An adaptive numerical method for semi-infinite elliptic control problems based on error estimates

Pedro Merino
Research Center on Mathematical Modelling,
MODEMAT, Escuela Politécnica Nacional,
Departamento de Matemática,
Ladrón de Guevara E11-253 Quito, Ecuador

Ira Neitzel
Technische Universität München,
Centre for Mathematical Sciences M17,
Boltzmannstr. 3, D-85748 Garching b. Munich, Germany

Fredi Tröltzsch
Technische Universität Berlin,
Fakultät II - Mathematik und Naturwissenschaften,
Str. des 17.Juni 136, D-10623 Berlin, Germany.

April 11, 2014

Abstract

We discuss numerical reduction methods for an optimal control problem of semi-infinite type with finitely many control parameters but infinitely many constraints. We invoke known a-priori error estimates to reduce the number of constraints. In a first strategy, we apply uniformly refined meshes, whereas in a second more heuristic strategy we use adaptive mesh-refinement and provide an a-posteriori error estimate for the control based on perturbation arguments.

Keywords: Optimal control, elliptic partial differential equation, FEM, semi-infinite programming, numerical methods

AMS subject classification: 49J20, 80M10, 49N05,90C34 41A25, 90C34

1 Introduction

In this paper, we discuss numerical reduction methods for the optimal control problem (2.1a)–(2.1d) that is introduced in the next section. This is a

control problem of semi-infinite type, where we have finitely many control parameters but infinitely many constraints. Many papers dealing with aspects of state-constrained optimal control problems are concerned with control functions rather than finitely many control parameters, even though the latter setting is very important for many practical situations. Consider for example the cooling of steel profiles, where the intensity of finitely many spray nozzles can be controlled or the case of local hyperthermia in cancer treatment, where the intensity of finitely many microwave antennas is to be controlled. More applications are summarized for instance in the introductory part of [4].

The presence of pointwise state constraints introduces mathematically challenging questions with regard to e.g. optimality conditions, see [4], or finite element error analysis, see [15, 16]. In past years, error estimates for optimal control problems attracted a lot of interest. Many papers on a-priori error estimates were published. For an extensive list of references, we refer the reader to the survey paper [11]. Although a-priori error estimates are interesting in their own, the question arises how they can be used in numerical applications. This is one of our main objectives in this paper. To our best knowledge, a-priori error estimates where not yet used to improve numerical methods for solving distributed optimal control problems.

Interestingly, the finite dimensional structure of the control space often leads to the typical situation of having only finitely many isolated active points, where the state touches the bound. In this paper, we use this typical structure and known a-priori error estimates for control problems of semi-infinite type to drastically reduce the number of constraints. Whenever very high precision of the numerical solution of the optimal control problem is needed, fine discretizations are necessary, which in turn leads to a large scale mathematical programming problem. Without special reduction techniques, the storage capacity of the computer will easily be exceeded. In particular, this might be important for 3D problems.

We propose two different methods. The main idea of the first one is to use the a-priori error estimates from our papers [15], [16] to exclude those node points, in which the state constraints cannot be active for the optimal solution. We solve a sequence of finite dimensional optimization problems on (uniformly) finer and finer grids. Nevertheless, although the number of grid points increases rapidly, the size of the optimization problems remains moderate, because the constraints have to be considered only in a small number of nodes. This technique is applied on regular grids that are refined by bisection in each refinement step.

In the second method, we apply an adaptive mesh refinement strategy. Again, we solve a sequence of finite dimensional optimization problems, but this time an adaptive grid generator is used. This idea is combined with our first technique to consider the constraints only in nodes that can be active. For this ad-hoc method, we do not have a convergence proof. Therefore, we also discuss an a-posteriori estimate of optimality of the computed solution. This method partially relies on well-known perturbation arguments due to [5], [14] that were later used as a theoretical tool for proving a-priori error estimates in PDE constrained optimization, cf. for instance [1]. Numerically, this technique

was applied in [25], [24]. To our best knowledge this idea is also new in the case of semi-infinite problems. Moreover, we apply a theorem by Miranda on the existence of solutions to nonlinear equations, [18], [23]. Here, we extend this idea to our case of semi-infinite optimization.

By the proposed methods, the discretized problem can be kept comparably small, since our estimates allow to neglect the state constraints in many points where the bound will never be active. In addition to the finite number of control parameters, which does not increase with the discretization, the resulting optimization problem only contains a rather small number of constraints.

In principle, such ideas are not new in the numerical treatment of semi-infinite optimization problems, we refer to the survey in [9] or [21]. However, in former contributions to the numerical analysis of semi-infinite optimization, the constraints were given by known analytical expressions. In contrast to this, our constraints are defined implicitly via solutions of an elliptic PDE. Therefore, they must be found numerically and the application of the finite element method leads to new questions of error analysis as in [15],[16]. The use of error estimates as justification of reduction techniques is a special feature of our paper.

2 The semi-infinite elliptic optimal control problem

Let $\Omega \subset \mathbb{R}^2$ be a convex bounded polygonal domain with boundary Γ. In Ω, we consider the following optimal control problem of semi-infinite type

$$(\mathbf{OCP}) \begin{cases} \min\limits_{(y,u)\in H_0^1(\Omega)\times\mathbb{R}^m} J(y,u) = \dfrac{1}{2}\|y-y_d\|^2 + \dfrac{\nu}{2}|u-u_d|^2 & (2.1a) \\ \text{subject to :} & \\ -\sum\limits_{i,j=1}^{2} \partial_i\left(a_{ij}(x)\partial_j y(x)\right) + c_0 y(x) = \sum\limits_{i=1}^{m} u_i e_i(x) \quad \text{in } \Omega & \\ \qquad\qquad y(x) = 0 \quad \text{on } \Gamma & (2.1b) \\ u_a \le u_i \le u_b, \quad i=1,\dots,m, & (2.1c) \\ y(x) \le b \quad \forall x \in \bar{\Omega}. & (2.1d) \end{cases}$$

For this problem, the following data are given: $\nu > 0$, $b > 0$, $u_a < u_b$ are real constants. We fix $1 \le m \in \mathbb{N}$ and functions $y_d \in L^2(\Omega)$, $e_i \in C^{0,\beta}(\Omega)$, $i = 1\dots,m$, with $0 < \beta < 1$. In addition, the coefficients a_{ij} are assumed to belong to $C^{1+\alpha}(\Omega)$, $0 < \alpha < 1$, $i,j = 1,2$, and to fulfill the uniform ellipticity condition

$$\sum_{i,j=1}^{2} a_{ij}(x)\xi_i\xi_j \ge \sigma_0 |\xi|^2, \quad \forall \xi \in \mathbb{R}^2,\ \forall x \in \Omega,$$

for some constant $\sigma_0 > 0$. It is well known that, for every function of $L^2(\Omega)$ on the right-hand side of (2.1b), there exists a unique weak solution $y \in$

$H_0^1(\Omega) \cap H^2(\Omega)$. Moreover, thanks to the Hölder regularity assumptions on e_i, standard regularity arguments for elliptic equations imply that $y \in C^{2,\beta}(\Omega)$ [7, Theorem 6.13]. This function y is the state associated with the control vector $u = [u_1, \dots, u_m]^\top \in \mathbb{R}^m$. The correspondence between u and y is indicated by an associated subscript; i.e. we denote by y_u the solution of the equation (2.1b) with right-hand side $\sum_{i=1}^m u_i e_i$.

Thanks to the linearity of the elliptic equation, we are able to simplify the control problem by using the superposition principle. Indeed, for all $u \in \mathbb{R}^m$ it holds that

$$y_u(x) = \sum_{i=1}^m u_i y_i(x),$$

where the functions $y_i \in H_0^1(\Omega) \cap H^2(\Omega) \cap C^{2,\beta}(\Omega)$ are the solutions of the state equation (2.1b) with corresponding right-hand side e_i. In view of this, problem **(OCP)** can be rewritten as a semi-infinite programming problem

$$(\mathbf{P}) \begin{cases} \min\limits_{u \in U_{ad}} f(u) = \dfrac{1}{2} \| \sum\limits_{i=1}^m u_i y_i - y_d \|^2 + \dfrac{\nu}{2} |u - u_d|^2 & \text{(2.2a)} \\ \text{subject to: } \sum\limits_{i=1}^m u_i y_i(x) \le b, \quad \forall x \in \bar{\Omega}, & \text{(2.2b)} \end{cases}$$

where $U_{ad} := \{u \in \mathbb{R}^m \ : \ u_a \le u_i \le u_b, \quad i = 1, \dots, m\}$ denotes the set of admissible controls.

Here and throughout the paper, we denote by $\| \cdot \|$ and $(\cdot, \cdot)$ the standard norm and the scalar product in $L^2(\Omega)$, respectively. The Euclidean norm in $\mathbb{R}^m$ is denoted by $| \cdot |$.

***Assumption* 1.** *(Slater condition) There exist a control vector $\tilde{u} \in U_{ad}$ and a constant $\varepsilon > 0$ such that*

$$y_{\tilde{u}}(x) \le b - \varepsilon \quad \forall x \in \bar{\Omega}. \tag{2.3}$$

***Remark* 1.** *By the Slater condition, the feasible set is not empty. Therefore, existence and uniqueness of an optimal control $\bar{u} \in U_{ad}$ follow from the classical Weierstraß theorem. Moreover, since b is positive and $y(x) = 0$ holds on Γ, for all feasible u the active set of y_u, i.e. the set of all $x \in \bar{\Omega}$ with $y_u(x) = b$, is contained in a compact subset $K \subset \Omega$. Notice that y_u is continuous, hence the active set is closed.*

To set up the optimality system for the optimal control $\bar{u}$, we introduce the Lagrangian

$$\mathcal{L} : \mathbb{R}^m \times M(\bar{\Omega}) \to \mathbb{R}: \quad \mathcal{L}(u, \mu) := f(u) + \int_{\bar{\Omega}} \Big(\sum_{i=1}^m u_i y_i(\cdot) - b \Big) \, d\mu(\cdot), \tag{2.4}$$

where $M(\bar{\Omega})$ is the space of all regular Borel measures defined on $\bar{\Omega}$. Note that $y_i \in C(\bar{\Omega})$, $i = 1, \dots, m$, so that the integral in (2.4) is well defined.

***Theorem* 1.** *Let $\bar{u}$ be the solution for (P). If the Slater condition (2.3) is satisfied, then there exists a non-negative Lagrange multiplier $\bar{\mu} \in M(\bar{\Omega})$ such that the conditions*

$$\nabla_u \mathcal{L}(\bar{u}, \bar{\mu})^\top (u - \bar{u}) \geq 0, \quad \forall u \in U_{ad}, \tag{2.5a}$$

$$\int_{\bar{\Omega}} \Big(\sum_{i=1}^{m} \bar{u}_i y_i(\cdot) - b \Big) d\bar{\mu} = 0 \tag{2.5b}$$

are satisfied.

This is a standard result of nonlinear optimization in Banach spaces, cf. [13]. In (2.5a), we have

$$\nabla_u \mathcal{L}(\bar{u}, \bar{\mu}) = \nabla f(\bar{u}) + \int_{\bar{\Omega}} (y_1(\cdot), \ldots, y_m(\cdot))^\top d\bar{\mu}(\cdot).$$

By convexity of f, the conditions (2.5), complemented by $\bar{u} \in U_{ad}$ and $\bar{y} \leq b$, are also sufficient for optimality.

For future reference, note that the variational inequality (2.5a) can be replaced in the usual way by the equation

$$\nabla_u \mathcal{L}(\bar{u}, \bar{\mu}) + \bar{\eta}_b - \bar{\eta}_a = 0 \tag{2.6}$$

if the componentwise nonnegative Lagrange multipliers $\bar{\eta}_a, \bar{\eta}_b \in \mathbb{R}^m$ are introduced, that fulfill

$$\bar{\eta}_a^T (u_a - \bar{u}) = \bar{\eta}_b^T (\bar{u} - u_b) = 0.$$

In view of (2.6), the Lagrange multipliers associated with the control constraints are given by

$$(\bar{\eta}_b)_i = [\frac{\partial \mathcal{L}}{\partial u_i}(\bar{u}, \bar{\mu})]_-, \qquad (\bar{\eta}_a)_i = [\frac{\partial \mathcal{L}}{\partial u_i}(\bar{u}, \bar{\mu})]_+ \tag{2.7}$$

3 Finite element approximation of (P)

Let $\mathcal{T}_h$ be a uniform and shape regular triangulation of Ω with mesh size $h > 0$, such that

$$\bar{\Omega} = \cup_{T \in \mathcal{T}_h} \bar{T}.$$

For future reference, we explicitely point out that we consider the triangles T to be open sets. On $\mathcal{T}_h$, we consider the space of standard piecewise linear and continuous finite element functions,

$$Y_h = \{\varphi_h \in C(\bar{\Omega}) : \varphi^h|_T \in \mathcal{P}_1(T) \, \forall T \in \mathcal{T}_h, \, \varphi^h = 0 \text{ on } \Gamma\} \subset H_0^1(\Omega) \cap C(\bar{\Omega}),$$

where $\mathcal{P}_1(T)$ denotes the set of polynomials of degree at most one defined on the triangle T. We consider the finite element scheme for the unknown functions

$y_i^h \in Y_h$, $i = 1, \ldots, m$,

$$a[y_i^h, \varphi^h] := \int_\Omega \sum_{k,l=1}^{2} \left(a_{kl}(x) \partial_l y_i^h(x) \partial_k \varphi^h(x) \right) + c_0 y_i^h(x) \varphi^h(x)\, dx$$
$$= \int_\Omega e_i(x) \varphi^h(x)\, dx, \quad \forall \varphi^h \in Y_h. \tag{3.1}$$

Analogously to y_u, we denote by $y_u^h = \sum_{i=1}^m u_i y_i^h$ the approximated state associated with u. Now we consider the following approximated control problem of $(\mathbf{P})$

$$(\mathbf{P_h}) \begin{cases} \min\limits_{u \in U_{ad}} f_h(u) = \dfrac{1}{2} \| \sum\limits_{i=1}^m u_i y_i^h - y_d \|^2 + \dfrac{\nu}{2} |u - u_d|^2 & \text{(3.2a)} \\ \sum\limits_{i=1}^m u_i y_i^h(x) \le b \quad \forall x \in \bar{\Omega}. & \text{(3.2b)} \end{cases}$$

Thanks to the high regularity of y_u in the compact set K, a maximum norm error estimate for $\|y_u - y_u^h\|_{C(K)}$ of [20] can be applied:

Theorem 2. *(Rannacher & Vexler 2005) Let K be a compact subset of Ω. Then there exists a constant $c_K > 0$ such that*

$$\|y_u - y_u^h\|_{C(K)} \le c_K\, h^2 |\log(h)|, \quad \forall u \in U_{ad},\ \forall h > 0. \tag{3.3}$$

If $h > 0$ is sufficiently small, the Slater condition (2.3) together with (3.3) implies that the feasible set of $(\mathbf{P_h})$ is not empty. Therefore, $(\mathbf{P_h})$ has a unique optimal control vector denoted by $\bar{u}^h$.

Under certain assumptions sketched below, the order of the approximation error from Theorem 2 transfers to the optimal controls. Roughly speaking, these assumptions are:

- The optimal state $\bar{y} := y_{\bar{u}}$ has exactly n pairwise different active points $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n \in \Omega$, where $1 \le n \le m$. Moreover, $m - n$ components of $\bar{u}$ are strongly active with $\bar{u}_i = u_a$ or $\bar{u}_i = u_b$, while the other n components of $\bar{u}$ are inactive. W.l.o.g. we assume that the components $u_{n+1}, \ldots, u_m$ are active.
- In $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$, the optimal state $\bar{y}$ is strongly active.
- The function $\bar{y}$ is strictly concave in neighborhoods around the active points $\bar{x}_j$.
- The matrix $(y_i(\bar{x}_j))_{i,j}$, $i, j = 1, \ldots, n$, is non-singular.

These assumptions are stated precisely as Assumptions 2 – 4 in the next section.

Under these assumptions, applying Theorem 3.23 in [16] to our semi-infinite programming problem yields:

***Theorem* 3.** *Let $\bar{u}$ be the optimal solution of Problem* $(\mathbf{P})$*, let $\bar{u}^h$ be optimal for* $(\mathbf{P_h})$*, and let Assumptions 2–4 to be specified later be satisfied. Then, there exists $h_0 > 0$ such that the following estimate is true for a constant $c_u > 0$ independent of h:*

$$|\bar{u} - \bar{u}^h| \le c_u h^2 |\log h| \quad \forall h \le h_0. \tag{3.4}$$

By Theorems 2 and 3 combined with the Lipschitz continuity of the control-to-state operator, there is also a constant $c_y > 0$ such that the optimal error estimate

$$\|\bar{y} - \bar{y}^h\|_{C(\bar{\Omega})} \le \|\bar{y} - y_{\bar{u}^h}\|_{C(\bar{\Omega})} + \|y_{\bar{u}^h} - \bar{y}^h\|_{C(\bar{\Omega})} \le c_y\, h^2 |\log(h)| \tag{3.5}$$

holds true.

If these assumptions are not met, but the linearized Slater condition is still satisfied, then the standard error estimate of half the optimal order,

$$|\bar{u} - \bar{u}^h| \le \tilde{c}_u\, h \sqrt{|\log(h)|} \tag{3.6}$$

can be derived that is not optimal under the assumptions formulated above. In this case, the main idea of our paper can still be applied.

Note that the number and location of the active points $\bar{x}_1, \ldots, \bar{x}_n$ is not known in advance. If the location of the active points would be a-priori known, then we might fix them and consider the inequality constraints only in these finitely many points. In this case, the Problem $(\mathbf{P})$ would be essentially finite dimensional cf. [17]. In that case, the order of convergence is the optimal one expressed by (3.4) without any assumption on the number of active constraints.

4 An adaptive method based on the a-priori error estimate

Let us now present our first reduction strategy based on the a-priori error estimate (3.4). We fix some sufficiently small initial mesh size $h_0 > 0$ and define a sequence of decreasing mesh sizes

$$h_k := h_0 \left(\frac{1}{2}\right)^k, \quad k = 1, 2, \ldots. \tag{4.1}$$

Our optimization method of iterative grid refinement is some type of nested iteration. It starts by solving problem $(\mathbf{P_{h_0}})$. Next, we take $\bar{u}^{h_0}$ as initial vector for the solution of $(\mathbf{P_{h_1}})$. Hereafter, $(\mathbf{P_{h_2}})$ is solved, starting with the solution $\bar{u}^{h_1}$ of $(\mathbf{P_{h_1}})$.

We need these initialization steps to compute an approximation of the unknown constant c_y that appears in the estimate (3.5). Next, we apply this error estimate for determining points of the current grid where the exact optimal state $\bar{y}$ cannot be active.

Thanks to our regularity assumptions, there exists a constant $c_{y''} > 0$ such that it holds

$$\|\nabla^2 y_u(x)\|_{\mathbb{R}^{2\times 2}} \le c_{y''} \quad \forall x \in K,\ \forall u \in U_{ad}. \tag{4.2}$$

Here, $\nabla^2 y_u(x)$ denotes the Hessian matrix of y_u at x.

Assumption 2. *The optimal state $\bar{y}$ is (strongly) active in exactly n points $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$, where $1 \le n \le m$. In addition, there are $m-n$ components $\bar{u}_i$ of $\bar{u}$ which are strongly active with $\bar{u}_i = u_a$ or $\bar{u}_i = u_b$. Strong activity means here that all Lagrange multipliers $(\eta_b)_i$ or $(\eta_a)_i$ associated with these active control constraints are positive. The other components of u are inactive. W.l.o.g. let $u_1, \dots, u_n$ be inactive.*

Assumption 3. *There is a constant $\sigma > 0$ such that*

$$\xi^\top \nabla^2 \bar{y}(\bar{x}_j)\xi \le -\sigma|\xi|^2 \quad \forall \xi \in \mathbb{R}^2,\ \forall j = 1, \dots, n. \tag{4.3}$$

Remark 2. *The fact that the state constraint is strongly active in the points $\bar{x}_1, \dots, \bar{x}_n$ guarantees the existence of associated active points $\bar{x}_1^h, \dots \bar{x}_n^h$ of the discretized problem in the neighborhood of the continuous active points, cf. [15, 16].*

For the optimal error estimate, we finally require the following

Assumption 4. *The matrix*

$$\big(y_i(\bar{x}_j)\big), \quad i, j = 1, \dots, n, \tag{4.4}$$

is non-singular.

Theorem 4. *(Inactivity Criterion) Assume that the a-priori estimate (3.5) and Assumption 3 are met and define $\bar{y}^{h_k} := y^{h_k}_{\bar{u}^{h_k}}$. Let $T \in \mathcal{T}_{h_k}$ be a triangle, where in each of the corners x_j, $j = 1, 2, 3$, the inequality*

$$\bar{y}^{h_k}(x_j) < b - c_{y''} h_k^2 - c_y h_k^2 |\log(h_k)| \tag{4.5}$$

is fulfilled. Then $\bar{y}(x) < b$ holds for all $x \in T$, i.e. the optimal state $\bar{y}$ cannot be active in T.

Proof. Let $\bar{x} \in \arg\max\{\bar{y}(x) : x \in T\}$. Assume first that $\bar{x}$ is not a corner of T. Then either $\bar{x} \in \text{ int } T$ or $\bar{x} \in \text{ int } [x_j, x_k]$, i.e. $\bar{x}$ is located in the relative interior of an edge $[x_j, x_k]$ of T.

In either case, we connect $\bar{x}$ by a straight line with x_j, where x_j is an arbitrary corner of T in the first case and one end of $[x_j, x_k]$ in the second. Consider for sufficiently small $\varepsilon > 0$ the function $g : (-\varepsilon, 1] \to T$,

$$g(t) = \bar{y}(\bar{x} + t(x_j - \bar{x})).$$

Then g attains its maximum at $t = 0$, hence by $g'(t) = \nabla\bar{y}(\bar{x}+t(x_j-\bar{x}))\cdot(x_j-\bar{x})$ we observe

$$0 = g'(0) = \nabla y(\bar{x}) \cdot (x_j - \bar{x}).$$

By a second-order Taylor approximation of g, $g(1) = g(0) + g'(0) + \frac{1}{2}g''(\vartheta)$ holds with some $\vartheta \in (0,1)$, thus

$$\begin{aligned} \bar{y}(x_j) &= \bar{y}(\bar{x}) + \nabla\bar{y}(\bar{x}) \cdot (x_j - \bar{x}) + \tfrac{1}{2}(x_j - \bar{x})^\top \nabla^2 \bar{y}(\xi)(x_j - \bar{x}) \\ &= \bar{y}(\bar{x}) + \tfrac{1}{2}(x_j - \bar{x})^\top \nabla^2 \bar{y}(\xi)(x_j - \bar{x}) \end{aligned}$$

holds with some vector $\xi \in (x_j, \bar{x})$. Re-arranging, we find

$$\begin{aligned} \bar{y}(\bar{x}) &= \bar{y}(x_j) - \tfrac{1}{2}(x_j - \bar{x})^\top \nabla^2 \bar{y}(\xi)(x_j - \bar{x}) \\ &= \bar{y}^{h_k}(x_j) - \tfrac{1}{2}(x_j - \bar{x})^\top \nabla^2 \bar{y}(\xi)(x_j - \bar{x}) + \bar{y}(x_j) - \bar{y}^{h_k}(x_j). \end{aligned}$$

Using estimate (4.2) for $\nabla^2 y$ and the error estimate (3.5), we obtain

$$\bar{y}(\bar{x}) \leq \bar{y}^{h_k}(x_j) + \frac{1}{2} c_{y''}\, h_k^2 + c_y\, h_k^2 |\log(h_k)|.$$

Clearly, this implies $\bar{y}(\bar{x}) < b$, if inequality (4.5) is satisfied. Therefore, $\bar{y}(x) < b$ holds also for all $x \in T$.

If the maximum of $\bar{y}$ is attained in a corner x_j, then we can directly compare $\bar{y}(x_j)$ with $\bar{y}^{h_k}(x_j)$ and obtain instantly the desired result. □

In the numerical implementation of the inactivity criterion (4.5), we are faced with two obstacles: In general, the constants $c_{y''}$ and c_y are not known. For $c_{y''}$, the following observation helps: If $h \to 0$, then the term $|\log(h)|$ tends to infinity and hence it holds $c_y |\log(h)| \geq c_{y''}$ for all sufficiently small h. Then, we observe that

$$\bar{y}^{h_k}(x_j) < b - 2c_y h_k^2 |\log h_k| \;\Rightarrow \bar{y}^{h_k}(x_j) < b - c_{y''} h_k^2 - c_y h_k^2 |\log h_k|.$$

In view of this, we apply instead of (4.5) the inequality

$$\bar{y}^{h_k}(x_j) < b - 2\, c_y\, h_k^2 |\log(h_k)| \tag{4.6}$$

as criterion of inactivity. The constant c_y will be estimated during the algorithm as explained below.

We will now define the subset of Ω where the state constraints might still be active. Due to the linear approximation of the state functions we can again formulate conditions on the nodes, only.

For that reason, we introduce the set $\mathcal{N}_{h_k}$ of nodes of a mesh $\mathcal{T}_{h_k}$ with mesh size h_k, i.e.

$$\mathcal{N}_{h_k} = \{x \in \Omega :\ \exists T \in \mathcal{T}_{h_k} \text{ such that } x \text{ is a vertex of } T\}, \quad k = 1, 2, \ldots, \tag{4.7}$$

as well as the sets

$$\mathcal{T}_I^k := \{T \in \mathcal{T}_{h_k} :\ \text{all nodes satisfy (4.6)}\}, \tag{4.8}$$

$$\mathcal{I}^k := \overline{\bigcup_{T \in \mathcal{T}_I^k} T} \tag{4.9}$$

$$\mathcal{A}^{k+1} := \overline{\bar{\Omega} \setminus \mathcal{I}^k} \cap \mathcal{N}_{h_{k+1}}. \tag{4.10}$$

The set $\mathcal{A}^{k+1}$ contains all nodes of the refined mesh where the constraints still need to be prescribed in order to exclude only the region $\mathcal{T}_I^k$. Note that the above construction may lead to some inactive triangles included in the set $\mathcal{A}^{k+1}$, but the constraints need to be prescribed in triangles where not all nodes satisfy (4.6). Due to the linear approximation of the state and the constant bound, this is equivalent to prescribing the constraints in the respective nodes, only.

Motivated by the last theorem, we consider the *reduced problem*

$$(\mathbf{P}^{\mathbf{r}}_{\mathbf{h_k}}) \begin{cases} \min\limits_{u\in U_{ad}} f(u) = \dfrac{1}{2}\|\sum\limits_{i=1}^{m} u_i y_i^{h_k} - y_d\|^2 + \dfrac{\nu}{2}|u-u_d|^2 & \text{(4.11a)} \\ \text{subject to: } \sum\limits_{i=1}^{m} u_i y_i^{h_k}(x) \le b \quad \forall x \in \mathcal{A}^k. & \text{(4.11b)} \end{cases}$$

To apply the error estimate (4.6) that is needed to set up problem $(\mathbf{P}^{\mathbf{r}}_{\mathbf{h}})$, we have to estimate the constant c_y in (3.5), which is not known. For this purpose, we propose the following method:

We start our adaptive technique by computing the solutions for $(\mathbf{P}_{\mathbf{h_0}})$, $(\mathbf{P}_{\mathbf{h_1}}), (\mathbf{P}_{\mathbf{h_2}})$ Assume now that, for some $k \ge 2$, the optimal states $\bar{y}^{h_{k-2}}$, $\bar{y}^{h_{k-1}}, \bar{y}^{h_k}$ are determined, i.e. $(\mathbf{P}^{\mathbf{r}}_{\mathbf{h_{k-2}}})$, $(\mathbf{P}^{\mathbf{r}}_{\mathbf{h_{k-1}}})$, $(\mathbf{P}^{\mathbf{r}}_{\mathbf{h_k}})$ were solved.

By the triangle inequality, we split

$$\begin{aligned} \|\bar{y}^{h_{k-2}} - \bar{y}^{h_k}\|_\infty &\le \|\bar{y}^{h_{k-2}} - \bar{y}\|_\infty + \|\bar{y} - \bar{y}^{h_k}\|_\infty \\ &\sim c_y \left\{ h_{k-2}^2 |\log h_{k-2}| + h_k^2 |\log h_k| \right\}. \end{aligned}$$

Note that here and for the remainder of this paper we abbreviate $\|\cdot\|_{C(\bar{\Omega})}$ by $\|\cdot\|_\infty$. Resolving for c_y, we find the following empirical approximation of c_y to be used in the next iteration:

$$c_{y,k+1} := \frac{\|\bar{y}^{h_{k-2}} - \bar{y}^{h_k}\|_\infty}{h_{k-2}^2 |\log h_{k-2}| + h_k^2 |\log h_k|}. \tag{4.12}$$

Note that the constant c_y is approximated using the information from two different refinement levels rather than only the last. One might compare the solutions of more than two subsequent optimization problems. In our examples, this was not necessary. We point out that an inacceptable approximation of the constant c_y could be detected by checking whether $y^{h_{k+1}}$ admits unfeasible values.

Summarizing, we implemented the following algorithm that depends on the choice of the constants $c_{y,k}$:

Algorithm 1

Initialization

1. Define initial grids with mesh sizes $h_k = (1/2)^k\, h_0$, $k = 0, 1, 2$.

2. Compute the functions $y_i^{h_k}$, $k = 0, 1, 2$, $i = 1, \dots, m$, by solving equation (3.1).

3. Solve the problems $(\mathbf{P_{h_k}})$, $k = 0, 1, 2$.

4. Fix a maximal number of iterations $N \in \mathbb{N} \cup \{\infty\}$.

Iterative steps

5. **For** $k = 3, 4, \dots, N$:

(i) Compute $c_{y,k}$ according to formula (4.12)

(ii) Determine the sets $\mathcal{T}_I^{k-1}$ and $\mathcal{I}^{k-1}$ by (4.8) and (4.9)

(iii) Set up the regular grid $\mathcal{T}_{h_k}$

(iv) Determine the set $\mathcal{A}^k$ by (4.10)

(v) Solve the reduced problem $(\mathbf{P^r_{h_k}})$.

***Remark* 3.** *Note that the main tool of the algorithm above is the a-priori error estimate* (3.5)*. We just assume that the underlying assumptions are satisfied so that the estimate holds true. In particular, we need that the number m of active points in the state constraints is equal to $k - m_c$, where m_c is the number of active control variables. If these assumptions are not met, then we can still use our algorithm based on the standard (non-optimal) estimate $\|\bar{y} - \bar{y}^h\|_\infty \leq c_y\, h\sqrt{|\log(h)|}$ to define the sets $\mathcal{A}^k$ and apply similar ideas.*

***Theorem* 5.** *(Convergence) Let the Assumptions 1 – 4 be satisfied and assume that there is a positive constant $\alpha > 0$ such that $\alpha c_y \leq c_{y,k} \leq c_{max}$ holds in all steps of Algorithm 1. Let $\bar{u}^{h_k}$ be the optimal control vectors of $(\mathbf{P^r_{h_k}})$, $k = 1, 2, \dots$.*

If $N = \infty$, i.e. if the algorithm is not terminated after a finite number of steps, then $\bar{u}^{h_k} \to \bar{u}$, $k \to \infty$, where $\bar{u}$ is the optimal control of $(\mathbf{P})$. Moreover, the vectors of active points $(\bar{x}_1^k, \dots, \bar{x}_n^k)$ of $\bar{y}^{h_k}$ converge to $(\bar{x}_1, \dots, \bar{x}_n)$, the vector of active points of $\bar{y}$.

Proof. In view of the known a-priori error estimate and thanks to the assumption on $c_{y,k}$, we have

$$|\bar{u}^{h_k} - \bar{u}| \leq c_y\, h_k^2 |\log h_k| \leq \alpha^{-1} c_{y,k}\, h_k^2 |\log h_k| \leq c_{max}\, h_k^2 |\log h_k|.$$

This implies the statement of the theorem, because $h_k \to 0$ as $k \to \infty$ follows from the definition of the h_k. The convergence of the active points was shown in [16], Lemma 3.19. □

Let us confirm the efficiency of the method using uniform mesh refinement by some computational examples. We start our numerical experiments with an example with known analytic solution.

h	$\lvert\bar{u}-\bar{u}_h\rvert$	$\lVert\bar{y}-\bar{y}_h\rVert_\infty$	$c_{y,k}$	♯ nodes	♯ constraints
0.7071	2.24E-000	3.42E-062	-	9	9
0.3536	1.30E-000	8.88E-001	-	25	25
0.1768	2.40E-001	4.36E-001	-	81	81
0.0884	1.21E-001	1.45E-001	3.93	289	29
0.0442	3.32E-002	3.89E-002	3.48	1089	74
0.0221	8.48E-003	9.90E-003	6.06	4225	90
0.0110	2.13E-003	2.49E-003	5.61	16641	90
0.0055	5.33E-004	6.13E-004	4.78	66049	102
0.0028	1.33E-004	1.56E-004	3.99	263169	102

Table 1: Example 1, Algorithm 1

***Example* 1.** *Let the following data in* (2.1) *be given:* $U_{ad} = \mathbb{R}^2$, $c = 0$, $\nu = 1$, $b = 1$, $u_d = [0\ 0]^\top$, $y_d = y_1 + 2y_2 + 100$, *where* $y_1 = \sin(2\pi x_1)\sin(2\pi x_2)$ *and*

$$y_2 = \begin{cases} -\sin(2\pi x_1)^3 \sin(2\pi x_2)^3 & \text{in } [\frac{1}{2}, 1] \times [0, \frac{1}{2}], \\ -\frac{1}{4}\sin(2\pi x_1)^3 \sin(2\pi x_2)^3 & \text{in } [\frac{1}{2}, 1] \times [\frac{1}{2}, 1], \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, we define $e_1(x) = -\Delta y_1(x)$ *and* $e_2(x) = -\Delta y_2(x)$. *The optimal solution is then given by* $\bar{u} = [1\ 2]^\top$ *and its optimal state is* $\bar{y} = y_1 + 2y_2$.

Note that the graph of $\bar{y}$, respectively its numerical approximation is presented in Figure 3. Table 4 shows the efficiency of the reduction method for this problem. For each h, we state the L^∞-error for the state as well as the error for the control that both tend to zero, the approximated constant c_y, as well as the number of nodes and the number of constraints prescribed on each level of refinement.

Note that the number of constraints is considerably smaller than the number of nodes. To illustrate this reduction graphically, Figure 1 depicts the potentially active region on two different grids.

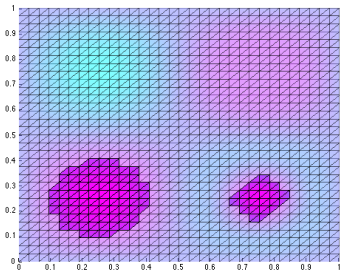
We repeat these experiments for a problem without explicitely known solution.

***Example* 2.** *We consider the following data in* (2.1)*:* $U_{ad} = \mathbb{R}^2$, $c = 0$, $\nu = 1$, $b = 0.01$, $u_d = [2\ 2]^\top$, $y_d \equiv 2000$.

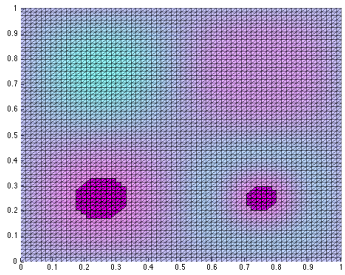
$$\begin{aligned} e_1(x) &= e^{-100|x-\tilde{x}_1|^2} + \tfrac{1}{5}e^{-200|x_1-0.8|^2} \\ e_2(x) &= e^{-100|x-\tilde{x}_2|^2} \end{aligned}$$

with $\tilde{x}_1 = [0.25\ 0.25]$, $\tilde{x}_2 = [0.75\ 0.75]$.

As a substitute for $\bar{u}$ and $\bar{y}$ we compute a reference solution $\tilde{u}_h$ with associated state $\tilde{y}_h$ on a fine grid for $h \approx 6.9e-04$. Table 4 summarizes the results analogously to the first example. Figure 2 shows the sets $\mathcal{A}^k$ of node points for two different values of h.



(a) $\bar{y}^h$ at $h \approx 0.044$

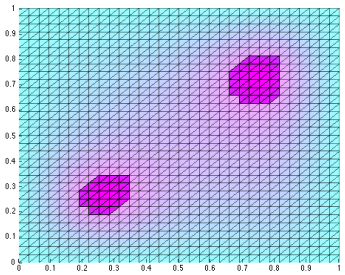


(b) $\bar{y}^h$ at $h \approx 0.022$

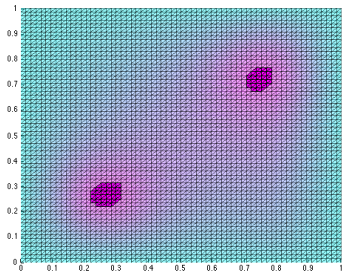
Figure 1: Example 1, sets $\mathcal{A}^k$ of potentially active points for two levels of refinement of Algorithm 1.

h	$\lvert\tilde{u} - \bar{u}_h\rvert$	$\lVert\tilde{y}_h - \bar{y}_h\rVert_\infty$	$c_{y,k}$	♯ nodes	♯ constraints
0.7071	1,40E+000	6.07E-003	-	9	9
0.3536	4,05E-001	3.43E-003	-	25	25
0.1768	2,02E-001	1.12E-003	-	81	81
0.0884	8,38E-002	4.99E-004	0.0418	289	121
0.0442	2,59E-002	1.68E-004	0.0163	1089	31
0.0221	4,77E-003	3.00E-005	0.0106	4225	38
0.0110	1,30E-003	8.19E-006	0.0168	16641	41
0.0055	2,98E-004	1.86E-006	0.0215	66049	24

Table 2: Example 2, Algorithm 1



(a) $\bar{y}^h$ at $h \approx 0.044$



(b) $\bar{y}^h$ at $h \approx 0.022$

Figure 2: Example 2, sets $\mathcal{A}^k$ of potentially active points for two levels of refinement of Algorithm 1.

5 Adaptive mesh refinement and a-posteriori estimation

5.1 The method of adaptive mesh refinement

In the former section, our numerical method was based on a uniform triangulation with mesh size $h > 0$. The state functions y_i where pre-computed on a sequence of successively refined triangulations. We bounded the size of the finite dimensional approximating problems by selecting only those nodes for the constraints, where the optimal state function $\bar{y}$ has still a chance to be active. In contrast to this, we now allow for a non-uniform triangulation of y_u. Instead of pre-computing the functions y_i, the non-uniform grids are defined by a PDE-related a-posteriori error estimator for certain suboptimal states y_u generated by our optimization method.

The main idea is the following: Having solved the approximated optimal control problem on a uniform initial grid of moderate size, a first approximation for the optimal control is obtained. Next, to avoid too fine uniform grids in the following steps, a technique of a-posteriori error estimation is applied for the next mesh refinement.

A posteriori error estimators are already well investigated in the numerical analysis of partial differential equations. We refer exemplarily to Nochetto et al. [19] and the references therein, where a maximum-norm error estimator is suggested, as well as [2] and [6]. The application of a-posteriori error estimations to optimal control problems is a new field of research since the recent past.

In particular, a-posteriori error estimators are fairly well understood for elliptic control problems. We mention the papers [12] or [10] for problems including control constraints. For state-constrained problems we point out [8] and [3] for error estimates with respect to the objective function.

In contrast to these papers, our a-posteriori estimation technique is not related to first order optimality conditions. This analysis would go beyond the scope of this paper. We consider a simplified approach. For a given control u that was generated during the optimization process, the mesh refinement is defined upon the associated state y_u. Since the mesh is related to a single u, we need to be aware of the fact that the estimator might not be valid for states associated with other controls.

We proceed as follows: For the numerical solution of the elliptic PDE, we use a maximum norm error estimator for elliptic partial differential equations suggested by Nochetto et al. in [19]. We need an estimator in the maximum norm, since the pointwise state constraints have to be considered in $C(\bar{\Omega})$ to guarantee the existence of Lagrange multipliers. In an initial step, we solve the approximated problem $(\mathbf{P_{h_0}})$ on a uniform and regular grid of moderate mesh size h_0 and obtain a first estimate $\bar{u}^0$ for the optimal control vector.

Then, we use the a-posteriori error estimator of [19]. This estimator is of the form $\|y - y_h\|_\infty \leq C(\eta_\infty + \eta_d)$, where the quantity η_∞ estimates the residual in an appropriate dual norm, η_d measures the contribution of the integration

error, and C is a positive constant. We are not considering the contribution of integration errors here. The estimator evaluates each triangle with respect to its contribution to the maximum norm error. It then successively refines those triangles with the largest local error until approximately twice the number of elements of the previous grid or a certain tolerance is reached. Note that the estimator tends to refine the mesh close to points where the state has large curvature. Since this is to be expected for single active points, the mesh will most likely be refined close to the active points.

Let us denote the state obtained by means of the error estimator of Nochetto et al. by $y_{h,u}$, where u stands for the underlying control vector. Notice that here we do not have a representative mesh size h, but we continue to denote the discretized states this way for convenience.

Thanks to the a-posteriori error estimator, for y_{h,u^0} we obtain some estimate

$$\|y_{u^0} - y_{h,u^0}\|_\infty \leq \delta_0.$$

Now we tacitly assume that this value δ_0 is useful for estimating the distance to the unknown exactly optimal state $\bar{y}$, i.e. we assume

$$\|\bar{y} - y_{h,u^0}\|_\infty \leq \delta_0.$$

Hereafter, analogously to the activity indicator of Theorem 4, we do not consider the state constraints in all nodes of the grid, where the inequality

$$\|y_{h,u^0}\|_\infty < b - \delta_0. \tag{5.1}$$

is satisfied.

We do not have a guarantee that this inequality is sufficient for non-activity of these nodes at the optimal state $\bar{y}$. However, if the initial uniform grid of mesh size h_0 was not too coarse, then we can expect to have sufficiently precise information on the location of active points and the optimal control. Therefore, the further adaptive mesh refinement should not lead to wrong solutions.

Nevertheless, these arguments would need a justification by an a-posteriori error estimator for elliptic control problems of semi-infinite type. To our knowledge, this problem that is out of the scope of this paper, was not yet considered in literature.

Next, we solve the resulting smaller finite dimensional optimization problem to obtain the new suboptimal control $\bar{u}^1$. Once again, we apply the a-posteriori error estimation technique to define a new nonuniform grid with approximately twice the number of elements according to the error estimator of [19]. The new error estimate is

$$\|y_{h,u^1} - y_{u^1}\|_\infty \leq \delta_1.$$

We repeat our procedure in the further steps. In summary, we use the following algorithm:

Algorithm 2

Initialization

1. Define an initial uniform grid $\mathcal{T}_0$ with mesh size h_0 and solve the problem $(\mathbf{P_{h_0}})$ to obtain initial approximations $u^k = u^0$ and $y^k = y^0$

2. Fix a maximal number of new elements $N \in \mathbb{N} \cup \{\infty\}$ for the adapted mesh.

Iterative steps

3. **For** $k = 1, \ldots, N$:

(i) Evaluate the a-posteriori error estimator from [19],

$$\|\bar{y} - y_{h,u^{k-1}}\|_\infty \leq \delta_{k-1}$$

(ii) Determine the sets

$$\mathcal{T}_I^{k-1} := \{T \in \mathcal{T}_{k-1} : \text{ all nodes satisfy } y_{h,u^{k-1}}(x) < b - \delta_{k-1}\}$$
$$\mathcal{I}^{k-1} := \overline{\bigcup_{T \in \mathcal{T}_I^{k-1}} T}$$

(iii) Set up the adapted mesh $\mathcal{T}_k$ according to the a posteriori estimator

(iv) Determine the set

$$\mathcal{A}^k := \overline{\bar{\Omega} \setminus \mathcal{I}^{k-1}} \cap \mathcal{N}_k$$

(v) Solve the discrete optimization problem

$$(\mathbf{P_k^r}) \quad \begin{cases} \min_{u \in U_{ad}} f(u) = \frac{1}{2}\|y_{h,u} - y_d\|^2 + \frac{\nu}{2}|u - u_d|^2 \\ \text{subject to: } y_{h,u}(x) \leq b \quad \forall x \in \mathcal{A}^k. \end{cases}$$

in the new adapted grid $\mathcal{T}_k$ to obtain a new suboptimal solution (u^k, y^k).

Note that analogously to Algorithm 1, $\mathcal{N}_k$ denotes the set of nodes in the (nonuniform) mesh $\mathcal{T}_k$. Since we do not have a representative mesh size h_k, we refer to the mesh on each refinement level as $\mathcal{T}_k$ instead of $\mathcal{T}_{h_k}$.

This method turned out to be very efficient in our numerical tests. However, we do not have a convergence analysis. The mesh refinement is only related to the residual of the state equation for the sequence of computed suboptimal controls. Residuals of the whole optimality system are not taken into account. Therefore, the method might end up in a non-optimal point, provided that it converges at all.

In view of that, after numerical convergence to a limit control, an a-posteriori optimality estimator should be invoked that evaluates the distance to the unknown optimal control $\bar{u}$. We discuss this issue in the next section as a first

♯ nodes	♯ constraints	$\|\bar{u} - \bar{u}_h\|$	$\|\bar{y} - \bar{y}_h\|$
25	25	1.30E+000	8.88E-001
57	57	1.09E+000	9.04E-001
137	112	1.28E-002	3.15E-001
429	97	1.43E-002	8.73E-002
1099	82	3.33E-003	3.88E-002
2956	130	1.95E-003	1.76E-002
7847	98	9.20E-004	5.58E-003
18693	118	4.40E-004	2.34E-003
46022	106	1.55E-004	1.04E-003
96539	86	7.75E-005	5.08E-004

Table 3: Example 1, Algorithm 2

step towards a complete a-posteriori error analysis for optimal control problems with finite-dimensional control space.

Let us first demonstrate the performance of this adaptive method by numerical examples. We apply the adaptive mesh refinement to Examples 1 and 2. Analogously to Tables 4 and 4, we sum up the results obtained with the adaptive strategy for different levels of refinement in Tables 5.1 and 5.1. Since there is no representative mesh size h for the adaptive method, we only show the number of nodes of the grid, which is also an indicator for the size of the problem. For each refinement level, the L^∞-error for the optimal control vector are reported, as well as the number of constraints after identifying inactive regions according to formula (5.1). For Example 1 with known solution we also include the L^∞-error for the state in Table 5.1. We observe that this leads to better precision of the optimal solution than that achieved by the first method. For instance, Table 4 shows a control error of $5.33 \cdot 10^{-4}$ obtained on a mesh with roughly 66000 nodes, whereas the adaptive method needs only roughly 18000 nodes to obtain a similar control error of $4.4 \cdot 10^{-4}$, cf. Table 5.1. Similar behavior is observed by comparing Tables 4 and 5.1. Note that for the second example we use again a reference solution obtained on a very fine grid. Figures 3 and 4 show the mesh and state, respectively, obtained by the adaptive procedure for both Examples 1 and 2.

6 A-posteriori error estimation of optimality

6.1 Introduction

Let us assume that we have determined a numerical approximation of the optimal control $\bar{u}$, which we will refer to as u^s for the remainder of this section. The index s stands for suboptimality, since the numerical approximation will not in general be optimal for the original problem (**P**). We assume that this numerical approximation was obtained on the basis of a finite element approximation of

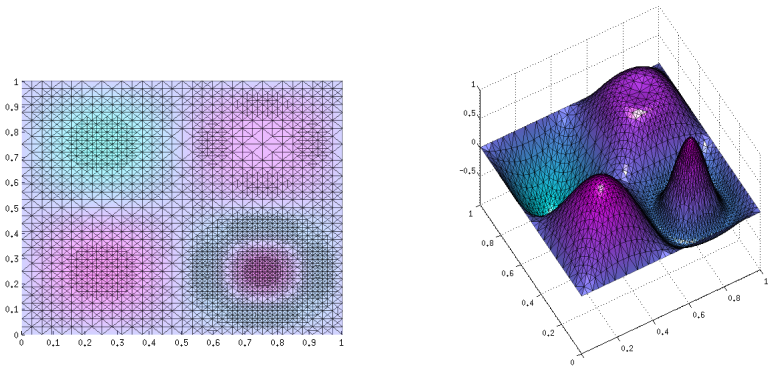


Figure 3: Example 1, adaptive mesh and approximated state after 5 steps of Algorithm 2.

♯ nodes	♯ constraints	$\lvert\tilde{u} - \bar{u}_h\rvert$
25	25	4.05E-001
45	45	5.58E-002
148	91	2.43E-002
357	93	1.32E-002
832	88	1.34E-003
1915	98	9.76E-004
4213	114	3.56E-004
10835	96	8.66E-005
23556	94	7.24E-005
50923	126	4.10E-005
123003	146	1.63E-005

Table 4: Example 2, Algorithm 2

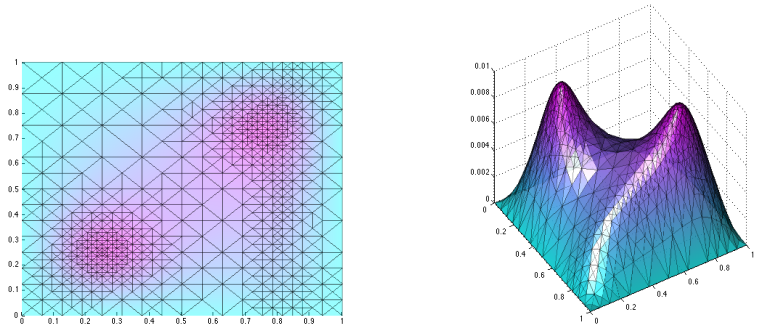


Figure 4: Example 2, adaptive mesh and approximated state after 5 steps of Algorithm 2.

the equation as in the former sections. Let y^s denote the finite element solution associated with u^s. We rely on the same assumptions as before. Thanks to Assumption 2 on activity, we assume that this is reflected by the result of the numerical computation: We assume to have activity exactly in some points $x_1^s, \ldots, x_n^s$, i.e.

$$y^s(x_j^s) = b \quad \forall j \in 1, \ldots, n$$

and that $y^s(x_j) < b$ holds in all other points x_j of the underlying grid. In particular, we point out that the number of active points of y^s is assumed to be the same as the number of active points of $\bar{y}$. Because the approximated state y^s is piecewise linear, it fulfills the state constraints in all points of Ω. However, u^s is most likely not feasible for the original optimal control problem. For this, the exact state y_{u^s} associated with u^s would have to be feasible.

In general, we do not know exact states and we will not be able to compute them. However, it is common practice to solve the FE state $y_{u^s}^h$ on a very fine uniform grid and to consider this as "exact state" y_{u^s}. We follow this approach and assume to know the exact state y_u for a given u, and in particular the state y_{u^s} associated with u^s, in this way.

Our further analysis is based on the following structural assumptions in addition to the ones introduced so far for the analysis of optimality conditions.

***Assumption* 5.**

- *The computed number n of active points $\{x_1, \ldots, x_n\}$ is correct.*
- *The exact optimal control is active in exactly the same $n-m$ components as u^s and equals the same bounds as u^s in these components. Consequently, the other n (optimal) control components are inactive as the ones of u^s.*
- *Assumption 6 below is satisfied.*

We want to check whether the control u^s is optimal or not. Moreover, if not, we are interested in an estimate of the distance of u^s to the unknown exact optimal control. If u^s, together with y^s and the Lagrange multiplier μ^s obtained from (6.7), satisfies the necessary optimality conditions, and y^s is also feasible, then u^s is optimal.

In view of this, we distinguish between two cases of non-optimality.

Case a) *The state y_{u^s} is is feasible, i.e. $y_{u^s}(x) \leq b$ holds for all $x \in \Omega$.*

Then, in view of Assumption 5, the necessary optimality conditions are not satisfied. This means that at least one component of the Lagrange multiplier μ^s is negative or the sign conditions (6.5) are not satisfied. The latter means that at least one of the Lagrange multipliers (2.7) associated with the control constraints does not have a sign that complies with both the nonnegativity property and the activity of the control component.

***Remark* 4.** *Having u^s and y^s, the vector μ^s is determined as unique solution to the system* (6.7) *below. The Lagrange multipliers associated with the control*

constraints are given by

$$\left[\frac{\partial f}{\partial u_i}(u^s) + \int_\Omega y_i(x)\, d\mu\right]_i \tag{6.1}$$

for the i-th control constraint. They must be nonnegative, if $u_i^s = u_a$*, nonpositive, if* $u_i^s = u_b$ *and zero, if* u_i^s *is inactive.*

Case b) *The Lagrange multiplier* μ^s *is nonnegative and the sign conditions* (6.5) *are satisfied, i.e. the Lagrange multipliers* (6.1) *associated with the control constraints have the correct sign.*

In this case, non-optimality can only be due to infeasibility of y_{u^s}. The state exceeds b in some points x of Ω. If we can trust in our numerical method, then these infeasible points should be close to the active grid points $x_1^s, \dots x_n^s$ of y^s, say in the interior of some of the neighboring triangles. For this case, we discuss the application of a generalized intermediate value theorem by Miranda [18] to deduce that the exact vector of active points $(\bar{x}_1, \dots, \bar{x}_n)$ exists in a certain small region.

6.2 Optimality conditions and Case a)

In this case, the state y_{u^s} satisfies the state constraints and is active in exactly n points $x_1^s, \dots, x_n^s$, i.e.

$$\sum_{i=1}^m u_i^s y_i(x) < b, \quad \text{for all } x \in \Omega,\ x \neq x_j^s,\ j = 1, \dots, n, \tag{6.2a}$$

$$\sum_{i=1}^m u_i^s y_i(x_j^s) = b, \quad j = 1, \dots, n. \tag{6.2b}$$

Moreover, we have assumed that $m - n$ components of u^s are active, i.e. equal to u_a or u_b. We may assume after a re-numbering, if necessary, that these are the components $u_{n+1}^s, \dots, u_m^s$.

If u^s were optimal, it would satisfy the variational inequality

$$\left\langle \nabla f(u^s) + \int_\Omega (y_1, \dots, y_m)^\top \, d\mu^s \, , \, u - u^s \right\rangle \geq 0 \quad \forall u \in U_{ad}, \tag{6.3}$$

where the nonnegative regular Borel measure μ^s is the Lagrange multiplier associated with the state constraints. It is obtained via (6.7) below.

This is equivalent to the sign conditions

$$\frac{\partial f}{\partial u_i}(u^s) + \int_\Omega y_i(x)\, d\mu^s \begin{cases} = 0 & \text{if } u_a < u_i^s < u_b, \\ \geq 0 & \text{if } u_i^s = u_a, \\ \leq 0 & \text{if } u_i^s = u_b. \end{cases} \tag{6.4}$$

Moreover, the complementary slackness condition

$$\int_{\bar{\Omega}} (y_{u^s}(\cdot) - b)\ d\mu^s = 0$$

must be satisfied.

Then (6.4) implies in particular

$$\frac{\partial f}{\partial u_i}(u^s) + \int_\Omega y_i(x)\, d\mu^s \begin{cases} = 0 & \text{if } i \in \{1,\dots,n\}, \\ \geq 0 & \text{if } i \in \{n+1,\dots,m\} \text{ and } u_i^s = u_a, \\ \leq 0 & \text{if } i \in \{n+1,\dots,m\} \text{ and } u_i^s = u_b. \end{cases} \tag{6.5}$$

In view of the complementarity conditions (6.2), μ^s is of the form

$$\mu^s = \sum_{j=1}^n \mu_j^s \delta_{x_j^s}, \quad \mu_j^s \geq 0, \quad j = 1,\dots,n, \tag{6.6}$$

with Dirac's measures $\delta_{x_j^s}$ concentrated at the active points x_j^s, for $j = 1,\dots,n$, and real numbers μ_j^s. Therefore, (6.5) yields in particular

$$\frac{\partial f}{\partial u_i}(u^s) + \sum_{j=1}^n \mu_j^s y_i(x_j^s) \; = 0, \quad i = 1,\dots,n. \tag{6.7}$$

For our a-posteriori estimation, we assume:

***Assumption* 6.** *The matrix*

$$\big(y_i(x_j^s)\big), \quad i,j = 1,\dots,n, \tag{6.8}$$

is non-singular.

Then the linear system (6.7) has a unique solution $\mu^s = (\mu_1^s,\dots,\mu_n^s)$.

This assumption is fulfilled in particular, if the points x_j^s are close enough to the unknown active points $\bar{x}_j$ and Assumption 2 is satisfied.

If $\mu^s \geq 0$ and the sign conditions (6.5) are fulfilled, then u^s satisfies the first order necessary optimality conditions of the problem (**P**). By convexity of the problem, they are also sufficient and hence u^s would be the optimal control.

In the opposite case, at least one component μ_j^s of μ^s is negative or the sign conditions (6.5) do not hold true. Then we proceed as follows:

First, we split the sum in (6.7)

$$\frac{\partial f}{\partial u_i}(u^s) + \sum_{\mu_j^s \geq 0} \mu_j^s y_i(x_j^s) + \sum_{\mu_j^s < 0} \mu_j^s y_i(x_j^s) = 0, \qquad i = 1,\dots,n, \tag{6.9}$$

and define the first n components of a perturbation vector $\zeta \in \mathbb{R}^m$ by the sum associated with the negative components of μ^s,

$$\zeta_i = \sum_{\mu_j^s < 0} \mu_j^s y_i(x_j^s), \qquad i = 1,\dots,n.$$

Then we have

$$\frac{\partial f}{\partial u_i}(u^s) + \zeta_i + \sum_{j=1}^n \tilde{\mu}_j y_i(x_j^s) = 0, \qquad i = 1,\dots,n, \tag{6.10}$$

where $\tilde{\mu}$ is the vector defined as the nonnegative part of μ^s by $\tilde{\mu}_j = (|\mu_j^s| + \mu_j^s)/2$ for $j = 1, \ldots, n$.

Now the Lagrange multiplier μ is updated to obey nonnegativity conditions, but the sign conditions of (6.5) must be checked as well. Following [1] we define the remaining components $\zeta_{n+1}, \ldots, \zeta_m$ of ζ by

$$\zeta_i = \begin{cases} \Big[\dfrac{\partial f}{\partial u_i}(u^s) + \displaystyle\sum_{j=1}^{n} \tilde{\mu}_j y_i(x_j^s)\Big]_- & \text{if } i \in \{n+1, \ldots, m\} \text{ and } u_i^s = u_a, \\ \Big[\dfrac{\partial f}{\partial u_i}(u^s) + \displaystyle\sum_{j=1}^{n} \tilde{\mu}_j y_i(x_j^s)\Big]_+ & \text{if } i \in \{n+1, \ldots, m\} \text{ and } u_i^s = u_b. \end{cases} \tag{6.11}$$

Then it holds

$$\frac{\partial f}{\partial u_i}(u^s) + \zeta_i + \sum_{j=1}^{n} \tilde{\mu}_j y_i(x_j^s) = \begin{cases} = 0 & \text{if } i \in \{1, \ldots, n\}, \\ \geq 0 & \text{if } i \in \{n+1, \ldots, m\} \text{ and } u_i^s = u_a, \\ \leq 0 & \text{if } i \in \{n+1, \ldots, m\} \text{ and } u_i^s = u_b, \end{cases}$$

i.e. the sign conditions (6.5) are satisfied after adding the perturbation vector ζ.

Therefore, u^s fulfills the optimality conditions for the following perturbed problem:

$$(\mathbf{P}_\zeta) \begin{cases} \min_{u \in U_{ad}} f(u) + \zeta^\top u \\ \displaystyle\sum_{i=1}^{m} u_i y_i(x) \leq b. \end{cases}$$

Now it is easy to estimate the distance to the unknown exact optimal control $\bar{u}$. This is explained in the next section.

6.3 A-posteriori error estimation in Case a)

Define the feasible set $\mathcal{M}$ of $(\mathbf{P})$ by

$$\mathcal{M} = \{u \in U_{\text{ad}} : \sum_{i=1}^{m} u_i y_i(x) \leq b \quad \forall x \in \Omega\}.$$

Then $\bar{u}$ solves $\min_{u \in \mathcal{M}} f(u)$, while u^s solves $\min_{u \in \mathcal{M}} f(u) + \zeta^\top u$. Both vectors have to satisfy their associated variational inequalities, hence

$$\nabla f(\bar{u})^\top (u^s - \bar{u}) \geq 0, \tag{6.12a}$$

$$(\nabla f(u^s) + \zeta)^\top (\bar{u} - u^s) \geq 0. \tag{6.12b}$$

To evaluate the variational inequalities (6.12), we have to expand $f(u)$:

$$\begin{aligned} f(u) &= \frac{1}{2} \| \sum_{i=1}^{m} u_i y_i - y_d \|^2 + \frac{\nu}{2} |u|^2 \\ &= \frac{1}{2} u^\top C u - a^\top u + \|y_d\|^2 + \frac{\nu}{2} |u|^2, \end{aligned} \tag{6.13}$$

where $C \in \mathbb{R}^{m\times m}$ is the matrix with entries

$$c_{ij} = \int_\Omega y_i(x) y_j(x)\, dx$$

for $i, j = 1, \ldots, m$, and $a \in \mathbb{R}^m$ is the vector with the components

$$a_i = \int_\Omega y_i(x) y_d(x)\, dx.$$

We now formulate our main result on a-posteriori error estimation.

Theorem 6. *Let $\lambda_{min} \geq 0$ be the smallest eigenvalue of C. If $\bar{u}$ is optimal for $(\mathbf{P})$ and u^s is optimal for $(\mathbf{P}_\zeta)$, then*

$$|u^s - \bar{u}| \leq \frac{|\zeta|}{\nu + \lambda_{min}}. \tag{6.14}$$

Proof. From (6.13) we have that $\nabla f(u) = Cu - a + \nu u$. Adding both inequalities of (6.12), we find

$$\left(C\left(u^s - \bar{u}\right) + \nu(u^s - \bar{u}) + \zeta\right)^\top (\bar{u} - u^s) \geq 0,$$

and hence

$$\zeta^\top(\bar{u} - u^s) \geq (\bar{u} - u^s)^\top C(\bar{u} - u^s) + \nu|u^s - \bar{u}|^2.$$

Since $v^\top C v \geq \lambda_{min}|v|^2$ holds for all $v \in \mathbb{R}^m$, we deduce

$$\lambda_{min}|u^s - \bar{u}|^2 + \nu|u^s - \bar{u}|^2 \leq \zeta^\top(\bar{u} - u^s),$$

which implies that

$$|u^s - \bar{u}| \leq \frac{|\zeta|}{\nu + \lambda_{min}}. \tag{6.15}$$

□

6.4 Verification of a feasible control with n active points in Case b)

As we pointed out in the last section, in Case b) we want to verify the existence of a feasible control that has exactly n active points close to the ones obtained from a numerically computed control u^s. Thanks to our assumptions, there exists at least one control of this type, namely the unknown optimal control $\bar{u}$. To study this question, we consider first the mapping defined in a neighborhood of $\bar{u}$ that assigns to a vector u the local maxima of the function y_u located in a neighborhood of the optimal active points $\bar{x}_1, \ldots, \bar{x}_n$.

Let us assume in this section that the number of active points is equal to m, the number of all (unknown) controls. This is a natural assumption for the

following reason: If $n < m$, then the remaining $m - n$ control components were assumed to be strictly active equal to u_a or u_b. If the precision is sufficiently high, then we can assume to know these components. Then only the remaining n control components must be determined. We need this reasonable simplification for the application of Miranda's theorem.

Lemma 1. *Under the Assumptions 1–4, there are constants $r_u > 0$ and $r_x > 0$ with the following properties: For all $u \in B(\bar{u}, r_u)$ and all $j \in \{1, \dots, m\}$ there exists a unique point $x_j \in B(\bar{x}_j, r_x)$ such that $\nabla y_u(x_j) = 0$. Moreover, y_u is strictly concave on $B(\bar{x}_j, r_x)$ for $j = 1, \dots, m$, and hence $x_1, \dots, x_m$ are strict maxima of y_u in $B(\bar{x}_1, r_x), \dots, B(\bar{x}_m, r_x)$, respectively. The mapping $u \mapsto (x_1, \dots, x_m)$ is of class C^2 from $B(\bar{u}, r_u)$ to $\mathbb{R}^m$.*

Proof. Under our assumptions, we have that $y_i \in C^2(\Omega)$ for all $i = 1, \dots, m$. Let us define the mapping $F : \mathbb{R}^m \times \Omega \to \mathbb{R}^m$ by

$$F(u, x) = \nabla y_u(x) = \sum_{i=1}^{m} u_i \nabla y_i(x).$$

Next, we select a fixed $j \in \{1, \dots, m\}$; we know that $F(\bar{u}, \bar{x}_j) = 0$. Moreover, the partial Jacobian of F with respect to x is

$$\frac{\partial F}{\partial x}(u, x) = y_u''(x),$$

i.e. the Hessian of y_u with respect to x. Thanks to our assumptions, the matrix $y_u''(\bar{x}_j)$ is non-singular. Now the implicit function theorem ensures the existence of open neighborhoods $B(\bar{u}, r_u)$, $B(\bar{x}_j, r_x)$ and the existence of a mapping $\varphi_j : B(\bar{u}, r_u) \to B(\bar{x}_j, r_x)$ such that $\varphi_j(\bar{u}) = \bar{x}_j$ and $F(u, \varphi_j(u)) = 0$ for all $u \in B(\bar{u}, r_u)$. Therefore, it holds

$$\nabla y_u(\varphi_j(u)) = 0, \quad \forall u \in B(\bar{u}, r_u). \tag{6.16}$$

The mapping φ_j is of class C^2. If r_u is taken sufficiently small, then $y_u''(\varphi_j(u))$ is uniformly negative definite in $B(\bar{x}_j, r_x)$, i.e. there exists $\delta > 0$ such that for all $x \in B(\bar{x}_j, r_x)$ it holds

$$-\xi^\top y_u''(x)\xi \geq \delta |\xi|^2 \quad \forall \xi \in \mathbb{R}^m. \tag{6.17}$$

This follows from Assumption 3 and the fact that the function y_u is of class C^2 for all $u \in \mathbb{R}^m$. Therefore, the function y_u attains its strict maximum in $B(\bar{x}_j, r_x)$ in the point $x_j = \varphi_j(u)$. Locally, there cannot be any other local extremum of y_u. The same arguments apply to all $j \in \{1, \dots, m\}$. The mapping $u \mapsto (\varphi_1(u), \dots, \varphi_m(u))$ is of class C^2 on $B(\bar{u}, r_u)$. □

Lemma 2. *There exist positive constants ρ_u, ρ_x, and $\delta > 0$ with the following property: If $u \in B(\bar{u}, \rho_u)$ and $j \in \{1, \dots, m\}$, then there holds*

$$y_u(x) \leq b - \delta \quad \forall\, x \in \Omega \setminus B(\bar{x}_j, \rho_x). \tag{6.18}$$

This result follows from Assumption 3, and can be shown in the same way as Lemma 3.15 in [16]: In a ball around $\bar{x}_j$, $j \in \{1, \dots, m\}$, $y_{\bar{u}}$ decays quadratically , while $\bar{y}(x) \leq b - 2\delta$ holds for all x outside these balls. By continuity, this property is stable with respect to small perturbations of $\bar{u}$ so that still $y_u(x) \leq b - \delta$ holds.

In view of this lemma, sufficiently small changes of u around $\bar{u}$ do not destroy the feasibility of u outside $B(\bar{x}_j, \rho_x)$, if $y_u(\varphi_i(u)) \leq b$ is satisfied for all $i = 1, \dots, m$. We consider now the mapping $g : B(\bar{u}, \rho_u) \to \mathbb{R}^m$ defined by

$$g(u) = (y_u(\varphi_1(u)), \dots, y_u(\varphi_m(u))) \,. \tag{6.19}$$

This is the mapping that assigns to u the values of y_u in its m maximum points $x_1 = \varphi_1(u)$, ..., $x_m = \varphi_m(u)$ that were discussed in the preceding lemma.

To apply Miranda's theorem, we study numerically the following problem: For varying vectors $\beta \in \mathbb{R}^m$ close to $(b, \dots, b)$, we want to compute controls u in $B(\bar{u}, \rho_u)$ such that

$$y_u(\varphi_i(u)) = \beta_i, \quad i = 1, \dots, m. \tag{6.20}$$

In other words, we want to determine control vectors u such that the values of y_u in its k maximum points are equal to β_i. By a suitable variation of the vectors β, we want to prove the existence of a vector u^s in a rectangle around $\bar{u}$ such that

$$y_{u^s}(\varphi_i(u^s)) = b, \quad i = 1, \dots, m. \tag{6.21}$$

Such a vector u^s would be active in $\hat{x}_1 = \varphi_1(u^s), \dots, \hat{x}_m = \varphi_m(u^s)$ and would also be feasible, i.e. $y_{u^s}(x) \leq b$ holds for all $x \in \Omega$.

To this aim, we use a theorem of [23] that generalizes a theorem by Miranda [18]. Miranda's result can be considered as a generalization of the intermediate value theorem to vector-valued continuous functions. The theorem uses the m-dimensional cuboid $\mathcal{R} \subset \mathbb{R}^m$ defined by

$$\mathcal{R} = \{u \in \mathbb{R}^m : |u_i - u_i^\circ| \leq L_i,\, i = 1, \dots, m\}$$

where $2L_i > 0$, $i = 1, \dots, k$, are the lengths of the rectangle, and $u^\circ \in \mathbb{R}^m$ is its center. We define the sets

$$\begin{aligned} \mathcal{R}_i^+ &= \{u \in \mathcal{R} : u_i = u_i^\circ + L_i\}, \text{ and} \\ \mathcal{R}_i^- &= \{u \in \mathcal{R} : u_i = u_i^\circ - L_i\}. \end{aligned}$$

Theorem 7. *([18],[23]) Let $f : \mathbb{R}^m \to \mathbb{R}^m$ be a continuous function. Assume that*

$$f_i(u) f_i(v) \leq 0 \quad \forall (u, v) \in \mathcal{R}_i^+ \times \mathcal{R}_i^- \tag{6.22}$$

holds for all $i \in \{1, \dots, m\}$. Then there exists at least one $u^s \in \mathcal{R}$ such that $f(u^s) = 0$.

We apply this theorem in the following situation: After the numerical solution by the a-posteriori error estimation method, we have a control u^s that is suboptimal. In points x_j^s, the FE approximation y_h^s of the associated state satisfies the equations

$$y_h^s(x_j^s) = b, \quad j = 1, \ldots, m$$

but most likely

$$y_{u^s}(x_j^s) \neq b \qquad \text{and} \qquad y_{u^s} \not\leq b.$$

Our aim is to find the location of a point $(x_1, \ldots, x_m)$ and a control u such that

$$y_u(x_j) = b, \quad j = 1, \ldots, m, \qquad y_u(x) < b \quad \forall x \neq x_j.$$

***Remark* 5.** *In general, we do not know y_u exactly. Therefore, we can only show a strong indication that the relations above are fulfilled.*

We select a cuboid $\mathcal{R}$ around u^s, where u^s is the suboptimal control with associated FE-state y_h^s satisfying $y_h^s(x_j^s) = b$, $j = 1, \ldots, m$. Then it will most likely hold

$$y_{u^s}(x_j^s) = \sum_{i=1}^{m} u_i^s y_i(x_j^s) \neq b.$$

Now we define $2k$ control values

$$u_{i\pm} := u_i^s \pm L_i, \quad i = 1, \ldots, m$$

and select all possible 2^m combinations of them that form pairwise different vectors of $\mathbb{R}^m$. These are the corners of the cuboid $\mathcal{R}$ with midpoint u^s.

We consider the mapping $g : u \mapsto (\varphi_1(u), \ldots, \varphi_m(u))$ that assigns to u the m maximum values of y_u according to Lemma 2. The selection of the values L_i should be such that the mapping

$$f : u \mapsto g(u) - (b, \ldots, b)$$

satisfies the assumptions of Theorem 7. We assume that this is possible. A sufficient condition is

$$g_i(u) \begin{cases} \geq b & \forall u \in \mathcal{R}_i^+ \\ \leq b & \forall u \in \mathcal{R}_i^-, \end{cases}$$

for all $i = 1, \ldots, m$. Then there exists a solution $u^s \in \mathcal{R}$ with $g_i(u^s) = b$ for all $i = 1, \ldots, m$.

***Remark* 6.** *For large dimension m, the selection of the L_i is certainly a tedious procedure. Basically, the conditions of Theorem 7 have to be checked for different combinations of u, v, and L_i. It is clear that this can only give an indication that these conditions are actually fulfilled. However, this issue is somehow intrinsic to all control problems in function spaces. Compare for instance results related to second order sufficient conditions for optimal control problems, where a certain positivity condition for the reduced Hessian associated with the undiscretized problem is assumed. Yet, the only information available*

	A	B	C	D
g_1	0.01107	0.01099	0.00986	0.00993
g_2	0.01086	0.00979	0.00938	0.01045
μ_1	36.9424	29.7339	47.8148	29.7334
μ_2	137.4811	156.8503	155.2044	156.8502

Table 5: Maximum values of g and their corresponding approximated Lagrange multipliers

is for discretized versions of the problem. As a rule of thumb, it is fairly impossible to conclude certain properties of the infinite dimensional setting from finite dimensions, see e.g. [22]. Here, we are faced with a similar problem. We do not claim that the process described above is a practicable test. Instead, the use of the Miranda theorem provides at least a theoretical tool that allows to address this question of predicting a feasible point in the neighborhood of a computed unfeasible one.

We apply this with $m = 2$ to the "maximum point mapping" $f := g - b$ in our Example 2. There, our optimal result obtained from the a-posteriori technique was

$$u_s = [1.1776870 \;\; 0.8661457]$$

We select $u^\circ = [1.2346 \;\; 0.8601]$ as midpoint. Perturbing the components of u° by $L = \pm 0.07$, we obtain a rectangle with the 4 corners $A = [1.3046 \;\; 0.9301]$, $B = [1.3046 \;\; 0.7901]$, $C = [1.1646 \;\; 0.7901]$ and $D = [1.1646 \;\; 0.9301]$.

Next, we approximate $y_A, \ldots, y_D$ by solving the PDE on a very fine uniform mesh with mesh size $h \sim 0.006$ to generate "exact states". Finally, we determine the 2 local maxima g_1, g_2 for $y_A, \ldots, y_D$. Recall that $b = 0.01$.

All associated Lagrange multipliers, which were determined by formula (6.7), are positive. This indicates that the optimal control $\bar{u}$ is included in the rectangle with the corners A, B, C, D .

References

[1] N. Arada, E. Casas, and F. Tröltzsch. Error estimates for the numerical approximation of a semilinear elliptic control problem. *Computational Optimization and Applications*, 23:201–229, 2002.

[2] R. Becker and R. Rannacher. An optimal control approach to a-posteriori error estimation in finite element methods. *Acta Numerica*, 10:1–102, 2001.

[3] Olaf Benedix and Boris Vexler. A posteriori error estimation and adaptivity for elliptic optimal control problems with state constraints. *Comput. Optim. Appl.*, 44(1):3–25, 2009.

[4] J.C. De los Reyes, P. Merino, J. Rehberg, and F. Tröltzsch. Optimality conditions for state-constrained pde control problems with time-dependent controls. *Control and Cybernetics*, 37:7–38, 2008.

[5] A. L. Dontchev, W. W. Hager, A. B. Poore, and B. Yang. Optimality, stability, and convergence in nonlinear control. *Applied Math. and Optimization*, 31:297–326, 1995.

[6] Kenneth Eriksson, Don Estep, Peter Hansbo, and Claes Johnson. Introduction to adaptive methods for differential equations. In *Acta numerica, 1995*, Acta Numer., pages 105–158. Cambridge Univ. Press, Cambridge, 1995.

[7] David Gilbarg and Neil S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Springer, 3rd edition, 1998.

[8] A. Günther, M. Hinze, and M.H. Tber. A posteriori error representation for elliptic optilmal control problems with control and state constraints. In G. Leugering, S. Engell, A. Griewank, M. Hinze, R. Rannacher, V. Schulz, M. Ulbrich, and S. Ulbrich, editors, *Constrained Optimization and Optimal Control for Partial Differential Equations*, volume 160 of *ISNM*, pages 303–318, Basel, 2012. Birkhäuser Verlag.

[9] R. Hettich and K. O. Kortanek. Semi-infinite programming: theory, methods, and applications. *SIAM Rev.*, 35(3):380–429, 1993.

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

[11] M. Hinze and A. Rösch. Discretization of optimal control problems. In G. Leugering, S. Engell, A. Griewank, M. Hinze, R. Rannacher, V. Schulz, M. Ulbrich, and S. Ulbrich, editors, *Constrained Optimization and Optimal Control for Partial Differential Equations*, volume 160 of *ISNM*, pages 391–430, Basel, 2012. Birkhäuser Verlag.

[12] K. Kohls, A. Rösch, and K.G. Siebert. A posteriori error estimators for control constrained optimal control problems. In G. Leugering, S. Engell, A. Griewank, M. Hinze, R. Rannacher, V. Schulz, M. Ulbrich, and S. Ulbrich, editors, *Constrained Optimization and Optimal Control for Partial Differential Equations*, volume 160 of *ISNM*, pages 431–444, Basel, 2012. Birkhäuser Verlag.

[13] D. G. Luenberger. *Optimization by Vector Space Methods*. Wiley, New York, 1969.

[14] K. Malanowski, Ch. Büskens, and H. Maurer. Convergence of approximations to nonlinear optimal control problems. In *Mathematical Programming*

with Data Perturbations, volume 195 of *Lecture Notes in Pure and Applied Mathematics*, pages 253–284, New York, 1997. Marcel-Dekker, Inc.

[15] P. Merino, I. Neitzel, and F. Tröltzsch. Error estimates for the finite-element discretization of semi-infinite elliptic control problems. *Discussiones Mathematicae*, 30:221–236, 2010.

[16] P. Merino, I. Neitzel, and F. Tröltzsch. On linear-quadratic elliptic control problems of semi-infinite type. *Applicable Analysis*, 90:1047–1074, 2011.

[17] P. Merino, F. Tröltzsch, and B. Vexler. Error estimates for the finite element approximation of a semilinear elliptic control problem with state constraints and finite dimensional control space. *ESAIM: Mathematical Modelling and Numerical Analysis*, 44:167 – 188, 2010.

[18] Carlo Miranda. Un'osservazione su un teorema di Brouwer. *Boll. Un. Mat. Ital. (2)*, 3:5–7, 1940.

[19] Ricardo H. Nochetto, Alfred Schmidt, Kunibert G. Siebert, and Andreas Veeser. Pointwise a-posteriori error estimates for monotone semi-linear equations. *Numer. Math.*, 104(4):515–538, 2006.

[20] Rolf Rannacher and Boris Vexler. A priori error estimates for the finite element discretization of elliptic parameter identification problems with pointwise measurements. *SIAM Journal on Control and Optimization*, 44:1844–1863, 2005.

[21] Rembert Reemtsen and Stephan Görner. Numerical methods for semi-infinite programming: a survey. In *Semi-infinite programming*, volume 25 of *Nonconvex Optim. Appl.*, pages 195–275. Kluwer Acad. Publ., Boston, MA, 1998.

[22] A. Rösch and D. Wachsmuth. Numerical verification of optimality conditions. *SIAM J. Control Optimization*, 47(5):2557–2581, 2008.

[23] Uwe Schäfer. A fixed point theorem based on Miranda. *Fixed Point Theory Appl.*, pages Art. ID 78706, 5, 2007.

[24] T. Tonn, K. Urban, and S. Volkwein. Optimal control of parameter-dependent convection-diffusion problems around rigid bodies. *SIAM J. Scientific Computing*, 32:1237–1260, 2010.

[25] F. Tröltzsch and S. Volkwein. The SQP method for control constrained optimal control of the Burgers equation. *ESAIM: Control, Optimisation and Calculus of Variations*, 6:649–674, 2001.