No. 681 **March 2025**

A POSTERIORI ERROR ANALYSIS FOR OPTIMIZATION WITH PDE CONSTRAINTS

F. Gaspoz, C. Kreuzer,
A. Veeser, W. Wollner

ISSN: 2190-1767

A POSTERIORI ERROR ANALYSIS FOR OPTIMIZATION WITH PDE CONSTRAINTS

FERNANDO GASPOZ, CHRISTIAN KREUZER, ANDREAS VEESER, AND WINNIFRIED WOLLNER

Abstract. We consider finite element solutions to optimization problems, where the state depends on the possibly constrained control through a linear partial differential equation. Basing upon a reduced and rescaled optimality system, we derive a posteriori bounds capturing the approximation of the state, the adjoint state, the control and the observation. The upper and lower bounds show a gap, which grows with decreasing cost or Tikhonov regularization parameter. This growth is mitigated compared to previous results and can be countered by refinement if control and observation involve compact operators. Numerical results illustrate these properties for model problems with distributed and boundary control.

1. Introduction

A basic example for optimization problems constrained by partial differential equations (PDEs) is

$$\min_{(q,u)\in K\times \mathring{H}^1(\Omega)} \frac{1}{2}\,\|u-u_d\|^2_{L^2(\Omega)}+\frac{\alpha}{2}\,\|q\|^2_{L^2(\Omega)} \quad \text{subject to} \quad -\Delta u = f+q \text{ in } \Omega, \tag{1.1}$$

where $\Omega \subset \mathbb{R}^d$ is a suitable domain, u_d is the desired state and we assume box constraints, i.e. $K := \{q \in L^2(\Omega) \mid a \le q \le b\}$ with $a < b$. Such problems are ubiquitous in the optimal control of PDEs. They appear also as Tikhonov regularizations of inverse problems. In the former case the parameter $\alpha > 0$ scales the cost of the control, while in the latter case it is the regularization parameter, which may be chosen quite small. Hence, in any result about such problems, the dependencies on α are critical.

This article concerns the a posteriori error analysis for problems like (1.1). An a posteriori error analysis aims at deriving computable quantities that, ideally, bound a suitable error from above and below. These quantities can then be used to assess the approximate solution, and as input for an adaptive strategy refining the mesh. For both applications, close bounds will be advantageous and thus, in the case of problems like (1.1), their dependencies on α are of particular interest.

To state our main results for the model problem (1.1), we consider the linear finite element solution, with or without discretized control; cf. [14]. We denote its

(Fernando Gaspoz) Universidad Nacional del Litoral, Facultad de Ingeniería Química, Santiago del Estero 2829, 3000 Santa Fe, Argentina.

(Christian Kreuzer) Technische Universität Dortmund, Fakultät für Mathematik, Vogelpothsweg 87, 44227 Dortmund, Germany.

(Andreas Veeser) Dipartimento di Matematica 'F. Enriques', Università degli Studi di Milano, Via C. Saldini, 50, 20133 Milano, Italy.

(Winnifried Wollner) Universität Hamburg, MIN Fakultät, Fachbereich Mathematik, Bundesstr. 55, 20146 Hamburg, Germany.

E-mail addresses: fgaspoz@fiq.unl.edu.ar, christian.kreuzer@tu-dortmund.de, andreas.veeser@unimi.it, winnifried.wollner@uni-hamburg.de.

Date: March 17, 2025.

2020 *Mathematics Subject Classification.* 49M25,49M41,65N15,65N20,65N30.

error by err, which suitably combines the $\mathring{H}^1$-errors of state and adjoint state and the L^2-errors of control and observation. Furthermore, let $\eta^2 = \sum_{z\in\mathcal{V}} \eta_z^2$ be an estimator, whose local indicators η_z quantify the local residuals in Theorem 4.1 below in the spirit of [6, 19]. Then Theorems 4.1 and 4.4 imply, as $\alpha \to 0$,

$$c_\Delta \eta \le \mathsf{err} \le C_\Delta \min\left\{1 + O\left(\frac{1}{\sqrt{\alpha}}\right), 1 + O\left(\frac{1}{\alpha}\right) \frac{\left(\sum_{z\in\mathcal{V}} h_z^2 \eta_z^2\right)^{\frac{1}{2}}}{\eta}\right\} \eta. \tag{1.2}$$

Here h_z denotes the local meshsize around a vertex $z \in \mathcal{V}$ and the constants c_Δ, C_Δ are independent of α. More precisely, they depend on the technique quantifying the local residuals, the shape regularity of the mesh cells, and the Poisson problem, which is associated with the state equation. Furthermore, the two constants ensure the equivalence

$$c_\Delta \eta_\Delta \le \|\nabla(v - V)\|_{L^2(\Omega)} \le C_\Delta \eta_\Delta, \tag{1.3}$$

where η_Δ is the counterpart of η for the $\mathring{H}^1$-error between any solution v to the Poisson problem and its linear finite element approximation V.

For equivalences like (1.3), the ratio $C_\Delta/c_\Delta \ge 1$ of the involved constants provides information about the quality of the estimator η_Δ. The structure of (1.2) is slightly more involved because the min depends on the indicators η_z, $z \in \mathcal{V}$, and therefore is no part of the constant. Nevertheless, we can measure the quality of the error quantification (1.2) similarly by the ratio of upper to lower bound, which we call gap. In this context, the error bounds (1.2) lead to the two following interrelated conclusions:

- Taking (1.3) as a benchmark for the used estimation technique, the gap in (1.2) becomes the one associated with η_Δ for $h := \max_{z\in\mathcal{V}} h_z \to 0$ and fixed $\alpha > 0$.
- Ensuring $\sum_{z\in\mathcal{V}} h_z^2 \eta_z^2 \le \alpha^2 \eta^2$, the gap in the error bounds (1.2) remains uniformly bounded for all $\alpha > 0$.

Apart from these two features, the error quantification (1.2) improves existing results. To be more precise, we compare with [16, 18], which is the first abstract a posteriori analysis and reviews previous a posteriori bounds. Therein, as $\alpha \to 0$, the gap grows with $O(1/\alpha)$ if the control is not discretized, otherwise with $O(1/\alpha^2)$; see Remark 3.4 for more details. In both cases, the error bounds (1.2) mitigate those growths to $O(1/\sqrt{\alpha})$.

The gap in (1.2) is closely related to the a priori results in our previous article [10]. To see this, let us suppose a variational discretization [14] for simplicity and denote by $\mathsf{best\text{-}err}$ the best approximation error in the underlying discrete spaces. Then [10, Section 5.2] ensures the following variant of Céa's lemma:

$$\mathsf{err} \le \mu_\Delta \min\left\{1 + O\left(\frac{1}{\sqrt{\alpha}}\right), 1 + O\left(\frac{h}{\alpha}\right)\right\} \mathsf{best\text{-}err} \quad \text{as } \alpha, h \to 0, \tag{1.4}$$

where μ_Δ is the near-best approximation constant associated with the discretization of the state equation. Inspecting Céa's lemma and the derivation of (1.3), we see that μ_Δ corresponds to the ratio C_Δ/c_Δ. Furthermore, we observe that the min in (1.4) is an upper bound for the one in (1.2).

The similarities between (1.2) and (1.4) result from a 'duality' in the respective derivations. To illustrate this, we outline the key ingredients of both derivations. In both cases, we first consider the possibly nonlinear optimality system

$$-\Delta u - \Pi_K\left(-\tfrac{1}{\alpha} p\right) = f, \qquad -\Delta p - u = -u_d, \tag{1.5}$$

where the control is implicitly given by $q = \Pi_K(-\frac{1}{\alpha}p)$, and then divide the adjoint equation by $\sqrt{\alpha}$ and replace the adjoint state p by $z = p/\sqrt{\alpha}$. The latter has the

effect that the perturbations of the Laplacian in both equations scale like $1/\sqrt{\alpha}$. Analyzing the properties of the possibly nonlinear operator $\mathcal{B}_\alpha$ associated with the resulting system prepares the ground for both (1.2) and (1.4).

More precisely, the continuity and generalized coercivity properties of $\mathcal{B}_\alpha$ are, respectively, crucial ingredients for the lower bound and for the first option in the upper bound of (1.2). For the first option in (1.4), one combines the continuity properties of $\mathcal{B}_\alpha$ with the coercivity properties of its discretization.

The second options instead hinge on the compactness of the perturbations of the Laplacian in (1.5), which arises from the embedding $\mathring{H}^1(\Omega) \subset L^2(\Omega)$. In both cases, the error is decomposed into a main part and a 'compact' part. To this end, we use an auxiliary function in the spirit of the elliptic reconstruction [22] for (1.2) and a generalized Ritz projection for (1.4); see also Remark 3.5. The actual benefit from the available compactness is limited by regularity theorems for the Poisson equation on polygonal domains and/or properties of the discretization.

To conclude this introduction, the following remarks are in order:

- To quantify the local residuals in Theorem 4.1, one can use classical techniques, see, e.g. [1,28], instead of [6,19]. In this case, the error quantification (1.2) holds then only up to so-called oscillation terms.
- In the absence of control constraints, Theorem 3.10 provides a variant of (1.2) with an improved gap.
- Although our approach is based upon the reduced optimality system (1.5), it is not restricted to variational control discretizations as in [14] and covers also discretized controls and bounds for their error; see Corollary 3.3.

This article is organized as follows. Section 2 recalls the continuity and coercivity properties of $\mathcal{B}_\alpha$ from [10], along with their proofs due to their importance and for the sake of a self-contained presentation. The principal part of our a posteriori analysis is then developed in Section 3. Finally, Section 4 illustrates the obtained results by applying them to (1.1) and a Neumann boundary control problem, as well as by numerical tests in both cases.

2. Optimization problem and reduced optimality system

This section presents the abstract optimization problem to be considered and recalls from [10] its reduced and rescaled optimality system, along with key ingredients for its well-posedness. These ingredients are also crucial for the subsequent a posteriori error analysis.

To introduce the abstract optimization problem, we take the viewpoint of an optimal control problem and start with the *state equation*. We assume that the control variable q is taken from a real Hilbert space $(Q, (\cdot,\cdot)_Q)$ with induced norm $\|\cdot\|_Q$. The relation between control $q \in Q$ and state $u \in V_1$ is given by a linear boundary value problem of the form

$$Au = f + Cq, \tag{2.1}$$

with the following properties:

- The differential operator $A : V_1 \to V_2^*$ is defined between the *state space* V_1, which is a Hilbert space with scalar product $(\cdot,\cdot)_1$ and the dual V_2^* of a second Hilbert space $(V_2, (\cdot,\cdot)_2)$. For $i = 1, 2$, the induced norms on V_i and V_i^* and the dual pairing are denoted by $\|\cdot\|_i$, $\|\cdot\|_{i,*}$, and $\langle\cdot,\cdot\rangle_i$, respectively. We assume that the operator A is a linear isomorphism, i.e. the bilinear

form $a : V_1 \times V_2 \to \mathbb{R}$ defined by $(v_1, v_2) \mapsto \langle Av_1, v_2\rangle_2$ satisfies

$$M_a := \sup_{\|v_1\|_1=1} \sup_{\|v_2\|_2=1} a(v_1, v_2) < \infty, \tag{2.2a}$$

$$\forall v_1 \in V_1 \quad \Big(\forall v_2 \in V_1 \; a(v_1, v_2) = 0\Big) \implies v_1 = 0, \tag{2.2b}$$

$$m_a := \inf_{\|v_2\|_2=1} \sup_{\|v_1\|_1=1} a(v_1, v_2) > 0; \tag{2.2c}$$

compare, e.g., with [23].

- The operator $C : Q \to V_2^*$ is linear and bounded with constant M_C.
- The load term satisfies $f \in V_2^*$.

Our goal is then to numerically solve the constrained optimization problem

$$\min_{(q,u)\in K\times V_1} \frac{1}{2} \|Iu - u_d\|_W^2 + \frac{\alpha}{2} \|q\|_Q^2 \quad \text{subject to} \quad Au = f + Cq, \tag{2.3}$$

and we suppose in addition:

- The set $K \subset Q$ of *admissible controls* is nonempty, closed and convex.
- The *cost of the control* is scaled with a parameter $\alpha > 0$, which can also be viewed as a Tikhonov regularization.
- The *desired state* u_d lies in the *target space* W, which is a Hilbert space with scalar product $(\cdot, \cdot)_W$ and induced norm $\|\cdot\|_W$.
- The *observation operator* $I : V_1 \to W$ is linear and bounded with constant M_I.

Problem (2.3) is a quadratic minimization problem with a possibly non-linear constraint (in the case when $K \neq Q$). As the set of admissible controls is convex and closed, standard arguments ensure the existence of a unique solution; see, e.g., [21, Theorem 1.1] or [27, Chapter 2.5].

To formulate the optimality system for (2.3), we introduce the adjoint operators A^*, C^*, I^* of A, C, I by

$$A^*v_2 = a(\cdot, v_2), \quad (q, C^*v_2)_Q = \langle Cq, v_2\rangle_2, \quad \langle I^*w, v_1\rangle_1 = (Iv, w)_W$$

for all $v_1 \in V_1$, $v_2 \in V_2$, $q \in Q$, $w \in W$. The unique solution (q, u) of (2.3) is equivalently characterized by the existence of a $p \in V_2$, called *adjoint state*, such that the following system of optimality conditions is satisfied:

$$Au = f + Cq, \quad A^*p = I^*(Iu - u_d), \quad q = \Pi_K(-\tfrac{1}{\alpha}C^*p); \tag{2.4}$$

compare with [27]. Here $\Pi_K \mid Q \to K \subset Q$ is the projection operator onto the admissible set K, which is characterized by $\|q - \Pi_K q\|_Q = \inf_{p\in K} \|q - p\|_Q$ or, equivalently, by

$$\forall p \in K \quad (q - \Pi_K q, \Pi_K q - p)_Q \geq 0. \tag{2.5}$$

We notice that Π_K is Lipschitz continuous with constant 1 and satisfies the inequality

$$(\Pi_K(q_1) - \Pi_K(q_2), q_1 - q_2)_Q \geq \|\Pi_K(q_1) - \Pi_K(q_2)\|_Q^2. \tag{2.6}$$

If $K = \{q \in L^2(\Omega) \mid a \leq q \leq b\}$, $V_1 = V_2 = \mathring{H}^1(\Omega)$, $A = -\Delta$ is the weak Laplacian, $Q = L^2(\Omega) = W$, C and I are the canonical compact immersions $L^2(\Omega) \to H^{-1}(\Omega)$ and $H_0^1(\Omega) \to L^2(\Omega)$, then (2.3) simplifies to the optimization problem (1.1) in the introduction. Notice that, in this case, $\Pi_K q = \max\{\min\{b, q\}, a\}$ and the operators C and I are related by $C^* = I$.

As in [10], we rescale the adjoint variable by

$$z = \frac{1}{\sqrt{\alpha}} p \in V_2. \tag{2.7}$$

This will turn out advantageous when considering the limit of vanishing Tikhonov regularization; see in particular Remark 3.4. We thus obtain the rescaled system

$$Au = f + Cq, \quad A^*z = \frac{1}{\sqrt{\alpha}} I^*(Iu - u_d), \quad q = \Pi_K(-\tfrac{1}{\sqrt{\alpha}} C^* z), \tag{2.8}$$

and inserting the last equation into the first one, we end up with the reduced optimality system

$$\begin{pmatrix} -\frac{1}{\sqrt{\alpha}} I^* I & A^* \\ A & -C\Pi_K(-\frac{1}{\sqrt{\alpha}} C^* \cdot) \end{pmatrix} \begin{pmatrix} u \\ z \end{pmatrix} = \begin{pmatrix} -\frac{1}{\sqrt{\alpha}} I^* u_d \\ f \end{pmatrix}. \tag{2.9}$$

In the typical case, when the operators C and I are compact (see (1.1)), we observe that the operator on the left-hand side of (2.9) is a compact perturbation of the control to state operator A and its adjoint A^*. This was exploited in [10] to show that the near-best approximation constant of Galerkin approximations for the system (2.9) asymptotically tends to the near-best approximation constant of the Galerkin approximation of the state equation; cf. (1.4). In this work, we aim to exploit this observation in the a posteriori analysis for (2.9).

The variational formulation of (2.9) reads

$$\forall \varphi_1 \in V_1 \qquad a(\varphi_1, z) - \tfrac{1}{\sqrt{\alpha}} (Iu, I\varphi_1)_W = -\tfrac{1}{\sqrt{\alpha}} (u_d, I\varphi_1)_W , \tag{2.10a}$$

$$\forall \varphi_2 \in V_2 \qquad a(u, \varphi_2) - \left(\Pi_K(-\tfrac{1}{\sqrt{\alpha}} C^* z), C^* \varphi_2\right)_Q = \langle f, \varphi_2 \rangle_2. \tag{2.10b}$$

This suggests to introduce the Hilbert space

$$V := V_1 \times V_2 \quad \text{with} \quad \|v\| := \left(\|v_1\|_1^2 + \|v_2\|_2^2\right)^{\frac{1}{2}}, \quad v = (v_1, v_2) \in V \tag{2.11}$$

with dual space $V^* = V_1^* \times V_2^*$ and induced dual norm $\|\cdot\|_*$, as well as the form $b_\alpha \mid V \times V \to \mathbb{R}$ given by

$$b_\alpha(v, \varphi) := \mathbf{a}(v, \varphi) + c_\alpha(v, \varphi) \tag{2.12a}$$

where

$$\mathbf{a}(v, \varphi) := a(v_1, \varphi_2) + a(\varphi_1, v_2) \tag{2.12b}$$

$$c_\alpha(v, \varphi) := -\left(\Pi_K\left(-\tfrac{1}{\sqrt{\alpha}} C^* v_2\right), C^* \varphi_2\right)_Q - \tfrac{1}{\sqrt{\alpha}} (Iv_1, I\varphi_1)_W \tag{2.12c}$$

for $v = (v_1, v_2), \varphi = (\varphi_1, \varphi_2) \in V$. Although $\mathbf{a}$ is bilinear, b_α is in general only linear in the second argument due to the presence of Π_K in c_α. In the introduction, we have mentioned the operator $\mathcal{B}_\alpha : V \to V^*$ given by $\mathcal{B}_\alpha v := b_\alpha(v, \cdot)$. The bilinear form $\mathbf{a} : V \times V \to \mathbb{R}$ inherits its continuity and nondegeneracy properties from a. More precisely, we have

$$\sup_{\|v\|=1} \sup_{\|\varphi\|=1} |\mathbf{a}(v, \varphi)| = M_a \quad \text{and} \quad \inf_{\|v\|=1} \sup_{\|\varphi\|=1} \mathbf{a}(v, \varphi) = m_a \tag{2.13}$$

with M_a and m_a from (2.2). While the first identity is straight-forward, the second one hinges on the inf-sup-duality, cf. Babuška [2],

$$\inf_{\|v_1\|_1=1} \sup_{\|\varphi_2\|_2=1} \mathbf{a}(v_1, \varphi_2) = \inf_{\|v_2\|_2=1} \sup_{\|\varphi_1\|_1=1} \mathbf{a}(\varphi_1, v_2). \tag{2.14}$$

In this notation, (2.9) reads

$$\text{find } x \in V \text{ such that } \ \forall \varphi \in V \quad b_\alpha(x, \varphi) = \langle f, \varphi_2 \rangle_2 - \tfrac{1}{\sqrt{\alpha}} (u_d, I\varphi_1)_W . \tag{2.15}$$

A pair $x = (u, z) \in V$ solves the variational formulation (2.15) of the reduced and rescaled optimality system if and only if the triple $\left(u, z, \Pi_K\left(-C^* z/\sqrt{\alpha}\right)\right) \in V \times Q$ satisfies the optimality system (2.8). Consequently, thanks to the convexity of (2.3),

$x = (u, z) \in X$ is a solution of (2.15) if and only if $\left(\Pi_K\left(-C^* z/\sqrt{\alpha}\right), u\right) \in Q \times V_1$ is a solution of the optimization problem (2.3).

Although b_α is not bilinear in general, we have derived in [10, Theorem 5.1] properties that generalize the continuity and inf-sup stability of bilinear forms. As this is fundamental for the following a posteriori analysis, we state them and repeat their proofs. To this end, we introduce on V the seminorm

$$|v| := \left(\|Iv_1\|_W^2 + \|C^* v_2\|_Q^2\right)^{1/2} \tag{2.16}$$

and its relative the pseudometric

$$\delta_\alpha(v,w)^2 := \alpha \left\|\Pi_K\left(-\tfrac{1}{\sqrt{\alpha}} C^* v_2\right) - \Pi_K\left(-\tfrac{1}{\sqrt{\alpha}} C^* w_2\right)\right\|_Q^2 + \|I(v_1 - w_1)\|_W^2 . \tag{2.17}$$

For $v, w \in V$ choosing $\varphi = (-(v_1 - w_1), v_2 - w_2)$, we have

$$c_\alpha(v,\varphi) - c_\alpha(w,\varphi) \geq \frac{1}{\sqrt{\alpha}} \delta_\alpha(v,w)^2, \tag{2.18a}$$

while, for any $v, w, \varphi \in V$, we have

$$|c_\alpha(v,\varphi) - c_\alpha(w,\varphi)| \leq \frac{1}{\sqrt{\alpha}} \delta_\alpha(v,w)\, |\varphi| \tag{2.18b}$$

and, thanks to (2.6),

$$\delta_\alpha(v,w) \leq |v - w| . \tag{2.19}$$

The continuity bound (2.18b) and a Cauchy-Schwarz inequality lead to

$$|b_\alpha(v,\varphi) - b_\alpha(w,\varphi)| \leq M_a\, d_\alpha(v,w)\, \|\varphi\| , \tag{2.20}$$

where the metric d_α is defined by

$$d_\alpha(v,w) := \|v - w\| + \frac{1}{\sqrt{\alpha}} \frac{M}{M_a} \delta_\alpha(v,w), \quad v, w \in V, \tag{2.21}$$

with

$$M := \max\{M_I, M_C\},$$

and $\|\cdot\|$ is from (2.11). This brings us in the position to state and prove the announced properties of the operator associated with the reduced and rescaled optimality system (2.9).

Theorem 2.1 (Continuity and inf-sup stability of form b_α)**.** *For any $v, w, \varphi \in V$, we have*

$$|b_\alpha(v,\varphi) - b_\alpha(w,\varphi)| \leq M_a\, d_\alpha(v,w)\, \|\varphi\| \tag{2.22a}$$

and there exists $0 \neq \psi \in V$ such that

$$b_\alpha\big(v,\psi\big) - b_\alpha\big(w,\psi\big) \geq \frac{m_a}{\kappa}\, d_\alpha(v,w)\, \|\psi\| , \tag{2.22b}$$

where κ is defined by

$$\kappa = \frac{1 + 2L}{1 + L}\left(1 + \frac{M}{m_a}(1 + 2L)\right) \quad \textit{with} \quad L = \frac{M}{\sqrt{\alpha}}. \tag{2.23}$$

Proof. The first inequality is (2.20). In order to prove the second one, we choose, for fixed $v, w \in V$,

$$\psi = m_a(A^{-1} J_2(v_2 - w_2), A^{-*} J_1(v_1 - w_1)) + \gamma(-(v_1 - w_1), (v_2 - w_2)),$$

for some $\gamma > 0$ to be determined later. Here the evaluation maps $J_i : V_i \to V_i^*$ are defined by $\langle J_i\psi_i, \cdot\rangle_i := (\psi_i, \cdot)_i$ for $i = 1, 2$. With this choice, we obtain

$$\begin{aligned}
&b_\alpha(v,\psi) - b_\alpha(w,\psi)\\
&\quad= m_a\left(a(v_1 - w_1, A^{-*}J_1(v_1 - w_1)) + a(A^{-1}J_2(v_2 - w_2), v_2 - w_2)\right)\\
&\qquad - m_a\left(\Pi_K\left(-\tfrac{1}{\sqrt{\alpha}}C^*v_2\right) - \Pi_K\left(-\tfrac{1}{\sqrt{\alpha}}C^*w_2\right), C^*A^{-*}J_1(v_1 - w_1)\right)_Q\\
&\qquad - m_a\tfrac{1}{\sqrt{\alpha}}\left(I(v_1 - w_1), IA^{-1}J_2(v_2 - w_2)\right)_W\\
&\qquad - \gamma\, m_a\left(\Pi_K\left(-\tfrac{1}{\sqrt{\alpha}}C^*v_2\right) - \Pi_K\left(-\tfrac{1}{\sqrt{\alpha}}C^*w_2\right), C^*(v_2 - w_2)\right)_Q\\
&\qquad + \gamma\tfrac{1}{\sqrt{\alpha}}\left(I(v_1 - w_1), I(v_1 - w_1)\right)_W\\
&\quad\geq m_a\|v - w\|^2 - \frac{M}{\sqrt{\alpha}}\,\delta_\alpha(v,w)\|v - w\| + \frac{\gamma}{\sqrt{\alpha}}\,\delta_\alpha(v,w)^2\\
&\quad\geq m_a\left(\|v - w\| + \frac{M}{M_a}\frac{1}{\sqrt{\alpha}}\,\delta_\alpha(v,w)\right)\|v - w\| - \frac{2M}{\sqrt{\alpha}}\,\delta_\alpha(v,w)\|v - w\|\\
&\qquad + \frac{\gamma}{\sqrt{\alpha}}\delta_\alpha(v,w)^2,
\end{aligned}$$

where we used $m_a \leq M_a$ as well as the continuity of C and I. Using Young's inequality $2st \leq \epsilon s^2 + t^2/\epsilon$ with $\epsilon = \frac{L}{1+2L}m_a > 0$, we may bound the critical term by

$$\frac{2M}{\sqrt{\alpha}}\,\delta_\alpha(v,w)\|v - w\| \leq \frac{L}{1+2L}m_a\,\|v - w\|^2 + \frac{1+2L}{L}\frac{M^2}{m_a\alpha}\delta_\alpha(v,w)^2.$$

Consequently, choosing

$$\gamma = \frac{M}{m_a}\,(1 + 2L)\,,$$

we arrive at

$$\begin{aligned}
b_\alpha(v,\psi) - b_\alpha(w,\psi) &\geq \frac{1+L}{1+2L}\,m_a\,d_\alpha(v,w)\|v - w\|\\
&\geq \frac{1}{\kappa}\,m_a\,d_\alpha(v,w)\|\psi\|.
\end{aligned}$$

Here the last inequality follows from

$$\|\psi\| \leq \left(1 + \frac{M}{m_a}(1 + 2L)\right)\|v - w\|. \qquad \square$$

3. Relating error and residual

This section constitutes the principal part of our a posteriori analysis for the abstract optimal control problem (2.3). A typical approach to such an analysis can be subdivided into the following three steps; cf., e.g., [6, §4] or [28, §1.4]: given an approximate solution,

- relate the error (norm) to a suitable norm of the so-called residual, a quantity that depends only on data and the approximate solution,
- split the residual norm, which is typically of dual character, into local contributions,
- further split the local contributions into a computable part involving the approximate solution and an oscillatory part depending only on data.

This section addresses the first step. It then turns out that the following two steps hinge only on the particular structure of the state and adjoint equations. Therefore, they are not addressed in general and postponed to the applications in Section 4.

We shall base our a posteriori analysis of the optimization problem (2.3) on the variational formulation (2.15) of the reduced and rescaled optimality system (2.9). Let $\tilde{x} = (\tilde{u}, \tilde{z}) \in V$ be some approximation of $x = (u, z)$, where u is the exact state and z the (rescaled) exact adjoint state. We define the *residual* in $\tilde{x}$ by

$$\mathrm{Res}(\tilde{x}) := \begin{pmatrix} -\frac{1}{\sqrt{\alpha}} I^* u_d \\ f \end{pmatrix} - \begin{pmatrix} -\frac{1}{\sqrt{\alpha}} I^* I & A^* \\ A & -C\Pi_K(-\frac{1}{\sqrt{\alpha}} C^* \cdot) \end{pmatrix} \begin{pmatrix} \tilde{u} \\ \tilde{z} \end{pmatrix},$$

or, equivalently in variational form, by

$$\langle \mathrm{Res}(\tilde{x}), \varphi \rangle = \left\langle f - A\tilde{u} + C\Pi_K \left(-\tfrac{1}{\sqrt{\alpha}} C^* \tilde{z} \right), \varphi_2 \right\rangle_2 + \left\langle \tfrac{1}{\sqrt{\alpha}} I^* (I\tilde{u} - u_d) - A^* \tilde{z}, \varphi_1 \right\rangle_1 .$$

In what follows, we shall offer three approaches to relate the error to the residual $\mathrm{Res}(\tilde{x})$, strengthening the relationship under increasingly stronger assumptions. For comparison, it is useful to recall that the assumptions on the state equation imply the following error-residual relationship: if $v \in V_1$ verifies $Av = g$ and $\tilde{v} \in V_1$ is some approximation of v, then error and residual norm $\|g - A\tilde{v}\|_{2,*}$ are equivalent:

$$\frac{1}{M_a} \|g - A\tilde{v}\|_{2,*} \leq \|v - \tilde{v}\|_1 \leq \frac{1}{m_a} \|g - A\tilde{v}\|_{2,*}. \tag{3.1}$$

(The proof of this follows the lines of the proof of Theorem 3.1 below, replacing the form b_α by the bilinear form a.) Notice that there is a *gap* between the upper and lower bound for $M_a \gg m_a$, which can be measured by the ratio $M_a/m_a \geq 1$ of upper to lower bound.

3.1. **Using continuity of control and observation.** In contrast to the subsequent subsections, here we shall not assume compactness of the control operator C and the observation operator I in addition to the conditions in Section 2, i.e. they are just linear and bounded operators.

The residual and error are related through a possibly nonlinear operator. In fact, since $x = (u, z)$ is the exact solution of (2.9), we have the identity

$$\mathrm{Res} \begin{pmatrix} \tilde{u} \\ \tilde{z} \end{pmatrix} = \begin{pmatrix} -\frac{1}{\sqrt{\alpha}} I^* I & A^* \\ A & -C\Pi_K(-\frac{1}{\sqrt{\alpha}} C^* \cdot) \end{pmatrix} \begin{pmatrix} u \\ z \end{pmatrix} - \begin{pmatrix} -\frac{1}{\sqrt{\alpha}} I^* I & A^* \\ A & -C\Pi_K(-\frac{1}{\sqrt{\alpha}} C^* \cdot) \end{pmatrix} \begin{pmatrix} \tilde{u} \\ \tilde{z} \end{pmatrix},$$

which in variational form reads

$$\langle \mathrm{Res}(\tilde{x}), \varphi \rangle = b_\alpha(x, \varphi) - b_\alpha(\tilde{x}, \varphi), \quad \varphi \in V. \tag{3.2}$$

Following standard arguments, this identity can be combined with the properties of the form b_α in Theorem 2.1. This *direct approach* leads to the following relationship between the residual in the dual norm $\|\cdot\|_*$, see (2.11), and the d_α*-error*, i.e. the distance of the states and their approximations in the metric (2.21).

Theorem 3.1 (Bounding the d_α-error – general case)**.** *Let $x = (u, z) \in V$ be the solution to* (2.15) *and $\tilde{u} \in V_2$ some approximate state and $\tilde{z} \in V_1$ some approximate rescaled adjoint state. Writing $\tilde{x} = (\tilde{u}, \tilde{z})$, their d_α-error is equivalent to the dual norm of the residual:*

$$\frac{1}{M_a} \|\mathrm{Res}(\tilde{x})\|_* \leq d_\alpha(x, \tilde{x}) \leq \frac{\kappa}{m_a} \|\mathrm{Res}(\tilde{x})\|_*$$

with κ from (2.23).

Proof. The identity (3.2) and the Lipschitz continuity (2.22a) imply that, for any $\varphi \in V$,

$$|\langle \mathrm{Res}(\tilde{x}), \varphi \rangle| = |b_\alpha(x, \varphi) - b_\alpha(\tilde{x}, \varphi)| \leq M_a \, d_\alpha(x, \tilde{x}) \, \|\varphi\| \, .$$

Dividing by $\|\varphi\|$ and taking the supremum over all $0 \neq \varphi \in V$ proves the lower bound.

For the upper bound, it follows from (2.22b) and (3.2) that there exists $0 \neq \psi \in V$ such that

$$\frac{m_a}{\kappa} d_\alpha(x, \tilde{x}) \|\psi\| \leq b_\alpha(x, \psi) - b_\alpha(\tilde{x}, \psi) = \langle \mathrm{Res}(\tilde{x}), \psi \rangle \leq \| \mathrm{Res}(\tilde{x}) \|_* \|\psi\|.$$

Thus dividing by $\|\psi\| > 0$ finishes the proof. □

Remark 3.2 (Gap in bounding d_α-error – general case)**.** *The upper and lower bound in Theorem 3.1 present the gap $(M_a \kappa)/m_a$. With respect to* (3.1) *concerning the state equation, this gap is amplified by the factor κ from Theorem 2.1.*

For the limit $\alpha \to 0$ of the Tikhonov regularization parameter, with I and C fixed, we have $L = \frac{M}{\sqrt{\alpha}} \to \infty$ and, therefore, the amplification factor κ verifies

$$\kappa = \left(\frac{4M^2}{m_a} + o(1) \right) \frac{1}{\sqrt{\alpha}} \quad \text{as} \quad \alpha \to 0. \tag{3.3}$$

Theorem 3.1 considers only approximations of the state and the (rescaled) adjoint state. As outlined above, this arises from the direct application of Theorem 2.1, which analyses features of the rescaled and reduced optimality system (2.9). The absence of an explicit control in a discretization of (2.9) then corresponds to assuming that the approximate control is given by $\Pi_K\big(-\frac{1}{\sqrt{\alpha}} C^* \tilde{z}\big)$. Therefore, we can combine Theorem 3.1 with suitable triangle inequalities to consider the complete optimality system (2.8) with the *augmented d_α-error*

$$d_\alpha(x, \tilde{x}) + \frac{M}{M_a} \|\tilde{q} - q\|_Q \, ,$$

where $\tilde{q}$ is some approximation of the exact control

$$q := \Pi_K\big(-\frac{1}{\sqrt{\alpha}} C^* z \big). \tag{3.4}$$

Of course, the bounds then involve an additional residual of the approximate control.

Corollary 3.3 (Bounding the augmented d_α-error – general case)**.** *In addition to the assumptions of Theorem 3.1 and* (3.4)*, let $\tilde{q} \in K$ be some approximate control. Then the augmented d_α-error of $(\tilde{u}, \tilde{z}, \tilde{q})$ is bounded from above and below by*

$$\begin{aligned} d_\alpha(x, \tilde{x}) + \frac{M}{M_a} \|q - \tilde{q}\|_Q \eqsim \Bigg(\|f - A\tilde{u} + C\tilde{q}\|_{2,*}^2 + \Big\| A^*\tilde{z} - \tfrac{1}{\sqrt{\alpha}} I^*(I\tilde{u} - u_d) \Big\|_{1,*}^2 \Bigg)^{\frac{1}{2}} \\ + \frac{M}{M_a} \left\| \tilde{q} - \Pi_K \left(-\tfrac{1}{\sqrt{\alpha}} C^* \tilde{z} \right) \right\|_Q , \end{aligned}$$

where the constants hidden in $\eqsim$ present the same asymptotics for $\alpha \to 0$ as those in Theorem 3.1.

Proof. We verify only the lower bound; the upper one follows from similar arguments upon noting $(\kappa M_a)/m_a \geq 1$. First notice that the triangle inequality ensures

$$\|f - A\tilde{u} + C\tilde{q}\|_{2,*} \leq \left\| f - A\tilde{u} + C\Pi_K \left(-\tfrac{1}{\sqrt{\alpha}} C^* \tilde{z} \right) \right\|_{2,*} + M \left\| \tilde{q} - \Pi_K \left(-\tfrac{1}{\sqrt{\alpha}} C^* \tilde{z} \right) \right\|_Q$$

and, combined with the definition (3.4) of q,

$$\begin{aligned} M \left\| \tilde{q} - \Pi_K \left(-\tfrac{1}{\sqrt{\alpha}} C^* \tilde{z} \right) \right\|_Q & \\ &\leq M \left\| \tilde{q} - q \right\|_Q + M \left\| \Pi_K \left(-\tfrac{1}{\sqrt{\alpha}} C^* \tilde{z} \right) - \Pi_K \left(-\tfrac{1}{\sqrt{\alpha}} C^* z \right) \right\|_Q \\ &\leq M \left\| \tilde{q} - q \right\|_Q + \frac{M}{\sqrt{\alpha}} \, \delta_\alpha(x, \tilde{x}) \leq M \left\| \tilde{q} - q \right\|_Q + M_a d_\alpha(x, \tilde{x}). \end{aligned}$$

Hence, the lower bound follows from the one in Theorem 3.1 without additional appearance of the parameter α. □

It is instructive to compare our approach based upon the reduced and rescaled optimality system with the previous a posteriori analysis in [17, 18].

Remark 3.4 (Comparison with [17, 18])**.** *We refer in particular to [18, Theorem 2.2], the proof of which presents the explicit dependencies between error terms and residual. Compared to our approach, there are two main differences regarding the error notion: First, in contrast to* (2.7)*, the adjoint state in [18] is not rescaled and the same holds true for the residual of the adjoint equation. Second, the error notion in [18] does not explicitly involve the observation error.*

For $\alpha \to 0$, the gap between upper and lower bound in [18, Theorem 2.2] grows like $O(1/\alpha^2)$ in general and reduces to $O(1/\alpha)$ if the residual of the approximate control vanishes as in the reduced optimality system. In contrast, the respective gaps in Corollary 3.3 and Theorem 3.1 are only $O(1/\sqrt{\alpha})$. This slower growth arises from the properties of the form b_α in Theorem 2.1 about the reduced and rescaled optimality system. The amplified growth in [18, Theorem 2.2] in the general case results from the fact that the absence of the observation error in the error notion has to be compensated by bounding in terms of the stability of the observation operator and the error of the adjoint state.

3.2. **Using compactness of control and observation.** In this subsection, we assume that both the control operator C and the observation operator I are bounded and compact. This situation is often encountered in applications; see Section 4 for two examples.

Let us first outline the idea how to take advantage of the compactness assumptions. To this end, we may split the operator of the reduced optimality system (2.9) as follows:

$$\begin{pmatrix} -\frac{1}{\sqrt{\alpha}} I^* I & A^* \\ A & -C\Pi_K(-\frac{1}{\sqrt{\alpha}} C^* \cdot) \end{pmatrix} \\ = \begin{pmatrix} 0 & A^* \\ A & 0 \end{pmatrix} + \begin{pmatrix} -\frac{1}{\sqrt{\alpha}} I^* I & 0 \\ 0 & -C\Pi_K(-\frac{1}{\sqrt{\alpha}} C^* \cdot) \end{pmatrix},$$

where the first term on the right-hand side is α-independent, while the second can be viewed as an α-dependent, compact perturbation. Since the first term is invertible, we can apply this operator splitting to the residual $\mathrm{Res}(\tilde{x})$ and translate it into a splitting of the error $x - \tilde{x}$. More precisely, applying the inverse of the first operator on the right-hand side to the residual, we define an auxiliary function $R\tilde{x} \in V$ by

$$R\tilde{x} - \tilde{x} = \begin{pmatrix} 0 & A^{-1} \\ A^{-*} & 0 \end{pmatrix} \mathrm{Res}(\tilde{x}) \tag{3.5}$$

and write

$$x - \tilde{x} = (x - R\tilde{x}) + (R\tilde{x} - \tilde{x}),$$

where Theorem 3.1 and (3.1) together with its 'adjoint' variant, respectively, imply

$$\frac{1}{M_a}\left\|\mathrm{Res}(R\tilde{x})\right\|_* \le d_\alpha(x,R\tilde{x}) \le \frac{\kappa}{m_a}\left\|\mathrm{Res}(R\tilde{x})\right\|_*, \tag{3.6}$$

$$\frac{1}{M_a}\left\|\mathrm{Res}(\tilde{x})\right\|_* \le \|\tilde{x}-R\tilde{x}\| \le \frac{1}{m_a}\left\|\mathrm{Res}(\tilde{x})\right\|_*. \tag{3.7}$$

Observe that (3.7) relies only on properties of the state equation and involves the known residual $\mathrm{Res}\,\tilde{x}$. In contrast, (3.6) involves κ and so the regularization parameter α as well as the unknown residual $\mathrm{Res}(R\tilde{x})$. However, applying the operator of the optimality system (2.9) to the decomposition of the error reveals

$$\mathrm{Res}(R\tilde{x}) = \begin{pmatrix} \frac{1}{\sqrt{\alpha}}I^*I & 0 \\ 0 & C\Pi_K(-\frac{1}{\sqrt{\alpha}}C^*\cdot) \end{pmatrix}(R\tilde{x}-\tilde{x}). \tag{3.8}$$

Therefore, we can bound the critical error $x - R\tilde{x}$ in terms of the error $\tilde{x} - R\tilde{x}$, which is accessible through the known residual $\mathrm{Res}(\tilde{x})$. More importantly, in view of the compactness properties of the operator matrix in (3.8), we may take a norm of $R\tilde{x} - \tilde{x}$ that is weaker than $\|\cdot\|$ and, thus, allows for faster convergence.

Let us put this *approach based on compactness* into the context of previous and related techniques.

Remark 3.5 (Connection with [18], elliptic reconstruction, Wheeler's trick, and Schatz's argument)**.** *The above construction of the auxiliary function $R\tilde{x}$ is also implicitly used in [18], but as simple perturbation argument without exploiting the compactness assumptions through* (3.8)*. The use with* (3.8) *is quite similar to the so-called elliptic reconstruction of [20, 22] in the context of parabolic equations. The error decomposition with the elliptic reconstruction can be viewed as the a posteriori counterpart of the error decomposition in Wheeler's trick [29] by means of the Ritz projection. Similarly, but incorporating the role of compactness, the above approach with $R\tilde{x}$ is the a posteriori counterpart of the a priori analysis in [10] based upon Schatz's argument [24].*

We proceed in variational terms and, in line with (3.5), define the auxiliary operator $R = (R_1, R_2) : V \to V$ for $\tilde{x} = (\tilde{u}, \tilde{z})$ by

$$\forall \varphi_2 \in V_2 \qquad a(R_1\tilde{x},\varphi_2) = \langle f,\varphi_2\rangle_2 + \left(\Pi_K\left(-\tfrac{1}{\sqrt{\alpha}}C^*\tilde{z}\right), C^*\varphi_2\right)_Q, \tag{3.9a}$$

$$\forall \varphi_1 \in V_1 \qquad a(\varphi_1,R_2\tilde{x}) = \tfrac{1}{\sqrt{\alpha}}\left(I\tilde{u}-u_d, I\varphi_1\right)_W. \tag{3.9b}$$

The crucial relationship (3.8) then amounts to

$$\langle \mathrm{Res}(R\tilde{x}),\varphi\rangle = b_\alpha(x,\varphi) - b_\alpha(R\tilde{x},\varphi) = c_\alpha(R\tilde{x},\varphi) - c_\alpha(x,\varphi), \tag{3.10}$$

whence the critical error part is bounded with the additional help of (3.6) by

$$\begin{aligned} \frac{m_a^2}{\kappa^2}\, d_\alpha(x,R\tilde{x})^2 &\le \|\,\mathrm{Res}(R\tilde{x})\|_*^2 \\ &= \left\| C\Pi_K\left(-\frac{C^*R_2\tilde{x}}{\sqrt{\alpha}}\right) - C\Pi_K\left(-\frac{C^*\tilde{z}}{\sqrt{\alpha}}\right)\right\|_{2,*}^2 + \left\|\frac{1}{\sqrt{\alpha}}I^*I(\tilde{u}-R_1\tilde{x})\right\|_{1,*}^2 \\ &=: \frac{M^2}{\alpha}\delta_\alpha^*(R\tilde{x},\tilde{x})^2. \end{aligned} \tag{3.11}$$

The approach based on compactness then leads to the following alternative to Theorem 3.1 of the direct approach.

Theorem 3.6 (Bounding the d_α-error – compact case)**.** *Let $x = (u,z) \in V$ be the solution to* (2.15) *and $\tilde{u} \in V_2$ be some approximate state and $\tilde{z} \in V_1$ some*

approximate rescaled adjoint state. Writing $\tilde{x} = (\tilde{u}, \tilde{z})$, their error is bounded from above and below as follows:

$$d_\alpha(x, \tilde{x}) \leq \frac{1}{m_a} \left(1 + \left(\kappa + \frac{m_a}{M_a}\right) \frac{M}{\sqrt{\alpha}} \frac{\delta_\alpha^*(R\tilde{x}, \tilde{x})}{\|\operatorname{Res}(\tilde{x})\|_*}\right) \|\operatorname{Res}(\tilde{x})\|_*$$

and

$$\frac{1}{M_a} \|\operatorname{Res}(\tilde{x})\|_* \leq d_\alpha(x, \tilde{x}).$$

Proof. The lower bound is a repetition from Theorem 3.1 and it remains to show the upper bound. The triangle inequality and the definition (2.21) of d_α yield

$$\begin{aligned} d_\alpha(x, \tilde{x}) &\leq d_\alpha(x, R\tilde{x}) + d_\alpha(R\tilde{x}, \tilde{x}) \\ &= d_\alpha(x, R\tilde{x}) + \|R\tilde{x} - \tilde{x}\| + \frac{M}{M_a} \frac{1}{\sqrt{\alpha}} \delta_\alpha^*(R\tilde{x}, \tilde{x}). \end{aligned}$$

Consequently, (3.11) and (3.7) ensure the upper bound and the proof is finished. □

The following three remarks elucidate why Theorem 3.6 is a valid alternative to Theorem 3.1 of the direct approach.

Remark 3.7 (Using compactness)**.** *The definition* (3.11) *of $\delta_\alpha^*(R\tilde{x}, \tilde{x})$ offers all available compactness, given by the operators C, I and their adjoints. Recall that this compactness will allow to choose a norm for $R\tilde{x} - \tilde{x}$ that is weaker than $\|\cdot\|$ and enjoys faster convergence. Such a faster convergence is usually ensured with the help of a duality argument depending on regularity properties of the state and adjoint equation as well as on approximation properties of the discretization providing the discrete solution. In particular, it may happen that if we insert the bound*

$$\delta_\alpha^*(R\tilde{x}, \tilde{x}) \leq \delta_\alpha(R\tilde{x}, \tilde{x})$$

in the upper bound of Theorem 3.6 and exploit only the compactness in the latter term, the faster convergence of $\delta_\alpha^*(R\tilde{x}, \tilde{x})$ is already fully captured. See also Remark 4.5 below.

Remark 3.8 (Gap in bounding d_α-error – compact case)**.** *If we do not exploit the compactness assumptions by inserting the bound*

$$\delta_\alpha^*(R\tilde{x}, \tilde{x}) \leq \delta_\alpha(R\tilde{x}, \tilde{x}) \leq M\|\tilde{x} - R\tilde{x}\|$$

in the upper bound of Theorem 3.6, the gap between its upper and lower bound is $O(\kappa/\sqrt{\alpha}) = O(1/\alpha)$ as $\alpha \to 0$; compare with Remark 3.2. This illustrates that the upper bound of Theorem 3.6 improves on the one in Theorem 3.1 only if $\delta_\alpha^*(R\tilde{x}, \tilde{x})$ is relatively small with respect to $\|\operatorname{Res}(\tilde{x})\|$. The significance of Theorem 3.6 thus hinges on the aforementioned effect of the compactness assumptions. Indeed, if

$$\frac{\delta_\alpha^*(R\tilde{x}, \tilde{x})}{\|\tilde{x} - R\tilde{x}\|} \to 0 \qquad \text{for } \tilde{x} \to x \text{ and fixed } \alpha > 0,$$

the gap converges to M_a/m_a, viz. the gap in the error-residual relation (3.1) *for the state equation. Finally, let us consider the situation when the Tikhonov regularization α varies and suppose that we can construct $\tilde{x}$ such that*

$$\frac{\delta_\alpha^*(R\tilde{x}, \tilde{x})}{\|\operatorname{Res}(\tilde{x})\|_*} \lesssim \alpha. \tag{3.12}$$

Then the gap is $O(1)$, uniformly in α.

Remark 3.9 (Compactness and lower bound for d_α-error)**.** *The lower bound in Theorem 3.6 does not exploit the compactness assumptions on control and observation. Note that it does not dependent on α and its form already coincides with the lower bound in* (3.1) *for the state equation. These two properties suggest that*

it cannot be improved by involving terms like δ_α^*. Numerical results corroborate this suspicion; see Figure 2 (A).

3.3. **Using compactness of unconstrained control and observation.** In this subsection, we assume, in addition to the compactness of the control operator C and the observation operator I, that there are no constraints for the control, i.e. we have $K = Q$ and therefore $\Pi_K = \mathrm{id}_Q$.

In view of these assumptions, the abstract optimal control problem (2.3) is quadratic and the reduced and rescaled optimality system (2.15) is linear. Correspondingly, the form

$$b_\alpha(v,\varphi) = a(v_1,\varphi_2) + a(\varphi_1,v_2) + \tfrac{1}{\sqrt{\alpha}}\,(C^*v_2, C^*\varphi_2)_Q - \tfrac{1}{\sqrt{\alpha}}\,(Iv_1, I\varphi_1)_W$$

is bilinear and symmetric and

$$\|v-w\|_\alpha := d_\alpha(v,w) = \|v-w\| + \frac{1}{\sqrt{\alpha}}\frac{M}{M_a}|v-w| \tag{3.13}$$

is a norm, where $|\cdot|$ is given in (2.16). Hence, combing Theorem 2.1 and the inf-sup duality (2.14) for b_α instead of $\mathbf{a}$, we obtain

$$\frac{1}{M_a}\,\|\mathrm{Res}(R\tilde{x})\|_{\alpha,*} \le \|x-R\tilde{x}\| \le \frac{\kappa}{m_a}\,\|\mathrm{Res}(R\tilde{x})\|_{\alpha,*} \tag{3.14}$$

with

$$\|\mathrm{Res}(R\tilde{x})\|_{\alpha,*} := \sup_{\varphi\in V\setminus\{0\}} \frac{\langle \mathrm{Res}(R\tilde{x}),\varphi\rangle}{\|\varphi\|_\alpha}.$$

Notice that, in contrast to (3.6), this equivalence involves like (3.7) the norm $\|\cdot\|$, which does not depend on α. This fact leads to the following alternative with $\|\cdot\|$ as error norm of Theorem 3.6.

Theorem 3.10 (Bounding the $\|\cdot\|$-error – compact and unconstrained case)**.** *Let $x=(u,z)\in V$ be the solution to* (2.15) *and $\tilde{u}\in V_2$ be some approximate state and $\tilde{z}\in V_1$ some approximate rescaled adjoint state. Writing $\tilde{x}=(\tilde{u},\tilde{z})$, their error in the α-independent norm $\|\cdot\|$ is bounded from above and below as follows:*

$$\|x-\tilde{x}\| \le \frac{1}{m_a}\left(1+\kappa\,\frac{M_a}{M}\,\frac{|\tilde{x}-R\tilde{x}|}{\|\,\mathrm{Res}(\tilde{x})\|_*}\right)\|\,\mathrm{Res}(\tilde{x})\|_*$$

and

$$\frac{1}{M_a}\max\left\{\frac{M_a\sqrt{\alpha}}{M_a\sqrt{\alpha}+M},\, 1-\kappa\,\frac{M_a}{m_a}\frac{M_a}{M}\,\frac{|\tilde{x}-R\tilde{x}|}{\|\,\mathrm{Res}(\tilde{x})\|_*}\right\}\,\|\,\mathrm{Res}(\tilde{x})\|_* \le \|x-\tilde{x}\|.$$

Proof. To verify the upper bound, we start by applying the triangle inequality

$$\|x-\tilde{x}\| \le \|x-R\tilde{x}\| + \|R\tilde{x}-\tilde{x}\|.$$

For the second term on the right-hand side, we use (3.7) as in the proof of Theorem 3.6. For the first term, we employ (3.14) and (3.10) to obtain

$$\frac{m_a}{\kappa}\,\|x-R\tilde{x}\| \le \|\mathrm{Res}(R\tilde{x})\|_{\alpha,*} = \sup_{\varphi\in V}\frac{c_\alpha(R\tilde{x}-\tilde{x},\varphi)}{\|\varphi\|_\alpha} \le \frac{M_a}{M}|\tilde{x}-R\tilde{x}|. \tag{3.15}$$

This alternative to (3.11) concludes the proof of the upper bound.

We turn to the lower bound. On the one hand, Theorem 3.1 from the direct approach and the definition (3.13) of $\|\cdot\|_\alpha$ imply

$$\frac{1}{M_a}\|\,\mathrm{Res}(\tilde{x})\|_* \le \|x-\tilde{x}\|_\alpha \le \left(1+\frac{M}{M_a}\frac{1}{\sqrt{\alpha}}\right)\|x-\tilde{x}\|.$$

This establishes the first option of the max. On the other hand, starting with (3.7), applying a triangle inequality and then (3.15), we can deduce

$$\begin{aligned}\frac{1}{M_a}\|\operatorname{Res}(\tilde{x})\|_* &\le \|R\tilde{x}-\tilde{x}\| \le \|x-\tilde{x}\| + \|x-R\tilde{x}\| \\ &\le \|x-\tilde{x}\| + \frac{\kappa}{M}\frac{M_a}{m_a}|\tilde{x}-R\tilde{x}|.\end{aligned}$$

This shows the second option of the max and the proof is finished. □

Remark 3.11 (Gap in bounding the $\|\cdot\|$-error – compact and unconstrained case)**.** *If we do not exploit the compactness by inserting the bound $|\tilde{x}-R\tilde{x}| \le M\,\|\tilde{x}-R\tilde{x}\|$ in the bounds of Theorem 3.10, only the first option in the* max *of the lower bound applies and the gap is $O(\kappa/\sqrt{\alpha}) = O(1/\alpha)$ as $\alpha \to 0$, the same order as in the corresponding case of Remark 3.8. Recall however that Theorem 3.10 considers an error notion that is independent of α.*

Similarly to Remark 3.8, if the compactness assumptions lead to

$$\frac{|\tilde{x}-R\tilde{x}|}{\|\tilde{x}-R\tilde{x}\|} \to 0 \qquad \text{for } \tilde{x} \to x \text{ and fixed } \alpha > 0,$$

the gap between upper and lower bound converges to M_a/m_a, viz. the gap in the error-residual relation (3.1) *for the state equation. Finally, if the Tikhonov regularization α varies and suppose that we can construct $\tilde{x}$ such that*

$$\frac{|\tilde{x}-R\tilde{x}|}{\|\operatorname{Res}(\tilde{x})\|_*} \lesssim \sqrt{\alpha}, \tag{3.16}$$

then the gap is $O(1)$, uniformly in α. This uses the second option in the max *of the lower bound in Theorem 3.10 and condition* (3.16) *can be less restrictive than its counterpart* (3.12) *in Remark 3.8.*

4. Applications to optimal control problems

This section proceeds with the a posteriori analysis for the abstract optimization problem (2.3). Recall that the remaining tasks are to split the error bounds of Section 3 into local contributions and, then, to split the latter into computable and oscillatory parts. To this end, one can apply

- classical techniques, see, e.g., [1, 28], leading to an equivalence of error and estimator up to so-called oscillation, or
- more recent techniques, see [19, Sections 3 and 4], [6, Section 4], leading to a strict equivalence.

In both cases, various estimator types can be chosen for the computable parts. In any case, these tasks are specific to the functional setting of the state equation, the control and observation operators and the discretization. We therefore exemplify these tasks by considering finite element discretizations of optimal control problems with distributed and boundary control. Doing so, we focus on the possible interplay between $\|\operatorname{Res}(\tilde{x})\|_*$ and $\delta_\alpha^*(R\tilde{x},\tilde{x})$, $\delta_\alpha(R\tilde{x},\tilde{x})$ or $|R\tilde{x}-\tilde{x}|$. Moreover, we numerically study the behavior of the derived a posteriori bounds. The adaptive simulations where carried out in the DUNE framework [3].

In what follows, we shall use the following notation. For a Lebesgue measurable set $\omega \subset \mathbb{R}^d$, we denote by $L^2(\omega)$ the space of square integrable functions on ω and define its norm by $\|\cdot\|_{L^2(\omega)}^2 := \int_\omega |\cdot|^2$. The space of functions having also weak first derivatives in $L^2(\omega)$ is denoted by $H^1(\omega)$ with norm defined by $\|\cdot\|_{H^1(\omega)}^2 := \|\nabla\cdot\|_{L^2(\omega)}^2 + \|\cdot\|_{L^2(\omega)}^2$ and $\mathring{H}^1(\omega)$ is its closed subspace of functions with zero trace. Recalling Poincaré's inequality $\|v\|_{L^2(\omega)} \le C_P\|\nabla v\|_{L^2(\omega)}$, $v \in \mathring{H}^1(\omega)$, an alternative norm on $\mathring{H}^1(\omega)$ is given by $\|\nabla\cdot\|_{L^2(\omega)}$. The dual space of $\mathring{H}^1(\omega)$ is denoted by

$H^{-1}(\omega)$ and equipped with the operator norm $\|\cdot\|_{H^{-1}(\omega)}$. By virtue of the Riesz map, $L^2(\omega)$ is identified with its dual space.

4.1. **Distributed control.** In this section, we show how the a posteriori bounds from Section 3 are applied to optimal control problems with possibly constrained distributed control.

Let $\Omega \subset \mathbb{R}^d$, $d > 1$, be a domain with polyhedral boundary Γ. For a subdomain $\Omega_{\mathsf{Q}} \subseteq \Omega$ and some bounds $a \in \mathbb{R} \cup \{-\infty\}$ and $b \in \mathbb{R} \cup \{\infty\}$ with $a < b$, let

$$K = \{q \in L^2(\Omega_{\mathsf{Q}}) \mid a \le q \le b \text{ a.e. in } \Omega_{\mathsf{Q}}\} \tag{4.1}$$

be the set of admissible controls. For a target function $u_d \in L^2(\Omega_{\mathsf{W}})$ supported in the subdomain $\Omega_{\mathsf{W}} \subseteq \Omega$ and cost parameter $\alpha > 0$, we consider

$$\begin{aligned} &\min_{(q,u)\in K\times \mathring{H}^1(\Omega)} \frac{1}{2}\|u-u_d\|^2_{L^2(\Omega_{\mathsf{W}})} + \frac{\alpha}{2}\|q\|^2_{L^2(\Omega_{\mathsf{Q}})} \\ &\text{subject to} \quad -\Delta u = f + q\chi_{\Omega_{\mathsf{Q}}} \text{ in } \Omega \quad \text{and} \quad u = 0 \text{ on } \Gamma. \end{aligned} \tag{4.2}$$

Here, $\chi_{\Omega_{\mathsf{Q}}}$ denotes the indicator function on Ω_{Q} and $q\chi_{\Omega_{\mathsf{Q}}}$ is considered the zero extension of $q \in L^2(\Omega_{\mathsf{Q}})$ to Ω. This problem fits into the framework of Section 2 with the Hilbert spaces

$$\begin{aligned} V_1 = V_2 = \mathring{H}^1(\Omega), &\quad (v_1,v_2)_{V_i} = (\nabla v_1, \nabla v_2)_{L^2(\Omega)}, \quad i = 1,2, \\ Q = L^2(\Omega_{\mathsf{Q}}), &\quad (q_1,q_2)_Q = (q_1,q_2)_{L^2(\Omega_{\mathsf{Q}})}, \\ W = L^2(\Omega_{\mathsf{W}}), &\quad (w_1,w_2)_W = (w_1,w_2)_{L^2(\Omega_{\mathsf{W}})}. \end{aligned}$$

The remaining ingredients are given by (4.1),

$$\begin{aligned} a(v,\varphi) = \int_\Omega \nabla v\cdot\nabla\varphi = (\nabla v,\nabla\varphi)_{L^2(\Omega)}, &\qquad m_a = 1 = M_a, \\ \langle Cq,\varphi\rangle = \int_{\Omega_{\mathsf{Q}}} q\varphi = (q,\varphi_{|\Omega_{\mathsf{Q}}})_{L^2(\Omega_{\mathsf{Q}})} = (q,C^*\varphi)_{L^2(\Omega_{\mathsf{Q}})}, &\qquad M = C_P, \\ Iv = v_{|\Omega_{\mathsf{W}}}, \qquad \langle I^*w,v\rangle = \int_{\Omega_{\mathsf{W}}} vw = (v_{|\Omega_{\mathsf{W}}},w)_{L^2(\Omega_{\mathsf{W}})}, & \\ \Pi_K q = \min\{\max\{q,\, a\},\, b\} \quad \text{a.e. in } \Omega_Q, & \end{aligned}$$

where $\langle\cdot,\cdot\rangle$ denotes the duality pairing in $H^{-1}(\Omega)\times\mathring{H}^1(\Omega)$.

The variational formulation of the reduced and rescaled optimality system (2.10) reads: find $(u,z)\in\mathring{H}^1(\Omega)\times\mathring{H}^1(\Omega)$ such that

$$\forall\varphi_1\in\mathring{H}^1(\Omega) \qquad \int_\Omega \nabla\varphi_1\cdot\nabla z - \tfrac{1}{\sqrt{\alpha}}\int_{\Omega_{\mathsf{W}}} u\varphi_1 = -\tfrac{1}{\sqrt{\alpha}}\int_{\Omega_{\mathsf{W}}} u_d\varphi_1, \tag{4.3a}$$

$$\forall\varphi_2\in\mathring{H}^1(\Omega) \qquad \int_\Omega \nabla u\cdot\nabla\varphi_2 - \int_{\Omega_{\mathsf{Q}}} \Pi_K(-\tfrac{1}{\sqrt{\alpha}}z)\varphi_2 = \langle f,\varphi_2\rangle\,. \tag{4.3b}$$

For its discretization, we use Lagrange finite elements. To this end, let $\mathcal{M}$ be a simplicial face-to-face (conforming) mesh of the domain Ω. Denoting by $\mathcal{V}$ the vertices of $\mathcal{M}$, we define the star around a vertex $z\in\mathcal{V}$ by

$$\omega_z := \bigcup\{K\in\mathcal{M}\mid z\in K\} \quad \text{with diameter} \quad h_z = \operatorname{diam}(\omega_z).$$

The discrete spaces associated with Lagrange finite elements of degree $\ell > 0$ are then given by

$$S^1_\ell(\mathcal{M}) := \big\{v\in H^1(\Omega)\mid v_{|K}\in\mathbb{P}_\ell(K), \forall K\in\mathcal{M}\big\}$$

and

$$\mathring{S}^1_\ell(\mathcal{M}) := S^1_\ell(\mathcal{M})\cap\mathring{H}^1(\Omega),$$

where $\mathbb{P}_\ell(K)$ denotes the set of polynomials up to degree ℓ on K. The variational discretization of (4.3) then reads as follows: find $(U, Z) \in \mathring{S}^1_\ell(\mathcal{M}) \times \mathring{S}^1_\ell(\mathcal{M})$ such that

$$\forall \Phi_1 \in \mathring{S}^1_\ell(\mathcal{M}) \qquad \int_\Omega \nabla\Phi_1 \cdot \nabla Z - \tfrac{1}{\sqrt{\alpha}} \int_{\Omega_\mathsf{W}} U\Phi_1 = -\tfrac{1}{\sqrt{\alpha}} \int_{\Omega_\mathsf{W}} u_d\Phi_1, \tag{4.4a}$$

$$\forall \Phi_2 \in \mathring{S}^1_\ell(\mathcal{M}) \qquad \int_\Omega \nabla U \cdot \nabla\Phi_2 - \int_{\Omega_\mathsf{Q}} \Pi_K(-\tfrac{1}{\sqrt{\alpha}} Z)\Phi_2 = \langle f, \Phi_2\rangle\,. \tag{4.4b}$$

Consequently, in the solution $X = (U, Z)$ of (4.4), the residual

$$\begin{aligned} \big\langle \operatorname{Res}(X),\, (\varphi_1, \varphi_2)\big\rangle := \langle f, \varphi_2\rangle_{\mathring{H}^1(\Omega)} - \int_\Omega \nabla U \cdot \nabla\varphi_2 + \int_{\Omega_\mathsf{Q}} \Pi_K(-\tfrac{1}{\sqrt{\alpha}} Z)\varphi_2 \\ - \tfrac{1}{\sqrt{\alpha}} \int_{\Omega_\mathsf{W}} u_d\varphi_1 - \int_\Omega \nabla\varphi_1 \cdot \nabla Z + \tfrac{1}{\sqrt{\alpha}} \int_{\Omega_\mathsf{W}} U\varphi_1, \end{aligned} \tag{4.5}$$

for $(\varphi_1, \varphi_2) \in \mathring{H}^1(\Omega) \times \mathring{H}^1(\Omega)$, satisfies the orthogonality condition

$$\langle \operatorname{Res}(X),\, \Phi\rangle = 0 \qquad \text{for all } \Phi \in \mathring{S}^1_\ell(\mathcal{M}) \times \mathring{S}^1_\ell(\mathcal{M}). \tag{4.6}$$

Using standard arguments, see, e.g., [19, Lemma 4], we can split the residual norm into local contributions such that

$$\begin{aligned} \frac{1}{d+1} \sum_{z\in\mathcal{V}} \|\operatorname{Res}(X)\|^2_{H^{-1}(\omega_z)} &\le \|\operatorname{Res}(X)\|^2_{H^{-1}(\Omega)} \\ &\le C_\mathcal{M} \sum_{z\in\mathcal{V}} \|\operatorname{Res}(X)\|^2_{H^{-1}(\omega_z)}, \end{aligned} \tag{4.7}$$

where the constant $C_\mathcal{M}$ only depends on the shape regularity of the mesh $\mathcal{M}$. Notice that each contribution $\|\operatorname{Res}(X)\|^2_{H^{-1}(\omega_z)}$, $z \in \mathcal{V}$, is a local quantity once the finite element solution $X = (U, Z)$ from (4.4) is available by means of a global solve. Combining this 'localization' with Theorem 3.1 of the direct approach readily provides the following a posteriori bounds.

Theorem 4.1 (Bounding d_α-error for distributed control – general case)**.** *Let $x = (u, z)$ be the exact states of the optimal control problem* (4.2)*, where the adjoint state is rescaled, cf.* (4.3)*. Furthermore, let $X = (U, Z)$ be their finite element approximations from* (4.4)*. Then, we have for the residual defined in* (4.5) *the equivalence*

$$\frac{1}{(d+1)} \sum_{z\in\mathcal{V}} \|\operatorname{Res}(\tilde{x})\|^2_{H^{-1}(\omega_z)} \le d_\alpha(x, X)^2 \le \kappa\, C_\mathcal{M} \sum_{z\in\mathcal{V}} \|\operatorname{Res}(X)\|^2_{H^{-1}(\omega_z)},$$

where the constant $C_\mathcal{M}$ depends on the shape regularity of the mesh and κ is defined in (2.23)*.*

Thanks to the compact embedding $\mathring{H}^1(\Omega) \subset L^2(\Omega)$, the operators C^* and I are compact. This allows applying the results of the compact approach in Section 3.2. To this end, we use the Lipschitz continuity of Π_K with Lipschitz constant 1 to deduce that

$$\delta^*_\alpha(RX, X)^2 \le \delta_\alpha(RX, X)^2 \le \|R_2X - Z\|^2_{L^2(\Omega)} + \|R_1X - U\|^2_{L^2(\Omega)}\,, \tag{4.8}$$

where $R = (R_1, R_2) : \mathring{S}^1_\ell(\mathcal{M}) \times \mathring{S}^1_\ell(\mathcal{M}) \to \mathring{H}^1(\Omega) \times \mathring{H}^1(\Omega)$ is the auxiliary operator defined in (3.9). Combining the definitions of the auxiliary operator R and the finite element solution X, we find the orthogonality relationships

$$\int_\Omega \nabla(R_1X - U) \cdot \nabla\Phi_2 = 0 = \int_\Omega \nabla(R_2X - U) \cdot \nabla\Phi_1 \quad \forall \Phi_1, \Phi_2 \in \mathring{S}^1_\ell(\mathcal{M}) \tag{4.9}$$

In other words, U and Z are, respectively, the Ritz projections in $\mathring{S}^1_\ell(\mathcal{M})$ of R_1X and R_2X with respect to the bilinear form $(\nabla\cdot,\nabla\cdot)_{L^2(\Omega)}$. Taking into account also the orthogonality (4.6) of $\mathrm{Res}(X)$, we can thus use a well-known argument, see, e.g., [1, Section 2.4], based upon duality and regularity, to bound the L^2-errors in (4.8) in terms of the residual $\mathrm{Res}(X)$. For simplicity, we shall assume that the domain Ω is convex and resort to the following well-known H^2-regularity result for the Poisson problem; compare with [13, (3,1,2,2) and Lemma 3.2.1.2].

Proposition 4.2 (Extra regularity for distributed control)**.** *Let $\Omega\subset\mathbb{R}^d$ be a convex domain. For any source $g\in L^2(\Omega)$, the unique solution $v_g\in\mathring{H}^1(\Omega)$ of the Poisson problem*

$$\forall v\in\mathring{H}^1(\Omega)\qquad \int_\Omega \nabla v_g\cdot\nabla v=\int_\Omega gv$$

satisfies

$$v_g\in H^2(\Omega)\qquad\text{and}\qquad |v_g|_{H^2(\Omega)}\le\|g\|_{L^2(\Omega)},$$

where $|\cdot|_{H^2(\Omega)}$ denotes the $H^2(\Omega)$-seminorm.

Using Proposition 4.2 in the cited duality argument then leads to the following a posteriori upper bound.

Lemma 4.3 (Upper bound for compact error – distributed control)**.** *Let $\Omega\subset\mathbb{R}^d$ be a convex polyhedral domain. The L^2-errors in* (4.8) *are bounded in terms of the residual of $X=(U,Z)$:*

$$\|R_2X-Z\|^2_{L^2(\Omega)}+\|R_1X-U\|^2_{L^2(\Omega)}\le C^2_{\mathcal{M}}\sum_{z\in\mathcal{V}}h_z^2\|\,\mathrm{Res}(X)\|^2_{H^{-1}(\omega_z)},$$

where $C_{\mathcal{M}}$ is the constant from (4.7)*.*

Proof. We sketch the proof only for the second term $\|R_1X-U\|^2_{L^2(\Omega)}$; the same argument applies also to the first term. According to Proposition 4.2, there is $\psi\in H^2(\Omega)\cap\mathring{H}^1(\Omega)$ with

$$-\Delta\psi=R_1X-U\quad\text{in }\Omega\qquad\text{and}\qquad |\psi|_{H^2(\Omega)}\le\|R_1X-U\|_{L^2(\Omega)}. \tag{4.10}$$

We denote by $\mathcal{I}_{\mathsf{sz}}:\mathring{H}^1(\Omega)\to\mathring{S}^1_\ell(\mathcal{M})$ the Scott-Zhang quasi-interpolation operator [25]. Thanks to the definition of R and the orthogonality (4.6) of the residual, we deduce

$$\begin{aligned}\|R_1X-U\|^2_{L^2(\Omega)}&=\int_\Omega\nabla(R_1X-U)\cdot\nabla\psi=\langle\mathrm{Res}_2(X),\psi\rangle\\&=\langle\mathrm{Res}_2(X),\psi-\mathcal{I}_{\mathsf{sz}}\psi\rangle=\sum_{z\in\mathcal{V}}\langle\mathrm{Res}_2(X),(\psi-\mathcal{I}_{\mathsf{sz}}\psi)\phi_z\rangle\\&\le\sum_{z\in\mathcal{V}}\|\,\mathrm{Res}_2(X)\|_{H^{-1}(\omega_z)}\|\nabla((\psi-\mathcal{I}_{\mathsf{sz}}\psi)\phi_z)\|_{L^2(\omega_z)},\end{aligned}$$

where we have used for the last equality that the Lagrange basis functions ϕ_z, $z\in\mathcal{V}$, of $S^1_1(\mathcal{M})$ form a partition of unity and that $\operatorname{supp}\phi_z=\omega_z$, $z\in\mathcal{V}$. In view of $\|\phi_z\|_{L^\infty(\omega_z)}=1$ and $\|\phi_z\|_{L^\infty(\omega_z)}\le C_{\mathcal{M}}h_z^{-1}$, standard interpolation estimates imply

$$\begin{aligned}&\|\nabla((\psi-\mathcal{I}_{\mathsf{sz}}\psi)\phi_z)\|_{L^2(\omega_z)}\\&\qquad\le\|\nabla(\psi-\mathcal{I}_{\mathsf{sz}}\psi)\|_{L^2(\omega_z)}+\|\nabla\phi_z\|_{L^\infty(\omega_z)}\|\psi-\mathcal{I}_{\mathsf{sz}}\psi\|_{L^2(\omega_z)}\\&\qquad\le C_{\mathcal{M}}h_z|\psi|_{H^2(\tilde\omega_z)}.\end{aligned}\tag{4.11}$$

Note that the constant $C_{\mathcal{M}}$ may vary from occurence to occurence but each time only depends on the shape regularity of $\mathcal{M}$. Here the domains $\tilde\omega_z=\bigcup_{y\in\mathcal{V}\cap\omega_z}\omega_y$ are

neighbourhoods of ω_z that only overlap finitely often depending on the regularity of $\mathcal{M}$. This together with (4.10) implies

$$\|R_1X - U\|^2_{L^2(\Omega)} \leq C_{\mathcal{M}} \left(\sum_{z\in\mathcal{V}} h_z^2 \|\operatorname{Res}_2(X)\|^2_{H^{-1}(\omega_z)} \right)^{\frac{1}{2}} \|R_1X - U\|_{L^2(\Omega)}. \qquad \square$$

Inserting Lemma 4.3 as well as the localization (4.7) into Theorem 3.6, we obtain the following alternative to Theorem 4.1.

Theorem 4.4 (Distributed control – compact case)**.** *Suppose in addition to the setting of Theorem 4.1 that $\Omega \subset \mathbb{R}^d$ is convex. Then we have*

$$\begin{aligned}
&\frac{1}{\sqrt{d+1}} \left(\sum_{z\in\mathcal{V}} \|\operatorname{Res}(X)\|^2_{H^{-1}(\omega_z)} \right)^{\frac{1}{2}} \leq d_\alpha(x, X) \\
&\leq C \left(1 + \frac{\kappa}{\sqrt{\alpha}} \left(\frac{\sum_{z\in\mathcal{V}} h_z^2 \|\operatorname{Res}(X)\|^2_{H^{-1}(\omega_z)}}{\sum_{z\in\mathcal{V}} \|\operatorname{Res}(X)\|^2_{H^{-1}(\omega_z)}} \right)^{\frac{1}{2}} \right) \left(\sum_{z\in\mathcal{V}} \|\operatorname{Res}(X)\|^2_{H^{-1}(\omega_z)} \right)^{\frac{1}{2}}.
\end{aligned}$$

The constant C depends on the Poincaré constant C_P and the shape regularity of the mesh $\mathcal{M}$; κ is defined in (2.23)*.*

Remark 4.5 (Limitations in exploiting compactness)**.** *In Theorem 4.4 we have exploited the compactness of C^* and I to obtain the accelerating factors h_z in front of the local contributions $\|\operatorname{Res}(X)\|_{H^{-1}(\omega_z)}$, $z \in \mathcal{V}$. Notice that the use of* (4.8) *entails that Theorem 4.4 does not exploit the compactness of C and I^* associated with the embedding $L^2(\Omega) \subset H^{-1}(\Omega)$. Thus, the question arises whether the upper bound in Theorem 4.4 can be improved by directly bounding the potentially smaller quantity $\delta^*_\alpha(RX, X)$. The line of argument allows for such an improvement in principle, but hinges on the combination of regularity properties for the state equation and its adjoint as well as on the order ℓ of their finite element solutions. The former obstructs an improvement of Theorem 4.4 in the case at hand.*

To illustrate this, let us consider the case with $\Omega = \Omega_{\mathsf{Q}} = \Omega_{\mathsf{W}}$ and $a = -\infty$ and $b = \infty$, i.e., $\Pi_K = \mathrm{id}$, leading to

$$M^2 \delta^*_\alpha(RX, R)^2 = \|R_2X - Z\|^2_{H^{-1}(\Omega)} + \|R_1X - U\|^2_{H^{-1}(\Omega)}.$$

As in the proof of Lemma 4.3, let us focus on the second term on the right-hand side. In view of

$$\|R_1X - U\|_{H^{-1}(\Omega)} = \sup_{\varphi\in\mathring{H}^1(\Omega)} \frac{\langle R_1X - U, \varphi\rangle}{\|\nabla\varphi\|_{L^2(\Omega)}},$$

we consider

$$-\Delta\psi = \varphi \in \mathring{H}^1(\Omega) \text{ in } \Omega, \qquad \psi = 0 \text{ on } \partial\Omega.$$

If we had $\psi \in H^3(\Omega)$ with $|\psi|_{H^3(\Omega)} \leq C\|\nabla\varphi\|_{L^2(\Omega)}$ and $\ell > 1$, then minor modifications in the proof of Lemma 4.3 would imply

$$\|R_1X - U\|^2_{H^{-1}(\Omega)} \leq C^2_{\mathcal{M}} \sum_{z\in\mathcal{V}} h_z^4 \|\operatorname{Res}(X)\|^2_{H^{-1}(\omega_z)}.$$

However, the supposed regularity theorem is not true for polyhedral domains or would not be useful in the case $\ell = 1$ of linear finite elements. As an alternative, one could invoke also more sophisticated regularity theorems with weights. We do not consider this option here for simplicity.

For the numerical comparison of bounds as in Theorem 4.1 and Theorem 4.4, we consider

$$\min_{(q,u)\in K\times H^1(\Omega)} \frac{1}{2}\,\|u-u_d\|^2_{L^2(\Omega_{\mathsf{W}})} + \int_\Omega g_1 u + \frac{\alpha}{2}\,\|q\|^2_{L^2(\Omega_{\mathsf{Q}})} \quad \text{subject to} \quad -\Delta u = f+q \text{ in } \Omega \quad \text{and} \quad u = g_2 \text{ on } \partial\Omega, \tag{4.12}$$

where the domains Ω, Ω_{W}, Ω_{Q} as in Figure 1, $K = \{q \in L^2(\Omega) \mid -1 \le q \le 1\}$, $u = 3r^{\frac{4}{3}}\sin(\frac{4}{3}\theta)$, $z = 4(y-y^2)(1-x)(x+y)$, $q = \chi_{\Omega_{\mathsf{Q}}}\Pi_{[-1,1]}(\frac{-1}{\sqrt{\alpha}}z)$, $f = -q$, $u_d = u + \Delta z$, $g_2 = u$, and $g_1 = \chi_{\Omega_{\mathsf{Q}}}\sqrt{\alpha}z$.

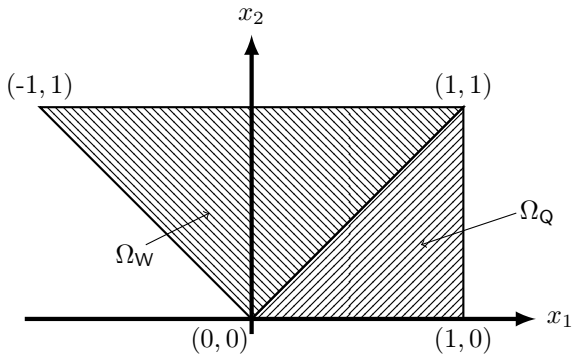


FIGURE 1. Domain Ω and subdomains $\Omega_{\mathsf{W}}, \Omega_{\mathsf{Q}}$ for Example (4.12).

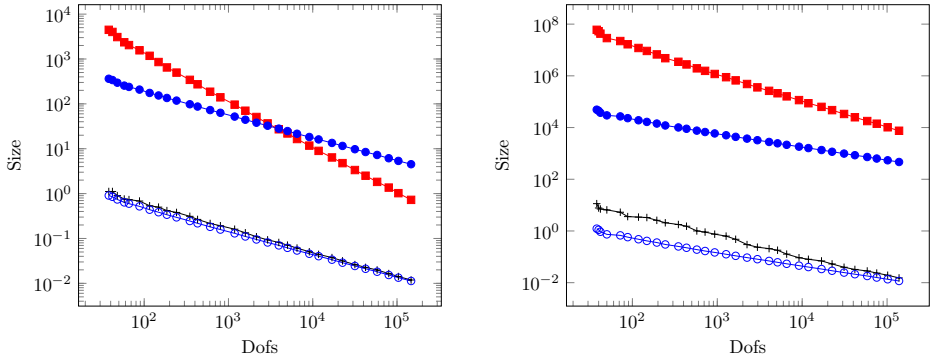
The numerical simulations are carried out with linear elements. Adaptive mesh refinement is driven by a standard residual error estimator, see, e.g., [28], that quantifies the local residual norms in Theorem 4.1. The estimator is scaled such that it coincides with the error for large $\alpha = 10^6$ and a fine, adaptive grid, providing a benchmark close to the situation of a pure Poisson problem. To mark elements for refinement, Dörfler's strategy [9] is used with parameter 0.6.

Figure 2 displays the d_α-error and the bounds in Theorem 4.1 and Theorem 4.4 using compactness, while Figure 3 gives an idea of the underlying adaptive mesh refinement. First, let us observe that the error may or may not be close to the α-independent lower bound on coarse meshes. Next, in line with the Remarks 3.2 and 3.8, we see that the upper bound using compactness is worse than the general one from Theorem 4.1 on coarse meshes, but can provide a much smaller gap on fine meshes; see part (A) with $\alpha = 10^{-4}$. However, this improvement hinges on the relationship of α and the available computational resources; see part (B) with $\alpha = 10^{-8}$.

We turn to applying the results of Section 3.3 and suppose that there are no control constraints, i.e. $a = -\infty$ and $b = \infty$. Theorem 3.10, combined with Lemma 4.3 and the localization (4.7), immediately yields the following a posteriori bounds for the combined $\mathring{H}^1(\Omega)$-error of the states.

Theorem 4.6 (Bounding the $\|\cdot\|$-error for distributed control)**.** *Suppose in addition to the setting in Theorem 4.4 that no control constraints apply. Then we have*

$$\max\left\{\frac{\sqrt{\alpha}}{\sqrt{\alpha}+C_P}, 1 - C\kappa\, F\big(\operatorname{Res}(X)\big)\right\}\left(\frac{1}{d+1}\sum_{z\in\mathcal{V}} \|\operatorname{Res}(X)\|^2_{H^{-1}(\omega_z)}\right)^{\frac{1}{2}}$$
$$\le \|\nabla(x-X)\|_{L^2(\Omega)} \le C\left(1+\kappa\, F\big(\operatorname{Res}(X)\big)\right)\left(\sum_{z\in\mathcal{V}} \|\operatorname{Res}(X)\|^2_{H^{-1}(\omega_z)}\right)^{\frac{1}{2}}$$



(A) $\alpha = 10^{-4}$ (B) $\alpha = 10^{-8}$

FIGURE 2. d_α-error (+) and associated general upper (●) and lower (○) bound, as well as upper bound with compactness (■) versus DOFs for Example (4.12).

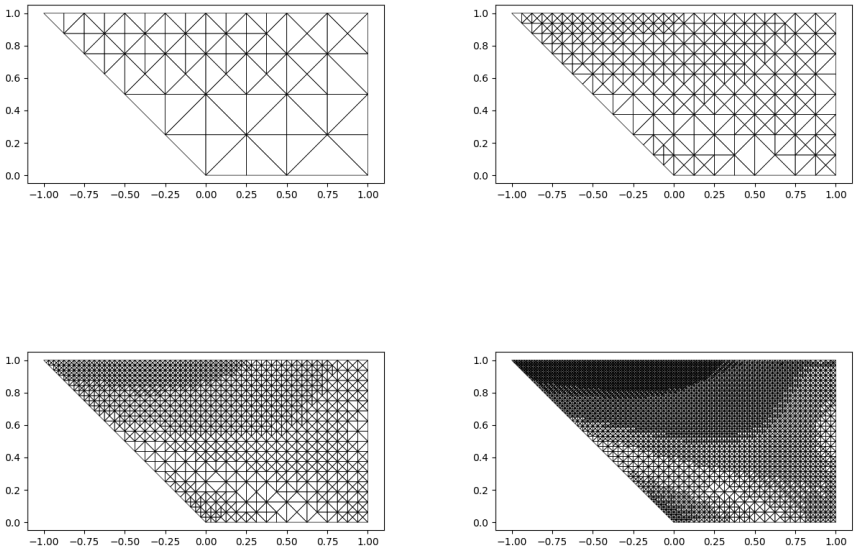


FIGURE 3. Adaptive mesh refinement history for Example (4.12) with $\alpha = 10^{-4}$.

with

$$F\big(\operatorname{Res}(X)\big) = \left(\frac{\sum_{z\in\mathcal{V}} h_z^2 \|\operatorname{Res}(X)\|^2_{H^{-1}(\omega_z)}}{\sum_{z\in\mathcal{V}} \|\operatorname{Res}(X)\|^2_{H^{-1}(\omega_z)}}\right)^{\frac{1}{2}}.$$

The constant C depends on the Poincaré constant C_P and the shape regularity of the mesh $\mathcal{M}$; κ is defined in (2.23).

4.2. **Boundary control.** This section illustrates how to apply the a posteriori bounds from Section 3.2 to optimal control problems with constrained Neumann boundary control.

Let $\Omega \subset \mathbb{R}^d$ be a domain with polyhedral boundary Γ and outward-pointing normal n. Given some lower bound $a \in \mathbb{R} \cup \{-\infty\}$, let

$$K = \left\{ q \in L^2(\Gamma) \mid \int_\Gamma q \geq a \right\}, \tag{4.13}$$

denote the set of admissible controls. For targets $u_d^\Omega \in L^2(\Omega)$, $u_d^\Gamma \in L^2(\Gamma)$ and cost parameter $\alpha > 0$, we consider

$$\begin{aligned} \min_{(q,u)\in K\times H^1(\Omega)} & \frac{1}{2}\|u - u_d^\Omega\|^2_{L^2(\Omega)} + \frac{1}{2}\|u - u_d^\Gamma\|^2_{L^2(\Gamma)} + \frac{\alpha}{2}\|q\|^2_{L^2(\Gamma)} \\ \text{subject to} \quad & -\Delta u + u = 0 \text{ in } \Omega \quad \text{and} \quad \partial_{\mathsf{n}} u = q \text{ on } \Gamma, \end{aligned} \tag{4.14}$$

where ∂_{n} denotes the normal derivative. This problem fits into the framework of Section 2 with the Hilbert spaces

$$\begin{aligned} V_1 = V_2 &= H^1(\Omega), && (v_1, v_2)_{V_i} = (v_1, v_2)_{H^1(\Omega)}, \quad i = 1, 2, \\ Q &= L^2(\Gamma), && (q_1, q_2)_Q = (q_1, q_2)_{L^2(\Gamma)}, \\ W &= L^2(\Omega) \times L^2(\Gamma), && (w_1, w_2)_W = (w_1^\Omega, w_2^\Omega)_{L^2(\Omega)} + (w_1^\Gamma, w_2^\Gamma)_{L^2(\Gamma)}, \end{aligned}$$

writing $w_i = (w_i^\Omega, w_i^\Gamma) \in W$, $i = 1, 2$. The other ingredients are given by (4.13),

$$\begin{gathered} a(v, \varphi) = \int_\Omega \nabla v \cdot \nabla \varphi + v\varphi = (v, \varphi)_{H^1(\Omega)}, \qquad m_a = 1 = M_a, \\ \langle Cq, \varphi \rangle = \int_\Gamma q\varphi = (q, \varphi_{|\Gamma})_{L^2(\Gamma)} = (q, C^*\varphi)_{L^2(\Gamma)}, \qquad M_C = C_\Gamma, \\ Iv = (v, v_{|\Gamma}), \qquad \langle I^*(w^\Omega, w^\Gamma), v \rangle = \int_\Omega v w^\Omega + \int_\Gamma v_{|\Gamma} w^\Gamma, \qquad M_I = (1 + C_\Gamma), \\ \Pi_K v = v + \frac{1}{|\Gamma|} \max\left\{ 0, a - \int_\Gamma v \right\}, \end{gathered}$$

where C_Γ is the embedding constant $H^1(\Omega) \subset L^2(\Gamma)$ and we write $\langle \cdot, \cdot \rangle$ for the duality pairing in $H^1(\Omega)^* \times H^1(\Omega)$.

The variational formulation of the reduced and rescaled optimality system (2.10) reads: find $(u, z) \in H^1(\Omega) \times H^1(\Omega)$ such that, for all $\varphi_1, \varphi_2 \in H^1(\Omega)$,

$$\begin{aligned} \int_\Omega \nabla\varphi_1 \cdot \nabla z + \varphi_1 z - \tfrac{1}{\sqrt{\alpha}} \left(\int_\Omega u\varphi_1 + \int_\Gamma u\varphi_1 \right) \\ = -\tfrac{1}{\sqrt{\alpha}} \left(\int_\Omega u_d^\Omega \varphi_1 + \int_\Gamma u_d^\Gamma \varphi_1 \right), \end{aligned} \tag{4.15a}$$

$$\int_\Omega \nabla u \cdot \nabla \varphi_2 + u\varphi_2 - \int_\Gamma \Pi_K(-\tfrac{1}{\sqrt{\alpha}} z)\varphi_2 = 0. \tag{4.15b}$$

Using the finite element framework of Section 4.1, its discretisation reads as follows: find $(U, Z) \in S^1_\ell(\mathcal{M}) \times S^1_\ell(\mathcal{M})$ such that, for all $\Phi_1, \Phi_2 \in S^1_\ell(\mathcal{M})$,

$$\begin{aligned} \int_\Omega \nabla\Phi_1 \cdot \nabla Z + \Phi_1 Z - \tfrac{1}{\sqrt{\alpha}} \left(\int_\Omega U\Phi_1 + \int_\Gamma U\Phi_1 \right) \\ = -\tfrac{1}{\sqrt{\alpha}} \left(\int_\Omega u_d^\Omega \Phi_1 + \int_\Gamma u_d^\Gamma \Phi_1 \right), \end{aligned} \tag{4.16a}$$

$$\int_\Omega \nabla U \cdot \nabla \varphi_2 + U\varphi_2 - \int_\Gamma \Pi_K(-\tfrac{1}{\sqrt{\alpha}} Z)\varphi_2 = 0. \tag{4.16b}$$

Consequently, in the solution $X = (U, Z)$ of (4.16), the residual

$$\begin{aligned}\big\langle \operatorname{Res}(X),\, (\varphi_1,\varphi_2)\big\rangle := &-\int_\Omega \nabla U\cdot\nabla\varphi_2 - U\varphi_2 + \int_\Gamma \Pi_K(-\tfrac{1}{\sqrt{\alpha}}Z)\varphi_2\\ &-\tfrac{1}{\sqrt{\alpha}}\int_\Omega u_d^\Omega\varphi_1 - \tfrac{1}{\sqrt{\alpha}}\int_\Gamma u_d^\Gamma\varphi_1\\ &-\int_\Omega \nabla\varphi_1\cdot\nabla Z - \varphi_1 Z + \tfrac{1}{\sqrt{\alpha}}\int_\Omega U\varphi_1 + \tfrac{1}{\sqrt{\alpha}}\int_\Gamma U\varphi_1,\end{aligned} \tag{4.17}$$

$(\varphi_1,\varphi_2)\in H^1(\Omega)\times H^1(\Omega)$, satisfies the orthogonality condition

$$\langle \operatorname{Res}(X),\,\Phi\rangle = 0 \qquad \text{for all } \Phi\in S^1_\ell(\mathcal{M})\times S^1_\ell(\mathcal{M}). \tag{4.18}$$

Similarly as in (4.7), we can localize the norm of the residual by

$$\frac{1}{d+1}\sum_{z\in\mathcal{V}} \|\operatorname{Res}(X)\|^2_{H^*_z} \le \|\operatorname{Res}(X)\|^2_{(H^1(\Omega))^*} \le C_{\mathcal{M}}\sum_{z\in\mathcal{V}} \|\operatorname{Res}(X)\|^2_{H^*_z} \tag{4.19}$$

with $H_z := H^{-1}(\omega_z)$ for interior vertices $z\in\mathcal{V}\cap\Omega$ and $\{v\in H^1(\omega_z)\mid v = 0 \text{ on } \partial\omega_z\setminus\partial\Omega\}$ for $z\in\mathcal{V}\cap\Gamma$. For the proof, we refer to Lemma 4.9, where similar arguments are used. Inserting the localization (4.19) into Theorem 3.1 yields the following result.

Theorem 4.7 (Bounding d_α-error for boundary control – general case)**.** *Let $x = (u,z)$ be the exact states of the optimal control problem* (4.14)*, where the adjoint state is rescaled; cf.* (4.15)*. Furthermore, let $X = (U,Z)$ be their finite element approximations from* (4.16) *and define its residual by* (4.17)*. Then we have the equivalence*

$$\frac{1}{d+1}\sum_{z\in\mathcal{V}} \|\operatorname{Res}(X)\|^2_{H^*_z} \le d_\alpha(x,X)^2 \le \kappa\, C_{\mathcal{M}}\sum_{z\in\mathcal{V}} \|\operatorname{Res}(X)\|^2_{H^*_z},$$

where the constant $C_{\mathcal{M}}$ depends on the shape regularity of the mesh and κ is defined in (2.23)*.*

Next, we shall use the fact, that the operators C^* and I involved in the definition (3.9) of the reconstruction are compact. Indeed, the compactness of C^* originates in the trace evaluation $H^1(\Omega)\ni\varphi\mapsto\varphi_\Gamma\in L^2(\Gamma)$, while the one of the observation operator I arises also with the help of the embedding $H^1(\Omega)\subset L^2(\Omega)$. We therefore can apply the results of Section 3.2 and need to quantify the compactness. To prepare this, we use the Lipschitz continuity of Π_K with Lipschitz constant 1 to obtain

$$\begin{aligned}\delta^*_\alpha(RX,X)^2 &\le \delta_\alpha(RX,X)^2\\ &\le \|R_2X - Z\|^2_{L^2(\Gamma)} + \|R_1X-U\|^2_{L^2(\Omega)} + \|R_1X-U\|^2_{L^2(\Gamma)},\end{aligned} \tag{4.20}$$

where $R = (R_1,R_2) : S^1_\ell(\mathcal{M})\times S^1_\ell(\mathcal{M}) \to H^1(\Omega)\times H^1(\Omega)$ is the auxiliary operator defined in (3.9) and, as before, $X = (U,Z)$ is the discrete solution of (4.16). Combining their definitions reveals the following orthogonality relationships: for all $\Phi_1,\Phi_2\in S^1_\ell(\mathcal{M})$, we have

$$\begin{aligned}&\int_\Omega \nabla(R_1X-U)\cdot\nabla\Phi_2 + (R_1X-U)\Phi_2\\ &\quad = 0 = \int_\Omega \nabla(R_2X-U)\cdot\nabla\Phi_1 + (R_2X-U)\Phi_1.\end{aligned} \tag{4.21}$$

In other words, U and Z are, respectively the Ritz projections in $S^1_\ell(\mathcal{M})$ of R_1X and R_2X with respect to the bilinear form $(\cdot,\cdot)_{H^1(\Omega)}$. Consequently, similarly to the preceding section, we can quantify the available compactness by means of a duality argument thanks to the orthogonality (4.18) of $\operatorname{Res}(X)$.

To this end, we restrict ourselves to polyhedral convex domains $\Omega \subset \mathbb{R}^d$ and analyze the regularity of the solution of the following Neumann problem: given $g = (g^\Omega, g^\Gamma) \in L^2(\Omega) \times L^2(\Gamma)$, find $v_g \in H^1(\Omega)$ such that

$$-\Delta v_g + v_g = g^\Omega \quad \text{in } \Omega \quad \text{and} \quad \partial_{\mathsf{n}} v_g = g^\Gamma,$$

which weakly reads as

$$\forall \varphi \in H^1(\Omega) \qquad \int_\Omega \nabla v_g \cdot \nabla \varphi + v_g \varphi = \int_\Omega g^\Omega \varphi + \int_\Gamma g^\Gamma \varphi. \tag{4.22}$$

Notice that the critical term on the right-hand side involves an $L^2(\Gamma)$-trace of the test function φ. Hence, in order to precisely measure its regularity, we shall need a sharp trace theorem for $L^2(\Gamma)$. Using fractional Sobolev spaces, the trace operator is bounded as a map from $H^{\frac{1}{2}+\epsilon}(\Omega)$ to $H^\epsilon(\Gamma)$ only for $\epsilon > 0$, and therefore is not sharp for $L^2(\Gamma)$. For a sharp trace theorem and thus avoiding ϵ, we invoke Besov spaces. Given $s > 0$, $p, q \in [1, \infty]$, we define the Besov space $B^s_q(L^p(\Omega))$ and its norm $\| \cdot \|_{B^s_q(L^p(\Omega))}$ as in [8, Section 2] through intrinsic moduli of smoothness. Furthermore, we need the real interpolation method of Peetre based upon the so-called K-functional; see, e.g., [7]. Given two Banach spaces X_1, X_2 with $X_1 \subset X_2$ and parameters $\theta \in (0,1)$, $q \in [1, \infty]$, we denote its interpolation by $(X_1, X_2)_{\theta,q}$ and its norm by $\| \cdot \|_{(X_1,X_2)_{\theta,q}}$.

Proposition 4.8 (Extra regularity for boundary control)**.** *There is a constant C_Ω depending only on the convex domain $\Omega \subset \mathbb{R}^d$ such that, for any $g = (g^\Omega, g^\Gamma) \in L^2(\Omega) \times L^2(\Gamma)$, the unique solution $v_g \in H^1(\Omega)$ of (4.22) satisfies*

$$\|v_g\|_{B^{\frac{3}{2}}_\infty(L^2(\Omega))} \leq C_\Omega \left(\|g^\Omega\|_{L^2(\Omega)} + \|g^\Gamma\|_{L^2(\Gamma)} \right).$$

Proof. 1 We start by measuring the regularity of the right-hand side

$$\langle F_g, \varphi \rangle := \int_\Omega g^\Omega \varphi + \int_\Gamma g^\Gamma \varphi, \quad \varphi \in H^1(\Omega),$$

in (4.22). In view of [8, Section 4] and [26, Sections 1.2.5 and 1.3.4], the above definition of $B^s_q(L^p(\Omega))$ coincides with [4, Definition (2.52)] in the sense of equivalent norms. Hence, we can use the sharp trace theorem [4, Proposition 3.5] to derive

$$\begin{aligned} |\langle F_g, \varphi \rangle| &\leq \|g^\Omega\|_{L^2(\Omega)} \|\varphi\|_{L^2(\Omega)} + C \|g^\Gamma\|_{L^2(\Gamma)} \|\varphi\|_{B^{\frac{1}{2}}_1(L^2(\Omega))} \\ &\leq C \left(\|g^\Omega\|_{L^2(\Omega)} + \|g^\Gamma\|_{L^2(\Gamma)} \right) \|\varphi\|_{B^{\frac{1}{2}}_1(L^2(\Omega))}. \end{aligned}$$

Thanks to [15], we have $B^{\frac{1}{2}}_1(L^2(\Omega)) = (H^1(\Omega), L^2(\Omega))_{\frac{1}{2},1}$ and, in view of the duality theorem [7, (14.1.8)], $(H^1(\Omega), L^2(\Omega))^*_{\frac{1}{2},1} = (L^2(\Omega), H^1(\Omega)^*)_{\frac{1}{2},\infty}$. Consequently,

$$\|F_g\|_{(L^2(\Omega), H^1(\Omega)^*))_{\frac{1}{2},\infty}} \leq C \left(\|g^\Omega\|_{L^2(\Omega)} + \|g^\Gamma\|_{L^2(\Gamma)} \right). \tag{4.23}$$

2 We next specify the corresponding regularity gain in the solution v_g. Replacing the right-hand side of (4.22) by a generic functional $G \in H^1(\Omega)^*$, we readily observe

$$\|v_G\|_{H^1(\Omega)} \leq \|G\|_{H^1(\Omega)^*}$$

for the corresponding solution v_G. Next, let us consider $G \in L^2(\Omega)$ and notice that this corresponds to a homogeneous Neumann problem with source term in $L^2(\Omega)$, i.e. (4.22) with $g^\Omega = G$ and $g^\Gamma = 0$. Since Ω is convex, we then have

$$\|v_G\|_{H^2(\Omega)} \leq C \|G\|_{L^2(\Omega)};$$

cf. [12, Theorem 3.2.1.3]. Interpolating these two inequalities with [7, (14.1.5)] gives

$$\|v_G\|_{B^{\frac{3}{2}}_\infty(L^2(\Omega))} \leq C \|v_G\|_{(H^2(\Omega), H^1(\Omega))_{\frac{1}{2},\infty}} \leq C \|G\|_{(L^2(\Omega), H^1(\Omega)^*))_{\frac{1}{2},\infty}},$$

where the first inequality follows from $(H^2(\Omega), H^1(\Omega))_{\frac{1}{2},\infty} = B^{\frac{3}{2}}_{\infty}(L^2(\Omega))$, see [5, 6.2.4], again taking into account [8, Section 4] and [26, Sections 1.2.5 and 1.3.4]. Hence, inserting (4.23) in the last inequality with $G = F_g$ finishes the proof. □

Proposition 4.8 puts us in the position to prove the following bound, which, thanks to the inequality (4.20), yields a bound for the compact part of the error.

Lemma 4.9 (Upper bound for compact error – boundary control)**.** *Let $\Omega \subset \mathbb{R}^d$ be a convex domain with polyhedral boundary. Then the L^2-errors in* (4.20) *are bounded in terms of the residual of X:*

$$\|R_2X - Z\|^2_{L^2(\Gamma)} + \|R_1X - U\|^2_{L^2(\Gamma)} + \|R_1X - U\|^2_{L^2(\Omega)} \\ \leq C_\Omega^2 C_{\mathcal{M}}^2 \sum_{z\in\mathcal{V}} h_z \|\operatorname{Res}(X)\|^2_{H^*_z}.$$

Here $C_{\mathcal{M}}, C_\Omega$ are essentially the constants from (4.19) *and Proposition 4.8, respectively.*

Proof. We provide only a sketch of the proof, which is very similar to the one of Lemma 4.3 but involves some additional technicality due to the Besov regularity in Proposition 4.8. We start with the term $\|R_1X - U\|^2_{L^2(\Gamma)} + \|R_1X - U\|^2_{L^2(\Omega)}$. According to Proposition 4.8, there exists $\psi \in B^{\frac{3}{2}}_{\infty}(L^2(\Omega))$ weakly solving

$$-\Delta\psi + \psi = R_1X - U_h \quad \text{in } \Omega \qquad \text{and} \qquad \partial_{\mathsf{n}}\psi = R_1X - U_h \quad \text{on } \Gamma$$

and

$$\|\psi\|_{B^{\frac{3}{2}}(L^2(\Omega))} \leq C_\Omega(\|R_1X - U_h\|_{L^2(\Omega)} + \|R_1X - U_h\|_{L^2(\Gamma)}).$$

Again, $\mathcal{I}_{\mathsf{sz}} : H^1(\Omega) \to S^1_\ell(\mathcal{M})$ denots the Scott-Zhang quasi-interpolation operator [25]. Recalling (4.22) and the definition of R, we have thanks to the orthogonality (4.18) of $\operatorname{Res}(X)$ that

$$\begin{aligned}\|R_1X - U\|^2_{L^2(\Omega)} + \|R_1X - U\|^2_{L^2(\Gamma)} &= \int_\Omega \nabla(R_1X - U)\cdot\nabla\psi + (R_1X - U)\psi \\ &= \langle \operatorname{Res}_2(X), \psi\rangle = \sum_{z\in\mathcal{V}} \langle \operatorname{Res}_2(X), (\psi - \mathcal{I}_{\mathsf{sz}}\psi)\phi_z\rangle \\ &\leq \sum_{z\in\mathcal{V}} \|\operatorname{Res}_i(X)\|_{H^*_z} \|\nabla((\psi - \mathcal{I}_{\mathsf{sz}}\psi)\phi_z)\|_{L^2(\omega_z)},\end{aligned}$$

where we have used that the Lagrange basis functions $\{\phi_z : z \in \mathcal{V}\}$ of $S^1_1(\mathcal{M})$ form a partition of unity and that $\operatorname{supp}\phi_z = \omega_z$, $z \in \mathcal{V}$. Similarly to (4.11), we derive

$$\|\nabla((\psi - \mathcal{I}_{\mathsf{sz}}\psi)\phi_z)\|_{L^2(\omega_z)} \leq C_{\mathcal{M}}\|\psi\|_{H^1(\tilde{\omega}_z)}.$$

Interpolating with [7, (14.1.5)] both inequalities yield

$$\|\nabla((\psi - \mathcal{I}_{\mathsf{sz}}\psi)\phi_z)\|_{L^2(\omega_z)} \leq C_{\mathcal{M}} h_z^{\frac{1}{2}} \|\psi\|_{B^{\frac{3}{2}}_{\infty}(L^2(\tilde{\omega}_z))}.$$

Employing averaged moduli, cf. [11, Lemma 4.10] and using that the overlapping of the domains $\tilde{\omega}_z = \bigcup_{y\in\mathcal{V}\cap\omega_z} \omega_y$ is controlled by the shape regularity of $\mathcal{M}$, we conclude with (4.10) that

$$\|R_1X - U\|^2_{L^2(\Omega)} + \|R_1X - U\|^2_{L^2(\Gamma)} \\ \leq C_{\mathcal{M}} C_\Omega \left(\sum_{z\in\mathcal{V}} h_z \|\operatorname{Res}_i(X)\|^2_{H^*_z}\right)^{\frac{1}{2}} \left(\|R_1X - U\|^2_{L^2(\Omega)} + \|R_1X - U\|^2_{L^2(\Gamma)}\right)^{\frac{1}{2}}.$$

Applying similar arguments to $\|R_2X - Z\|^2_{L^2(\Gamma)}$, the assertion follows. □

In combination with Theorem 3.6, we thus get the following a posteriori bounds.

Theorem 4.10 (Bounding the d_α-error for boundary control – compact case)**.** *Suppose in addition to the setting in Theorem 4.7 that $\Omega \subset \mathbb{R}^d$ is convex with a polyhedral boundary. Then, we have*

$$\left(\frac{1}{d+1}\sum_{z\in\mathcal{V}} \|\operatorname{Res}(X)\|_{H_z^*}^2\right)^{\frac12} \le d_\alpha(x,X)$$
$$\le C\left(1+\frac{\kappa}{\sqrt{\alpha}}\left(\frac{\sum_{z\in\mathcal{V}} h_z\|\operatorname{Res}(X)\|_{H_z^*}^2}{\sum_{z\in\mathcal{V}}\|\operatorname{Res}(X)\|_{H_z^*}^2}\right)^{\frac12}\right)\left(\sum_{z\in\mathcal{V}}\|\operatorname{Res}(X)\|_{H_z^*}^2\right)^{\frac12}$$

The constant C depends on the Poincaré constant C_P, the constant C_Ω from Proposition 4.8 and the shape regularity of the mesh $\mathcal{M}$; the constant κ is defined in (2.23)*.*

In the absence of control constraints, we have from Theorem 3.10 the following a posteriori bounds for the combined $H^1(\Omega)$-errors of the states.

Theorem 4.11 (Bounding the $\|\cdot\|$-error for boundary control)**.** *Suppose in addition to the setting in Theorem 4.10 that $a=-\infty$, i.e. no control constraints apply. Then we have*

$$\max\left\{\frac{\sqrt{\alpha}}{\sqrt{\alpha}+(1+C_\Gamma)}, 1-C\kappa\, G\big(\operatorname{Res}(X)\big)\right\}\left(\frac{1}{d+1}\sum_{z\in\mathcal{V}}\|\operatorname{Res}(X)\|_{H_z^*}^2\right)^{\frac12}$$
$$\le \|x-X\|_{H^1(\Omega)} \le C\left(1+\kappa\, G\big(\operatorname{Res}(X)\big)\right)\left(\sum_{z\in\mathcal{V}}\|\operatorname{Res}(X)\|_{H_z^*}^2\right)^{\frac12},$$

with

$$G\big(\operatorname{Res}(X)\big) = \left(\frac{\sum_{z\in\mathcal{V}} h_z\|\operatorname{Res}(X)\|_{H_z^*}^2}{\sum_{z\in\mathcal{V}}\|\operatorname{Res}(X)\|_{H_z^*}^2}\right)^{\frac12}.$$

The constant C depends on the Poincaré constant C_P, the constant C_Ω from Proposition 4.8 and the shape regularity of the mesh $\mathcal{M}$; the constant κ is defined in (2.23)*.*

In view of $\|x-X\| \le d_\alpha(x,X)$, the upper bound in Theorem 4.7 can be used also for the $\|\cdot\|$-error. Note that the first option in the max in the lower bound of Theorem 4.11 does not involve compactness. We thus have upper and lower bounds for $\|\cdot\|$-error which hold in general, i.e. do not need compactness. This pair of general bounds can be compared with the bounds in Theorem 4.11 using compactness. For the numerical comparison of such pairs, we consider

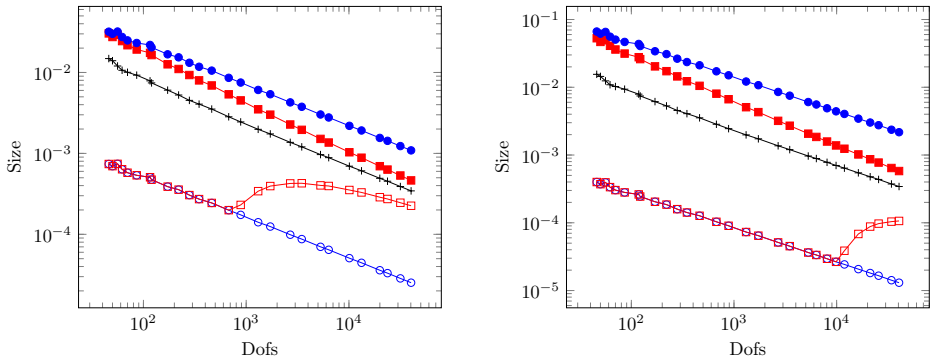
$$\min_{(q,u)\in L^2(\Gamma)\times H^1(\Omega)} \frac12\|u-u_d\|_{L^2(\Omega)}^2 + \int_\Gamma g_1 u + \frac{\alpha}{2}\|q\|_{L^2(\Gamma)}^2 \quad \text{subject to} \quad -\Delta u + u = f \text{ in } \Omega \quad \text{and} \quad \partial_{\mathsf{n}} u = g_2 + q \text{ on } \Gamma, \tag{4.24}$$

where Ω is the convex domain that is meshed in Figure 5 and has an internal maximum angle $\frac{35\pi}{36}$ at the origin, $u=0$, $z = r^{\frac{36}{35}}\cos(\frac{36}{35}\theta)$, $u_d = -\sqrt{\alpha}z$, $g_1 = \frac{\partial z}{\partial n}$, and $g_2 = \frac{1}{\sqrt{\alpha}}z$.

As in Section 4.1, the numerical simulations are carried out with linear finite elements, a standard residual estimator scaled by the same procedure, and the adaptive mesh refinement is based on Dörfler's marking strategy [9] with parameter 0.6.

Figure 4 depicts the $\|\cdot\|$-error and the aforementioned associated bounds, while Figure 5 gives an idea for the underlying adaptive mesh refinement. The following

differences to the discussion of the bounds for the d_α-error in Section 4.1 are noteworthy. Compactness is useful in both upper and lower bound. The upper bound with compactness is advantageous from the start. This is related with the fact that both upper bounds depend on α only through the factor κ from (2.23). For the lower bounds of the $\|\cdot\|$-error, we have a similar situation as for the upper bounds of the d_α-error. Indeed, the lower bounds with compactness improves the general one only for fine meshes and the necessary fineness increases with decreasing α.



(A) $\alpha = 10^{-1}$ (B) $\alpha = 5 \cdot 10^{-2}$

FIGURE 4. $\|\cdot\|$-error (+), general upper (•) and lower (○) bound, as well as upper (■) and lower bound with compactness versus DOFs for Example (4.24).

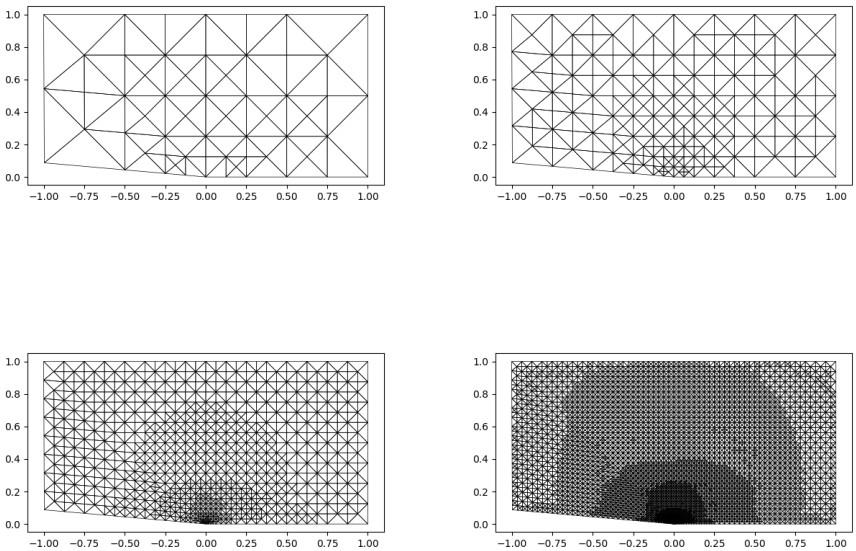


FIGURE 5. Adaptive mesh refinement history for Example (4.24) with $\alpha = 10^{-1}$.

Acknowledgment. Fernando Gaspoz is partially supported by Consejo Nacional de Investigaciones Científicas y Técnicas through grant PIP 11220200101180CO, by Agencia Nacional de Promoción Científica y Tecnológica through grants PICT-2020-SERIE A-03820 and 03267, and by Universidad Nacional del Litoral through grant CAI+D-2020 50620190100136LI.

References

[1] M. Ainsworth and J. T. Oden, *A posteriori error estimation in finite element analysis*, Pure and Applied Mathematics (New York), Wiley-Interscience [John Wiley & Sons], New York, 2000.
[2] I. Babuška, *Error-bounds for finite element method*, Numer. Math., 16 (1971), pp. 322–333.
[3] P. Bastian, M. Blatt, A. Dedner, N.-A. Dreier, C. Engwer, R. Fritze, C. Gräser, C. Grüninger, D. Kempf, R. Klöfkorn, M. Ohlberger, and O. Sander, *The Dune framework: Basic concepts and recent developments*, Computers & Mathematics with Applications, 81 (2021), pp. 75–112.
[4] J. Behrndt, F. Gesztesy, and M. Mitrea, *Sharp Boundary Trace Theory and Schrödinger Operators on Bounded Lipschitz Domains*, Sept. 2022. arXiv:2209.09230 [math].
[5] J. Bergh and J. Löfström, *Interpolation spaces: an introduction*, no. 223 in Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, Springer, Berlin, 1976. OCLC: 2373287.
[6] A. Bonito, C. Canuto, R. H. Nochetto, and A. Veeser, *Adaptive finite element methods*, 2024.
[7] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, vol. 15 of Texts in Applied Mathematics, Springer New York, New York, NY, 2008.
[8] R. A. DeVore and R. C. Sharpley, *Besov Spaces on Domains in Rd*, Transactions of the American Mathematical Society, 335 (1993), pp. 843–864.
[9] W. Dörfler, *A convergent adaptive algorithm for Poisson's equation*, SIAM J. Numer. Anal., 33 (1996), pp. 1106–1124.
[10] F. Gaspoz, C. Kreuzer, A. Veeser, and W. Wollner, *Quasi-best approximation in optimization with PDE constraints*, Inverse Problems, 36 (2020), pp. 014004, 29.
[11] F. D. Gaspoz and P. Morin, *Approximation classes for adaptive higher order finite element approximation*, Mathematics of Computation, 83 (2013), pp. 2127–2160.
[12] P. Grisvard, *Elliptic problems in nonsmooth domains*, no. 24 in Monographs and studies in mathematics, Pitman Advanced Pub. Program, Boston, 1985.
[13] P. Grisvard, *Elliptic problems in nonsmooth domains*, vol. 69 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011. Reprint of the 1985 original [MR0775683], With a foreword by Susanne C. Brenner.
[14] M. Hinze, *A variational discretization concept in control constrained optimization: The linear-quadratic case*, Comp. Optim. Appl., 30 (2005), pp. 45–61.
[15] H. Johnen and K. Scherer, *On the equivalence of the K-functional and moduli of continuity and some applications*, in Constructive Theory of Functions of Several Variables, W. Schempp and K. Zeller, eds., Berlin, Heidelberg, 1977, Springer, pp. 119–140.
[16] K. Kohls, C. Kreuzer, A. Rösch, and K. G. Siebert, *Convergence of adaptive finite elements for optimal control problems with control constraints*, North-West. Eur. J. Math., 4 (2018), pp. 157–184, i.
[17] K. Kohls, A. Rösch, and K. G. Siebert, *A posteriori error estimators for control constrained optimal control problems*, in Constrained optimization and optimal control for partial differential equations, vol. 160 of Internat. Ser. Numer. Math., Birkhäuser/Springer Basel AG, Basel, 2012, pp. 431–443.
[18] ———, *A posteriori error analysis of optimal control problems with control constraints*, SIAM J. Control Optim., 52 (2014), pp. 1832–1861.
[19] C. Kreuzer and A. Veeser, *Oscillation in a posteriori error estimation*, Numer. Math., 148 (2021), pp. 43–78.
[20] O. Lakkis and C. Makridakis, *Elliptic reconstruction and aposteriori error estimates for fully discrete linear parabolic problems*, Math. Comp., 75 (2006), pp. 1627–1658.
[21] J.-L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, Die Grundlehren der mathematischen Wissenschaften, Springer, Berlin – Heidelberg – New York, 1. ed., 1971.
[22] C. Makridakis and R. H. Nochetto, *Elliptic reconstruction and a posteriori error estimates for parabolic problems*, SIAM J. Numer. Anal., 41 (2003), pp. 1585–1594.

[23] J. Nečas, *Sur une méthode pour resoudre les équations aux dérivées partielles du type elliptique, voisine de la variationnelle*, Ann. Sc. Norm. Super. Pisa, Sci. Fis. Mat., III. Ser., 16 (1962), pp. 305–326.
[24] A. H. Schatz, *An observation concerning Ritz-Galerkin methods with indefinite bilinear forms*, Math. Comp., 28 (1974), pp. 959–962.
[25] L. R. Scott and S. Zhang, *Finite element interpolation of nonsmooth functions satisfying boundary conditions*, Math. Comp., 54 (1990), pp. 483–493.
[26] H. Triebel, *Theory of Function Spaces II*, Springer Basel, Basel, 1992.
[27] F. Tröltzsch, *Optimal control of partial differential equations*, vol. 112 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2010. Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels.
[28] R. Verfürth, *A posteriori error estimation techniques for finite element methods*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 2013.
[29] M. F. Wheeler, *A priori L_2 error estimates for Galerkin approximations to parabolic partial differential equations*, SIAM J. Numer. Anal., 10 (1973), pp. 723–759.