathscr{T}} \cdot \nu]\!]_\gamma, w \right)_{L^2(\gamma)} + \left(\bar{p} - \bar{p}_{\mathscr{T}}, [\![\nabla w \cdot \nu]\!]_\gamma \right)_{L^2(\gamma)} \right]$$

$$= - \sum_{K \in \mathscr{T}} \left(\left(\bar{p} - \bar{p}_{\mathscr{T}}, \Delta w \right)_{L^2(K)} + \left(\bar{y} - \bar{y}_{\mathscr{T}}, w \right)_{L^2(K)} - \left(y_\Omega - \Pi_K^1(y_\Omega), w \right)_{L^2(K)} \right),$$

for all $w \in H_0^1(\Omega)$ such that $\nu|_K \in C^2(K)$. Similar estimates to (5.65)–(5.68) can be thus obtained.

We are now in position to derive the following local efficiency result.

THEOREM 5.5 (local efficiency of E) Let $(\bar{y}, \bar{p}, \bar{u}) \in H_0^1(\Omega) \times H_0^1(\Omega) \times \mathbb{U}_{ad}$ be the solution to the optimality system (3.8) and $(\bar{y}_{\mathscr{T}}, \bar{p}_{\mathscr{T}}, \bar{u}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \times \mathbb{V}(\mathscr{T}) \times \mathbb{U}_{ad,1}(\mathscr{T})$ its numerical approximation obtained as the solution to (4.7). Then, for $K \in \mathscr{T}$, the local error indicators $E_{y,K}$, $E_{p,K}$, $E_{u,K}$, and $E_{\lambda,K}$, defined as in (5.11), (5.12), (5.16), and (5.17), respectively, satisfy that

$$E_{y,K} \lesssim \|\bar{y} - \bar{y}_{\mathscr{T}}\|_{L^2(\Omega_K)} + h_K^2 \|\bar{u} - \bar{u}_{\mathscr{T}}\|_{L^2(\Omega_K)} + \mathrm{osc}_{\mathscr{T},1}(f; \Omega_K), \tag{5.69}$$

$$E_{p,K} \lesssim \|\bar{p} - \bar{p}_{\mathscr{T}}\|_{L^2(\Omega_K)} + h_K^2 \|\bar{y} - \bar{y}_{\mathscr{T}}\|_{L^2(\Omega_K)} + \mathrm{osc}_{\mathscr{T},1}(y_\Omega; \Omega_K), \tag{5.70}$$

$$E_{u,K} \leqslant \|\bar{u} - \bar{u}_{\mathscr{T}}\|_{L^2(K)} + 2\alpha^{-1} \|\bar{p} - \bar{p}_{\mathscr{T}}\|_{L^2(K)}, \tag{5.71}$$

and

$$E_{\lambda,K} \leqslant \|\bar{\lambda} - \bar{\lambda}_{\mathscr{T}}\|_{L^2(K)} + \beta^{-1} \|\bar{p} - \bar{p}_{\mathscr{T}}\|_{L^2(K)}, \tag{5.72}$$

where Ω_K is defined as in (5.2) and the hidden constants are independent of the optimal variables, their approximations, the size of the elements in the mesh $\mathscr{T}$ and $\#\mathscr{T}$.

Proof. The control of the interior residual in (5.11) is a consequence of the triangle inequality and the estimate (5.65). The jump or interelement residual in (5.11) is bounded in (5.68). The collection of these estimates yield (5.69). Similar arguments yield (5.70). The local efficiency estimates (5.71) and (5.72) correspond to local versions of the estimates (5.59)–(5.62). □

5.3.3 *Efficiency properties of* $\mathfrak{E}$. The results of Section 5.3.2 immediately imply the following local efficiency result for the error indicator $\mathfrak{E}_K$, which is defined as in (5.42).

THEOREM 5.6 (local efficiency of $\mathfrak{E}$) Let $(\bar{y}, \bar{p}, \bar{u}) \in H_0^1(\Omega) \times H_0^1(\Omega) \times \mathbb{U}_{ad}$ be the solution to the optimality system (3.8) and $(\bar{y}_{\mathscr{T}}, \bar{p}_{\mathscr{T}}, \bar{u}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \times \mathbb{V}(\mathscr{T}) \times \mathbb{U}_{ad}$ its numerical approximation obtained as the solution to (4.12). Then, for $K \in \mathscr{T}$, the local error indicators $E_{y,K}$ and $E_{p,K}$, defined as in (5.11) and (5.12), satisfy that

$$E_{y,K} \lesssim \|\bar{y} - \bar{y}_{\mathscr{T}}\|_{L^2(\Omega_K)} + h_K^2 \|\bar{u} - \bar{u}_{\mathscr{T}}\|_{L^2(\Omega_K)} + \mathrm{osc}_{\mathscr{T},1}(f; \Omega_K), \tag{5.73}$$

$$E_{p,K} \lesssim \|\bar{p} - \bar{p}_{\mathscr{T}}\|_{L^2(\Omega_K)} + h_K^2 \|\bar{y} - \bar{y}_{\mathscr{T}}\|_{L^2(\Omega_K)} + \mathrm{osc}_{\mathscr{T},1}(y_\Omega; \Omega_K), \tag{5.74}$$

where Ω_K is defined as in (5.2) and the hidden constants are independent of the optimal variables, their approximations, the size of the elements in the mesh $\mathscr{T}$ and $\#\mathscr{T}$.

Proof. The proof of the estimates (5.73) and (5.74) can be found in (5.69) and (5.70), respectively. □

6. Numerical Results

In this section we conduct a series of numerical examples that illustrate the performance of the error estimators that we designed and analyzed in the previous sections. In Example 2 below, we go beyond the presented analysis and perform a numerical experiment where we violate the assumption of the convexity of the domain; such an assumption is needed in Theorems 5.1 and 5.3. All the numerical experiments have been carried out with the help of a code that we implemented using C++. All matrices have been assembled exactly with the exception of the ones needed in the implementation of the variational discretization approach, where we have used a quadrature formula that is exact for polynomials of degree 19. The right hand sides as well as the approximation errors are computed with the help of the aforementioned quadrature formula. The global linear systems were solved using the multifrontal massively parallel sparse direct solver (MUMPS) Amestoy *et al.* (2000, 2001).

For a given partition $\mathscr{T}$, we seek $(\bar{y}_{\mathscr{T}}, \bar{p}_{\mathscr{T}}, \bar{u}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \times \mathbb{V}(\mathscr{T}) \times \mathbb{U}_{ad}(\mathscr{T})$, that solves the discrete optimality system (4.2) if $\mathbb{U}_{ad}(\mathscr{T}) = \mathbb{U}_{ad,0}(\mathscr{T})$, (4.7) if $\mathbb{U}_{ad}(\mathscr{T}) = \mathbb{U}_{ad,1}(\mathscr{T})$, and (4.12) if $\mathbb{U}_{ad}(\mathscr{T}) = \mathbb{U}_{ad}$. The nonlinear systems obtained when $\mathbb{U}_{ad}(\mathscr{T}) = \mathbb{U}_{ad,0}(\mathscr{T})$ and $\mathbb{U}_{ad}(\mathscr{T}) = \mathbb{U}_{ad,1}(\mathscr{T})$, are solved by using the Newton–type primal–dual active set strategy of (Stadler, 2009, Section 4). The nonlinear systems obtained for the variational discretization approach is solved by using a semi-smooth Newton method, as is described in (Brett *et al.*, 2015, Section 6). Once the discrete solutions $(\bar{y}_{\mathscr{T}}, \bar{p}_{\mathscr{T}}, \bar{u}_{\mathscr{T}})$ are obtained, we calculate the error indicators E_K, $\mathscr{E}_K$ and $\mathfrak{E}_K$, defined by (5.22), (5.34), and (5.42), respectively, to drive the adaptive procedures described in **Algorithms 1** and **2**.

Algorithm 1: Adaptive Primal-Dual Active Set Algorithm.

Input: Initial mesh $\mathscr{T}_0$, desired state y_{Ω}, constraints a and b, regularization parameter α, sparsity parameter β and external source f.

Set: $i = 0$.

Active set strategy:

1 : Choose an initial guess for the adjoint variable $p^0_{\mathscr{T}} \in \mathbb{V}(\mathscr{T})$.

2 : Compute $[\bar{y}_{\mathscr{T}}, \bar{p}_{\mathscr{T}}, \bar{u}_{\mathscr{T}}, \bar{\lambda}_{\mathscr{T}}] = \textbf{Active-Set}[\mathscr{T}, p^0_{\mathscr{T}}, \alpha, \beta, a, b, y_{\Omega}, f]$, which implements the active set strategy of (Stadler, 2009, Algorithm 2). In this step, the characterizations, given in Lemmas 4.1 and 4.2, for the discrete variables $\bar{u}_{\mathscr{T}}$ and $\bar{\lambda}_{\mathscr{T}}$, are used.

A posteriori error estimation:

3 : For each $K \in \mathscr{T}$ compute the local error indicator E_K ($\mathscr{E}_K$) given in (5.22) ((5.34)).

4 : Mark an element K for refinement if $E_K > \dfrac{1}{2} \max\limits_{K' \in \mathscr{T}} E_{K'} \left(\mathscr{E}_K > \dfrac{1}{2} \max\limits_{K' \in \mathscr{T}} \mathscr{E}_{K'} \right)$.

5 : From step **4**, construct a new mesh, using a longest edge bisection algorithm. Set $i \leftarrow i + 1$, and go to step **1**.

For the numerical results, we define the total numbers of degrees of freedom as

$$
\begin{aligned}
\mathsf{Ndof}_0 &:= 2\dim(\mathbb{V}(\mathscr{T})) + \#\mathscr{T}, \text{ if } \mathbb{U}_{ad}(\mathscr{T}) = \mathbb{U}_{ad,0}(\mathscr{T}), \\
\mathsf{Ndof}_1 &:= 3\dim(\mathbb{V}(\mathscr{T})), \text{ if } \mathbb{U}_{ad}(\mathscr{T}) = \mathbb{U}_{ad,1}(\mathscr{T}), \\
\mathsf{Ndof}_2 &:= 2\dim(\mathbb{V}(\mathscr{T})), \text{ if } \mathbb{U}_{ad}(\mathscr{T}) = \mathbb{U}_{ad}.
\end{aligned}
$$

Algorithm 2: Adaptive Semi-Smooth Newton Algorithm.

Input: Initial mesh $\mathscr{T}_0$, desired state y_Ω, constraints a and b, regularization parameter α, sparsity parameter β and external source f.

Set: $i = 0$.

Active set strategy:

1 Choose initial guesses $y_{\mathscr{T}}^0 \in \mathbb{V}(\mathscr{T})$ and $p_{\mathscr{T}}^0 \in \mathbb{V}(\mathscr{T})$, for the state and adjoint variables, respectively.

2 Compute $[\bar{y}_{\mathscr{T}}, \bar{p}_{\mathscr{T}}, \bar{u}_{\mathscr{T}}, \bar{\lambda}_{\mathscr{T}}] = \textbf{Semi-Smooth}[\mathscr{T}, y_{\mathscr{T}}^0, p_{\mathscr{T}}^0, \alpha, \beta, a, b, y_\Omega, f]$, which implements an adaption of the semi-smooth strategy of (Brett *et al.*, 2015, Algorithm 1).

A posteriori error estimation:

3 For each $K \in \mathscr{T}$ compute the local error indicator $\mathfrak{E}_K$ given in (5.42).

4 Mark an element K for refinement if $\mathfrak{E}_K > \dfrac{1}{2} \max_{K' \in \mathscr{T}} \mathfrak{E}_{K'}$ or if $\bar{u}_{\mathscr{T}}$ have kinks on K.

5 From step **4**, construct a new mesh, using a longest edge bisection algorithm. Set $i \leftarrow i + 1$, and go to step **1**.

We recall that $e = (e_y, e_p, e_u, e_\lambda)$ and that

$$\|\!|e|\!\|_\Omega^2 = |e_y|_{H^1(\Omega)}^2 + |e_p|_{H^1(\Omega)}^2 + \|e_u\|_{L^2(\Omega)}^2 + \|e_\lambda\|_{L^2(\Omega)}^2,$$
$$\|e\|_\Omega^2 = \|e_y\|_{L^2(\Omega)}^2 + \|e_p\|_{L^2(\Omega)}^2 + \|e_u\|_{L^2(\Omega)}^2 + \|e_\lambda\|_{L^2(\Omega)}^2.$$

The initial meshes for our numerical examples are shown in Figure 1.

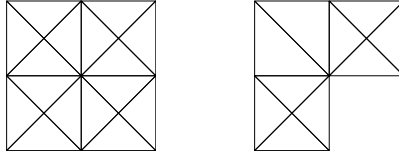Figure 1: The initial meshes used when the domain Ω is a square (Example 1) and a two–dimensional L–shaped domain (Example 2).

It is known that the visualization of the quadratic rate of convergence for the approximation error delivered by the variational discretization approach requires a complicated post processing step. We refer the reader to (Casas *et al.*, 2012a, Section 4), (Casas *et al.*, 2012c, Section 6), and (Meyer & Rösch, 2004, Section 5) for a thorough discussion about this matter. In order to overcome this difficulty and to maintain the proposed AFEM as simple as possible, we have included an extra marking criterion in Step **4** on **Algorithm 2** by also refining the elements of the mesh where $\bar{u}_{\mathscr{T}}$ presents kinks.

We now provide two numerical experiments. In the first one, we consider a problem where an exact solution can be obtained: we fix the optimal state and adjoint state variables and compute the exact optimal control, its associated subgradient, the desired state y_Ω and the source term f, by invoking the projection formulas (3.6) and (3.7) and the state and adjoint equations (3.2) and (3.3), respectively. In Example 2, the exact solutions are not known.

Example 1: We set $\Omega = (0,1)^2$, $a = -3$, and $b = 3$. The exact optimal state and adjoint state are given by

$$\bar{y} = x_1 x_2 (x_1 - 1)(x_2 - 1)\arctan((x_1 - 0.5)/0.01),$$
$$\bar{p} = -(\exp(5x_1 x_2) - 1)(1 - x_1)(1 - x_2).$$

The purpose of this example is to investigate the performance of the a posteriori error estimators when varying the parameters α and β. First, we investigate the effect of varying the regularization parameter α by considering

$$\beta = 7 \cdot 10^{-1} \quad \text{and} \quad \alpha \in \{10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}.$$

Second, we investigate the effect of varying the sparsity parameter β by considering

$$\alpha = 10^{-3} \quad \text{and} \quad \beta \in \{10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}.$$

Third, we study the influence of the sparsity parameter β in the adaptive procedures by presenting a series of adaptively refined meshes. We consider

$$\alpha = 10^{-3} \quad \text{and} \quad \beta \in \{0.1, 0.3, 0.5, 0.7\}.$$

The results are shown in Figures 2, 3, 4 and 5. We observe optimal experimental rates of convergence for the error estimators $\mathscr{E}$ and $\mathfrak{E}$, as well as optimal experimental decay for each of their contributions for all the values of the parameters α and β considered. Since we have access to the exact solution, we also present and observe optimal experimental rates of convergence for the errors $\|\!|e|\!\|_\Omega$ and $\|e\|_\Omega$ for, again, all the values of the parameter α and β considered. In Figure 3, we observe that the experimental rates of convergences for E and its contributions decay as $\mathsf{Ndof}_1^{-1/2}$. This reduced experimental rate of convergence can be explained by the fact that the operator Θ, which is defined as in (4.8) and is essential for the projection formulas of Lemma 4.2, only exhibits linear order of convergence. We also present, in Figure 5, some suitable meshes obtained with our AFEMs. We observe that most of the refinement is concentrated close to the places where the restrictions become active and also where the state variable $\bar{y}$ exhibits singular behavior.

Example 2: We set $\Omega = (0,1)^2 \setminus [0.5,1) \times (0,0.5]$, $a = -0.6$, $b = 0.4$, $\alpha = 10^{-8}$, $\beta = 8 \cdot 10^{-2}$ and $y_\Omega = 50\sin(3\pi x_2)(x_1 - 0.5)$.

In Figure 6 we report our findings. Similar conclusions as the ones presented for Example 1 can be drawn. We also report experimental rates of convergence for the devised error estimators when using uniform refinement. It can be observed that the AFEMs described in **Algorithms 1** and **2** outperform the FEMs described in Section 4.

From the presented numerical examples several general conclusions can be drawn:

• The error estimators $\mathscr{E}$ and $\mathfrak{E}$, as well as each of their contributions, exhibit optimal experimental rates of convergence for all the values of the parameters α and β considered in the experiments that we perform.

- We observe that, even when the assumption of convexity of Ω is violated, the designed AFEM driven by $\mathfrak{E}$ delivers optimal experimental rates of convergence; see Figure 6.

- The error estimator E exhibits a linear experimental rate of convergence. This can be explained with the fact that the operator Θ only delivers linear order of convergence. As it is shown in Figure 6, this is also achieved if the condition of convexity of Ω is violated. In spite of this fact, for some combinations of the parameters α and β, we can visualize a higher experimental rate of convergence; see Figure 3.

REFERENCES

AINSWORTH, M. (2010) A framework for obtaining guaranteed error bounds for finite element approximations. *J. Comput. Appl. Math.*, **234**, 2618–2632.

AINSWORTH, M. & ODEN, J. T. (2000) *A posteriori error estimation in finite element analysis*. Pure and Applied Mathematics (New York). Wiley-Interscience [John Wiley & Sons], New York, pp. xx+240.

AMESTOY, P. R., DUFF, I. S. & L'EXCELLENT, J.-Y. (2000) Multifrontal parallel distributed symmetric and unsymmetric solvers. *Comput. Methods in Appl. Mech. Eng.*, **184**, 501 – 520.

AMESTOY, P. R., DUFF, I. S., L'EXCELLENT, J.-Y. & KOSTER, J. (2001) A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM J. Matrix Anal. Appl.*, **23**, 15–41 (electronic).

BECKER, R., BRAACK, M., MEIDNER, D., RANNACHER, R. & VEXLER, B. (2007) Adaptive finite element methods for PDE-constrained optimal control problems. *Reactive Flows, Diffusion and Transport*. Springer.

BRETT, C., S. DEDNER, A. & M. ELLIOTT, C. (2015) Optimal control of elliptic pdes at points. *IMA Journal of Numerical Analysis*, **36**, 1015–1050.

CASAS, E., HERZOG, R. & WACHSMUTH, G. (2012a) Approximation of sparse controls in semilinear elliptic equations. *Large-scale scientific computing*. Lecture Notes in Comput. Sci., vol. 7116. Springer, Heidelberg, pp. 16–27.

CASAS, E., HERZOG, R. & WACHSMUTH, G. (2012b) Approximation of sparse controls in semilinear equations by piecewise linear functions. *Numerische Mathematik*, **122**, 645–669.

CASAS, E., HERZOG, R. & WACHSMUTH, G. (2012c) Optimality conditions and error analysis of semilinear elliptic control problems with L^1 cost functional. *SIAM J. Optim.*, **22**, 795–820.

CASAS, E., HERZOG, R. & WACHSMUTH, G. (2017) Analysis of spatio-temporally sparse optimal control problems of semilinear parabolic equations. *ESAIM Control Optim. Calc. Var.*, **23**, 263–295.

CASAS, E. (2017) A review on sparse solutions in optimal control of partial differential equations. *SeMA Journal*, **74**, 319–344.

CASAS, E. & KUNISCH, K. (2017) Stabilization by sparse controls for a class of semilinear parabolic equations. *SIAM J. Control Optim.*, **55**, 512–532.

CIARLET, P. G. (2002) *The finite element method for elliptic problems*. Philadelphia, PA: SIAM, pp. xxviii+530.

CLARKE, F. H. (1990) *Optimization and nonsmooth analysis*. Classics in Applied Mathematics, vol. 5, second edn. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, pp. xii+308.

ERN, A. & GUERMOND, J.-L. (2004) *Theory and practice of finite elements*. Applied Mathematical Sciences, vol. 159. New York: Springer-Verlag, pp. xiv+524.
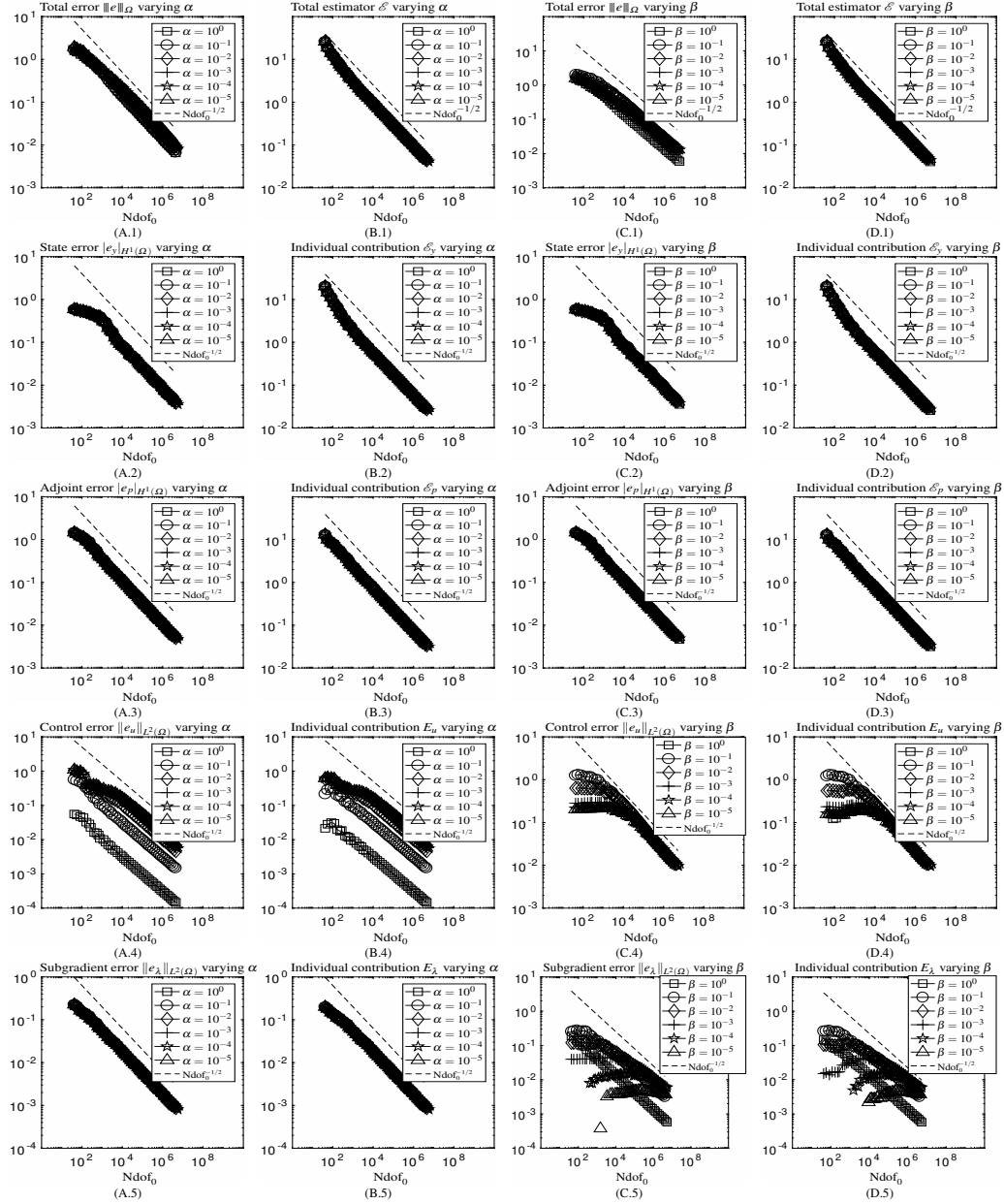
Figure 2: Example 1. Experimental rates of convergence for the piecewise constant discretization method described in Section 4.1. In (A.1)–(A.5) and (B.1)–(B.5) we have considered $\beta = 7 \cdot 10^{-1}$ and $\alpha \in \{10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$ while in (C.1)–(C.5) and (D.1)–(D.5) we have considered $\alpha = 10^{-3}$ and $\beta \in \{10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$.
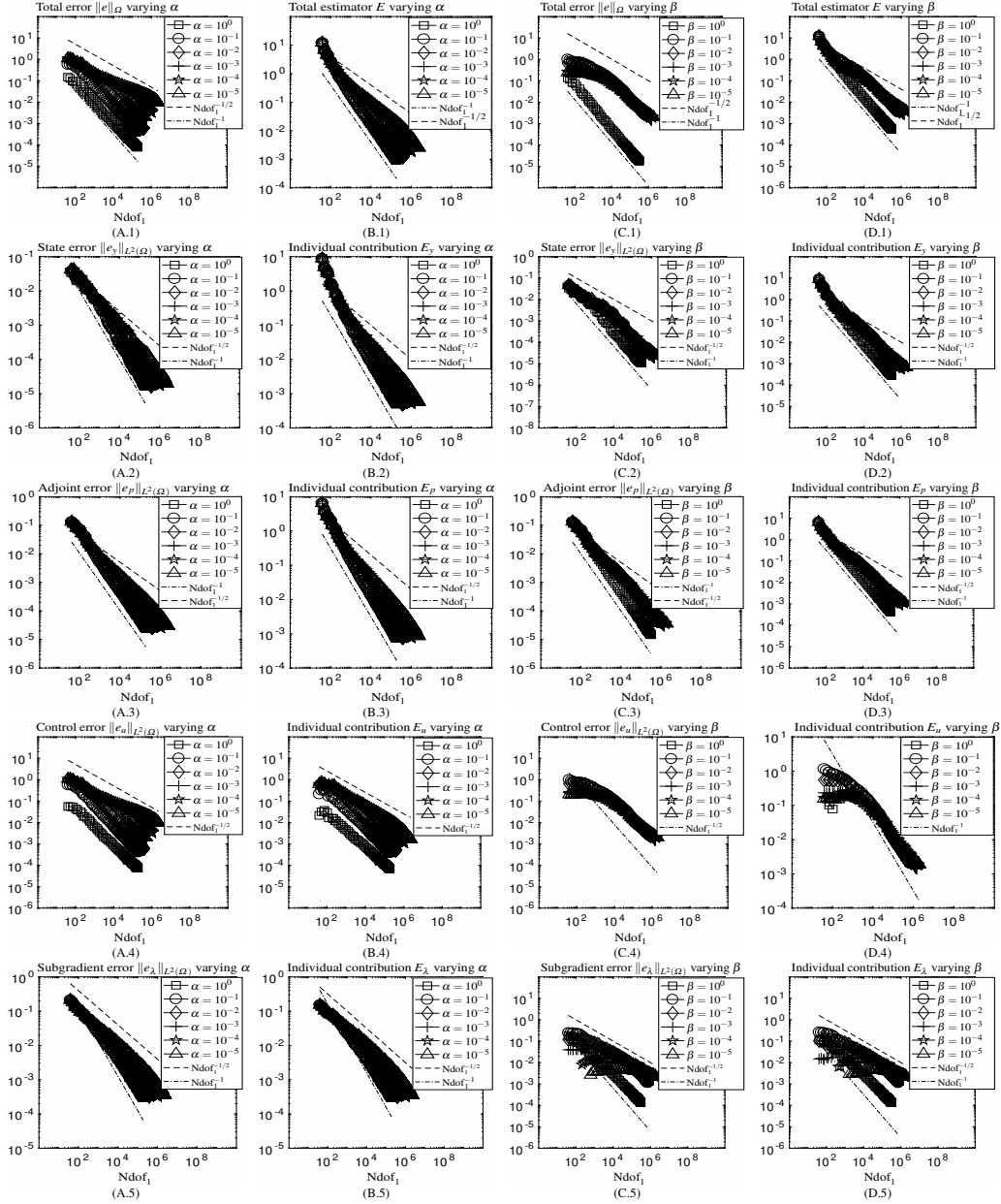
Figure 3: Example 1. Experimental rates of convergence for the piecewise linear discretization method described in Section 4.2. In (A.1)–(A.5) and (B.1)–(B.5) we have considered $\beta = 7 \cdot 10^{-1}$ and $\alpha \in \{10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$ while in (C.1)–(C.5) and (D.1)–(D.5) we have considered $\alpha = 10^{-3}$ and $\beta \in \{10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$.
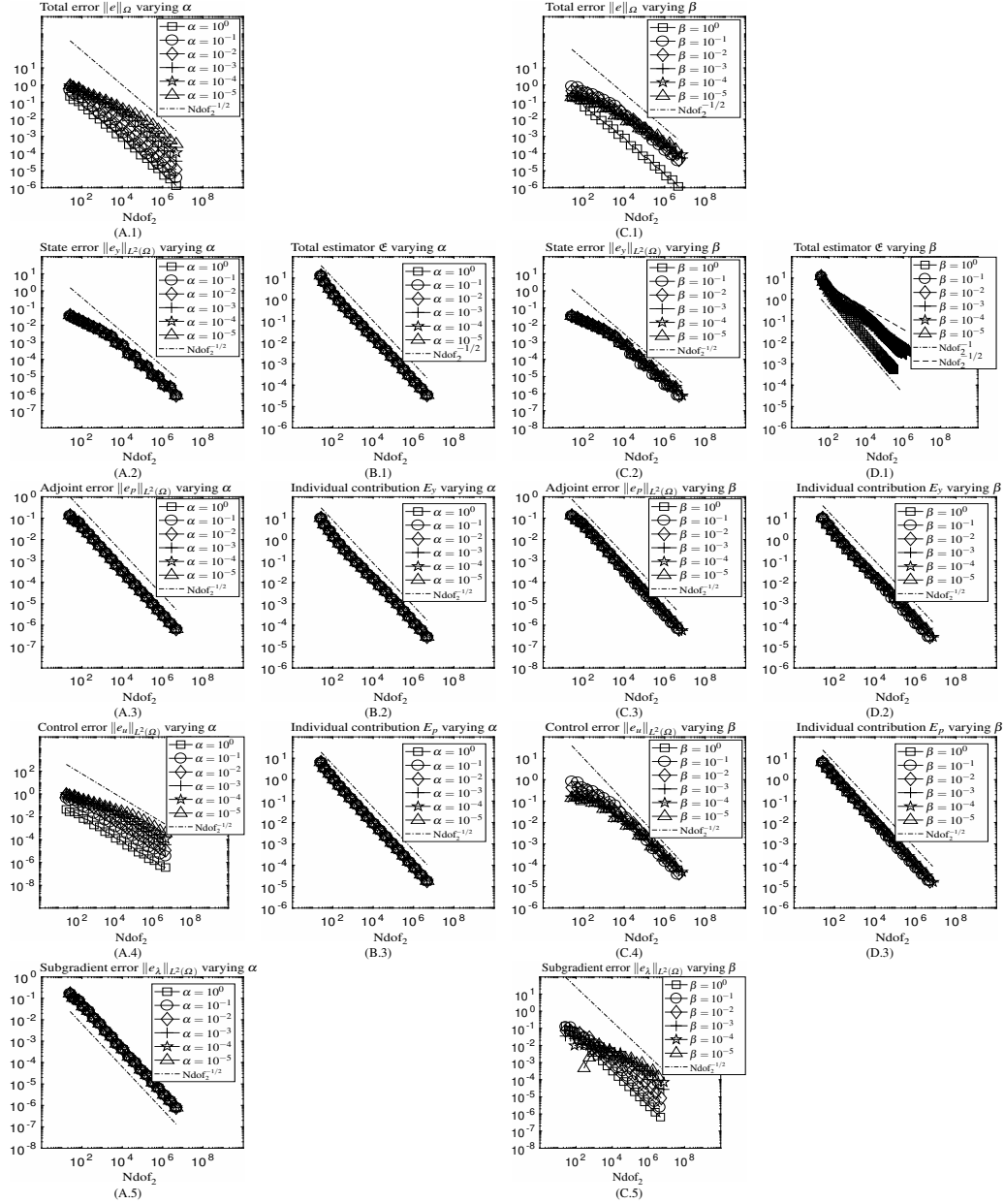
Figure 4: Example 1. Experimental rates of convergence for the variational discretization scheme of Section 4.3. In (A.1)–(A.5) and (B.1)–(B.3) we have considered $\beta = 7 \cdot 10^{-1}$ and $\alpha \in \{10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$ while in (C.1)–(C.5) and (D.1)–(D.3) we have considered $\alpha = 10^{-3}$ and $\beta \in \{10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$.

Piecewise constant:

Piecewise linear:

Variational discretization:

18*th* adaptively refined mesh.

10*th* adaptively refined mesh.

9*th* adaptively refined mesh.

(E.1)

(F.1)

(G.1)

20*th* adaptively refined mesh.

12*th* adaptively refined mesh.

9*th* adaptively refined mesh.

(E.2)

(F.2)

(G.2)

18*th* adaptively refined mesh.

12*th* adaptively refined mesh.

9*th* adaptively refined mesh.

(E.3)

(F.3)

(G.3)

20*th* adaptively refined mesh.

12*th* adaptively refined mesh.

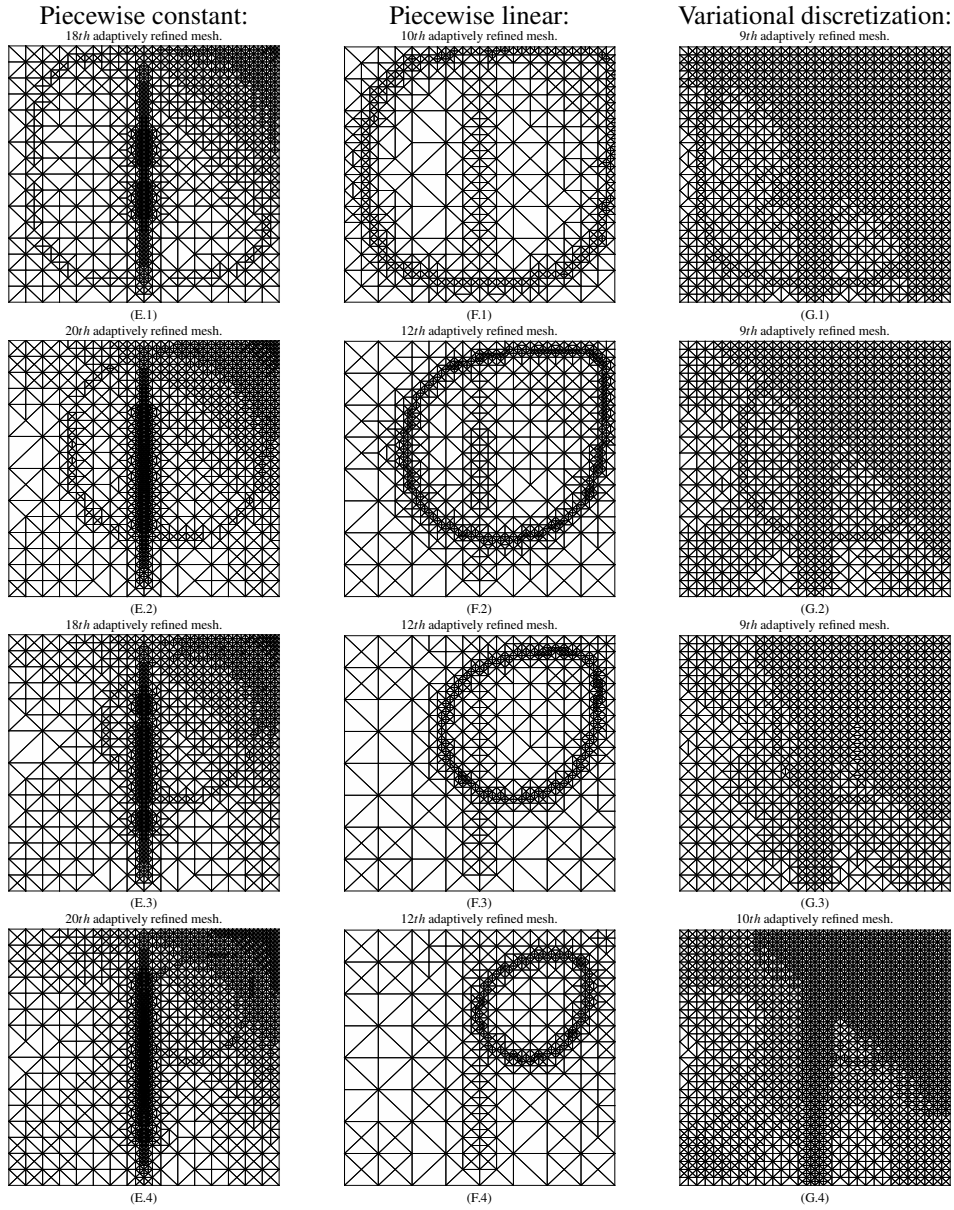10*th* adaptively refined mesh.

(E.4)

(F.4)

(G.4)

Figure 5: Example 1. Adaptively refined meshes for the piecewise constant discretization (first column), piecewise linear discretization scheme (second column), and variational discretization scheme (third column). Set $\alpha = 10^{-3}$: In (E.1), (F.1) and (G.1) we have considered $\beta = 0.1$, in (E.2), (F.2) and (G.2) $\beta = 0.3$; in (E.3), (F.3) and (G.3) $\beta = 0.5$, in (E.4), (F.4) and (G.4) $\beta = 0.7$.
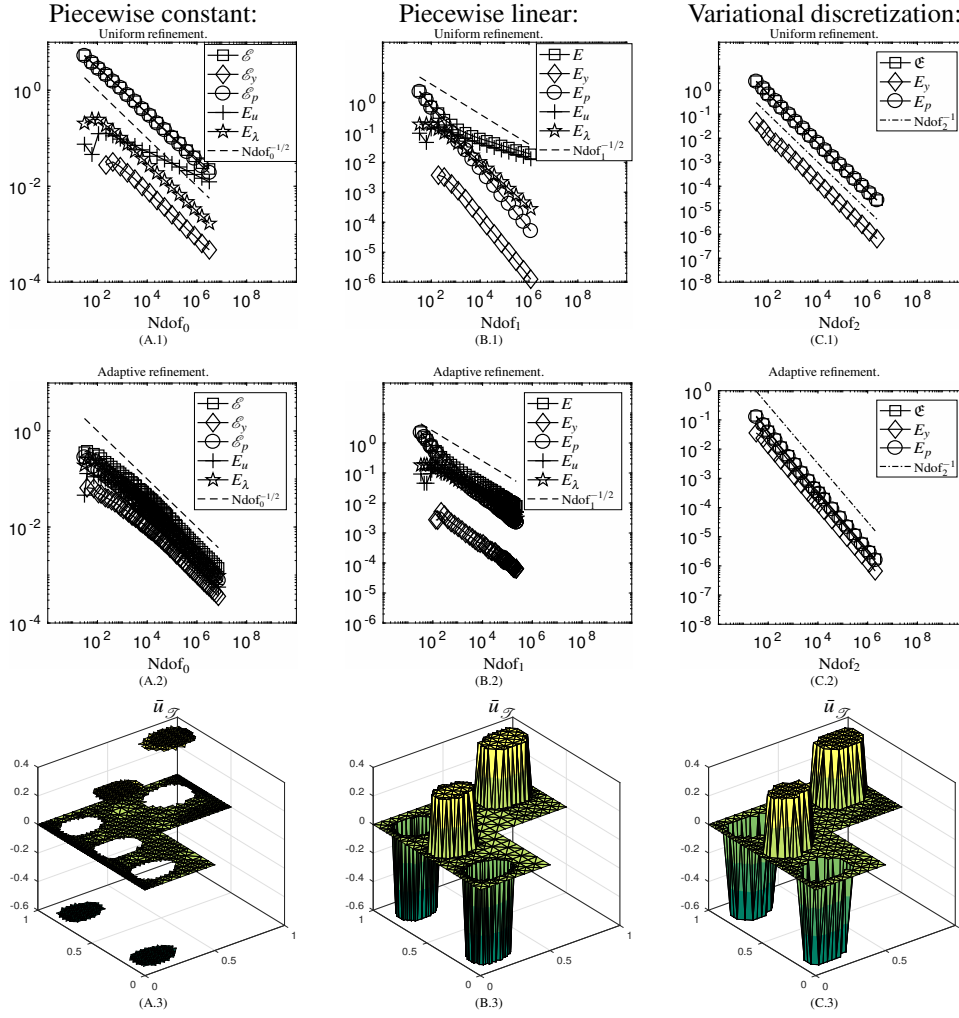
HINTERMÜLLER, M., HOPPE, R. H. W., ILIASH, Y. & KIEWEG, M. (2008) An a posteriori error analysis of adaptive finite element methods for distributed elliptic control problems with control constraints. *ESAIM:*

Figure 6: Example 2. Experimental rates of convergence for $\mathscr{E}$ and each of its contributions for uniform (A.1) and adaptive refinement (A.2), and the approximate discrete solution $\bar{u}_{\mathscr{T}} \in \mathbb{U}_{ad,0}(\mathscr{T})$ (A.3). Experimental rates of convergence for E and each of its contributions for uniform (B.1) and adaptive refinement (B.2), and the approximate discrete solution $\bar{u}_{\mathscr{T}} \in \mathbb{U}_{ad,1}(\mathscr{T})$ (B.3). Experimental rates of convergence for $\mathfrak{E}$ and each of its contributions for uniform (C.1) and adaptive refinement (C.2) and the approximate discrete solution $\bar{u}_{\mathscr{T}} \in \mathbb{U}_{ad}$ (C.3).

Control Optim. Calc. of Var., **14**, 540–560.

HINTERMÜLLER, M. & HOPPE, R. H. W. (2008) Goal-oriented adaptivity in control constrained optimal control of partial differential equations. *SIAM J. Control Optim.*, **47**, 1721–1743.

HINZE, M. (2005) *A variational discretization concept in control constrained optimization: the linear-quadratic case*. *Computational Optimization and Applications*, **30**, 45–63.

KOHLS, K., RÖSCH, A. & SIEBERT, K. G. (2014) A posteriori error analysis of optimal control problems with

control constraints. *SIAM J. Control Optim.*, **52**, 1832–1861.

LIU, W. & YAN, N. (2001) A posteriori error estimates for distributed convex optimal control problems. *Adv. in Comput. Math.*, **15**, 285–309.

MEYER, C., RADEMACHER, A. & WOLLNER, W. (2015) Adaptive optimal control of the obstacle problem. *SIAM J. Sci. Comput.*, **37**, 918–945.

MEYER, C. & RÖSCH, A. (2004) Superconvergence properties of optimal control problems. *SIAM Journal on Control and Optimization*, **43**, 970–985.

MIKHLIN, S. G. (1986) *Constants in some inequalities of analysis*. A Wiley-Interscience Publication, John Wiley and Sons, Ltd., Chichester.

NOCHETTO, R. H., SIEBERT, K. G. & VEESER, A. (2009) Theory of adaptive finite element methods: an introduction. *Multiscale, nonlinear and adaptive approximation*. Springer.

NOCHETTO, R. H. & VEESER, A. (2011) Primer of adaptive finite element methods. *Multiscale and Adaptivity: Modeling, Numerics and Applications, CIME Lectures*. Springer.

OTÁROLA, E. & SALGADO, A. J. (2018) Sparse optimal control for fractional diffusion. *Comput. Methods Appl. Math.*, **18**, 95–110.

REPIN, S. & SAUTER, S. (2006) Functional a posteriori estimates for the reaction-diffusion problem. *C. R. Math. Acad. Sci. Paris*, **343**, 349–354.

SCHIROTZEK, W. (2007) *Nonsmooth analysis*. Universitext. Springer, Berlin, pp. xii+373.

SCHNEIDER, R. & WACHSMUTH, G. (2016) A posteriori error estimation for control-constrained, linear-quadratic optimal control problems. *SIAM Journal on Numerical Analysis*, **54**, 1169–1192.

STADLER, G. (2009) Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices. *Comput. Optim. Appl.*, **44**, 159–181.

TRÖLTZSCH, F. (2010) *Optimal control of partial differential equations*. Graduate Studies in Mathematics, vol. 112. American Mathematical Society, Providence, RI, pp. xvi+399. Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels.

VERFÜRTH, R. (1989) A posteriori error estimators for the Stokes equations. *Numer. Math.*, **55**, 309–325.

VERFÜRTH, R. (2013) *A posteriori error estimation techniques for finite element methods*. Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford, pp. xx+393.

VEXLER, B. & WOLLNER, W. (2008) Adaptive finite elements for elliptic optimization problems with control constraints. *SIAM J. Control Optim.*, **47**, 509–534.

VOHRALÍK, M. (2008) A posteriori error estimation in the conforming finite element method based on its local conservativity and using local minimization. *C. R. Math. Acad. Sci. Paris*, **346**, 687–690.

WACHSMUTH, G. & WACHSMUTH, D. (2011) Convergence and regularization results for optimal control problems with sparsity functional. *ESAIM Control Optim. Calc. Var.*, **17**, 858–886.