Computational Optimization and Applications manuscript No.
(will be inserted by the editor)

A posteriori error estimation and adaptivity for elliptic optimal control problems with state constraints

Olaf Benedix*, Boris Vexler

Lehrstuhl für Mathematische Optimierung, Technische Universität München, Fakultät für Mathematik, Boltzmannstraße 3, Garching b. München, Germany, e-mail: {benedix,vexler}@ma.tum.de

Received: date / Revised version: date

Abstract In this paper optimal control problems governed by elliptic semilinear equations and subject to pointwise state constraints are considered. These problems are discretized using finite element methods and a posteriori error estimates are derived assessing the error with respect to the cost functional. These estimates are used to obtain quantitative information on the discretization error as well as for guiding an adaptive algorithm for local mesh refinement. Numerical examples illustrate the behavior of the method.

Key words optimal control, state constraints, semilinear equations, finite elements, a posteriori error estimation

1 Introduction

In this paper we consider optimal control problems governed by semilinear elliptic partial differential equations and subject to inequality constraints on the state variable, so called *state constraints*, formulated as follows:

$$\begin{cases} \text{minimize} \quad J(q,u), \quad q \in Q,\ u \in V, \\ u = S(q), \\ u_a(x) \le u(x) \le u_b(x) \quad \text{for } x \in \bar{\Omega}. \end{cases} \tag{1.1}$$

Here, the pair (q,u) consists of the control variable q and the state variable u from the corresponding function spaces Q and V to be specified later.

* The author's research was supported by the Austrian Science Fund FWF, project P18971-N18 "Numerical analysis and discretization strategies for optimal control problems with singularities"

The operator $S\colon Q \to V$ represents the solution operator of a semilinear elliptic equation, J denotes the cost functional to be minimized, and the inequality constraints on the state variable u are formulated pointwise on the computational domain $\Omega \subset \mathbb{R}^2$ using lower and upper bounds $u_a, u_b \in C(\bar{\Omega})$. Such problems are known to be difficult from the theoretical as well as from the numerical point of view, due to the fact that the Lagrange multipliers associated with the state constraints are in general regular Borel measures, see, e.g., [8]. There are several recent results on numerical solution algorithms, see, e.g., [19,20,29,30,32] as well as on a priori error analysis of finite element discretizations for such problems, see [11,12,25,27].

In this paper we are concerned with a posteriori error estimation for finite element discretization of (1.1). The use of adaptive techniques based on a posteriori error estimation is well accepted in the context of finite element discretization of partial differential equations, see, e.g., [13,35,2]. The application of these techniques to optimization problems with PDEs is an active area of research. A posteriori error estimators are developed for problems with control constraints assessing the error with respect to the natural norms of the corresponding spaces, see, e.g., [17,23]. In a recent preprint [22], these techniques are extended to an optimal control problem with state constraints governed by a linear elliptic equation.

An approach for estimating the error in terms of the cost functional was introduced in [1] for problems without inequality constraints and has been further developed for assessing the error with respect to a given quantity of interest in [3,4] and for treatment of parabolic problems in [26]. This approach is extended to problems with control constraints in [36,18]. In a recent preprint [16] a posteriori error estimates with respect to the cost functional are derived for a linear-quadratic optimal control problem with state constraints.

Our goal is to develop an a posteriori estimator for the error

$$J(q,u) - J(q_h,u_h),$$

where (q,u) denotes a (local) solution of the optimal control problem (1.1) governed by a semilinear elliptic equation and (q_h,u_h) is the corresponding solution of the discretized problem. Moreover, we discuss a strategy for evaluation of this error estimator which significantly differs from the approach presented in [16]. To the authors knowledge, this is the first paper where a posteriori error estimates for nonlinear optimal control problems with state constraints are developed.

The organization of the paper is as follows: In the next section we discuss the functional analytic setting of the problem under consideration and recall necessary optimality conditions. In Section 3 the finite element discretization of (1.1) is presented. Then, a posteriori error estimates are derived in Section 4. In the last section two numerical examples are considered illustrating the behavior of our method.

2 Problem formulation

In this section we discuss the precise formulation of the optimal control problem under consideration and the corresponding optimality conditions. We start with the introduction of some basic notation and with some assumptions on the quantities involved.

Let $\Omega \subset \mathbb{R}^2$ be a polygonal Lipschitz domain with boundary Γ, which is composed of the two disjoint parts Γ_1 and Γ_2 with $\Gamma = \Gamma_1 \cup \Gamma_2$ and $\mathrm{meas}(\Gamma_1) > 0$. Let moreover a symmetric 2×2-matrix $A(x) = (a_{ij}(x))$ be given with entries $a_{ij} \in L^\infty(\Omega)$ fulfilling

$$\sum_{i,j=1}^{2} a_{ij}(x)\xi_i\xi_j \geq \alpha_0|\xi|^2 \quad \forall \xi \in \mathbb{R}^2 \text{ and a.e. in } \Omega$$

for some $\alpha_0 > 0$. The corresponding uniformly elliptic differential operator is defined as

$$\mathcal{A}u(x) = -\sum_{i,j=1}^{2} \frac{\partial}{\partial x_i}\left(a_{ij}(x)\frac{\partial}{\partial x_j}u(x)\right). \tag{2.1}$$

Further $n(x)$ denotes the outer unit normal of Ω and $\partial_{\nu_\mathcal{A}}(x)$ is the conormal derivative to the operator $\mathcal{A}$ defined as the directional derivative in the direction $\nu_\mathcal{A}(x) := A(x) \cdot n(x)$.

For the in general nonlinear functions $d : \Omega \times \mathbb{R} \to \mathbb{R}$ and $b : \Gamma_2 \times \mathbb{R} \to \mathbb{R}$ we make throughout the following assumption

Assumption 1 *The functions d and b are measurable with respect to the first argument and $d(x,\cdot)$ and $b(x,\cdot)$ are monotone increasing and three times differentiable on $\mathbb{R}$ with respect to the second argument for each fixed $x \in \Omega$ or $x \in \Gamma_2$ respectively. Moreover, there exists a positive constant K such that*

$$|d(x,0)| + |d_u(x,0)| + |d_{uu}(x,0)| \leq K \quad \textit{a.e. in } \Omega$$

and

$$|b(x,0)| + |b_u(x,0)| + |b_{uu}(x,0)| \leq K \quad \textit{a.e. on } \Gamma_2.$$

Furthermore, for the Hilbert space Q, called the control space, with the norm $\|\cdot\|_Q$ and the inner product $(\cdot,\cdot)_Q$, let $B\colon Q \to L^2(\Omega)$ be a linear and continuous operator, called the control operator.
The space of state functions is defined by

$$V = C(\bar{\Omega}) \cap H^1_{\Gamma_1}(\Omega), \quad \text{with} \quad H^1_{\Gamma_1}(\Omega) = \left\{ v \in H^1(\Omega) \;\middle|\; v|_{\Gamma_1} = 0 \right\}. \tag{2.2}$$

With this notation, the state equation is given by

$$\begin{cases} \mathcal{A}u(x) + d(x,u(x)) = Bq(x) & \text{in } \Omega, \\ u(x) = 0 & \text{on } \Gamma_1, \\ \partial_{\nu_\mathcal{A}}u(x) + b(x,u(x)) = 0 & \text{on } \Gamma_2, \end{cases} \tag{2.3}$$

and belongs to the class of semilinear elliptic partial differential equations. Using the notation $(\cdot,\cdot)$ for the usual $L^2(\Omega)$-inner product and $\langle\cdot,\cdot\rangle_{\Gamma_2}$ for the $L^2(\Gamma_2)$-inner product we introduce the semilinear form $a\colon Q\times V\times V\to\mathbb{R}$ associated to the state equation by

$$a(q,u)(\varphi)=(A\nabla u,\nabla\varphi)+(d(\cdot,u),\varphi)+\langle b(\cdot,u),\varphi\rangle_{\Gamma_2}-(Bq,\varphi). \tag{2.4}$$

For given $q\in Q$ the weak formulation of the state equations reads: Find $u\in V$ such that

$$a(q,u)(\varphi)=0\quad\forall\varphi\in V. \tag{2.5}$$

Lemma 2.1 *Let Assumption 1 be fulfilled. Then, for every $q\in Q$ there exists a unique weak solution $u\in V$ of the state equation* (2.5)*. Moreover, there holds $u\in W^{1,r}(\Omega)$ for some $r>2$.*

Proof In fact, the unique existence in $H^1_{\Gamma_1}(\Omega)$ follows by standard arguments for monotone operators. For the additional proof of the continuity of u, one can directly follow the steps in [34, Theorem 4.7, 4.8]. It remains to proof, that $u\in W^{1,r}(\Omega)$ holds.

The solution u fulfills the linear elliptic equation

$$\begin{cases} \mathcal{A}u(x)=f(x) & \text{in } \Omega,\\ u(x)=0 & \text{on } \Gamma_1,\\ \partial_{\nu_{\mathcal{A}}}u(x)=g(x) & \text{on } \Gamma_2,\end{cases}$$

where $f(x)=Bq(x)-d(x,u(x))$ and $g(x)=-b(x,u(x))$. By Assumption 1 and the continuity of u we obtain $f\in L^2(\Omega)$. Using a trace theorem we get that $u\in H^{\frac12}(\Gamma_2)\cap C(\bar\Gamma_2)$. Then, we obtain due to the Lipschitz-continuity of $b(\cdot,\cdot)$ with respect to the second argument that $g\in H^{\frac12}(\Gamma_2)$. This implies by [15, Theorem 4.4.4.13, Corollary 4.4.4.14] that for all $s<2$

$$u-\sum c_i\psi_i\in W^{2,s}(\Omega),$$

where ψ_i are the functions describing the singular behaviour of u at the corners of the domain Ω. It can be directly checked, that $\psi_i\in W^{1,r}(\Omega)$ holds with some $r>2$. This completes the proof. □

This Lemma gives rise to the definition of the solution operator $S\colon Q\to V$ mapping a given control $q\in Q$ to the corresponding solution of (2.5). This operator is known to be twice continuously Fréchet differentiable, see, e.g., [34].

Remark 2.1 While the space of state functions V has been settled in (2.2), the control space Q will be left quite general, to allow for different situations like distributed control, i.e., $Q=L^2(\Omega)$, or finite dimensional control, i.e., $Q=\mathbb{R}^m$. This does not affect the analysis involved; for the numerical realization however it will be a difference whether Q is an infinite-dimensional space or not, see the discussion in Section 3.

To formulate the optimal control problem we introduce the cost functional $J\colon Q \times V \to \mathbb{R}$ to be of the following form:

$$J(q,u) = \Psi(u) + \frac{\alpha}{2}\|q\|_Q^2 \tag{2.6}$$

with a regularization parameter $\alpha > 0$ and a three times differentiable functional $\Psi\colon V \to \mathbb{R}$.

After these preliminaries, we formulate the optimal control problem as follows:

$$(P) \quad \begin{cases} \text{Minimize} \quad J(q,u), \quad q \in Q, u \in V, \\ a(q,u)(\varphi) = 0 \quad \forall \varphi \in V, \\ u_a(x) \le u(x) \le u_b(x) \quad \forall x \in \bar{\Omega}, \end{cases} \tag{2.7}$$

where the bounds $u_a, u_b \in C(\bar{\Omega})$ should match the rest of the setting by fulfilling $u_a < u_b$ in $\bar{\Omega}$ and $u_a \le 0 \le u_b$ on Γ_1. Thus the set of admissible state functions

$$V_{\mathrm{ad}} := \left\{ u \in V \;\middle|\; u_a \le u \le u_b \text{ in } \bar{\Omega} \right\}$$

has a nonempty interior $\mathrm{int}(V_{\mathrm{ad}})$.

The following assumptions will guarantee the existence of an optimal solution to (2.7) and the validity of optimality conditions.

Assumption 2 *There exists $\hat{q} \in Q$ so that $S(\hat{q}) \in V_{ad}$ and there holds*

$$\inf_{q \in Q, S(q) \in V_{ad}} J(q, S(q)) > -\infty.$$

Under this assumption the optimal control problem (2.7) is guaranteed to have a globally optimal solution. Due to the nonlinearity of the solution operator, however, uniqueness can not be assured.

To formulate necessary optimality conditions for a locally optimal solution pair (q,u) we require the following Slater condition.

Assumption 3 *For the locally optimal point (q,u) there exists $\tilde{q} \in Q$ such that*

$$S(q) + S'(q)(\tilde{q} - q) \in \mathrm{int}(V_{ad}).$$

In general, it is not easy to verify this assumption, since it contains the unknown solution. In the case of a linear state equation this assumption reduces to $S(\tilde{q}) \in \mathrm{int}(V_{\mathrm{ad}})$, which can simply be checked.

The formulation of the optimality system is based on the Lagrange functional defined by

$$\mathcal{L}(q,u,z,\mu^+,\mu^-) = J(q,u) - a(q,u)(z) - \langle \mu^+, u_b - u \rangle - \langle \mu^-, u - u_a \rangle, \tag{2.8}$$

where z denotes the adjoint state variable, μ^+ and μ^- are Lagrange multipliers for the upper and lower inequality constraints from the space $\mathcal{M}(\Omega) = C(\bar{\Omega})^*$ of regular Borel measures. The duality product between $\mathcal{M}(\Omega)$ and $C(\bar{\Omega})$ is denoted by $\langle \cdot, \cdot \rangle$. For a measure $\mu \in \mathcal{M}(\Omega)$ we will write as usual:

$$\mu \ge 0 \quad \Leftrightarrow \quad \langle \mu, f \rangle \ge 0 \quad \forall f \in C(\bar{\Omega}) \quad \text{with } f(x) \ge 0 \text{ in } \Omega.$$

According to, e.g., [8,9,28], we formulate the necessary optimality conditions in the following proposition.

Proposition 2.1 *Suppose that (q,u) is a locally optimal solution of (2.7) and that the Assumptions 1, 2, and 3 are fulfilled. Then, there exist an adjoint state $z \in W^{1,p}(\Omega)$ with $p < 2$ and Lagrange multipliers $\mu^+, \mu^- \in \mathcal{M}(\Omega)$, such that the following optimality system holds for $x = (q,u,z,\mu^+,\mu^-)$:*

$$\mathcal{L}_z'(x)(\varphi) = 0 \qquad \forall \varphi \in V, \tag{2.9a}$$

$$\mathcal{L}_u'(x)(\varphi) = 0 \qquad \forall \varphi \in V \cap W^{1,p'}(\Omega), \tag{2.9b}$$

$$\mathcal{L}_q'(x)(\xi) = 0 \qquad \forall \xi \in Q, \tag{2.9c}$$

$$\langle \mu^+, u_b - u \rangle = 0, \qquad \mu^+ \geq 0, \tag{2.9d}$$

$$\langle \mu^-, u - u_a \rangle = 0, \qquad \mu^- \geq 0, \tag{2.9e}$$

where $1/p + 1/p' = 1$.

The equation (2.9a) is equivalent to the state equation (2.5). The equation (2.9b) is called *adjoint equation* and can be explicitly rewritten as

$$a_u'(q,u)(\varphi,z) = J_u'(q,u)(\varphi) + \langle \mu^+ - \mu^-, \varphi \rangle \quad \forall \varphi \in V \cap W^{1,p'}(\Omega). \tag{2.10}$$

The regularity of the adjoint solution is natural due to the presence of regular Borel measures on the right-hand side, see, e.g., [8]. The equation (2.9c) is often called *gradient equation* and has the form

$$a_q'(q,u)(\xi,z) = J_q'(q,u)(\xi) \quad \forall \xi \in Q. \tag{2.11}$$

Remark 2.2 The Lagrange multipliers μ^+ and μ^- are searched for in the space of regular Borel measures which is dual to the space $C(\bar{\Omega})$, where the state constraints are formulated. Unlike in the case of problems with control or mixed control-state constraints one can in general not expect better regularity of the Lagrange multipliers, see, e.g., [29] for an example with a Dirac measure as a Lagrange multiplier.

Remark 2.3 Since the support of μ^+ and μ^- is disjoint, they could be replaced by a single measure $\mu := \mu^+ - \mu^-$, which is especially suitable for implementation purposes but will not be used in the analysis below. The conditions (2.9d) and (2.9e) can be replaced by

$$\langle \mu, u_b - u \rangle = \langle \mu, u - u_a \rangle = 0.$$

3 Discretization

For the efficient numerical solution of the optimal control problem under consideration, it is discretized on a sequence of locally refined meshes. The iterative construction of these meshes is guided by the error estimator to be derived in the next section. This section is devoted to the discretization of (2.7) on a single mesh, which is later on used on each locally refined mesh within the adaptive algorithm.

We consider a shape-regular mesh $\mathcal{T}_h$ consisting of quadrilateral cells K. The mesh parameter h is defined as a cellwise constant function by setting $h\big|_K = h_K$ and h_K is the diameter of K. On the mesh $\mathcal{T}_h$ we consider a conforming finite element space $V_h \subset V$ consisting of cellwise bilinear shape functions, see, e.g., [5,6] for details.

In order to ease the mesh refinement, we allow the cells to have nodes which lie on midpoints of edges of neighboring cells. But at most one of such *hanging nodes* is permitted per edge. Consideration of meshes with hanging nodes requires additional care. There are no degrees of freedom corresponding to these irregular nodes and the value of the finite element function is determined by pointwise interpolation. We refer, e.g., to [7] for implementation details.

Throughout we will denote the set of all nodes of $\mathcal{T}_h$ not belonging to Γ_1 by $\mathcal{N}_h = \{x_i\}$ and the corresponding nodewise basis functions of V_h by $\{\phi_i\}$. The space V_h will be used for the discretization of the state variable.

The discrete control variable will be searched for in a subspace $Q_h \subset Q$. Depending on the structure of the control space different possibilities can be considered: In the case of a finite dimensional control space Q one typically chooses $Q_h = Q$. If the space Q is defined as a space of functions on (a part of) Ω, e.g., $Q = L^2(\Omega)$, then the finite-dimensional space $Q_h \subset Q$ can be constructed as an analogue of V_h consisting of cellwise bilinear shape functions or as a space consisting of cellwise constant functions. Another possibility is to choose $Q_h = Q$ also in the case of an infinite-dimensional control space following the approach from [21]. We note however, that for the problem under consideration with $Q = L^2(\Omega)$, the latter approach is equivalent to the setting $Q_h = V_h$ due to the structure of the optimality system.

Remark 3.1 As pointed out, e.g., in [24,26], it might be desirable to use different meshes for the control and the state variable. The error estimator presented below can provide information for separate assessment of the errors due to the control and the state discretizations. The refinement then follows an equilibration strategy for both estimators, cf. [26].

For a given discrete control $q_h \in Q_h$ the solution $u_h \in V_h$ of the discrete state equation is determined by

$$a(q_h, u_h)(\varphi_h) = 0 \quad \forall \varphi_h \in V_h. \tag{3.1}$$

Due to Assumption 1 one can prove like on the continuous level that this equation has a unique solution $u_h \in V_h$ for each $q_h \in Q_h$. This fact defines the discrete solution operator $S_h \colon Q_h \to V_h$. This operator is twice continously differentiable.

For the discretization of the state constraints we use a nodewise interpolation operator $i_h \colon C(\bar{\Omega}) \to V_h$ and define the discrete admissible set as

$$V_{\mathrm{ad},h} = \left\{ u_h \in V_h \;\middle|\; i_h u_a \le u_h \le i_h u_b \text{ on } \bar{\Omega} \right\}.$$

Altogether, the discretized optimal control problem is formulated as follows:

$$(P_h)\quad \begin{cases} \text{Minimize} \quad J(q_h, u_h), \quad q_h \in Q_h, u_h \in V_h, \\ a(q_h, u_h)(\varphi_h) = 0 \quad \forall \varphi_h \in V_h, \\ i_h u_a(x) \le u_h(x) \le i_h u_b(x) \quad \forall x \in \bar{\Omega}. \end{cases} \tag{3.2}$$

Remark 3.2 Although the state constraints in (3.2) are formulated pointwise, they are equivalent to a finite number of inequality constraints:

$$u_{a,i} \le u_i \le u_{b,i}, \quad i = 1, \dots, \dim V_h,$$

where the coefficients $u_{a,i}$, u_i, and $u_{b,i}$ are determined by

$$i_h u_a = \sum_i u_{a,i}\phi_i, \quad i_h u_b = \sum_i u_{b,i}\phi_i, \quad \text{and} \quad u_h = \sum_i u_i \phi_i.$$

To ensure the existence of a solution to (3.2) we make an assumption similar to Assumption 2.

Assumption 4 *There exists $\hat{q}_h \in Q_h$ so that $S_h(\hat{q}_h) \in V_{ad,h}$ and there holds*

$$\inf_{q_h \in Q_h, S_h(q_h) \in V_{ad,h}} J(q_h, S_h(q_h)) > -\infty.$$

Remark 3.3 Although, in general, the above conditions have to be assumed to ensure the existence of a solution to (3.2), Assumption 4 can follow from Assumption 2 in several situations and for h small enough, see [28] for details.

As on the continuous level, Assumption 4 guarantees the existence of a solution, which is in general not unique. To pose necessary optimality conditions for a discrete locally optimal solution (q_h, u_h) an analogue of the continuous Slater condition (Assumption 3) has to be required:

Assumption 5 *For the discrete locally optimal point (q_h, u_h) there exists $\tilde{q}_h \in Q_h$ such that*

$$S_h(q_h) + S_h'(q_h)(\tilde{q}_h - q_h) \in \operatorname{int}(V_{ad,h}).$$

For the formulation of the optimality system we use the fact, that the inequality constraints in (3.2) are equivalent to a finite number of constraints formulated for all nodes $x_i \in \mathcal{N}_h$, cf. Remark 3.2. Therefore a finite number of Lagrange-multipliers $\{\mu_i^{\pm}\}$ has to be introduced. In order to write the optimality system similarly to the continuous case, we introduce a space of discrete Lagrange multipliers:

$$\mathcal{M}_h = \left\{ \mu_h = \sum_i \mu_i \delta_{x_i} \;\middle|\; x_i \in \mathcal{N}_h \right\} \subset \mathcal{M}(\Omega),$$

where δ_{x_i} denotes the Dirac measure concentrated at the point x_i. Then, the optimality conditions are formulated utilizing the same Lagrangian function (2.8) in the following Proposition.

Proposition 3.1 *Suppose that* (q_h, u_h) *is a locally optimal solution of* (3.2) *and that the Assumptions 1, 4, and 5 are fulfilled. Then, there exist a discrete adjoint state* $z_h \in V_h$ *and discrete Lagrange multipliers* $\mu_h^+, \mu_h^- \in \mathcal{M}_h$, *such that the following optimality system holds for* $x_h = (q_h, u_h, z_h, \mu_h^+, \mu_h^-)$*:*

$$\mathcal{L}_z'(x_h)(\varphi_h) = 0 \qquad \forall \varphi_h \in V_h, \tag{3.3a}$$
$$\mathcal{L}_u'(x_h)(\varphi_h) = 0 \qquad \forall \varphi_h \in V_h, \tag{3.3b}$$
$$\mathcal{L}_q'(x_h)(\xi_h) = 0 \qquad \forall \xi_h \in Q_h, \tag{3.3c}$$
$$\langle \mu_h^+, u_b - u_h \rangle = 0, \qquad \mu_h^+ \geq 0, \tag{3.3d}$$
$$\langle \mu_h^-, u_h - u_a \rangle = 0, \qquad \mu_h^- \geq 0. \tag{3.3e}$$

Proof The proof can be done like on the continuous level, cf. [11,22], leading to the equations (3.3a)–(3.3c) and the complementarity conditions of the following form

$$\langle \mu_h^+, i_h u_b - u_h \rangle = 0, \quad \mu_h^+ \geq 0,$$

which is equivalent to (3.3d) due to the structure of $\mathcal{M}_h$. Equation (3.3e) is obtained in the same way. □

Rewriting the equations (3.3a)–(3.3c) more explicitly we obtain the discrete state equation (3.1), the discrete adjoint equation for $z_h \in V_h$

$$a_u'(q_h, u_h)(\varphi_h, z_h) = J_u'(q_h, u_h)(\varphi_h) + \langle \mu_h^+ - \mu_h^-, \varphi_h \rangle \quad \forall \varphi_h \in V_h,$$

and the discrete gradient equation

$$a_q'(q_h, u_h)(\xi_h, z_h) = J_q'(q_h, u_h)(\xi_h) \quad \forall \xi_h \in Q_h. \tag{3.4}$$

4 A posteriori error estimation

In this section we derive an error estimator for the error between the solution (q, u) of (2.7) and the solution (q_h, u_h) of the discretized problem (3.2) with respect to the cost functional, i.e.

$$J(q, u) - J(q_h, u_h).$$

This error estimator will be used for two purposes: Firstly, to obtain quantitative information on the discretization error and secondly, to guide an adaptive mesh refinement process in order to reduce the error in an efficient way. To this end the error estimator can be localized to cellwise (or nodewise) contributions, called error indicators, see, e.g., [2] for details on adaptive algorithms.

As the first step we provide the following Lemma, which is an extension of the result in [2].

Lemma 4.1 *Let $x = (q, u, z, \mu^+, \mu^-)$ be a solution of the optimality system* (2.9a)–(2.9e) *and $x_h = (q_h, u_h, z_h, \mu_h^+, \mu_h^-)$ be a solution of the discrete optimality system* (3.3a)–(3.3e). *Then there holds:*

$$J(q,u) - J(q_h,u_h) = \frac{1}{2}\mathcal{L}'(x)(x - x_h) + \frac{1}{2}\mathcal{L}'(x_h)(x - x_h) + \mathcal{R},$$

where the remainder term $\mathcal{R}$ is of third order in the error $e = x - x_h$ and is given by

$$\mathcal{R} = \frac{1}{2}\int_0^1 \mathcal{L}'''(x_h + se)(e,e,e)\cdot s\cdot(s-1)\,ds.$$

Proof Using the fact that the pair (q, u) fulfills the state equation (2.5) and employing the complementarity conditions (2.9d)–(2.9e) we obtain

$$J(q,u) = \mathcal{L}(x).$$

Due to the Galerkin type discretization the same argument is possible on the discrete level leading to

$$J(q,u) - J(q_h,u_h) = \mathcal{L}(x) - \mathcal{L}(x_h) = \int_0^1 \mathcal{L}'(x_h + se)(e)\,ds.$$

Approximating this integral by the trapezoidal rule we complete the proof like in [2]. □

In the next lemma we give a more explicit form of the above error representation.

Lemma 4.2 *Let $x = (q, u, z, \mu^+, \mu^-)$ be a solution of the optimality system* (2.9a)–(2.9e) *and $x_h = (q_h, u_h, z_h, \mu_h^+, \mu_h^-)$ be a solution of the discrete optimality system* (3.3a)–(3.3e). *Then there holds:*

$$\begin{aligned} J(q,u) - J(q_h,u_h) =& \frac{1}{2}\Big(J_u'(q_h,u_h)(u-u_h) - a_u'(q_h,u_h)(u-u_h,z_h) \\ &+ J_q'(q_h,u_h)(q-\tilde{q}_h) - a_q'(q_h,u_h)(q-\tilde{q}_h,z_h) \\ &- a(q_h,u_h)(z-\tilde{z}_h) - \langle \mu^+ - \mu^-, u-u_h\rangle\Big) + \mathcal{R}, \end{aligned} \tag{4.1}$$

where $\tilde{q}_h \in Q_h$ and $\tilde{z}_h \in V_h$ can be arbitrarily chosen and $\mathcal{R}$ is given as in Lemma 4.1.

Proof Starting from the representation in Lemma 4.1 we obtain

$$\begin{aligned} \mathcal{L}'(x)(x-x_h) =& \mathcal{L}_z'(x)(z-z_h) + \mathcal{L}_u'(x)(u-u_h) + \mathcal{L}_q'(x)(q-q_h) \\ &+ \mathcal{L}_{\mu^+}'(x)(\mu^+ - \mu_h^+) + \mathcal{L}_{\mu^-}'(x)(\mu^- - \mu_h^-). \end{aligned} \tag{4.2}$$

The first, the second, and the third terms vanish due to the state equation (2.5), adjoint equation (2.10), and the gradient equation (2.11) respectively. Note that for employing (2.10) we have to check that $u - u_h \in$

$W^{1,p'}(\Omega)$, which is obvious for u_h due to the fact that $V_h \subset W^{1,\infty}(\Omega)$ and is shown for u in Lemma 2.1. Using the explicit form of $\mathcal{L}'_{\mu^\pm}$ we get

$$\mathcal{L}'(x)(x - x_h) = -\langle \mu^+ - \mu_h^+, u_b - u\rangle - \langle \mu^- - \mu_h^-, u - u_a\rangle.$$

For the second term in the error representation in Lemma 4.1 we proceed as follows:

$$\begin{aligned}\mathcal{L}'(x_h)(x - x_h) =& \mathcal{L}'_z(x_h)(z - z_h) + \mathcal{L}'_u(x_h)(u - u_h) + \mathcal{L}'_q(x_h)(q - q_h)\\ &+ \mathcal{L}'_{\mu^+}(x_h)(\mu^+ - \mu_h^+) + \mathcal{L}'_{\mu^-}(x_h)(\mu^- - \mu_h^-),\end{aligned} \tag{4.3}$$

where for the first term in (4.3) we obtain:

$$\mathcal{L}'_z(x_h)(z - z_h) = -a(q_h, u_h)(z - z_h) = -a(q_h, u_h)(z - \tilde{z}_h)$$

with an arbitrary $\tilde{z}_h \in V_h$. The replacement of z_h by $\tilde{z}_h$ in the residual of the state equation follows by the Galerkin orthogonality apparent from (3.1). The second term in (4.3) gives

$$\mathcal{L}'_u(x_h)(u-u_h) = J'_u(q_h, u_h)(u-u_h) - a'_u(q_h, u_h)(u-u_h, z_h) + \langle \mu_h^+ - \mu_h^-, u-u_h\rangle,$$

the third term results in

$$\begin{aligned}\mathcal{L}'_q(x_h)(q - q_h) =& J'_q(q_h, u_h)(q - q_h) - a'_q(q_h, u_h)(q - q_h, z_h)\\ =& J'_q(q_h, u_h)(q - \tilde{q}_h) - a'_q(q_h, u_h)(q - \tilde{q}_h, z_h)\end{aligned}$$

with an arbitrary $\tilde{q}_h \in Q_h$ due to the discrete gradient equation (3.4). For the last two terms in (4.3) we obtain directly

$$\begin{aligned}\mathcal{L}'_{\mu^+}(x_h)(\mu^+ - \mu_h^+) + \mathcal{L}'_{\mu^-}(x_h)(\mu^- - \mu_h^-) =&\\ - \langle \mu^+ - \mu_h^+, u_b - u_h\rangle - \langle \mu^- - \mu_h^-, u_h - u_a\rangle.&\end{aligned}$$

Summing over all terms involving μ^+ and μ_h^+ and exploiting the complementarity conditions (2.9d) and (3.3d) we have

$$\begin{aligned}-\langle \mu^+ - \mu_h^+, u_b - u_h\rangle - \langle \mu^+ - \mu_h^+, u_b - u\rangle + \langle \mu_h^+, u - u_h\rangle&\\ = -\langle \mu^+, u_b - u_h\rangle + \langle \mu_h^+, u_b - u\rangle + \langle \mu_h^+, u - u_h\rangle&\\ = -\langle \mu^+, u_b - u\rangle + \langle \mu^+, u_h - u\rangle + \langle \mu_h^+, u_b - u_h\rangle&\\ = \langle \mu^+, u_h - u\rangle.&\end{aligned}$$

Similarly we obtain for the terms involving μ^- and μ_h^-:

$$-\langle \mu^- - \mu_h^-, u_h - u_a\rangle - \langle \mu^- - \mu_h^-, u - u_a\rangle - \langle \mu_h^-, u - u_h\rangle = -\langle \mu^-, u_h - u\rangle.$$

Summing all these terms up yields the desired result. □

Remark 4.1 Note, that the errors $q - q_h$ and $z - z_h$ are replaced by $q - \tilde{q}_h$ and $z - \tilde{z}_h$ respectively, which can be seen as interpolation errors. However, the error in the state variable $u - u_h$ can not be directly replaced in this way due to the structure of the optimality system.

Remark 4.2 The residual of the gradient equation

$$J'_q(q_h, u_h)(q - \tilde{q}_h) - a'_q(q_h, u_h)(q - \tilde{q}_h, z_h)$$

in the above error representation can be shown to be zero, if either $Q_h = Q$ or the control operator B is chosen as the identity on $Q = L^2(\Omega)$ and $V_h \subset Q_h$.

Remark 4.3 The explicit form of the remainder term $\mathcal{R}$ in (4.1) is given as

$$\begin{aligned}\mathcal{R} = \frac{1}{2}\int_0^1 \Big(&\Psi'''(u_h+se_u)e_u^3 - (d'''(u_h+se_u)e_u^3, z_h+se_z) - 2(d''(u_h+se_u)e_u^2, e_z)\\ &- \langle b'''(u_h+se_u)e_u^3, z_h+se_z\rangle_{\Gamma_2} - 2\langle b''(u_h+se_u)e_u^2, e_z\rangle_{\Gamma_2}\Big) s(s-1)\,ds,\end{aligned}$$

where the error in the state variable $e_u = u - u_h$ and the error in the adjoint variable $e_z = z - z_h$ are involved. Note, that the errors in the Lagrange multipliers $\mu^{\pm}$ do not appear explicitly. Therefore the remainder term $\mathcal{R}$ can usually be neglected as a higher order term.

Remark 4.4 To illustrate the behavior of the remainder term $\mathcal{R}$ we discuss the following two special cases:

1. In the case of a linear-quadratic problem (i.e. $\Psi(\cdot)$ quadratic in u, $d(\cdot,\cdot)$ and $b(\cdot,\cdot)$ linear in the second argument) the remainder term $\mathcal{R}$ vanishes.
2. Let $\Psi(\cdot)$ be quadratic in u, $d(x,u) = u^3$ and $b(x,u) = u$. Then the remainder term has the form

$$\mathcal{R} = \frac{1}{4}(e_u^3, z + z_h) + \frac{1}{2}(e_u^2 e_z, u_h + u).$$

As in [11] we expect that $\|q_h\|_Q$ and $\|\mu_h\|_{\mathcal{M}(\Omega)}$ are bounded uniformly in h. Using stability of the Ritz-Projection in $W^{1,p}(\Omega)$, cf. [6, Corollary 8.6.3], we obtain that $\|u_h\|_{L^\infty(\Omega)}$ and $\|z_h\|_{W^{1,p}(\Omega)}$ are also bounded uniformly in h. A possible estimate is then

$$|\mathcal{R}| \le c\,\|e_u\|^3_{L^{3+\epsilon}(\Omega)} + c\,\|e_u\|^2_{L^4(\Omega)}\,\|e_z\|_{L^2(\Omega)}$$

with an arbitrary $\epsilon > 0$ and an h-independent constant c. To our knowledge there are no rigorous results on a priori error analysis for nonlinear optimal control problems with state constraints. Nevertheless the above bound for the remainder term is expected to be a higher order term with respect to the error in the cost functional.

The error representation formula (4.1) still contains continuous solutions (q, u, z, μ^+, μ^-), which have to be approximated in order to obtain a computable error estimator. To this end, we employ the technique of interpolation in higher order finite element spaces, which is observed to work very successfully in the context of a posteriori error estimation, see, e.g. [2, 3, 26, 36]. We use operators $\Pi_h \colon V_h \to \tilde{V}_h$ and $\Pi_h^q \colon Q_h \to \tilde{Q}_h$ with suitable finite element spaces $\tilde{V}_h \neq V_h$ and $\tilde{Q}_h \neq Q_h$ so that $\Pi_h u_h$ is assumed to be an asymptotically better approximation of u than u_h. Such operators can be constructed for example by the interpolation of the computed bilinear solutions into the space of biquadratic finite elements on patches consisting of four cells each. A typical structure of a patch and the degrees of freedom used for this biquadratic interpolation are illustrated in Figure 4.1.

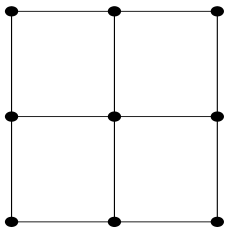


Fig. 4.1 Mesh points corresponding to the nine degrees of freedom for the construction of the patchwise biquadratic interpolation

Using the operators Π_h and Π_h^q we approximate all terms in (4.1) not involving the Lagrange multipliers as follows:

$$\begin{aligned} J'_u(q_h, u_h)(u - u_h) - a'_u(q_h, u_h)(u - u_h, z_h) \\ \approx J'_u(q_h, u_h)(\Pi_h u_h - u_h) - a'_u(q_h, u_h)(\Pi_h u_h - u_h, z_h), \end{aligned}$$

$$\begin{aligned} J'_q(q_h, u_h)(q - \tilde{q}_h) - a'_q(q_h, u_h)(q - \tilde{q}_h, z_h) \\ \approx J'_q(q_h, u_h)(\Pi_h^q q_h - \tilde{q}_h) - a'_q(q_h, u_h)(\Pi_h^q q_h - \tilde{q}_h, z_h), \end{aligned}$$

and

$$-a(q_h, u_h)(z - \tilde{z}_h) \approx -a(q_h, u_h)(\Pi_h z_h - \tilde{z}_h).$$

The construction of an approximation of the term $\langle \mu^+ - \mu^-, u - u_h \rangle$ is more involved, since no direct analog of higher order interpolation can be used for the multipliers $\mu^\pm$, being Borel measures. A simple approximation $\mu^\pm \approx \mu_h^\pm$ is not directly useful, since $\langle \mu_h^+ - \mu_h^-, \Pi_h u_h - u_h \rangle$ is identical to zero if the biquadratic interpolation is taken for Π_h. In [16] the authors use therefore another construction for Π_h. Instead, we express the term $\langle \mu^+ - \mu^-, u - u_h \rangle$ using the adjoint equation (2.10) leading to

$$\langle \mu^+ - \mu^-, u - u_h \rangle = -J'_u(q, u)(u - u_h) + a'_u(q, u)(u - u_h, z),$$

and then we have an expression which does not involve any Lagrange multiplier and can be approximated like above:

$$\begin{aligned}&\langle \mu^+ - \mu^-, u - u_h \rangle \\ &\approx -J'_u(\Pi_h^q q_h, \Pi_h u_h)(\Pi_h u_h - u_h) + a'_u(\Pi_h^q q_h, \Pi_h u_h)(\Pi_h u_h - u_h, \Pi_h z_h).\end{aligned}$$

Taking all these considerations into account, the error representation formula (4.1) motivates the definition of the error estimator as

$$\begin{aligned}\eta := \frac{1}{2}\Big(& J'_q(q_h, u_h)(\Pi_h^q q_h - q_h) - a'_q(q_h, u_h)(\Pi_h^q q_h - q_h, z_h) \\ & + J'_u(q_h, u_h)(\Pi_h u_h - u_h) - a'_u(q_h, u_h)(\Pi_h u_h - u_h, z_h) \\ & + J'_u(\Pi_h^q q_h, \Pi_h u_h)(\Pi_h u_h - u_h) - a'_u(\Pi_h^q q_h, \Pi_h u_h)(\Pi_h u_h - u_h, \Pi_h z_h) \\ & - a(q_h, u_h)(\Pi_h z_h - z_h)\Big).\end{aligned} \tag{4.4}$$

Remark 4.5 In order to use this error estimator as an indicator for mesh refinement, we have to localize it to cellwise or nodewise contributions. A direct localization of the terms like

$$J'_u(q_h, u_h)(\Pi_h u_h - u_h) - a'_u(q_h, u_h)(\Pi_h u_h - u_h, z_h)$$

leads, in general, to a wrong order of the local contributions (overestimation) due to oscillatory behavior of the residual terms. To overcome this, one may integrate the residual terms by part, see, e.g., [2], or use a filtering operator, see, e.g., [33] for details. In the numerical examples below, the latter possibility is used.

Remark 4.6 The use of operators Π_h and Π_h^q for estimation of local approximation errors can be rigorously justified only for smooth solutions q, u, z employing super-convergence effects. However, the adjoint solution z and consequently the control variable q possess in general only reduced regularity ($z \in W^{1,p}(\Omega)$, $p < 2$). Nevertheless, we expect a good behaviour of the proposed error estimator, since the operators Π_h and Π_h^q are defined locally and the regions, where the adjoint state z is not smooth, are usually strongly localized.

5 Numerical examples

In this section we present numerical results illustrating the behavior of the error estimator and of the adaptive algorithm developed in this paper. To this end we consider several example configurations which correspond to prototypical types of the structure of the solution to (2.7). For all examples the optimal control problems are solved on sequences of meshes produced either by uniform refinement or by the refinement according to the proposed error estimator. This allows us to investigate the saving in degrees of freedom

required to reach a given error tolerance if local refinement is employed. Moreover, we study the quality of the error estimation by calculating the effectivity index of the error estimator defined by

$$I_{\text{eff}} = \frac{J(q,u) - J(q_h,u_h)}{\eta}\,. \tag{5.1}$$

On each mesh the discrete optimal control problem (3.2) is solved using a primal-dual-active-set strategy implemented in the software packages RoDoBo [31] and Gascoigne [14]. For visualization we used the visualization tool VisuSimple [37].

5.1 Example 1

We consider the following optimal control problem governed by a linear state equation

$$(P_1)\quad \begin{cases} \text{Minimize} & J(u,q) = \dfrac{1}{2}\|u-u_d\|^2_{L^2(\Omega)} + \dfrac{1}{2}\|q\|^2_{L^2(\Omega)}, \\ -\Delta u = q + f & \text{in } \Omega, \\ u = 0 & \text{on } \Gamma_1, \\ \partial_n u = 0 & \text{on } \Gamma_2, \\ u \le u_b & \text{in } \bar{\Omega}, \end{cases}$$

where $\Omega = (0,1)^2$ is the unit square,

$$\Gamma_1 = \{\, x = (x_1,x_2) \in \partial\Omega \mid x_1 = 0 \,\} \quad \text{and} \quad \Gamma_2 = \partial\Omega \setminus \Gamma_1.$$

The choice of the data u_d, u_b and f is constructed in such a way, that the optimal solution (q,u), the adjoint state z and the Lagrange multiplier μ^+ are known and possess a typical structure described, e.g., in [20]. That is, the active set A^+ defined by

$$A^+ = \left\{\, x \in \bar{\Omega} \;\middle|\; u(x) = u_b(x) \,\right\},$$

has non empty interior and the Lagrange multiplier μ^+ consists of two parts $\mu^+ = \mu_1^+ + \mu_2^+$, where $\mu_2^+ \in L^\infty(\Omega)$ and μ_1^+ is a line measure concentrated on a part of the boundary of A^+. Moreover, we aim on obtaining this structure with smooth data $u_b, u_d \in C^2(\Omega)$, $f \in C^1(\Omega)$ resulting in the following construction: We introduce the parameters $s \in (0,1)$, $m < s^{-3}$, $b > 0$ and set

$$u_d(x_1,x_2) = \begin{cases} \frac{x_1^3}{s^3} - 3\frac{x_1^2}{s^2} + 3\frac{x_1}{s} + 2, & x_1 < s \\ -\frac{3m}{4(1-s)}(x_1-s)^4 + m(x_1-s)^3 + 3, & x_1 \ge s, \end{cases}$$

$$u_b(x_1,x_2) = \begin{cases} 1, & x_1 < s \\ -\frac{3m}{4(1-s)}(x_1-s)^4 + m(x_1-s)^3 + 1, & x_1 \ge s, \end{cases}$$

and

$$f(x_1,x_2)=\begin{cases}\frac{6}{s^2}-6mx_1+x_1(x_1-2)+b(1-s)x_1, & x_1<s\\ (1-r)x_1^2+(b-\frac{18ms}{1-s}-2-6m)x_1+\frac{6}{s^2}-rs^2, & x_1\geq s,\end{cases}$$

where $r=\frac{b}{2}-\frac{9m}{1-s}$. The corresponding solution (q,u), the adjoint state z and the Lagrange multiplier μ^+ have the following explicit form

$$q(x_1,x_2)=\begin{cases}-x_1(x_1-2)-(\frac{6}{s^3}-6m)x_1-b(1-s)x_1, & x_1<s\\ -x_1(x_1-2)-\frac{6}{s^2}+6ms+\frac{b}{2}x_1^2-bx_1+\frac{b}{2}s^2, & x_1\geq s,\end{cases}$$

$$u(x_1,x_2)=\begin{cases}\frac{x_1^3}{s^3}-3\frac{x_1^2}{s^2}+3\frac{x_1}{s}, & x_1<s\\ -\frac{3m}{4(1-s)}(x_1-s)^4+m(x_1-s)^3+1, & x_1\geq s,\end{cases}$$

$z=-q$, and $\mu^+=\mu_1^+ + \mu_2^+$ with

$$\langle\mu_1^+,\varphi\rangle=\left(\frac{6}{s^3}-6m\right)\int_0^1\varphi(s,x_2)\,dx_2\ \forall\varphi\in C(\bar\Omega),\quad \mu_2^+=\begin{cases}0, & x_1<s\\ b, & x_1\geq s.\end{cases}$$

All quantities involved in the above definitions are independent of x_2 leading to a one-dimensional structure of the solution. The active set A^+ is obviously given as

$$A^+=\left\{\,(x_1,x_2)\in\bar\Omega\;\middle|\;x_1\geq s\,\right\}.$$

The control operator B is in this example the identity on $L^2(\Omega)$ and the control discretization is chosen as the discretization of the state variable, i.e., $Q_h=V_h$. Therefore, the part of the error estimator consisting of the residual of the gradient equation (2.11) vanishes, cf. Remark 4.2. Due to the linearity of the state equation, the remainder term $\mathcal{R}$ in the error representation (4.1) vanishes too.

We present the results of the computations for two sets of parameters. The first parameter choice is $b=50$, $m=-2$, and $s=0.125$. The development of the discretization errors and the effectivity indices defined as in (5.1) are given in Table 5.1 for both uniform and local mesh refinement. In both cases we observe good agreement of the error and the error estimator after two refinement steps. In Figure 5.1 the dependence of the discretization error on the number of degrees of freedom N is shown for uniform and local refinements.

The above parameter choice however does not correspond to the most general case, since the line $\{x=s\}$ at which the measure is concentrated is always a grid line. Avoiding this effect, for a second calculation we choose $s=0.3$. The corresponding results can be seen in Table 5.2 and Figure 5.2.

Table 5.1 Development of discretization errors and of the efficiency indices for $s = 0.125$ for P_1

(a) local

N	$J(q,u) - J(q_h,u_h)$	I_{eff}
25	1.37e+3	4.54
55	-5.93e-2	0.00
113	-8.56e-03	0.41
189	-1.48e-02	0.94
403	-3.90e-03	0.94
1233	-1.72e-03	0.93
4241	-6.60e-04	0.96

(b) uniform

N	$J(q,u) - J(q_h,u_h)$	I_{eff}
25	1.37e+3	4.54
81	-6.62e-02	0.00
289	-1.59e-02	0.98
1089	-3.92e-03	0.96
4225	-9.70e-04	0.97

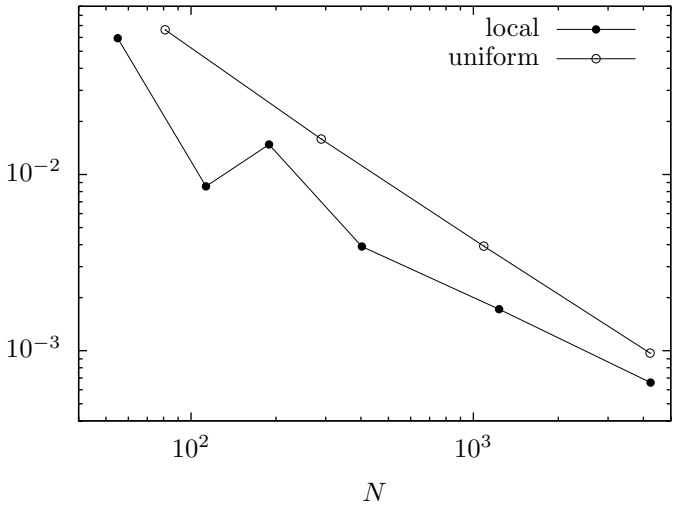


Fig. 5.1 Discretization error vs. degrees of freedom for the two refinement strategies for $s = 0.125$ for P_1

Table 5.2 Development of discretization errors and of the efficiency indices for $s = 0.3$ for P_1

(a) local refinement

N	$J(q,u) - J(q_h,u_h)$	$I_{\texttt{eff}}$
25	3.02e+00	0.65
55	1.33e+00	8.74
139	1.15e-01	1.71
403	2.56e-02	-4.68
955	-3.88e-03	0.96
2185	-1.24e-03	0.78
5125	-2.99e-05	0.81

(b) uniform refinement

N	$J(q,u) - J(q_h,u_h)$	$I_{\texttt{eff}}$
25	3.02e+00	0.65
81	1.32e+00	8.03
289	1.33e-01	1.52
1089	3.03e-02	-0.45
4225	6.10e-04	1.23

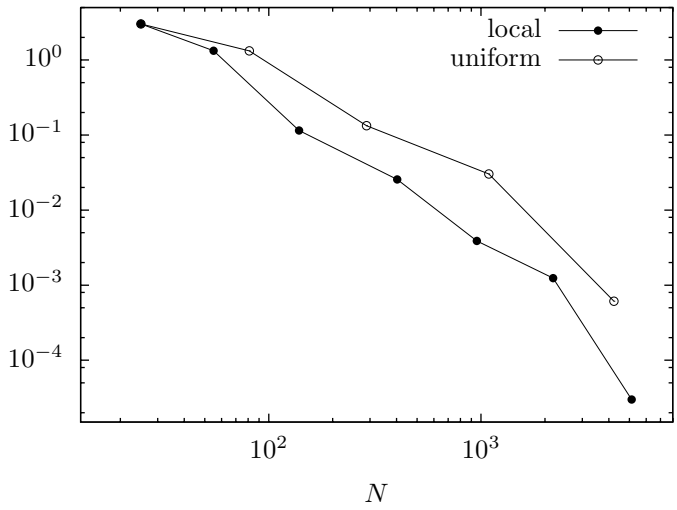


Fig. 5.2 Discretization error vs. degrees of freedom for the two refinement strategies for $s = 0.3$ for P_1

An example of a locally refined mesh for this case is shown in Figure 5.3. We observe stronger refinement close to the line where the measure part of the Lagrange multiplier μ^+ is concentrated.

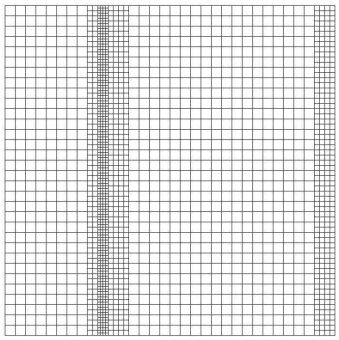


Fig. 5.3 An example of a locally refined mesh for $s = 0.3$ for P_1

In this example the choice of the parameter m secured that the multiplier part of the error representation (4.1), i.e., $-\frac{1}{2}\langle \mu^+ - \mu^-, u - u_h \rangle$, has a significant size compared to the other parts. The choice of the parameter b was less significant.

5.2 Example 2

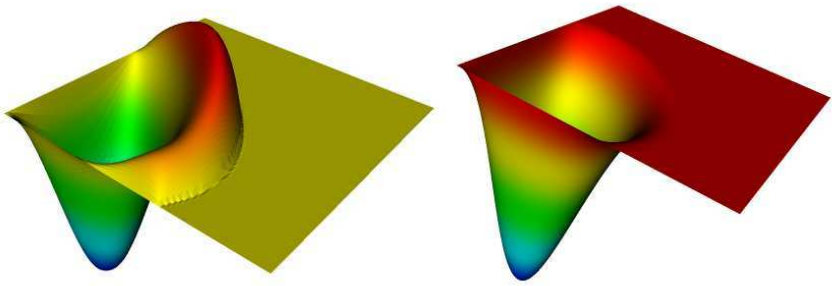
In this section we consider an optimal control problem governed by a semilinear elliptic equation on the unit square $\Omega = (0,1)^2$:

$$(P_2)\quad \begin{cases} \text{Minimize} & J(u,q) = \dfrac{1}{2}\|u-u_d\|^2_{L^2(\Omega)} + \dfrac{\alpha}{2}\|q\|^2_{L^2(\Omega)}, \\ -\Delta u + u^3 = q + f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \\ u_a \leq u \leq u_b & \text{in } \bar{\Omega}, \end{cases}$$

with $\alpha = 0.001$, $f = 0$, $u_b = 0$, and

$$u_d = 16x(1-x)^2(x-y) + \frac{3}{5}, \quad u_a = -0.08 - 4\left(x - \frac{1}{4}\right)^2 - 4\left(y - \frac{27}{32}\right)^2.$$

For this example problem no exact solution is avaliable. Therefore we throughout use the value $J(q_{h^*}, u_{h^*})$ computed on a very fine mesh $\mathcal{T}_{h^*}$ as an approximation of the exact value $J(q,u)$. The plots of the optimal control and state for this example are shown in Figure 5.4.



(a) control variable (b) state variable

Fig. 5.4 Plots of the optimal control and the optimal state for (P_2)

We observe that the active set A^+ corresponding to the upper bound is a two-dimensional set with nonempty interior and the active set A^- contains apparently only one point. A typical locally refined mesh which captures the feature of the problem under consideration is shown in Figure 5.5. The development of the discretization errors and the effectivity indices defined as in (5.1) are given in Table 5.3 for both uniform and local mesh refinement. The comparison of both refinement strategies with respect to the required number of degrees of freedom to reach a given error tolerance is done in Figure 5.6.

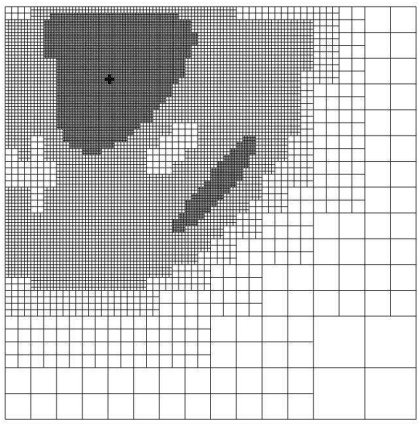


Fig. 5.5 Example of a locally refined mesh for (P_2)

Table 5.3 Development of discretization errors and of the efficiency indices for (P_2)

(a) local refinement

N	$J(q,u) - J(q_h,u_h)$	I_{eff}
25	5.38e-04	-1.41
41	-1.16e-04	0.43
99	-4.48e-05	0.33
245	-2.68e-05	0.60
541	-1.04e-05	0.56
1459	-6.04e-06	0.89
4429	-1.54e-06	0.83
13107	-5.01e-07	0.89

(b) uniform refinement

N	$J(q,u) - J(q_h,u_h)$	I_{eff}
25	5.38e-04	-1.41
81	-1.58e-04	0.62
289	-6.18e-05	0.87
1089	-1.58e-05	0.87
4225	-3.99e-06	0.89
16641	-7.45e-07	0.66

5.3 Example 3

Finally, we consider an example that has been presented in [10], and in similar form in [29] before. Here, on the unit circle $\Omega = B_1(0)$ as the domain, the problem reads

$$(P_3) \quad \begin{cases} \text{Minimize} & J(u,q) = \dfrac{1}{2}\|u - u_d\|^2_{L^2(\Omega)} + \dfrac{\alpha}{2}\|q\|^2_{L^2(\Omega)}, \\ -\Delta u = q & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \\ u \le u_b & \text{in } \bar{\Omega}, \end{cases}$$

with $\alpha = 0.01$, and the data given in polar coordinates

$$u_d = \frac{1}{2\pi\alpha}\left(\frac{1}{4} - \frac{1}{4}r^2 + \frac{1}{4}r^2\log(r)\right), \quad u_b = \frac{1}{2\pi\alpha}\left(\frac{1}{4} - \frac{r}{2}\right).$$

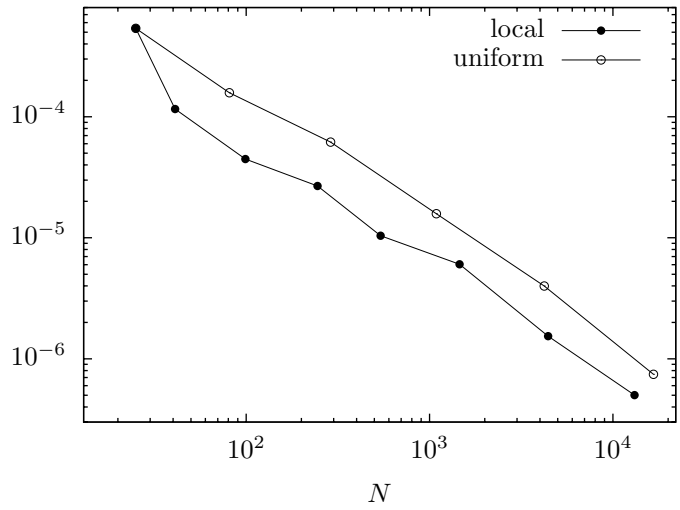


Fig. 5.6 Discretization error vs. degrees of freedom for the two refinement strategies for (P_2)

As for example 1, we are in the situation of knowing the exact solution in algebraic form:

$$u = \frac{1}{2\pi\alpha}\left(\frac{1}{4} - \frac{1}{4}r^2 + \frac{1}{4}r^2\log(r)\right), \quad q = \frac{1}{2\pi\alpha}\log(r), \quad z = -\frac{1}{2\pi}\log(r)$$

and $\mu = \delta_0$, where δ_0 denotes the Dirac measure concentrated in the origin. We can see that the active set consists of one point only, and the adjoint state has a regularity of $W^{1,p}(\Omega)$ with $p < 2$, but $z \notin H^1(\Omega)$. Note, that the data $(u_d,\, u_b)$ is smooth, and the singularity is caused by state constraints and not by some singularities of the data.

The convergence of the error in the functional value can be seen in Figure 5.7. The graphs point to a different order of convergence for the two discretization strategies.

Although the local discretization strategy leads to a much better convergence of the functional value, we observe that the effectivity indices are not close to 1. They are even smaller than zero, but stay bounded in their absolute values.

Conclusions

In this paper we derived a posteriori error estimates for the finite element discretization of a class of nonlinear optimal control problems with state constraints. The error estimator assesses the discretization error with respect to the cost functional. The terms in the estimator which involve Lagrange multipliers being regular Borel measures are reformulated utilizing the continuous formulation of the adjoint equation.

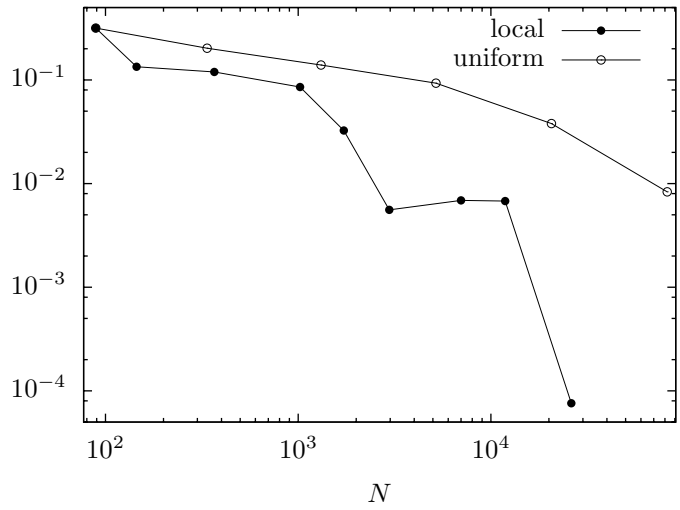


Fig. 5.7 Discretization error vs. degrees of freedom for the two refinement strategies for (P_3)

The numerical results show mostly good agreement between the estimated and real discretization errors. Moreover, the adaptive algorithm for local mesh refinement allows for substantial savings in degrees of freedom required for reaching a given tolerance level.

References

1. Roland Becker, Hartmut Kapp, and Rolf Rannacher. Adaptive finite element methods for optimal control of partial differential equations: Basic concepts. *SIAM J. Control Optim.*, 39(1):113–132, 2000.
2. Roland Becker and Rolf Rannacher. An optimal control approach to a posteriori error estimation. In Arieh Iserles, editor, *Acta Numerica 2001*, pages 1–102. Cambridge University Press, 2001.
3. Roland Becker and Boris Vexler. A posteriori error estimation for finite element discretizations of parameter identification problems. *Numer. Math.*, 96(3):435–459, 2004.
4. Roland Becker and Boris Vexler. Mesh refinement and numerical sensitivity analysis for parameter calibration of partial differential equations. *J. Comp. Physics*, 206(1):95–110, 2005.
5. Dietrich Braess. *Finite Elements: Theory, Fast Solvers and Applications in Solid Mechanics*. Cambridge University Press, Cambridge, 2007.
6. Susanne C. Brenner and L. Ridgway Scott. *The mathematical theory of finite element methods*. Springer Verlag, Berlin, Heidelberg, New York, 2002.
7. Graham F. Carey and J. Tinsley Oden. *Finite Elements. Computational Aspects*, volume 3. Prentice-Hall, 1984.
8. Eduardo Casas. Control of an elliptic problem with pointwise state constraints. *SIAM J. Control Optim.*, 24:1309–1318, 1986.

9. Eduardo Casas. Boundary control of semilinear elliptic equations with pointwise state constraints. *SIAM J. Control Optim.*, 34:933–1006, 1993.
10. Svetlana Cherednichenko, Klaus Krumbiegel, and Arnd Rösch. Error estimates for the lavrentiev regularization of elliptic optimal control problems. *Inverse Problems*, 2008. accepted.
11. Klaus Deckelnick and Michael Hinze. Convergence of a finite element approximation to a state constrained elliptic control problem. *SIAM J. Numer. Anal.*, 35:1937–1953, 2007.
12. Klaus Deckelnick and Michael Hinze. A finite element approximation to elliptic control problems in the presence of control and state constraints. 2007. submitted.
13. Kenneth Eriksson, Don Estep, Peter Hansbo, and Claes Johnson. Introduction to adaptive methods for differential equations. In Arieh Iserles, editor, *Acta Numerica 1995*, pages 105–158. Cambridge University Press, 1995.
14. The finite element toolkit GASCOIGNE. `http://www.gascoigne.uni-hd.de`.
15. Pierre Grisvard. *Singularities in Boundary Value Problems*. Springer-Verlag, Masson, Paris, Berlin, 1992.
16. Andreas Günther and Michael Hinze. A posteriori error control of a state constrained elliptic control problem. *J. Numer. Math.*, 2008. to appear.
17. Michael Hintermüller, Roland H.W. Hoppe, Yuri Iliash, and Michael Kieweg. An a posteriori error analysis of adaptive finite element methods for distributed elliptic control problems with control constraints. *ESIAM Control Optim. Calc. Var.*, 2006. to appear.
18. Michael Hintermüller and Ronald H.W. Hoppe. Goal-oriented adaptivity in control constrained optimal control of partial differential equations. *SIAM J. Control Optim.*, 2007. submitted.
19. Michael Hintermüller and Karl Kunisch. Feasible and noninterior path-following in constrained minimization with low multiplier regularity. *SIAM J. Control Optim.*, 45(4):1198–1221, 2006.
20. Michael Hintermüller and Karl Kunisch. Stationary optimal control problems with pointwise state constraints. 2007. to appear.
21. Michael Hinze. A variational discretization concept in control constrained optimization: The linear-quadratic case. *Comput. Optim. Appl.*, 30(1):45–61, 2005.
22. Ronald H.W. Hoppe and Michael Kieweg. A posteriori error estimation of finite element approximations of pointwise state constrained distributed control problems. 2007. submitted.
23. Ruo Li, Wenbin Liu, Heping Ma, and Tao Tang. Adaptive finite element approximation for distributed elliptic optimal control problems. *SIAM J. Control Optim.*, 41(5):1321–1349, 2002.
24. Wenbin Liu. Adaptive multi-meshes in finite element approximation of optimal control. *Contemporary Mathematics*, (383):113–132, 2005.
25. Wenbin Liu, Wei Gong, and Ningning Yan. A new finite element approximation of a state constrained optimal control problem. *Journal of Computational Mathematics*, 2008. accepted.
26. Dominik Meidner and Boris Vexler. Adaptive space-time finite element methods for parabolic optimization problems. *SIAM J. Control Optim.*, 46(1):116–142, 2007.
27. Christian Meyer. Error estimates for the finite-element approximation of an elliptic control problem with pointwise state and control constraints. *Control Cybern.*, 2008. to appear.

28. Christian Meyer and Michael Hinze. Stability of infinite dimensional control problems with pointwise state constraints. 2007. submitted.
29. Christian Meyer, Uwe Prüfert, and Fredi Tröltzsch. On two numerical methods for state-constrained elliptic control problems. *Optim. Methods Softw.*, 22:871–899, 2007.
30. Christian Meyer, Arnd Rösch, and Fredi Tröltzsch. Optimal control of PDEs with regularized pointwise state constraints. *Comput. Optim. Appl.*, 33:209–228, 2006.
31. RoDoBo: A C++ library for optimization with stationary and nonstationary PDEs with interface to Gascoigne [14]. `http://www.rodobo.uni-hd.de`.
32. Anton Schiela. Barrier methods for optimal control problems with state constraints. 2007. submitted.
33. Michael Schmich and Boris Vexler. Adaptivity with dynamic meshes for space-time finite element discretizations of parabolic equations. *SIAM J. Sci. Comput.*, 30(1):369–393, 2008.
34. Fredi Tröltzsch. *Optimale Steuerung partieller Differentialgleichungen*. Friedr. Vieweg & Sohn Verlag, Wiesbaden, 2005.
35. Rüdiger Verfürth. *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*. Wiley/Teubner, New York-Stuttgart, 1996.
36. Boris Vexler and Winnifried Wollner. Adaptive finite elements for elliptic optimization problems with control constraints. *SIAM J. Control Optim.*, 47(1):509–534, 2008.
37. VisuSimple: An interactive VTK-based visualization and graphics/mpeg-generation program. `http://www.visusimple.uni-hd.de`.