Hamburger Beiträge
zur Angewandten Mathematik

Adaptive optimal control of the obstacle problem

Christian Meyer, Andreas Rademacher
und Winnifried Wollner

Erschienen in: Ergebnisberichte des Instituts für Angewandte Mathematik,
Fakultät für Mathematik, TU Dortmund, Nummer 494

Nr. 2014-10
May 2014

Adaptive optimal control of the obstacle problem

Ch. Meyer, A. Rademacher, W. Wollner

April 30, 2014

Abstract

The article is concerned with the derivation of a posteriori error estimates for optimization problems subject to an obstacle problem. To circumvent the nondifferentiability inherent to this type of problem, we introduce a sequence of penalized but differentiable problem. We show differentiability of the central path and derive separate a posteriori dual weighted residual estimates for the errors due to penalization, discretization, and iterative solution of the discrete problems. The effectivity of the derived estimates is demonstrated on two numerical examples.

1. Introduction

This paper is concerned with the adaptive approximation of optimal control problems governed by the obstacle problem. The construction of the algorithm is based on a regularization approach in combination with an adaptive finite element discretization of the regularized problems. The two errors induced in this way, i.e., the regularization error and the discretization error, are equilibrated by means of suitable error estimators based on the dual weighted residual (DWR) method.

Regarding the adaptive approximation of the obstacle problem itself, there is a large amount of contributions regarding a posteriori error estimates available in the literature, see for instance [6, 28] for dual weighted error estimates, [15, 11, 30, 8, 9] for residual type estimates. In particular, we refer to [12] where residual type error estimates for a penalized obstacle problem where derived.

In contrast to the solution of the obstacle problem itself, consideration of optimization problems subject to the obstacle problem is complicated by the nondifferentiability of the solution operator of the obstacle problem, see e.g. [19]. To this end, we consider a sequence of penalized obstacle problems as constraints for our optimization problem. Such an approach is classical and has been investigated by various authors before. We only refer to [3, 13, 27] and the references therein. For the penalized but differentiable problems, we derive DWR error estimates following the pioneering work of [4],

see also [5, 2]. More precisely, we utilize the DWR estimates for control constrained problems proposed in [31]. To simultaneously control the error due to penalization and discretization, we extend the ideas of [34, 33] for optimization problems with regularized pointwise state constraints to regularization in the constraining equation, see also the survey [22]. Finally, we include the possibility to balance the former two errors with the error due to the iterative solution of the problems adapting the work of [24, 23].

To the best of our knowledge, the only contribution in the field of adaptive approximation of optimal control problems governed by variational inequalities is [14]. This work is also concerned with an optimal control problem subject to the obstacle problem, but the authors directly apply the DWR method to the original problem without regularization and penalization, respectively. In contrast to this, as mentioned before, our strategy is to regularize the problem by penalization, which allows to derive error estimates for the penalized problems by the classical DWR method. Afterwards, we will equilibrate the discretization error and the regularization error by means of a regularization error estimator which is based on the path derivative of the solution of the regularized problems w.r.t. the penalization parameter.

The paper is organized as follows: After introducing the specific optimal control problem under consideration and stating the standing assumptions in Section 2. We present the regularization, in Section 3, and perform a limit analysis for penalty parameter tending to infinity. Section 4 is then devoted to the estimation of the regularization error by means of the path derivative, while Section 5 deals with the a posteriori error estimation of the discretization error for the regularized problems. Numerical experiments illustrating the efficiency of our approach are presented in Section 6. The paper ends with some concluding remarks in Section 7.

2. Problem formulation and standing assumptions

Throughout this paper, we consider the following optimal control problem governed by the obstacle problem

$$\left.\begin{array}{ll} \min & J(u,q) \\ \text{s.t.} & a(u, v-u) \geq \langle q, v-u\rangle \quad \forall\, v \in K \\ & u \in K, \quad q \in L^2(\Omega) \end{array}\right\} \tag{P}$$

where $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, is a bounded domain.

We suppose the following standing assumptions on the data in (P):
The feasible set K is given by

$$K = \{v \in H_0^1(\Omega) : v \geq \psi \text{ a.e. in } \Omega\}$$

with $\psi \in H_0^1(\Omega)$ given. The dual pairing between $H_0^1(\Omega)$ and $H^{-1}(\Omega) := H_0^1(\Omega)^*$ is denoted by $\langle ., . \rangle$. Moreover, the bilinear form $a : H_0^1(\Omega) \times H_0^1(\Omega)$ is given by the

following second-order elliptic operator

$$a(u,v) = \int_\Omega \sum_{i=1}^{d} \Big(\sum_{j=1}^{d} a_{ij} \frac{\partial u}{\partial x_j} \frac{\partial v}{\partial x_j} dx + b_i \frac{\partial u}{\partial x_i} v \Big) + a_0 \, u \, v \, dx \tag{1}$$

where $a_{ij}, b_i, a_0 \in L^\infty(\Omega)$, $i,j = 1,..,d$, are such that a is coercive, i.e.,

$$a(u,u) \geq \beta \, \|u\|^2_{H^1(\Omega)} \quad \forall \, u \in H^1_0(\Omega) \tag{2}$$

with a constant $\beta > 0$. In addition, we require

$$a_0 \geq 0. \tag{3}$$

By $A : H^1_0(\Omega) \to H^{-1}(\Omega)$, we denote the operator induced by a, i.e., $\langle Au, v \rangle = a(u,v)$ for all $u, v \in H^1_0(\Omega)$. Finally,

$$J(u,q) = j(u) + g(q), \tag{4}$$

where $g : L^2(\Omega) \to \mathbb{R}$ and $j : H^1_0(\Omega) \to \mathbb{R}$ are supposed to be three times continuously differentiable. Moreover, j is assumed to be bounded from below and, further, that there is a constant $\alpha > 0$ such that

$$g''(u)h^2 \geq \alpha \, \|h\|^2_{L^2(\Omega)} \quad \forall \, u, h \in L^2(\Omega). \tag{5}$$

It is well known that the VI in (P), i.e.

$$u \in K, \quad a(u, v-u) \geq \langle q, v-u \rangle \quad \forall \, v \in K, \tag{6}$$

can equivalently be reformulated by a complementarity system, since $K - \psi$ is a convex cone. The optimal control problem then reads

$$\text{(P)} \quad \Leftrightarrow \quad \left\{ \begin{array}{ll} \min & J(u,q) \\ \text{s.t.} & Au = q + \lambda \\ & u \geq \psi \text{ a.e. in } \Omega, \quad \lambda \geq 0 \text{ in } H^{-1}(\Omega), \quad \langle \lambda, \psi - u \rangle = 0, \end{array} \right.$$

where $\lambda \in H^{-1}(\Omega)$ is the corresponding slack variable.

Based on the maximal monotony of $A + \partial I_K$, where I_K denotes the indicator functional associated with K, one shows by standard arguments that (6) admits for every $q \in H^{-1}(\Omega)$ a unique solution $u \in H^1_0(\Omega)$. Furthermore, it is easily seen that the corresponding solution operator $S : H^{-1}(\Omega) \ni q \mapsto u \in H^1_0(\Omega)$ is globally Lipschitz continuous. Based on this result and the special structure of J in (4) and (5), it is shown by standard arguments that (P) admits a least one globally optimal solution. However, due to the nonlinearity of S, the problem is not convex in general so that uniqueness of the global minimizer cannot be expected.

3. Regularization: Known and preliminary results

Although S is globally Lipschitz, it is not Gâteaux-differentiable, since the directional derivative at q in direction h is itself a solution of a VI of first kind, as shown by Mignot [19]. Therefore a standard adjoint approach to tackle (P) is not possible and various regularization approaches have been introduced to smooth the control-to-state map S. Our regularization of (P) is given by

$$\left.\begin{array}{ll} \min & J(u,q) \\ \text{s.t.} & Au + r(\gamma;u) = q \end{array}\right\} \tag{P_γ}$$

where $r : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$ is given by

$$r(\gamma;u) := -\left[\max\left(\gamma(\psi - u), 0\right)\right]^3. \tag{7}$$

This choice of r stems from a bi-quadratic penalization of the energy functional associated with (6). Of course, other choices of r are frequently in use such as e.g. $r(\gamma;u) = -\gamma \max_\gamma(\psi - u)$, where $\max_\gamma$ denotes a suitable smoothed version of the max-function, see for instance [17]. An advantage of our particular choice for the regularization is that the Nemyzkii operator associated with the nonlinearity in (7) is twice continuously Fréchet-differentiable in $L^\infty(\Omega)$, which allows to solve the regularized optimal control problems by standard second-order methods such as SQP. Since

$$H_0^1(\Omega) \hookrightarrow L^4(\Omega) \ni u \mapsto r(\gamma;u) \in L^{4/3}(\Omega) \hookrightarrow H^{-1}(\Omega)$$

is a locally Lipschitz continuous and monotone operator, Browder and Minty's theorem on monotone operators yields existence and uniqueness of a solution to the PDE in (P_γ), i.e.

$$Au + r(\gamma;u) = q \tag{8}$$

for every $\gamma > 0$. The associated solution operator is denoted by $S_\gamma : H^{-1}(\Omega) \to H_0^1(\Omega)$. Furthermore, using again the monotony of $t \mapsto \max\{t,0\}^3$, one easily deduces that S_γ is Lipschitz continuous with the same Lipschitz constant as S, hence independent of γ. Thus, by completely identical arguments as in case of (P), one deduces the existence of a global solution to (P_γ).

Owing to the monotonicity of $r(\gamma;\,.\,)$, we can apply Stampacchia's classical technique, cf. [16], to prove the following

Lemma 1. *For every $q \in L^2(\Omega)$ the unique solution u of (8) is essentially bounded.*

The differentiability of $r(\gamma;\,.\,)$ in $L^\infty(\Omega)$ for fixed γ, then allows to derive first-order necessary optimality conditions for the regularized problems in a standard way, see, e.g., [29]. In this way one obtains the following result:

Proposition 1. *Let q_γ be a local optimum of (P_γ) with associated state $u_\gamma = S_\gamma(q_\gamma)$.*

Then there exist $\lambda_\gamma, \mu_\gamma \in L^2(\Omega)$ *and* $p_\gamma, \theta_\gamma \in H_0^1(\Omega)$ *such that*

$$Au_\gamma = q_\gamma + \lambda_\gamma, \tag{9a}$$
$$\lambda_\gamma + r(\gamma; u_\gamma) = 0, \tag{9b}$$
$$A^* p_\gamma = \nabla j(u_\gamma) - \mu_\gamma, \tag{9c}$$
$$p_\gamma + \nabla g(q_\gamma) = 0, \tag{9d}$$
$$p_\gamma - \theta_\gamma = 0, \tag{9e}$$
$$\mu_\gamma - \partial_u r(\gamma; u_\gamma)\theta_\gamma = 0. \tag{9f}$$

Note that p_γ *and thus* θ_γ *and* μ_γ *are uniquely defined by* (9d).

Observe that p_γ is nothing else than the adjoint state. Note further that μ_γ and θ_γ can be eliminated directly from the system, but we introduced them for reasons of comparison with later optimality systems.

We now address the convergence for $\gamma \to \infty$. Concerning the state equation, the following approximation result holds true:

Lemma 2. *Let* $q \in H^{-1}(\Omega)$ *be given and denote by* $u, u_\gamma \in H_0^1(\Omega)$ *the solutions to* (6) *and* (8)*, respectively. Then* $u_\gamma \to u$ *strongly in* $H_0^1(\Omega)$ *as* $\gamma \to \infty$*. If we further assume that* $q, A\psi \in L^{4/3}(\Omega)$*, then there exists a constant* $c > 0$ *so that*

$$\|u - u_\gamma\|_{H^1(\Omega)} \le c \frac{1}{\sqrt{\gamma}} \|q - A\psi\|_{L^{4/3}(\Omega)}^{2/3}.$$

The proof follows by classical arguments and is therefore postponed to Appendix A. With the above result at hand, it is straightforward to prove the following first-order necessary optimality conditions for (P), the so-called Clarke(C)-stationarity conditions:

Theorem 1. *1. For every* $\gamma > 0$ *there is a globally optimal solution of* (P_γ)*, denoted by* q_γ*. If* $\gamma \to \infty$*, then every sequence* $\{q_\gamma\}$ *of global minimizers of* (P_γ) *admits a weak accumulation point* $\bar{q} \in L^2(\Omega)$*. Every weak accumulation point is also a strong accumulation point, i.e.,* $q_\gamma \to \bar{q}$ *strongly in* $L^2(\Omega)$*, and each of these accumulation points is a global minimizer of* (P)*.*

2. If $q_\gamma \to \bar{q}$ *in* $L^2(\Omega)$*, then the associated sequence of solutions to* (9) *fulfills*

$$u_\gamma \to \bar{u} \quad \text{in } H_0^1(\Omega), \tag{10}$$
$$\lambda_\gamma \to \bar{\lambda} \quad \text{in } H^{-1}(\Omega), \tag{11}$$
$$p_\gamma \rightharpoonup \bar{p} \quad \text{in } H_0^1(\Omega), \tag{12}$$
$$\theta_\gamma \rightharpoonup \bar{\theta} \quad \text{in } H_0^1(\Omega), \tag{13}$$
$$\mu_\gamma \rightharpoonup \bar{\mu} \quad \text{in } H^{-1}(\Omega), \tag{14}$$

and the limit satisfies the following optimality system:

$$A\bar{u} = \bar{q} + \bar{\lambda}, \tag{15a}$$

$$\bar{u} \geq \psi \text{ a.e. in } \Omega, \quad \bar{\lambda} \geq 0 \text{ in } H^{-1}(\Omega), \quad \langle \bar{\lambda}, \bar{u} - \psi \rangle = 0, \tag{15b}$$

$$A^* \bar{p} = \nabla j(\bar{u}) - \bar{\mu}, \tag{15c}$$

$$\bar{p} + \nabla g(\bar{q}) = 0, \tag{15d}$$

$$\bar{p} - \bar{\theta} = 0, \tag{15e}$$

$$\langle \bar{\theta}, \bar{\lambda} \rangle = 0, \quad \langle \bar{\mu}, \psi - \bar{u} \rangle = 0, \quad \langle \bar{\theta}, \bar{\mu} \rangle \geq 0. \tag{15f}$$

Proof. The proof follows by standard arguments known from other types of regularization, cf. e.g. [27]. For our particular regularization, the verification of the complementarity relations in (15f) becomes astonishingly easy, so we present the proof in detail for convenience of the reader.

Convergence of the primal variables:
Owing to (4), (5), and the Lipschitz continuity of S_γ from $H^{-1}(\Omega)$ to $H_0^1(\Omega)$, one deduces the existence of at least one global minimum of (P_γ) for every $\gamma > 0$. Moreover, due to their optimality and (5), every sequence of global minimizers $\{q_\gamma\}$ is bounded in $L^2(\Omega)$. Hence there exists a weakly converging subsequence, also denoted by $\{q_\gamma\}$. From the Lipschitz continuity of S_γ, independently of γ, and Lemma 2, we infer

$$\begin{aligned} \|S_\gamma(q_\gamma) - S(\bar{q})\|_{H^1(\Omega)} &\leq \|S_\gamma(q_\gamma) - S_\gamma(\bar{q})\|_{H^1(\Omega)} + \|S_\gamma(\bar{q}) - S(\bar{q})\|_{H^1(\Omega)} \\ &\leq L \|q_\gamma - \bar{q}\|_{H^{-1}(\Omega)} + c \frac{1}{\sqrt{\gamma}} \|\bar{q} - A\psi\|^{2/3}_{L^{4/3}(\Omega)}. \end{aligned}$$

Thus the compact embedding $L^2(\Omega) \hookrightarrow H^{-1}(\Omega)$ yields strong convergence of the state, i.e., (10). Thanks to (5), g is continuous and convex, hence weakly lower semicontinuous. This and the strong convergence of the states yield

$$\begin{aligned} J(S(\bar{q}), \bar{q}) \leq \liminf_{\gamma\to\infty} J(S_\gamma(q_\gamma), q_\gamma) &\leq \limsup_{\gamma\to\infty} J(S_\gamma(q_\gamma), q_\gamma) \\ &\leq \lim_{\gamma\to\infty} J(S_\gamma(q), q) = J(S(q), q) \quad \forall\, q \in L^2(\Omega), \end{aligned}$$

which is just global optimality of $\bar{q}$ for (P). Furthermore, inserting $q = \bar{q}$ in the above inequality implies convergence of the objective, which, together with the strong convergence of the states, gives in turn

$$g(q_\gamma) - g(\bar{q}) = J(S(q_\gamma), q_\gamma) - J(S(\bar{q}), \bar{q}) - \big(j(S(q_\gamma)) - j(S(\bar{q}))\big) \to 0 \tag{16}$$

as $\gamma \to \infty$. Since g is twice continuously differentiable, there is a $t \in [0, 1]$ so that

$$\begin{aligned} g(q_\gamma) - g(\bar{q}) &= g'(\bar{q})(q_\gamma - \bar{q}) + g''\big(\bar{q} + t(q_\gamma - \bar{q})\big)(q_\gamma - \bar{q})^2 \\ &\geq g'(\bar{q})(q_\gamma - \bar{q}) + \alpha \|q_\gamma - \bar{q}\|^2_{L^2(\Omega)}, \end{aligned}$$

where we used (5) for the last estimate. Thanks to $g'(\bar{q}) \in L^2(\Omega)^*$, the weak convergence $q_\gamma \rightharpoonup \bar{q}$ in $L^2(\Omega)$ and (16) imply strong convergence of q_γ to $\bar{q}$. Since $\bar{q}$ was an arbitrary weak accumulation point, this gives the first claim.

Convergence of the dual variables:
The convergence of the slack variables is an easy consequence of the continuity of $A : H_0^1(\Omega) \to H^{-1}(\Omega)$ and the compact embedding $L^2(\Omega) \hookrightarrow H^{-1}(\Omega)$:

$$\lambda_\gamma = -r(\gamma; u_\gamma) = Au_\gamma - q_\gamma \to A\bar{u} - \bar{q} = \bar{\lambda} \quad \text{in } H^{-1}(\Omega).$$

As $\bar{u} = S(\bar{q})$ is the solution of (6), $\bar{\lambda}$ is the associated slack variable fulfilling complementarity system in (15b).

To prove the weak convergence of the adjoint state, insert θ_γ and μ_γ in (9c) to obtain

$$A^* p_\gamma + \partial_u r(\gamma; u_\gamma) p_\gamma = \nabla j(u_\gamma)$$

with

$$\partial_u r(\gamma; u_\gamma) = 3\,\gamma \left[\max\big(\gamma(\psi - u_\gamma), 0\big)\right]^2. \tag{17}$$

Testing this equation with p_γ itself yields

$$\|p_\gamma\|_{H^1(\Omega)} \le \frac{1}{\beta} \|\nabla j(u_\gamma)\|_{H^{-1}(\Omega)} \tag{18}$$

$$\text{and} \quad \int_\Omega \left[\max\big(\gamma(\psi - u_\gamma), 0\big) p_\gamma\right]^2 dx \le \frac{1}{3\gamma} \langle \nabla j(u_\gamma), p_\gamma \rangle \to 0 \quad \text{as } \gamma \to \infty, \tag{19}$$

where we used (18) and the boundedness of $\{u_\gamma\}$ in $H_0^1(\Omega)$ for the passage to the limit. From (18) we infer the existence of a subsequence, weakly converging in $H_0^1(\Omega)$ to $\bar{p}$. For simplicity we denote this subsequence by p_γ, too. Moreover, the convergence of μ_γ and (15c) follow from

$$\mu_\gamma = \nabla j(u_\gamma) - A^* p_\gamma \rightharpoonup \nabla j(\bar{u}) - A^* \bar{p} =: \bar{\mu} \quad \text{in } H^{-1}(\Omega).$$

As g is assumed to be continuously differentiable, we can pass to the limit in (9d) to obtain (15d). The weak limit is therefore unique, namely $-\nabla g(\bar{q})$, and consequently the whole sequence $\{p_\gamma\}$ converges weakly to $\bar{p}$, i.e., (12) is shown. Hence, the whole sequence μ_γ converges weakly, too, which shows (14).

It remains to verify the complementarity relations in (15f). Due to $\lambda_\gamma = \partial_u r(\gamma; u_\gamma)$ and the construction of r in (7), we find

$$\begin{aligned} |\langle \lambda_\gamma, p_\gamma \rangle| &= \Big| \int_\Omega \left[\max\big(\gamma(\psi - u_\gamma), 0\big)\right]^3 p_\gamma \, dx \Big| \\ &\le \| \max\big(\gamma(\psi - u_\gamma), 0\big) \|^2_{L^4(\Omega)} \| \max\big(\gamma(\psi - u_\gamma), 0\big) p_\gamma \|_{L^2(\Omega)}. \end{aligned} \tag{20}$$

To estimate the L^4-norm, test (9a) with $\max(\gamma(\psi - u_\gamma), 0) \in H_0^1(\Omega)$ such that, similarly to the proof of Lemma 2 in Appendix A,

$$\begin{aligned} &\| \max(\gamma(\psi - u_\gamma), 0) \|^4_{L^4(\Omega)} \\ &\quad = a\big(u_\gamma - \psi, \max(\gamma(\psi - u_\gamma), 0)\big) - \langle q_\gamma - A\psi, \max(\gamma(\psi - u_\gamma), 0) \rangle \\ &\quad = -\gamma\, a(u_\gamma - \psi, u_\gamma - \psi) - \langle q_\gamma - A\psi, \max(\gamma(\psi - u_\gamma), 0) \rangle \\ &\quad \le \|q - A\psi\|_{L^{4/3}(\Omega)} \| \max(\gamma(\psi - u_\gamma), 0) \|_{L^4(\Omega)}. \end{aligned}$$

is obtained. Thus $\{\max(\gamma(\psi - u_\gamma), 0)\}_{\gamma>0}$ is bounded in $L^4(\Omega)$ and, in view of (20) and (19), the strong convergence of λ_γ in $H^{-1}(\Omega)$ and the weak convergence of p_γ in $H_0^1(\Omega)$ yield

$$\langle \bar{\lambda}, \bar{p} \rangle = \lim_{\gamma \to \infty} \langle \lambda_\gamma, p_\gamma \rangle = 0,$$

i.e., the first equation in (15f). To derive the second equation, observe that the definition of μ_γ in (9f) implies

$$\begin{aligned} \langle \mu_\gamma, \psi - u_\gamma \rangle &= 3 \int_\Omega \left[\max\left(\gamma(\psi - u_\gamma), 0\right) \right]^2 p_\gamma \, \gamma(\psi - u_\gamma) \, dx \\ &= 3 \int_\Omega \left[\max\left(\gamma(\psi - u_\gamma), 0\right) \right]^3 p_\gamma \, dx = 3 \, \langle \lambda_\gamma, p_\gamma \rangle \to 0. \end{aligned}$$

Since $\mu_\gamma \rightharpoonup \bar{\mu}$ in $H^{-1}(\Omega)$ and $u_\gamma \to \bar{u}$ in $H_0^1(\Omega)$, this gives the claim. In order to prove the sign condition in (15f), we test (9c) and (15c) each with $p_\gamma - \bar{p}$ and subtract the arising equations to obtain

$$\begin{aligned} \langle \mu_\gamma - \bar{\mu}, p_\gamma - \bar{p} \rangle &= \big(j'(u_\gamma) - j'(\bar{u}) \big)(p_\gamma - \bar{p}) - a(p_\gamma - \bar{p}, p_\gamma - \bar{p}) \\ &\leq \big(j'(u_\gamma) - j'(\bar{u}) \big)(p_\gamma - \bar{p}). \end{aligned}$$

Employing again the definition of μ_γ in (9f), we find

$$\langle \mu_\gamma, p_\gamma \rangle = 3 \int_\Omega \gamma \left[\max\left(\gamma(\psi - u_\gamma), 0\right) \right]^2 p_\gamma^2 \, dx \geq 0 \quad \forall \, \gamma > 0.$$

Thus we arrive at

$$\begin{aligned} \langle \bar{\mu}, \bar{p} \rangle &= \langle \mu_\gamma - \bar{\mu}, p_\gamma - \bar{p} \rangle - \langle \mu_\gamma, p_\gamma \rangle + \langle \mu_\gamma, \bar{p} \rangle + \langle \bar{\mu}, p_\gamma \rangle \\ &\leq \big(j'(u_\gamma) - j'(\bar{u}) \big)(p_\gamma - \bar{p}) + \langle \mu_\gamma, \bar{p} \rangle + \langle \bar{\mu}, p_\gamma \rangle. \end{aligned}$$

Because of $u_\gamma \to \bar{u}$ in $H_0^1(\Omega)$, $p_\gamma \rightharpoonup \bar{p}$ in $H_0^1(\Omega)$, and $\mu_\gamma \rightharpoonup \bar{\mu}$ in $H^{-1}(\Omega)$, the right hand side converges to $2\langle \bar{\mu}, \bar{p} \rangle$, which gives the desired sign condition. Introducing $\bar{\theta}$ by (15e) finally completes the proof. □

Remark 1. *Using a classical localization argument, see e.g. [10], one can show that every strict local minimizer of* (P) *can be approximated by local minimizers of* (P_γ). *Thus every strict local minimizer satisfies* (15). *Furthermore, by using the following obvious modification of the objective in* (P_γ)

$$\tilde{J}(u, q) = J(u, q) + \frac{1}{2} \, \|q - \bar{q}\|^2_{L^2(\Omega)},$$

one can show that even every *local optimum of* (P) *satisfies the optimality system* (15), *cf. [20]. Of course the associated regularized problems are only of academic interest, and cannot be used numerically, since they involve the unknown solution* $\bar{q}$.

We underline that the complementarity relations as well as the sign condition in (15f) *can be sharpened as shown in [27, Def. 1.1, Thm. 3.9, and 3.10].*

In view of (15e) the multiplier $\bar{\theta}$ can directly be eliminated from the system (15). We introduced this additional variable for reasons of comparison with C-stationarity conditions for finite dimensional MPECs, cf., e.g., [26]. This comparison shows that $\bar{\theta}$ is the MPEC-multiplier for the constraint $\bar{\lambda} \geq 0$, while $\bar{\mu}$ is the MPEC-multiplier associated with $\bar{u} \geq \psi$, cf., also the complementarity relations in (15f). A more rigorous first-order system is given by the strong stationarity conditions. These involve in addition to (15) the following sign conditions

$$\bar{\theta}(x) \geq 0 \quad \text{q.e., where } \bar{u}(x) = \psi(x) \tag{21a}$$

$$\langle \bar{\mu}, v \rangle \geq 0 \quad \forall\, v \in H_0^1(\Omega) :\ \langle \bar{\lambda}, v \rangle = 0,\ v(x) \geq 0 \text{ q.e., where } \bar{u}(x) = \psi(x). \tag{21b}$$

Note that the complementarity relations in (15f) together with (21a) and (21b) implies $\langle \bar{\theta}, \bar{\mu} \rangle \geq 0$. Indeed strong stationarity is the most rigorous stationarity concept. In case of (P) it can be verified to be necessary for local optimality as proven in [20], provided that $\bar{u} \in H_0^1(\Omega)$, which follows from (15d), if $g = \alpha/2 \|.\|^2_{L^2(\Omega)}$. However, if, for instance, additional control constraints are present or boundary control is considered, this is not true in general. For further details on this topic we refer to [32]. We point out that our construction of a posteriori error estimators for the regularization error is only based on the C-stationarity conditions in (15), which can be verified to be necessary for local optimality in very general situations allowing for instance also pointwise constraints on the control or boundary control problems.

To complete this introduction, we finally comment on a priori estimates concerning the regularization error that have been proven in the literature. We only refer to [27], where the error estimates are shown for the regularization of (P) based on the locally smoothed version of the max-function. To be more precise, if $\bar{u}$ is a strict local optimum of (P) so that it can be approximated by a sequence of local optima of (P_γ), then for all $\gamma > 0$ sufficiently large this sequence satisfies

$$|J(\bar{u}, \bar{q}) - J(u_\gamma, q_\gamma)| \lesssim \frac{1}{\gamma} \quad \text{and} \quad \|\bar{q} - q_\gamma\|_{L^2(\Omega)} \lesssim \frac{1}{\sqrt{\gamma}},$$

provided that (9) fulfills a certain regularity condition, see [27, 17] for details.

4. A posteriori estimation of the regularization error

In this section, we will derive an identity for the regularization error w.r.t. the objective, i.e., $J(\bar{u}, \bar{q}) - J(u_\gamma, q_\gamma)$. The idea is based on the DWR-method. We begin by defining the following MPEC-Lagrangian:

$$\begin{aligned} &\mathscr{L} : L^2(\Omega) \times H_0^1(\Omega) \times H^{-1}(\Omega) \times H_0^1(\Omega) \times H^{-1}(\Omega) \times H_0^1(\Omega) \to \mathbb{R}; \\ &\mathscr{L}(q, u, \lambda, p, \mu, \theta) := J(u, q) - \langle Au - q - \lambda, p \rangle + \langle \mu, \psi - u \rangle - \langle \theta, \lambda \rangle. \end{aligned} \tag{22}$$

We will sometimes consider $\mathscr{L}$ with different domain denoted by the same symbol for simplicity. Note that we do not introduce a multiplier associated with the complementarity relation $\langle \lambda, u - \psi \rangle = 0$, which is typical for MPECs. In the following, we

abbreviate $\xi := (q, u, \lambda, p, \mu, \theta)$. Due to (15a) and (15f), we have $J(\bar{u}, \bar{q}) = \mathscr{L}(\bar{\xi})$ such that (9a) yields

$$\begin{aligned} J(\bar{u}, \bar{q}) - J(u_\gamma, q_\gamma) &= \mathscr{L}(\bar{\xi}) - \mathscr{L}(\xi_\gamma) + \langle \mu_\gamma, \psi - u_\gamma \rangle - \langle \theta_\gamma, \lambda_\gamma \rangle \\ &= \frac{1}{2}\, \mathscr{L}'(\bar{\xi})(\bar{\xi} - \bar{\xi}_\gamma) + \frac{1}{2}\, \mathscr{L}'(\xi_\gamma)(\bar{\xi} - \bar{\xi}_\gamma) + R_{\mathrm{pen}} \\ &\quad + \langle \mu_\gamma, \psi - u_\gamma \rangle - \langle \theta_\gamma, \lambda_\gamma \rangle \end{aligned}$$

with

$$\begin{aligned} R_{\mathrm{reg}} &= \frac{1}{2} \int_0^1 \mathscr{L}'''(\xi_\gamma + t(\bar{\xi} - \xi_\gamma))(\bar{\xi} - \xi_\gamma)^3\, t(t-1)\, dt \\ &= \frac{1}{2} \int_0^1 \Big(j'''(u_\gamma + t(\bar{u} - u_\gamma))(\bar{u} - u_\gamma)^3 + g'''(q_\gamma + t(\bar{q} - q_\gamma))(\bar{q} - q_\gamma)^3 \Big) t(t-1)\, dt. \end{aligned}$$

Note that, due to our assumptions on J, the Lagrangian $\mathscr{L}$ is three times continuously Fréchet-differentiable, since every bounded bilinear form does so. If the trilinear forms $j'''(u)$ and $g'''(q)$ are uniformly bounded, say by constants c_j and c_g, then

$$|R_{\mathrm{reg}}| \leq \frac{1}{12} \Big(c_j \, \|u_\gamma - \bar{u}\|_{H^1(\Omega)}^3 + c_g \, \|q_\gamma - \bar{q}\|_{L^2(\Omega)}^3 \Big)$$

such that R_{reg} can be neglected in a neighborhood of $(\bar{u}, \bar{q})$. We point out that, due to the non-convexity of (P), multiple local minima can occur such that one can in general not expect that $q_\gamma \to \bar{q}$ and $u_\gamma \to \bar{u}$, cf. Theorem 1. However, if J is of tracking type, i.e., if j and g are squared norms, then $R_{\mathrm{reg}} = 0$.

In view of (22), (9), and (15), we arrive at

$$\begin{aligned} J(\bar{u}, \bar{q}) - J(u_\gamma, q_\gamma) &= \frac{1}{2} \langle \psi - \bar{u}, \bar{\mu} - \mu_\gamma \rangle - \frac{1}{2} \langle \bar{\lambda}, \bar{\theta} - \theta_\gamma \rangle \\ &\quad + \frac{1}{2} \langle \psi - u_\gamma, \bar{\mu} + \mu_\gamma \rangle - \frac{1}{2} \langle \lambda_\gamma, \bar{\theta} + \theta_\gamma \rangle + R_{\mathrm{reg}} \end{aligned}$$

Note that $\langle \mu_\gamma, \psi - u_\gamma \rangle$ and $\langle \theta_\gamma, \lambda_\gamma \rangle$ can be interpreted as complementarity errors, cf. (15f). Note further that we have not used any of the complementarity conditions in (15f) so far. Now using these complementarity conditions we can switch the sign of the multipliers $\bar{\mu}$ and $\bar{\theta}$ in the first two terms on the right to get

$$J(\bar{u}, \bar{q}) - J(u_\gamma, q_\gamma) = \left\langle \bar{u} - u_\gamma, \frac{1}{2} (\bar{\mu} + \mu_\gamma) \right\rangle + \left\langle \bar{\lambda} - \lambda_\gamma, \frac{1}{2} (\bar{\theta} + \theta_\gamma) \right\rangle + R_{\mathrm{reg}}. \tag{23}$$

In view of the convergence results in Theorem 1, this seems to be a reasonable splitting, since the differences on the left of the duality pairings converge strongly in $H_0^1(\Omega)$ and the sum on the right is bounded in $H^{-1}(\Omega)$.

In order to efficiently estimate the differences in the duality pairings in (23), we aim to exploit a Taylor expansion w.r.t. the regularization parameter γ. To this end, we show that, under suitable assumptions, the mapping $\gamma \mapsto (u_\gamma, \lambda_\gamma)$ is at least locally

differentiable. Comparable path differentiability results have already been proven in [17] for a different type of regularization. In [17], the authors exploit the equivalence of the linearized optimality system to a (under suitable assumptions) convex optimization problem. Here, we proceed along a different path by directly proving the solvability of the linearized system, leading to comparable results.

We will prove the differentiability of $\gamma \mapsto (u_\gamma, \lambda_\gamma)$ by means of the implicit function theorem. For this purpose, observe first that λ_γ, μ_γ, and θ_γ are given by simple algebraic relations and can thus be eliminated directly from optimality conditions (9). Moreover, because of (5), $\nabla g : L^2(\Omega) \to L^2(\Omega)$ is a strongly monotone and continuous operator. Thus, (9d) can be resolved for q_γ, i.e., there is a mapping $Q : L^2(\Omega) \to L^2(\Omega)$ such that

$$p_\gamma + \nabla g(q_\gamma) = 0 \quad \iff \quad q_\gamma = Q(p_\gamma).$$

This leaves us with the semilinear elliptic system for finding the state u_γ and adjoint p_γ solving

$$\begin{aligned} Au_\gamma - Q(p_\gamma) + r(\gamma; u_\gamma) &= 0, \\ A^* p_\gamma - \nabla j(u_\gamma) + \partial_u r(\gamma; u_\gamma) p_\gamma &= 0. \end{aligned} \tag{24}$$

Since the bilinear form $L^2(\Omega)^2 \ni (v, w) \mapsto g''(q_\gamma)[v, w] \in \mathbb{R}$ is bounded and coercive by (5), the operator $g''(q_\gamma) : L^2(\Omega) \to L^2(\Omega)^* = L^2(\Omega)$ is a homeomorphism so that the implicit function theorem yields that Q is continuously Fréchet differentiable with

$$Q'(p_\gamma)\dot{p} = -g''(Q(p_\gamma))^{-1}\dot{p}. \tag{25}$$

To apply the implicit function theorem to (24), we need to show solvability of the linearized system associated to (24), which is given by

$$A\dot{u} - Q'(p_\gamma)\dot{p} + \partial_u r(\gamma; u_\gamma)\dot{u} = z_1, \tag{26a}$$

$$A^*\dot{p} - j''(u_\gamma)\dot{u} + \partial_u^2 r(\gamma; u_\gamma) p_\gamma \dot{u} + \partial_u r(\gamma; u_\gamma)\dot{p} = z_2, \tag{26b}$$

with arbitrary right hand sides $z_1, z_2 \in H^{-1}(\Omega)$. To this end, we require the following

Assumption 1. *1. The first part j of the objective, acting on the state, is supposed to be convex.*

2. We assume that u_γ and p_γ are such that $\partial_u^2 r(\gamma; u_\gamma) p_\gamma \leq 0$ a.e. in Ω.

Lemma 3. *Given Assumption 1 the linearized system* (26) *admits a unique solution $\dot{u}$ and $\dot{p}$ for every right hand side $z_1, z_2 \in H^{-1}(\Omega)^2$.*

Proof. We start by reducing (26) to a single in equation in $\dot{p}$. For this purpose, observe that $\partial_u r(\gamma; u_\gamma) \geq 0$, see (17). Moreover, by Lemma 1, we have $\partial_u r(\gamma; u_\gamma) \in L^\infty(\Omega)$. Thus the bilinear form induced by $A + \partial_u r(\gamma; u_\gamma)$ is bounded and coercive in $H_0^1(\Omega)$ giving the existence of an operator $S'_\gamma : H^{-1}(\Omega) \to H_0^1(\Omega)$ such that

$$w = S'_\gamma f \quad \iff \quad Aw + \partial_u r(\gamma; u_\gamma) w = f.$$

Then (26a) is equivalent to

$$\dot{u} = S'_\gamma Q'(p_\gamma)\dot{p} + S'_\gamma z_1 \tag{27}$$

so that (26b) becomes a single equation in $\dot{p}$, namely

$$A^*\dot{p} + \partial_u r(\gamma; u_\gamma)\dot{p} = \big(j''(u_\gamma) - \partial_u^2 r(\gamma; u_\gamma)p_\gamma\big)S'_\gamma Q'(p_\gamma)\dot{p} + \tilde{z}$$

with

$$\tilde{z} = \big(j''(u_\gamma) - \partial_u^2 r(\gamma; u_\gamma)p_\gamma\big)S'_\gamma z_1 + z_2 \in H^{-1}(\Omega).$$

Utilizing that $(A^* + \partial_u r(\gamma; u_\gamma))^{-1} = (S'_\gamma)^* =: S^*_\gamma$, this equation is equivalent to

$$\dot{p} = K\dot{p} + S^*_\gamma \tilde{z}. \tag{28}$$

where

$$K = S^*_\gamma\big(j''(u_\gamma) - \partial_u^2 r(\gamma; u_\gamma)p_\gamma\big)S'_\gamma Q'(p_\gamma).$$

Since $S^*_\gamma : H^{-1}(\Omega) \to H^1_0(\Omega)$, there holds $K : L^2(\Omega) \to H^1_0(\Omega) \subset\subset L^2(\Omega)$ so that K regarded as an operator with range in $L^2(\Omega)$ is compact by Rellich's theorem. Thus we can apply Fredholm's alternative, i.e., either $(I-K)\dot{p} = z$ admits a unique solution $\dot{p} \in L^2(\Omega)$ for every $z \in L^2(\Omega)$ or the homogeneous equation

$$(I - K)\dot{p} = 0 \tag{29}$$

has non-trivial solutions in $L^2(\Omega)$.

By construction of K and S'_γ, (29) is equivalent to the homogeneous counterpart of (26), where $z_1 = z_2 = 0$. To see that this homogeneous system only admits the trivial solution, we test the first equation in (26) with $\dot{p}$ and the second one with $\dot{u}$ and subtract the arising equalities to obtain

$$-(Q'(p_\gamma)\dot{p}, \dot{p}) + j''(u_\gamma)\dot{u}^2 - \int_\Omega \partial_u^2 r(\gamma; u_\gamma)p_\gamma \dot{u}^2\, dx = 0.$$

Testing (25) with $Q'(p_\gamma)\dot{p}$ results in

$$g''\big(Q(p_\gamma)\big)[Q'(p_\gamma)\dot{p}]^2 = -(Q'(p_\gamma)\dot{p}, \dot{p}).$$

Thus (5) and Assumption 1, i.e., the convexity of j and the sign condition, give

$$\alpha \, \|Q'(p_\gamma)\dot{p}\|^2_{L^2(\Omega)} \le \int_\Omega \partial_u^2 r(\gamma; u_\gamma)p_\gamma \dot{u}^2\, dx \le 0.$$

Consequently $Q'(p_\gamma)\dot{p} = 0$ and thus $\dot{p} = -g''\big(Q(p_\gamma)\big)Q'(p_\gamma)\dot{p} = 0$. Therefore (29) indeed just admits the trivial solution, and thus Fredholm's alternative implies that (28) and hence also (26) are uniquely solvable. □

Remark 2. *Some comments concerning Assumption 1 are in order. Note first that our analysis allows, for instance, for objectives of the form* $j(u) = \|\nabla u - z\|^2_{L^2(\Omega;\mathbb{R}^d)}$ *with given* $z \in L^2(\Omega;\mathbb{R}^d)$ *or* $j(u) = \|u - w\|^2_{L^2(\tilde{\Gamma})}$*, where* $\tilde{\Gamma} \subset \Omega$ *is a smooth manifold. The sign condition in Assumption 1 is an analogon to [17, Assumption 2]. In view of*

$$\partial_u^2 r(\gamma; u_\gamma) = -6\,\gamma^2 \max\big(\gamma(\psi - u_\gamma), 0\big),$$

it is fulfilled, if

$$p_\gamma \geq 0 \quad \text{a.e. on} \quad \{u_\gamma < \psi\}.$$

Note that the strong stationarity conditions imply $\bar{p} \geq 0$ *q.e., where* $\bar{u} = \psi$. *As mentioned above, these conditions hold in case of* (P), *if* $\bar{u} \in H_0^1(\Omega)$, *cf., [20].*

Corollary 1. *Let* (q_γ, u_γ) *with corresponding multipliers* $(p_\gamma, \lambda_\gamma, \theta_\gamma)$ *satisfy the optimality system* (9). *Assume that Assumption 1 is satisfied, then* $(q_\gamma, u_\gamma, p_\gamma, \lambda_\gamma, \theta_\gamma)$ *is locally unique. Furthermore, the solution variables are differentiable with respect to* γ *and the derivatives* $(\dot{q}_\gamma, \dot{u}_\gamma, \dot{p}_\gamma, \dot{\lambda}_\gamma, \dot{\theta}_\gamma)$ *solve the following system of linearized equations:*

$$A\dot{u}_\gamma = \dot{q}_\gamma + \dot{\lambda}_\gamma, \tag{30a}$$

$$\dot{\lambda}_\gamma + \partial_\gamma r(\gamma, u_\gamma) + \partial_u r(\gamma; \gamma)\dot{u}_\gamma = 0, \tag{30b}$$

$$A^*\dot{p}_\gamma = j''(u_\gamma)\dot{u}_\gamma - \dot{\mu}_\gamma, \tag{30c}$$

$$\dot{p}_\gamma + g''(q_\gamma)\dot{q}_\gamma = 0, \tag{30d}$$

$$\dot{p}_\gamma - \dot{\theta}_\gamma = 0, \tag{30e}$$

$$\dot{\mu}_\gamma - \partial_u r(\gamma; u_\gamma)\dot{\theta}_\gamma - \partial^2_{u\gamma} r(\gamma; u_\gamma)\theta_\gamma - p_\gamma \partial^2_u r(\gamma; u_\gamma)\dot{u}_\gamma = 0. \tag{30f}$$

Proof. As demonstrated above, the optimality system (9) is equivalent to (24). Local uniqueness and differentiability for (24) follow by the implicit function theorem from invertibility of the linearized operator in (26) as shown in Lemma 3.

The derivative of $\mathbb{R} \ni \gamma \mapsto (u_\gamma, p_\gamma) \in H_0^1(\Omega)^2$ is given by the unique solution $(\dot{u}_\gamma, \dot{p}_\gamma)$ of the linearized version of (24):

$$\begin{aligned} A\dot{u}_\gamma - Q'(p_\gamma)\dot{p}_\gamma + \partial_u r(\gamma; u_\gamma)\dot{u}_\gamma + \partial_\gamma r(\gamma; u_\gamma) = 0, \\ A^*\dot{p}_\gamma - j''(u_\gamma)\dot{u}_\gamma + \partial^2_u r(\gamma; u_\gamma)p_\gamma \dot{u}_\gamma + \partial_u r(\gamma; u_\gamma)\dot{p}_\gamma + \partial^2_{u\gamma} r(\gamma; u_\gamma) = 0. \end{aligned}$$

By setting $\dot{q}_\gamma = Q'(p_\gamma)\dot{p}_\gamma = -g(q_\gamma)^{-1}\dot{p}_\gamma$, see (25), and introducing $\dot{\lambda}_\gamma$, $\dot{\theta}_\gamma$, and $\dot{\mu}_\gamma$ through (30b), (30e), and (30f), respectively, we obtain (30). □

To obtain an estimator for the regularization error, which can be evaluated a posteriori, we make the following ansatz for the unknown $\bar{u}$, $\bar{\lambda}$ based on a first-order Taylor expansion:

$$\bar{u} \approx u_\infty := u_\gamma + \gamma\, \dot{u}_\gamma, \tag{31a}$$

$$\bar{\lambda} \approx \lambda_\infty := \lambda_\gamma + \gamma\, \dot{\lambda}_\gamma. \tag{31b}$$

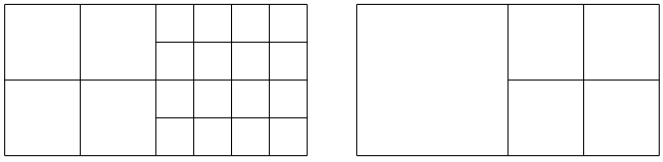
Then, by (23), we arrive at the following approximation of the difference in the objective

$$\begin{aligned} J(\bar{u}, \bar{q}) - J(u_\gamma, q_\gamma) &= \left\langle \bar{u} - u_\gamma, \frac{1}{2}(\bar{\mu} + \mu_\gamma) \right\rangle + \left\langle \bar{\lambda} - \lambda_\gamma, \frac{1}{2}(\bar{\theta} + \theta_\gamma) \right\rangle + R_{\text{reg}} \\ &\approx \gamma\big(\langle \dot{u}_\gamma, \mu_\gamma \rangle + \langle \dot{\lambda}_\gamma, \theta_\gamma \rangle\big) =: \eta_t. \end{aligned} \tag{32}$$

It is to be noted that the above approximation is by far not rigorous. First of all Assumption 1(2) is fairly ad hoc. Moreover, the minimum of (P) need not be unique so that convergence $u_\gamma \to \bar{u}$ and $\lambda_\gamma \to \bar{\lambda}$ cannot be expected. Therefore the Taylor approximations in (31) might be quite inaccurate. In addition the replacement of $1/2(\bar{\mu} + \mu_\gamma)$ and $1/2(\theta + \theta_\gamma)$ by μ_γ and θ_γ, respectively, is even more critical, since the dual variables only converge weakly, if they converge at all. Despite all these issues, the proposed error estimator performs very well in the numerical tests, as we will see in Section 6.

5. Adaptive finite element discretization

In this section, we describe the adaptive finite element discretization of the reduced regularized optimality system (24). Let $\mathcal{T}_h$ be a triangulation of Ω consisting of quadrilateral elements for $d = 2$ and of hexahedral ones for $d = 3$. In our approach, we allow hanging nodes of degree one to realize adaptive mesh refinement. For the evaluation of the later discussed a posteriori error estimate, a special structure of the adaptively refined finite element mesh is required. This so-called patch-structure is obtained through the refinement of all sons of a refined element, provided that one of these sons is actually marked for refinement. It is illustrated in Figure 1. The finite element



(a) Mesh with patch structure (b) Corresponding patch mesh

Figure 1: Illustration of the patch structure of the finite element mesh

ansatz space is given by

$$V_h := \left\{ v \in H_0^1(\Omega) : v|_T \in Q_1(T) \, \forall \, T \in \mathcal{T}_h \right\},$$

where $Q_1(T)$ consists of d-linear basis functions on the element T. This leads us to the discrete problem: Find the discrete state $u_{\gamma,h}$ and the discrete adjoint $p_{\gamma,h}$ as solution of the semilinear elliptic system

$$\begin{aligned} a(u_{\gamma,h}, \varphi_h) - Q(p_{\gamma,h})(\varphi_h) + (r(\gamma; u_{\gamma,h}), \varphi_h) = 0 \qquad & \forall \varphi_h \in V_h \\ a(\chi_h, p_{\gamma,h}) - \nabla j(u_{\gamma,h})(\chi_h) + (\partial_u r(\gamma; u_{\gamma,h}) p_{\gamma,h}, \chi_h) = 0 \qquad & \forall \chi_h \in V_h. \end{aligned} \tag{33}$$

Furthermore, $q_{\gamma,h}$ is given by $q_{\gamma,h} = Q(p_{\gamma,h})$. We use Newton's method to determine the discrete solution. However, this introduces an additional error such that we only compute approximate solutions $(\tilde{u}_{\gamma,h}, \tilde{q}_{\gamma,h}, \tilde{p}_{\gamma,h})$.

The detailed solution algorithm, Algorithm 1, will be outlined in at the end of this section. Beforehand, we discuss an a posteriori error estimate of the discretization error and the numerical error:

Proposition 2. *For the error in the cost functional J the following error representation holds:*

$$\begin{aligned} J(u_\gamma, q_\gamma) - & J(\tilde{u}_{\gamma,h}, \tilde{q}_{\gamma,h}) \\ &= \frac{1}{2}\rho^*(\tilde{u}_{\gamma,h}, \tilde{q}_{\gamma,h}, \tilde{p}_{\gamma,h})(u - \tilde{u}_{\gamma,h}) + \frac{1}{2}\rho^q(\tilde{q}_{\gamma,h}, \tilde{p}_{\gamma,h})(q - \tilde{q}_{\gamma,h}) \\ &\quad + \frac{1}{2}\rho(\tilde{u}_{\gamma,h}, \tilde{q}_{\gamma,h})(p - \tilde{p}_{\gamma,h}) - \rho(\tilde{u}_{\gamma,h}, \tilde{q}_{\gamma,h})(\tilde{p}_{\gamma,h}) + \mathcal{R}_h^{(3)}. \end{aligned} \tag{34}$$

Here, the dual residual is given by

$$\rho^*(\tilde{u}_{\gamma,h}, \tilde{q}_{\gamma,h}, \tilde{p}_{\gamma,h})(\cdot) := \nabla j(u_{\gamma,h})(\cdot) - a(\cdot, p_{\gamma,h}) - (\partial_u r(\gamma; u_{\gamma,h}) p_{\gamma,h}, \cdot),$$

the control residual by

$$\rho^q(\tilde{q}_{\gamma,h}, \tilde{p}_{\gamma,h})(\cdot) := (\tilde{q}_{\gamma,h}, \cdot) - Q(\tilde{p}_{\gamma,h})(\cdot),$$

and the primal residual by

$$\rho(\tilde{u}_{\gamma,h}, \tilde{q}_{\gamma,h})(\cdot) := -a(u_{\gamma,h}, \cdot) + (q_{\gamma,h}, \cdot) - (r(\gamma; u_{\gamma,h}), \cdot).$$

The remainder term $\mathcal{R}_h^{(3)}$ arising from the application of the trapezoidal rule is of the form

$$\begin{aligned} \mathcal{R}_h^{(3)} =& \frac{1}{2}\int_0^1 \Big(j'''(\tilde{u}_{\gamma,h} + te^u)(e^u)^3 + g'''(\tilde{q}_{\gamma,h} + te^q)(e^q)^3 \Big) t(t-1)\, dt \\ &+ \frac{1}{2}\int_0^1 \Big(\frac{\partial^3}{\partial u^3} r(\gamma; \tilde{u}_{\gamma_h} + te^u)(e^u)^3, \tilde{p}_{\gamma,h} + te^p \Big) t(t-1)\, dt \\ &+ \frac{3}{2}\int_0^1 \Big(\frac{\partial^2}{\partial u^2} r(\gamma; \tilde{u}_{\gamma_h} + te^u)(e^u)^2, e^p \Big) t(t-1)\, dt. \end{aligned} \tag{35}$$

with the discretization errors $e^u := u_\gamma - \tilde{u}_{\gamma,h}$, $e^q := q_\gamma - \tilde{q}_{\gamma,h}$, and $e^p := p_\gamma - \tilde{p}_{\gamma,h}$.

Proof. The assertion, except the explicit form of the remainder term, follows directly from Proposition 4.1 in [23] using the standard arguments of the DWR approach based on the Lagrangian

$$\begin{aligned} &\mathcal{L}_\gamma : L^2(\Omega) \times H_0^1(\Omega) \times H_0^1(\Omega) \to \mathbb{R}; \\ &\mathcal{L}_\gamma(q_\gamma, u_\gamma, p_\gamma) := J(u_\gamma, q_\gamma) - a(u_\gamma, p_\gamma) + (q_\gamma - r(\gamma; u_\gamma), p_\gamma). \end{aligned} \tag{36}$$

Following Proposition 4.1 in [23], the remainder term is given by

$$\mathcal{R}_h^{(3)} = \frac{1}{2}\int_0^1 \mathcal{L}_\gamma'''(\chi_{\gamma,h} + t(\chi_\gamma - \chi_{\gamma,h}))(\chi_\gamma - \chi_{\gamma,h})^3 t(t-1)\, dt$$

with $\chi_\gamma = (q_\gamma, u_\gamma, p_\gamma)$. Now, we explicitly calculate the remainder term for our concrete case. Since the third derivative of $a(u_\gamma, p_\gamma)$ and (q_γ, p_γ) is vanishing, the only terms, which have to be considered, are $J(u_\gamma, q_\gamma)$ and $(r(\gamma; u_\gamma), p_\gamma)$. The first one has been discussed in the beginning of Section 4. By the same approach, we deduce

$$\begin{aligned}\frac{1}{2}\int_0^1 & J'''(\tilde{u}_{\gamma,h} + te^u, \tilde{q}_{\gamma,h} + te^q)(e^u, e^q)^3 t(t-1)\,dt \\ =&\frac{1}{2}\int_0^1 \Big(j'''(\tilde{u}_{\gamma,h} + te^u)(e^u)^3 + g'''(\tilde{q}_{\gamma,h} + te^q)(e^q)^3\Big)t(t-1)\,dt.\end{aligned}$$

For the second term $(r(\gamma; u_\gamma), p_\gamma)$, we deduce

$$\begin{aligned}\frac{1}{2}\int_0^1 & \Big(r(\gamma; \tilde{u}_{\gamma,h} + te^u), \tilde{p}_{\gamma,h} + te^p\Big)'''(e^u, e^p)^3 t(t-1)\,dt \\ =&\frac{1}{2}\int_0^1 \Big(\frac{\partial^3}{\partial u^3} r(\gamma; \tilde{u}_{\gamma_h} + te^u)(e^u)^3, \tilde{p}_{\gamma,h} + te^p\Big)t(t-1)\,dt \\ &+ \frac{3}{2}\int_0^1 \Big(\frac{\partial^2}{\partial u^2} r(\gamma; \tilde{u}_{\gamma_h} + te^u)(e^u)^2, e^p\Big)t(t-1)\,dt.\end{aligned} \tag{37}$$

Using the Lemma of Stampacchia, we get an discontinuous but integrable third derivative of r,

$$\frac{\partial^3}{\partial u^3} r(\gamma; u) = \begin{cases} -6\gamma^3, & u < \psi \\ 0, & \text{else,} \end{cases} \tag{38}$$

and therewith an bounded remainder term for fixed γ. □

Remark 3. *If $j'''(u)$, $g'''(q)$, $\frac{\partial^2}{\partial u^2} r(\gamma; u)$, and $\frac{\partial^3}{\partial u^3} r(\gamma; u)$ are bounded uniformly w.r.t. x, say by constants c_j, c_g, $c_{r,1}$, and $c_{r,2}$, then we obtain*

$$\left|\mathcal{R}_h^{(3)}\right| \leq \frac{1}{12}\Big((c_j + c_{r,1})\|e^u\|^3_{L^2(\Omega)} + c_g\|e^q\|^3_{L^2(\Omega)} + c_{r,2}\|e^u\|^2_{L^2(\Omega)}\|e^p\|_{L^2(\Omega)}\Big),$$

i.e., the remainder term is of third order in the discretization error. Unfortunately, the constants $c_{r,1}$ and $c_{r,2}$ are depending on the regularization parameter γ, and especially $c_{r,1}$ might tend to infinity, if $\gamma \to \infty$, see (38)*. However, in view of* (38)*, the first addend on the right hand side of* (37) *only has to be estimated on the inactive set $\{x \in \Omega : (\tilde{u}_{\gamma_h} + te^u)(x) < \psi(x)\}$. One can easily see that the measure of this set tends to zero, if e^u tends to zero in $L^1(\Omega)$. It is unclear whether the convergences are such that the growth in* (38) *is compensated in such a way, that*

$$|J(u_\gamma, q_\gamma) - J(\tilde{u}_{\gamma,h}, \tilde{q}_{\gamma,h})| \gg |\mathcal{R}_h^{(3)}|.$$

However, our numerical results indicate, that neglecting the error term is acceptable, at least in our test cases.

Since the error representation formula (34) is numerically not evaluable, we approximate it using patchwise quadratic interpolation, c.f., e.g., [2, Section 4.1] for this well

known procedure. Let $i_{2h}^{(2)}$ be the corresponding interpolation operator. We obtain neglecting the remainder term $\mathcal{R}_h^{(3)}$,

$$J(u_\gamma, q_\gamma) - J(\tilde{u}_{\gamma,h}, \tilde{q}_{\gamma,h}) \approx \eta_h + \eta_n \tag{39}$$

with the spatial and the numerical error estimator:

$$\begin{aligned}
\eta_h &:= \frac{1}{2}\rho^*(\tilde{u}_{\gamma,h}, \tilde{q}_{\gamma,h}, \tilde{p}_{\gamma,h})(i_{2h}^{(2)}\tilde{u}_{\gamma,h} - \tilde{u}_{\gamma,h}) + \frac{1}{2}\rho^q(\tilde{q}_{\gamma,h}, \tilde{p}_{\gamma,h})(i_{2h}^{(2)}\tilde{q}_{\gamma,h} - \tilde{q}_{\gamma,h}),\\
&\quad + \frac{1}{2}\rho(\tilde{u}_{\gamma,h}, \tilde{q}_{\gamma,h})(i_{2h}^{(2)}\tilde{p}_{\gamma,h} - \tilde{p}_{\gamma,h})\\
\eta_n &:= -\rho(\tilde{u}_{\gamma,h}, \tilde{q}_{\gamma,h})(\tilde{p}_{\gamma,h}).
\end{aligned}$$

Alternatively, the control residual can also be approximated by

$$\frac{1}{2}\rho^q(\tilde{q}_{\gamma,h}, \tilde{p}_{\gamma,h})(q_d),$$

since it vanishes for $q_d \equiv 0$, c.f., [31]. Beside the remainder, the terms

$$\begin{aligned}
&\frac{1}{2}\rho^*(\tilde{u}_{\gamma,h}, \tilde{q}_{\gamma,h}, \tilde{p}_{\gamma,h})(u - i_{2h}^{(2)}\tilde{u}_{\gamma,h}) + \frac{1}{2}\rho^q(\tilde{q}_{\gamma,h}, \tilde{p}_{\gamma,h})(q - i_{2h}^{(2)}\tilde{q}_{\gamma,h})\\
&+ \frac{1}{2}\rho(\tilde{u}_{\gamma,h}, \tilde{q}_{\gamma,h})(p - i_{2h}^{(2)}\tilde{p}_{\gamma,h}),
\end{aligned}$$

are also neglected. In [2, Section 5.2], it is proven that the corresponding term for the Poisson problem is of higher order in h assuming smooth solutions and uniform meshes. In the general case, this is an open question. However, the numerical results in Section 6 suggest that this approach also works in the situation considered here, as it does in many other situations, see for instance [5].

Finally, the a posteriori error estimate of the regularization error (32) can not be evaluated numerically, since it is based on the analytical values. Thus, we use the discrete counterpart

$$\gamma\big(\langle \dot{u}_{\gamma,h}, \mu_{\gamma,h}\rangle + \langle \dot{\lambda}_{\gamma,h}, \theta_{\gamma,h}\rangle\big) =: \eta_t \tag{40}$$

as estimator. All in all, we have deduced the a posteriori error estimate

$$J(\bar{u}, \bar{q}) - J(\tilde{u}_{\gamma,h}, \tilde{q}_{\gamma,h}) \approx \eta := \eta_t + \eta_h + \eta_n.$$

We note that all of the three indicators are, in fact, functions evaluated in the current iterate. To ease notation, we neglected the explicit statement of this fact.

To utilize η_h in an adaptive refinement strategy, we have to localize the error contributions given by the primal, dual, and control residuals with respect to the single mesh elements $T \in \mathcal{T}_h$ leading to local error indicators $\eta_{h,T}$. Here, the filtering technique developed in [7] is applied, which implies less implementational effort than the standard approach using integration by parts outlined for instance in [2]. The optimal mesh strategy developed in [25] is used as adaptive refinement strategy.

The adaptive solution procedure is described in the following Algorithm:

Algorithm 1. *The adaptive solution algorithm consists of the following steps using a stopping tolerance* $\mathrm{tol} > 0$*, a safety factor* $c_s \in (0,1)$*, an equilibration constant* $c_e > 1$*, and an factor* $c_\gamma > 1$ *to steer the growth of* γ *during the algorithm:*

(1) Choose an initial triangulation $\mathcal{T}_h^0$ *and an initial regularization parameter* γ^0*. Set* $l = 0$*.*

(2) Choose an initial value $\left(\tilde{u}_{\gamma,h}^{l,0}, \tilde{q}_{\gamma,h}^{l,0}, \tilde{p}_{\gamma,h}^{l,0}\right)$ *and set* $n = 0$*.*

(3) Perform one step of Newton's method: $\left(\tilde{u}_{\gamma,h}^{l,n}, \tilde{q}_{\gamma,h}^{l,n}, \tilde{p}_{\gamma,h}^{l,n}\right) \to \left(\tilde{u}_{\gamma,h}^{l,n+1}, \tilde{q}_{\gamma,h}^{l,n+1}, \tilde{p}_{\gamma,h}^{l,n+1}\right)$*.*

(4) Determine η_h *and* η_n*.*

(5) If $c_s|\eta_h| < |\eta_n|$ *then increment* n *and go to (3).*

(6) Determine η_t*.*

(7) If $|\eta_t| + |\eta_h| + |\eta_n| < \mathrm{tol}$ *then quit.*

(8) If $|\eta_t| > c_e|\eta_h|$ *then adaptively refine,* $\mathcal{T}_h^l \to \mathcal{T}_h^{l+1}$*, and set* $\gamma^{l+1} = \gamma^l$*.*

(9) If $c_e|\eta_t| < |\eta_h|$ *then enhance the regularization,* $\gamma^{l+1} = c_\gamma \gamma^l$*, and set* $\mathcal{T}_h^l = \mathcal{T}_h^{l+1}$*.*

(10) If $c_e|\eta_t| \geq |\eta_h| \geq c_e^{-1}|\eta_t|$ *then adaptively refine,* $\mathcal{T}_h^l \to \mathcal{T}_h^{l+1}$*, and enhance the regularization,* $\gamma^{l+1} = c_\gamma \gamma^l$*.*

(11) Increment l *and go to (2).*

Some comments on the adaptive solution algorithm are in order. We use $c_s = 10^{-3}$, $c_e = 5$, and $c_\gamma = \sqrt{10}$ in our numerical experiments. Especially, the choice of c_γ is crucial, since one wants to stay in the quadratic convergence radius of Newton's method. In our numerical experiments the mentioned choice has worked well. Furthermore, the initial values for Newton's method are determined by extrapolation of the old solution, if we increase γ, or by interpolation on the new mesh, if an adaptive mesh refinement is conducted.

6. Numerical results

In this section, we consider two challenging examples to test the presented error estimator and the adaptive algorithm. For comparison, an heuristic patch recovery estimator of the form

$$\eta_{zz} := \|P(\nabla u_h) - \nabla u_h\| + \|P(\nabla q_h) - \nabla q_h\| + \|P(\nabla p_h) - \nabla p_h\|$$

for the discretization error in combination with the estimator

$$\eta_{e,t} := J(\tilde{u}_{\infty,h}, \tilde{q}_{\infty,h}) - J(\tilde{u}_{\gamma,h}, \tilde{q}_{\gamma,h}) = \gamma(\tilde{u}_{\gamma,h} - u_d, \dot{u}_{\gamma,h}) + \gamma\alpha(\tilde{q}_{\gamma,h} - q_d, \dot{q}_{\gamma,h})$$

of the regularization error is used. Here, P denotes the usual recovery operator described for instance in [1].

6.1. First example

We set $\Omega = [0,1]^2$ with homogeneous Dirichlet boundary conditions on $\partial\Omega$. The obstacle ψ is given by $\psi \equiv -0.25$ and a volume force f by $f = -2\pi^2 \sin(\pi x)\sin(\pi y)$. As desired state u_d, we choose $u_d = -\sin(\pi x)\sin(\pi y)$ in the subdomain $\Omega_1 = [0.375, 0.625]^2$, i.e.,

$$J(u,q) = \frac{1}{2}\|u - u_d\|^2_{L^2(\Omega_1)} + \frac{\alpha}{2}\|q\|_{L^2(\Omega)}.$$

Due to $u_d < \psi$ in Ω_1, the optimal state in Ω_1 is given by $u = \psi$. Because of the volume force f, the optimal state is achieved at least for all $q \leq 0$. Consequently, the optimal control is $q \equiv 0$ and for the optimal value it holds

$$J(u,q) = \frac{1}{2}\|\psi - u_d\|^2_{L^2(\Omega_1)} = \frac{5\pi^2 + 16\pi\sqrt{2} + 128\sqrt{2} - 224}{512\pi^2}.$$

In Figure 2, the solution is depicted. As the optimal control q is included in the discrete

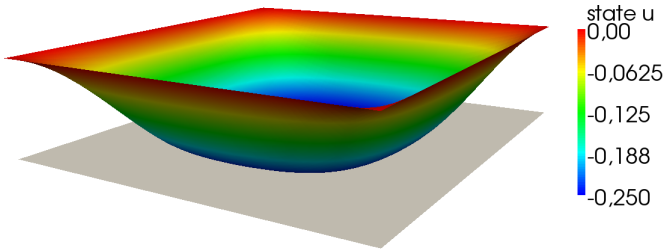


Figure 2: Setting of the first example.

ansatz space, i.e. $q \in V_h$, for all mesh sizes h, the discrete solution is $q_h = q$ and does not depend on the mesh size h. Consequently, the adaptive algorithm should only increase the regularization parameter without any adaptive mesh refinement. Table 1 shows that the presented error estimator based on the DWR method really gives the expected behavior, where

$$E_{\text{rel}} := \frac{J(u,q) - J\left(\tilde{u}_{\gamma,h}, \tilde{q}_{\gamma,h}\right)}{J(u,q)}$$

and

$$I_{\text{eff}} := \frac{J(u,q) - J\left(\tilde{u}_{\gamma,h}, \tilde{q}_{\gamma,h}\right)}{\eta}.$$

The only adaptive mesh refinements are carried out in the last steps of the solution algorithm. The corresponding adaptive mesh is depicted in Figure 3. In contrast to

l	N	γ	E_{rel}	I_{eff}
1	64	10^{1}	$5.744{\cdot}10^{-1}$	0.272
2	64	$10^{3/2}$	$2.238{\cdot}10^{-1}$	0.498
3	64	10^{2}	$7.446{\cdot}10^{-2}$	0.603
4	64	$10^{5/2}$	$2.375{\cdot}10^{-2}$	0.650
5	64	10^{3}	$7.517{\cdot}10^{-3}$	0.669
6	64	$10^{7/2}$	$2.379{\cdot}10^{-3}$	0.675
7	64	10^{4}	$7.529{\cdot}10^{-4}$	0.674
8	64	$10^{9/2}$	$2.376{\cdot}10^{-4}$	0.669
9	64	10^{5}	$7.439{\cdot}10^{-5}$	0.654
10	64	$10^{11/2}$	$2.275{\cdot}10^{-5}$	0.604
11	64	10^{6}	$6.415{\cdot}10^{-6}$	0.422
12	208	$10^{13/2}$	$2.250{\cdot}10^{-6}$	0.385
13	400	10^{7}	$6.353{\cdot}10^{-7}$	1.063
14	400	$10^{15/2}$	$1.159{\cdot}10^{-7}$	2.140

Table 1: Results of the adaptive algorithm based on the DWR error estimator for the first example with coarse initial mesh.

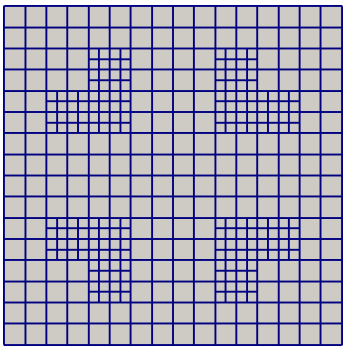


Figure 3: Adaptive mesh in the 13^{th} step of the algorithm based on the DWR estimator with coarse initial mesh.

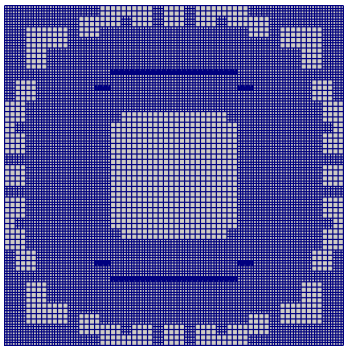


Figure 4: Adaptive mesh in the 5^{th} step of the algorithm based on the heuristic estimator.

the DWR estimator the heuristic estimator shows a completely different behavior as outlined in Table 2. We observe mainly mesh refinements and no reductions of the regularization parameter. Hence, the error is not efficiently reduced. The adaptive mesh is depicted in Figure 4. The effectivity indices for the presented error estimator given in Table 1 are not satisfactory due to the very coarse mesh. Using a finer initial mesh, we obtain much better effectivity indices, c.f., Table 3.

l	N	γ	E_{rel}
1	64	10^1	0.574
2	160	10^1	0.574
3	400	10^1	0.572
4	1216	10^1	0.571
5	4624	$10^{3/2}$	0.222
6	13408	$10^{3/2}$	0.222
7	50656	10^2	0.0741

Table 2: Results of the adaptive algorithm based on the heuristic error estimator for the first example.

l	N	γ	E_{rel}	I_{eff}
1	16384	10^1	$5.712{\cdot}10^{-1}$	0.467
2	16384	$10^{3/2}$	$2.218{\cdot}10^{-1}$	0.688
3	16384	10^2	$7.412{\cdot}10^{-2}$	0.821
4	16384	$10^{5/2}$	$2.381{\cdot}10^{-2}$	0.897
5	16384	10^3	$7.564{\cdot}10^{-3}$	0.939
6	16384	$10^{7/2}$	$2.396{\cdot}10^{-3}$	0.960
7	16384	10^4	$7.580{\cdot}10^{-4}$	0.971
8	16384	$10^{9/2}$	$2.397{\cdot}10^{-4}$	0.975
9	16384	10^5	$7.581{\cdot}10^{-5}$	0.977
10	16384	$10^{11/2}$	$2.397{\cdot}10^{-5}$	0.977
11	16384	10^6	$7.581{\cdot}10^{-6}$	0.977
12	16384	$10^{13/2}$	$2.397{\cdot}10^{-6}$	0.977
13	16384	10^7	$7.578{\cdot}10^{-7}$	0.978
14	16384	$10^{15/2}$	$2.395{\cdot}10^{-7}$	0.979
15	16384	10^8	$7.553{\cdot}10^{-8}$	0.982

Table 3: Results of the adaptive algorithm based on the DWR error estimator for the first example with fine initial mesh.

6.2. Second example

The second example is `mpccdist1` taken from the OPTPDE-problem collection [21], it was introduced originally in [18, Example 7.1]. We use the domain $\Omega := [0,1]^2$ as well as the subdomains $\Omega_2 := [0,0.5] \times [0,0.8]$ and $\Omega_3 := [0.5,1] \times [0,0.8]$. The subdomain Ω_1 is a square with midpoint $\hat{x} = (0.8, 0.9)$ and edge length 0.1, which has been rotated about its midpoint by 30 degrees in counter-clockwise direction. hence, its boundary

is not resolved by the mesh. The four vertices of Ω_1 can thus be obtained from

$$\begin{pmatrix} \hat{x} & \hat{x} & \hat{x} & \hat{x} \end{pmatrix} + Q \begin{pmatrix} -0.05 & 0.05 & 0.05 & -0.5 \\ -0.05 & -0.05 & 0.05 & 0.05 \end{pmatrix} \approx \begin{pmatrix} 0.7817 & 0.8683 & 0.8183 & 0.7317 \\ 0.8317 & 0.8817 & 0.9683 & 0.9183 \end{pmatrix}$$

with the rotation matrix

$$Q = \begin{pmatrix} \cos\frac{\pi}{6} & -\sin\frac{\pi}{6} \\ \sin\frac{\pi}{6} & \cos\frac{\pi}{6} \end{pmatrix}.$$

Note that Ω_1 does not intersect Ω_2 nor Ω_3. The desired state is given by

$$u_d(x) = \begin{cases} -400\,(q_1(y_1) + q_2(y_2))|_{y=Q^\top(x-\hat{x})+\hat{x}}\,, & x \in \Omega_1, \\ z_1(x_1)z_2(x_2), & x \in \Omega_2, \\ 0, & \text{elsewhere}, \end{cases}$$

and the desired control by

$$q_d(x) = \begin{cases} p_1(Q\top(x-\hat{x}), & x \in \Omega_1, \\ -z_1''(x_1) - z_2''(x_2), & x \in \Omega_2, \\ -z_1(x_1-0.5)z_2(x_2), & x \in \Omega_3, \\ 0, & \text{elsewhere}, \end{cases}$$

where the remaining pieces of data are

$$\begin{aligned} z_1(x_1) &= -4096x_1^6 + 6144x_1^5 - 3072x_1^4 + 512x_1^3, \\ z_2(x_2) &= -244140625x_2^6 + 585937500x_2^5 - 468750x_2^4 + 125x_2^3, \\ q_1(x_1) &= -200(x_1 - 0.8)^2 + 0.5, \\ q_2(x_2) &= -200(x_2 - 0.9)^2 + 0.5, \\ p_1(x_1, x_2) &= q_1(x_1)q_2(x_2). \end{aligned}$$

The main characteristics of this example are a biactive set of a measure larger than zero, low regularity and data not aligned along the finite element mesh. We will show that the adaptive algorithm can cope with all these aspects and lead to an efficient solution algorithm. The solution is illustrated in the Figures 5, 6, and 7. In Table 4, the results concerning a uniform mesh refinement are summarized. There, the regularization parameter γ is chosen as the smallest value, where the results do not change any more. We observe a coupled decrease of h and increase of γ to obtain reasonable results. Furthermore, the effectivity indices of the DWR error estimator are stated, which are close to one but with a minus sign. The results of the adaptive algorithm based on the DWR error estimator are given in Table 5. We achieve a similar accuracy as the uniform approach with far less mesh elements and smaller values of γ, where the error is estimated accurately as well as the estimated error constituents are properly balanced. The adaptively refined mesh in the 8^{th} iteration is shown in Figure 8. We observe that mainly the lower left part is refined, whereas

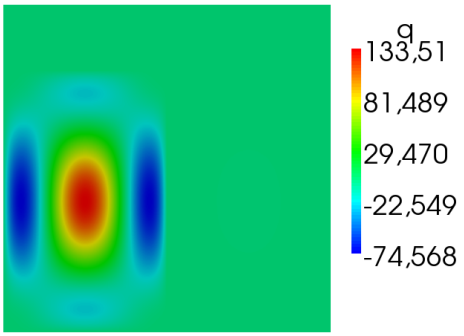


Figure 5: Optimal control in the second example.

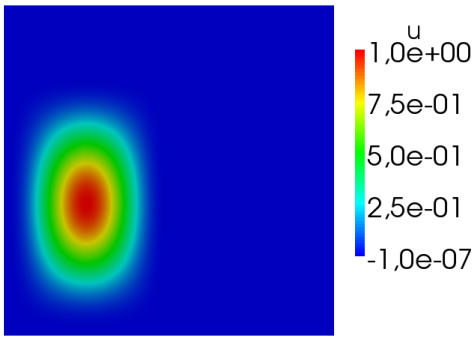


Figure 6: State in the second example.

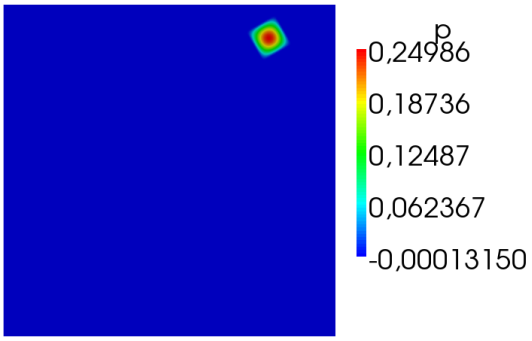


Figure 7: Adjoint state in the second example.

l	N	γ	E_{rel}	I_{eff}
1	100	10^1	-6.549·10^{-2}	-0.327
2	400	10^3	-3.092·10^{-3}	-0.879
3	1600	10^5	-1.537·10^{-4}	-0.985
4	6400	10^6	-7.709·10^{-6}	-1.042
5	25600	10^7	-4.237·10^{-7}	-1.047
6	102400	$10^{15/2}$	-3.030·10^{-8}	-0.847
7	409600	10^9	-1.372·10^{-9}	-1.120

Table 4: Results considering uniform mesh refinement.

the upper right part is much less refined, although one would at first sight expect refinements due to the problem data not aligned to the grid there. However, the use of special quadrature rules resolves this problem and so no additional refinements are needed. For comparison, the results of the adaptive algorithm based on the heuristic

l	N	γ	E_{rel}	I_{eff}	η_r	η_h
1	100	10^{1}	$-6.548{\cdot}10^{-2}$	-0.327	$1.865{\cdot}10^{-4}$	$8.370{\cdot}10^{-0}$
2	184	10^{1}	$-2.462{\cdot}10^{-2}$	-0.0985	$1.999{\cdot}10^{-4}$	$9.477{\cdot}10^{-1}$
3	340	10^{1}	$-2.592{\cdot}10^{-3}$	-0.512	$1.737{\cdot}10^{-4}$	$5.193{\cdot}10^{-1}$
4	472	10^{1}	$-3.641{\cdot}10^{-4}$	-0.840	$1.623{\cdot}10^{-4}$	$1.194{\cdot}10^{-1}$
5	916	10^{1}	$-4.882{\cdot}10^{-5}$	-0.954	$1.653{\cdot}10^{-4}$	$1.805{\cdot}10^{-2}$
6	2248	10^{1}	$1.390{\cdot}10^{-5}$	0.597	$1.714{\cdot}10^{-4}$	$3.075{\cdot}10^{-3}$
7	4444	10^{1}	$1.703{\cdot}10^{-5}$	0.112	$1.731{\cdot}10^{-4}$	$5.734{\cdot}10^{-4}$
8	9640	10^{1}	$2.345{\cdot}10^{-5}$	0.0316	$1.743{\cdot}10^{-4}$	$1.155{\cdot}10^{-4}$
9	20176	$10^{3/2}$	$2.107{\cdot}10^{-5}$	0.261	$2.133{\cdot}10^{-3}$	$2.070{\cdot}10^{-5}$
10	20176	10^{2}	$1.257{\cdot}10^{-5}$	0.6136	$2.992{\cdot}10^{-3}$	$2.294{\cdot}10^{-5}$
11	20176	$10^{5/2}$	$5.376{\cdot}10^{-6}$	0.875	$1.815{\cdot}10^{-3}$	$2.544{\cdot}10^{-5}$
12	20176	10^{3}	$1.736{\cdot}10^{-6}$	1.135	$7.444{\cdot}10^{-4}$	$2.665{\cdot}10^{-5}$
13	20176	$10^{7/2}$	$4.294{\cdot}10^{-7}$	1.509	$2.264{\cdot}10^{-4}$	$2.695{\cdot}10^{-5}$
14	20176	10^{4}	$6.175{\cdot}10^{-8}$	3.555	$5.886{\cdot}10^{-5}$	$2.700{\cdot}10^{-5}$
15	43840	$10^{9/2}$	$2.648{\cdot}10^{-8}$	2.077	$1.593{\cdot}10^{-5}$	$5.571{\cdot}10^{-6}$
16	83512	10^{5}	$9.670{\cdot}10^{-9}$	1.508	$4.379{\cdot}10^{-6}$	$1.326{\cdot}10^{-6}$
17	185572	$10^{11/2}$	$-3.079{\cdot}10^{-9}$	-1.221	$1.205{\cdot}10^{-6}$	$2.663{\cdot}10^{-7}$

Table 5: Results of the adaptive algorithm based on the DWR estimator for the second example.

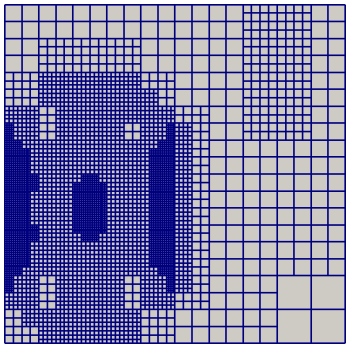


Figure 8: Adaptive mesh in the 8^{th} step of the algorithm based on the DWR estimator for the second example.

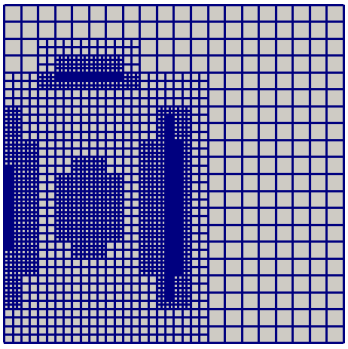


Figure 9: Adaptive mesh in the 5^{th} step of the algorithm based on the heuristic estimator for the second example.

error estimator are outlined in Table 6. As in the first example, the regularization parameter γ is not increased because of the overestimation of the discretization error by the heuristic estimator. Thus, the error is not optimally reduced. The resulting adaptive mesh, c.f., Figure 9, shows the same properties as the mesh based on the DWR error estimator.

l	N	γ	E_{rel}
1	100	10^1	$-6.549\cdot10^{-2}$
2	196	10^1	$-6.486\cdot10^{-3}$
3	616	10^1	$-3.537\cdot10^{-4}$
4	1516	10^1	$-5.509\cdot10^{-6}$
5	2500	10^1	$1.042\cdot10^{-5}$
6	4648	10^1	$1.709\cdot10^{-5}$
7	9580	10^1	$2.148\cdot10^{-5}$
8	22084	10^1	$2.296\cdot10^{-5}$
9	38692	10^1	$1.863\cdot10^{-5}$
10	88336	10^1	$1.869\cdot10^{-5}$
11	155956	10^1	$2.129\cdot10^{-5}$

Table 6: Results of the adaptive algorithm based on the heuristic estimator for the second example.

7. Conclusions

We presented an adaptive algorithm for the finite element approximation of optimal control problems governed by variational inequalities of obstacle type. The construction of the algorithm is based on a penalty-type regularization, whose a priori approximation properties can be analyzed by standard techniques. To obtain an efficient overall algorithm the most crucial point is to design an accurate a posteriori estimator for the regularization error, which can be evaluated with reasonable effort. Our construction relies on the path derivative, i.e., the derivative of the optimal solution w.r.t. the penalization parameter. Under additional assumptions on the regularized solutions, the path can indeed be shown to be differentiable. Unfortunately, these assumptions cannot be shown to be fulfilled in general. The major advantage of this regularization error estimator is its low computational costs, since one only has to solve the linearized system (30), which amounts to the same effort as an additional Newton step for the nonlinear regularized optimality system (9). The discretization error for the regularized problems is then estimated by the standard DWR method. For fixed values of the regularization parameter γ, the classical results are obtained, but the estimator depends on γ. It is an open question, whether the remainder term in (35) can be bounded independently of γ or one has to couple mesh size and γ suitably to keep this remainder term negligible. This gives rise to future research. Despite these open questions, the algorithm performs well in two challenging numerical tests. It reliably detects when to keep the mesh constant and to increase the penalty parameter as the first example shows. Furthermore, as seen in the second example, the algorithm can cope with a biactive set of positive measure, which induces a non-differentiability in the unregularized control-to-state mapping. Finally, both examples show convincing efficiency indices.

References

[1] M. Ainsworth and J. T. Oden. *A posteriori error estimation in finite element analysis*. Pure and Applied Mathematics. Wiley-Interscience, Chichester, 2000.

[2] W. Bangerth and R. Rannacher. *Adaptive finite element methods for differential equations*. Lectures in Mathematics, ETH Zürich. Birkhäuser, Basel, 2003.

[3] V. Barbu. *Optimal Control of Variational Inequalities*, volume 100 of *Research Notes in Mathematics*. Pitman, Boston, 1984.

[4] R. Becker and R. Rannacher. A feed-back approach to error control in finite element methods: Basic analysis and examples. *East-West J. Numer. Math.*, 4:237–264, 1996.

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

[6] H. Blum and F.-T. Suttmeier. Weighted error estimates for finite element solutions of variational inequalities. *Computing*, 65(2):119–134, 2000.

[7] M. Braack and A. Ern. A posteriori control of modeling errors and discretization errors. *Multiscale Model. Simul.*, 1(2):221–238, 2003.

[8] D. Braess. A posteriori error estimators for obstacle problems - another look. *Numer. Math.*, 101(3):415–421, 2005.

[9] D. Braess, C. Carstensen, and R. H. W. Hoppe. Convergence analysis of a conforming adaptive finite element method for an obstacle problem. *Numer. Math.*, 107:455–471, 2007.

[10] E. Casas and F. Tröltzsch. Error estimates for the finite-element approximation of a semilinear elliptic control problem. *Control and Cybernetics*, 31:695–712, 2002.

[11] Z. Chen and R. H. Nochetto. Residual type a posteriori error estimates for elliptic obstacle problems. *Numer. Math.*, 84(4):527–548, 2000.

[12] D. A. French, S. Larsson, and R. H. Nochetto. Pointwise a posteriori error analysis for an adaptive penalty finite element method for the obstacle problem. *Comput. Methods Appl. Math*, 1(1):18–39, 2001.

[13] M. Hintermüller. Inverse coefficient problems for variational inequalities: Optimality conditions and numerical realization. *ESAIM Mathematical Modelling and Numerical Analysis*, 35(1):129–152, 2001.

[14] M. Hintermüller, R.H.W. Hoppe, and C. Löbhard. A dual-weighted residual approach to goal-oriented adaptivity for optimal control of elliptic variational inequalities. *ESAIM: Control, Optimisation and Calculus of Variations*, 20(2):524–546, 2014.

[15] C. Johnson. Adaptive finite element method for the obstacle problem. *Math. Models Methods Appl. Sci.*, 2(4):483–487, 1992.

[16] D. Kinderlehrer and G. Stampacchia. *An Introduction to Variational Inequalities and Their Applications*. Academic Press, New York, 1980.

[17] K. Kunisch and D. Wachsmuth. Path-following for optimal control of stationary variational inequalities. *Computational Optimization and Applications*, pages 1–29, 2012.

[18] Ch. Meyer and O. Thoma. A priori finite element error analysis for optimal control of the obstacle problem. *SIAM Journal on Numerical Analysis*, 51(1):605–628, 2013.

[19] F. Mignot. Contrôle dans les inéquations variationelles elliptiques. *Journal of Functional Analysis*, 22(2):130–185, 1976.

[20] F. Mignot and J.-P. Puel. Optimal control in some variational inequalities. *SIAM Journal on Control and Optimization*, 22(3):466–476, 1984.

[21] OPTPDE — a collection of problems in PDE-constrained optimization. `http://www.optpde.net`.

[22] R. Rannacher, B. Vexler, and W. Wollner. A posteriori error estimation in PDE-constrained optimization with pointwise inequality constraints. In *Constrained Optimization and Optimal Control for Partial Differential Equations*, volume 160 of *International Series of Numerical Mathematics*, pages 349–373. Springer, 2012.

[23] R. Rannacher and J. Vihharev. Adaptive finite element analysis of nonlinear problems: balancing of discretization and iteration errors. *Journal on Numerical Mathematics*, 21(1):23–61, 2013.

[24] R. Rannacher, A. Westenberger, and W. Wollner. Adaptive finite element solution of eigenvalue problems: Balancing of discretization and iteration error. *J. Numer. Math*, 18(4):303–327, 2010.

[25] T. Richter. *Parallel Multigrid Method for Adaptive Finite Elements with Application to 3D Flow Problems*. PhD thesis, Ruprecht-Karls-Universität Heidelberg, (Diss.), 2005.

[26] H. Scheel and S. Scholtes. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, 25(1):1–22, 2000.

[27] A. Schiela and D. Wachsmuth. Convergence analysis of smoothing methods for optimal control of stationary variational inequalities. *ESAIM Math. Model. Numer. Anal.*, 47(3):771–787, 2013.

[28] F.-T. Suttmeier. General approach for a posteriori error estimates for finite element solutions of variational inequalities. *Comput. Mech.*, 27(4):317–323, 2001.

[29] F. Tröltzsch. *Optimal Control of Partial Differential Equations*, volume 112 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, 2010. Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels.

[30] A. Veeser. Efficient and reliable a posteriori error estimators for elliptic obstacle problems. *SIAM J. Numer. Anal.*, 39(1):146–167, 2001.

[31] B. Vexler and W. Wollner. Adaptive finite elements for elliptic optimization problems with control constraints. *SIAM J. Control Optim.*, 47(1):509–534, 2008.

[32] G. Wachsmuth. Strong stationarity for optimal control of the obstacle problem with control constraints. Preprint, Technical University Chemnitz, 2014.

[33] W. Wollner. *Adaptive Methods for PDE-based Optimal Control with Pointwise Inequality Constraints*. PhD thesis, Mathematisch-Naturwissenschaftliche Gesamtfakultät, Universität Heidelberg, 2010.

[34] W. Wollner. A posteriori error estimates for a finite element discretization of interior point methods for an elliptic optimization problem with state constraints. *Comput. Optim. Appl.*, 47(1):133–159, 2010.

A. Proof of Lemma 2

Proof of Lemma 2. Although the proof is standard, we present the arguments for convenience of the reader. By the Lipschitz continuity of S_γ, the sequence $\{u_\gamma\}$ is bounded in $H_0^1(\Omega)$ such that there is a weakly converging subsequence, w.l.o.g. the whole sequence itself. The weak limit is denoted by $u \in H_0^1(\Omega)$. Testing (8) with $\psi - u_\gamma$ yields

$$\int_\Omega \max\{\psi - u_\gamma, 0\}^4 \, dx = \frac{1}{\gamma^3}\Big(\langle q, u_\gamma - \psi\rangle + a(u_\gamma, \psi - u_\gamma)\Big).$$

Thanks to the weak convergence the term in brackets is bounded so that the right hand side converges to zero for $\gamma \to \infty$. The compactness of $H^1(\Omega) \to L^4(\Omega)$ implies that the left hand side converges to $\|\max\{\psi - u, 0\}\|^4_{L^4(\Omega)}$, giving in turn that $u \geq \psi$ a.e. in Ω, i.e., $u \in K$. Next we test (8) with $v - u_\gamma$, where $v \in K$ is arbitrary. Then, due to

$$\begin{aligned}\int_\Omega r(\gamma; u_\gamma)(v - u_\gamma)\, dx \\ = -\gamma^3\Big(\int_\Omega \max\{\psi - u_\gamma, 0\}^3 \underbrace{(v - \psi)}_{\geq 0}\, dx + \int_\Omega \max\{\psi - u_\gamma, 0\}^4\, dx\Big) \leq 0,\end{aligned}$$

we arrive at

$$a(u_\gamma, v - u_\gamma) \geq \langle q, v - u_\gamma\rangle \quad \forall\, v \in K$$

and the weak lower semicontinuity of a implies that the weak limit u is indeed the unique solution of (6). By weak lower semicontinuity, we further infer that

$$\begin{aligned} 0 \leq \liminf_{\gamma\to\infty} \big(a(u_\gamma, u_\gamma) - a(u,u)\big) &\leq \limsup_{\gamma\to\infty} \big(a(u_\gamma, u_\gamma) - a(u,u)\big) \\ &\leq \lim_{\gamma\to\infty} \big(a(u, u_\gamma - u) - \langle q, u - u_\gamma\rangle\big) = 0. \end{aligned}$$

Hence, due to coercivity of a, we have $\|u_\gamma\|_{H^1(\Omega)} \to \|u\|_{H^1(\Omega)}$, and norm and weak convergence imply strong convergence.

To verify the convergence rate, first note that, by the Stampacchia-Lemma, the pointwise projection on K yields an element of $H_0^1(\Omega)$, i.e., $\max\{v,\psi\} \in H_0^1(\Omega)$. If one tests (8) with $v = u - \max\{u_\gamma, \psi\}$ so that

$$\begin{aligned} &a(u_\gamma, u - \max\{u_\gamma,\psi\}) \\ &\quad - \gamma^3 \int_\Omega \max\{\psi - u_\gamma, 0\}^3 \big(u - (\psi + \max\{u_\gamma - \psi, 0\})\big)\, dx = \langle q, u - \max\{u_\gamma,\psi\}\rangle \end{aligned}$$

is obtained. Since $\max\{\psi - u_\gamma, 0\}^3 (u - \psi) \geq 0$ a.e. in Ω (due to $u \in K$) and $\max\{\psi - u_\gamma, 0\}^3 \max\{u_\gamma - \psi, 0\} = 0$ a.e. in Ω, this implies

$$a(u_\gamma, u - \max\{u_\gamma,\psi\}) \geq \langle q, u - \max\{u_\gamma,\psi\}\rangle.$$

Adding this inequality to (6) tested with $v = \max\{u_\gamma,\psi\} \in K$ gives

$$a(u - u_\gamma, \max\{u_\gamma,\psi\} - u) \geq 0$$

such that the coercivity of a yields

$$\begin{aligned} \beta \|u - u_\gamma\|^2_{H_0^1(\Omega)} &\leq a(u - u_\gamma, \max\{u_\gamma,\psi\} - u_\gamma) \\ &\leq c \|u - u_\gamma\|_{H_0^1(\Omega)} \|u_\gamma - \max\{u_\gamma,\psi\}\|_{H_0^1(\Omega)}. \end{aligned} \tag{41}$$

It remains to estimate the projection error $u_\gamma - \max\{u_\gamma,\psi\}$. For this purpose, test (8) with $u_\gamma - \max\{u_\gamma,\psi\}$ which gives

$$\begin{aligned} &\langle q - A\psi, u_\gamma - \max\{u_\gamma,\psi\}\rangle \\ &\qquad = a(u_\gamma - \psi, u_\gamma - \max\{u_\gamma,\psi\}) + \gamma^3 \|u_\gamma - \max\{u_\gamma,\psi\}\|^4_{L^4(\Omega)}. \end{aligned} \tag{42}$$

By defining –up to sets of zero measure– $\widehat{\Omega} := \{x \in \Omega : u_\gamma(x) < \psi(x)\}$, we obtain $u_\gamma - \max\{u_\gamma,\psi\} = 0$ a.e. in $\Omega \setminus \widehat{\Omega}$ and consequently

$$\begin{aligned} &a(u_\gamma - \psi, u_\gamma - \max\{u_\gamma,\psi\}) \\ &\quad = \int_{\widehat{\Omega}} \sum_{i=1}^d \Big(\sum_{j=1}^d a_{ij} \frac{\partial(u_\gamma - \psi)}{\partial x_j} \frac{\partial(u_\gamma - \psi)}{\partial x_j} dx + b_i \frac{\partial(u_\gamma - \psi)}{\partial x_i} (u_\gamma - \psi)\Big) + a_0 (u_\gamma - \psi)^2\, dx \\ &\quad = a(u_\gamma - \max\{u_\gamma,\psi\}, u_\gamma - \max\{u_\gamma,\psi\}). \end{aligned}$$

Thus, (42) together with the coercivity of a and Young's inequality yields

$$\begin{aligned}
&\beta \left\| u_\gamma - \max\{u_\gamma, \psi\} \right\|^2_{H^1(\Omega)} \\
&\qquad \leq a(u_\gamma - \max\{u_\gamma, \psi\}, u_\gamma - \max\{u_\gamma, \psi\}) \\
&\qquad = \int_\Omega (q - A\psi)\big(u_\gamma - \max\{u_\gamma, \psi\}\big) dx - \gamma^3 \left\| u_\gamma - \max\{u_\gamma, \psi\} \right\|^4_{L^4(\Omega)} \\
&\qquad \leq c \, \frac{1}{\gamma} \left\| q - A\psi \right\|^{4/3}_{L^{4/3}(\Omega)}.
\end{aligned}$$

Plugging this into (41) yields the assertion. □