

OPTIMAL CONTROL OF AN OBLIQUE DERIVATIVE PROBLEM*

Daniel Wachsmuth[†] Gerd Wachsmuth[‡]

Abstract

We investigate optimal control of an elliptic partial differential equation (PDE) with oblique boundary conditions. These boundary conditions do not lead directly to a weak formulation of the PDE. Thus, the equation is reformulated as a variational problem. Existence of optimal controls and regularity of solutions is proven. First-order optimality conditions are investigated. The adjoint state is interpreted as the solution of a boundary value problem with non-variational boundary conditions. Numerical results demonstrate the approximative solution of the optimal control problem by finite element discretization.

MSC: 49K20, 49M15, 65K10

Keywords: oblique boundary condition, non-variational boundary value problem, optimal control, optimality conditions.

1 Introduction

In this article we consider an optimal control problem for an elliptic partial differential equation with oblique boundary conditions. More precisely, we study the optimal control of the equation

$$-\partial_j(a_{ij} \partial_i y) + a_i \partial_i y + a_0 y = f \quad \text{in } \Omega, \quad (1.1a)$$

$$b_i \partial_i y + b_0 y = g \quad \text{on } \Gamma = \partial\Omega. \quad (1.1b)$$

*Accepted for publication in revised form on February 17, 2014

[†]Institut für Mathematik, Universität Würzburg, 97074 Würzburg, Germany,
daniel.wachsmuth@mathematik.uni-wuerzburg.de

[‡]Technische Universität Chemnitz, Faculty of Mathematics, 09107 Chemnitz, Germany,
gerd.wachsmuth@mathematik.tu-chemnitz.de

The control will act in the boundary condition. Here and throughout the paper we follow the Einstein summation convention. All the assumptions on the various coefficients will be made precise below.

In this model, the term $b_i \partial_i y$ is not a co-normal derivative of the elliptic differential operator. Thus the equation does not admit a weak formulation in the standard way: integration by parts of the strong formulation (1.1a) and inserting the boundary condition (1.1b) will not yield a variational formulation. This difficulty also influences the analysis of the optimal control problem: typically, necessary optimality conditions are expressed in terms of solutions of adjoint equations, which are naturally obtained in a weak form. Here, the question arises, whether the first-order necessary optimality conditions can be expressed by adjoint equations, and what is the corresponding weak and strong formulation of the adjoint equations. In the sequel we will use a well-known strategy to obtain a weak formulation of the equation by applying a suitable transformation of the differential operator, see [Troianiello, 1987, Proof of Lem. 3.18].

Oblique derivative problems have an abundance of applications, including geodesy, quantum gravity and portfolio optimization, see, e.g., Rozanov and Sansò [2002], Raskop and Grothaus [2006], Dowker and Kirsten [1997, 1999], Herzog et al. [2013]. For the mathematical theory of problems with those non-variational boundary conditions, we refer to Gilbarg and Trudinger [1983], Grisvard [1985], Troianiello [1987]. Optimal control problems for elliptic equations with boundary control are studied, e.g., in Tröltzsch [2010]. Control of semilinear and quasilinear equations is well studied, see, e.g., Casas and Dharmo [2012], Casas et al. [2005]. However, to the best of our knowledge, all the available results involve only PDEs with Dirichlet, Neumann, or Robin boundary conditions.

The investigation of the optimal control problem with oblique boundary conditions proceeds as follows. First, a reformulation is introduced, which turns the problem into a variational form. This variational formulation is equivalent to the strong formulation for $H^2(\Omega)$ -functions. Then, we prove existence and regularity of solutions of the weak formulation. Moreover, we show that the solution is independent of the choice of parameters introduced in the reformulation process.

Afterwards, we analyze the optimal control problem. The necessary optimality conditions are shown to involve an adjoint equation. Here, it is interesting to note that the strong formulation of the adjoint equation and the regularity of its solutions needs stronger smoothness assumptions on the coefficients of the differential operator.

Finally, we present some numerical results.

1.1 Notation

The partial derivative w.r.t. the coordinate x_i is denoted by ∂_i . We use Einstein's summation convention for repeated indices over $1, \dots, N$. If we state a condition involving one (or more) isolated indices, e.g., i , this condition is meant to hold for all possible values of these indices, e.g., $i = 1, \dots, N$. For example, $\nu_i \in C^{0,1}(\Gamma)$ means $\nu_i \in C^{0,1}(\Gamma)$ for all $i = 1, \dots, N$. By $C^{0,1}(\bar{\Omega})$, $C^{0,1}(\Gamma)$ we denote the Lipschitz continuous functions on $\bar{\Omega}$, Γ , respectively. Note that $C^{0,1}(\bar{\Omega}) = W^{1,\infty}(\Omega)$.

1.2 Standing assumptions

The domain $\Omega \subset \mathbb{R}^N$ is assumed to have a boundary $\partial\Omega$ of class $C^{1,1}$, see, e.g., [Troianiello, 1987, p. 13]. In (1.1), the coefficients satisfy $a_{ij} \in C^{0,1}(\bar{\Omega})$, $a_i, a_0 \in L^\infty(\Omega)$ and $b_i, b_0 \in C^{0,1}(\Gamma)$. Moreover, $a_{ij} = a_{ji}$ and

$$a_{ij}(x) \xi_i \xi_j \geq \underline{a} > 0 \quad \text{for all } x \in \Omega \text{ and } \xi \in \mathbb{R}^N. \quad (1.2)$$

Furthermore, we require the *oblique derivative condition* (1.1b) to be *regular*, i.e.,

$$b_i(x) \nu_i(x) \geq \underline{b} > 0 \quad \text{for all } x \in \Gamma, \quad (1.3)$$

where $\nu(x) \in \mathbb{R}^N$ is the outer unit normal vector at $x \in \Gamma$. Note that $\nu_i \in C^{0,1}(\Gamma)$.

We further assume

$$a_0 \geq 0, \quad b_0 \geq 0, \quad \text{ess sup}_\Omega a_0 + \max_\Gamma b_0 > 0. \quad (1.4)$$

1.3 Preliminary result: Multipliers on the boundary

We recall that the trace operator is a linear mapping that maps $H^1(\Omega)$ onto $H^{1/2}(\Gamma)$, see [Grisvard, 1985, Thm. 1.5.1.3]. The following lemma shows that the product of a function in $H^{1/2}(\Gamma)$ with a Lipschitz continuous one belongs to $H^{1/2}(\Gamma)$. That is, the Lipschitz continuous functions are multipliers in $H^{1/2}(\Gamma)$.

Lemma 1.1. *Let $u \in H^{1/2}(\Gamma)$ and $v \in C^{0,1}(\Gamma)$ be given. Then, the pointwise product uv belongs to $H^{1/2}(\Gamma)$ and*

$$\|uv\|_{H^{1/2}(\Gamma)} \leq C \|u\|_{H^{1/2}(\Gamma)} \|v\|_{C^{0,1}(\Gamma)},$$

where the constant C depends only on Ω .

Proof. We start by extending u and v to functions on Ω denoted by \tilde{u} and \tilde{v} , respectively. By applying [Troianiello, 1987, Thm. 1.2] and [Grisvard, 1985, Thm. 1.5.1.3], we obtain

$$\|\tilde{u}\|_{H^1(\Omega)} \leq C \|u\|_{H^{1/2}(\Gamma)} \quad \text{and} \quad \|\tilde{v}\|_{C^{0,1}(\bar{\Omega})} \leq C \|v\|_{C^{0,1}(\Gamma)}.$$

Now, it is easy to check, that

$$\|\tilde{u} \tilde{v}\|_{H^1(\Omega)} \leq C \|\tilde{u}\|_{H^1(\Omega)} \|\tilde{v}\|_{C^{0,1}(\bar{\Omega})}.$$

Applying [Grisvard, 1985, Thm. 1.5.1.3] again yields that the trace of $\tilde{u} \tilde{v}$ belongs to $H^{1/2}(\Omega)$ and

$$\|\tilde{u} \tilde{v}\|_{H^{1/2}(\Gamma)} \leq C \|\tilde{u}\|_{H^1(\Omega)} \|\tilde{v}\|_{C^{0,1}(\bar{\Omega})} \leq C \|u\|_{H^{1/2}(\Gamma)} \|v\|_{C^{0,1}(\Gamma)}.$$

Finally, it remains to prove that the trace of $\tilde{u} \tilde{v}$ coincides with uv . Since the product of the traces is the trace of the product for continuous functions, this can be established by approximating \tilde{u} with a continuous function. \square

2 The state equation

Albeit (1.1a) is in divergence form (and can be understood in the sense of distributions on Ω for $y \in H^1(\Omega)$), it is not straightforward to define the weak solution of (1.1) for $y \in H^1(\Omega)$, since (1.1b) is not a co-normal derivative. Therefore, we consider the case of regular solutions $y \in H^2(\Omega)$ first. Then, (1.1b) can be understood in the sense of traces since $\partial_i y \in H^1(\Omega)$. We call this a *strong solution* $y \in H^2(\Omega)$ of (1.1). We have the following result concerning existence and uniqueness.

Theorem 2.1 ([Troianiello, 1987, Thm. 3.29]). *For every $f \in L^2(\Omega)$ and $g \in H^{1/2}(\Gamma)$, there exists a unique strong solution $y = y(f, g) \in H^2(\Omega)$ of (1.1) and this solution satisfies*

$$\|y(f, g)\|_{H^2(\Omega)} \leq C (\|f\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\Gamma)}),$$

where $C > 0$ does not depend on f and g .

The same result, but with slightly stronger assumptions on the boundary data, can be found in [Grisvard, 1985, Thm. 2.4.2.6].

Following the approach of [Troianiello, 1987, Proof of Lem. 3.18], we are going to define *weak solutions* $y \in H^1(\Omega)$ of (1.1). Therefore, we derive a weak formulation of (1.1) such that the weak solutions coincide with the strong solutions of (1.1) in the regular case $f \in L^2(\Omega)$ and $g \in H^{1/2}(\Gamma)$.

To this end, let $y \in H^2(\Omega)$ and $\alpha_{ij}, \mu_i \in C^{0,1}(\bar{\Omega})$ be arbitrary. The symmetry of the Hessian matrix for smooth functions implies the symmetry of the weak Hessian matrix of y , i.e., $\partial_i \partial_j y = \partial_j \partial_i y$. Consequently, we obtain $(\alpha_{ij} - \alpha_{ji}) \partial_i \partial_j y = 0$. Together with the product rule we find that (1.1a) is equivalent to

$$-\partial_j [(a_{ij} + \alpha_{ij} - \alpha_{ji}) \partial_i y + \mu_j y] + (a_i + \partial_j(\alpha_{ij} - \alpha_{ji}) + \mu_i) \partial_i y + (a_0 + \partial_j \mu_j) y = f.$$

The co-normal derivative associated with this differential operator in divergence form is

$$\nu_j [(a_{ij} + \alpha_{ij} - \alpha_{ji}) \partial_i y + \mu_j y].$$

Hence, we will to construct α_{ij} and μ_j such that

$$\nu_j (a_{ij} + \alpha_{ij} - \alpha_{ji}) = \theta b_i \quad \text{and} \quad \nu_j \mu_j = \theta b_0 \quad (2.1)$$

hold on Γ , where $\theta \in C^{0,1}(\Gamma)$, $\theta \geq \underline{\theta} > 0$ is an appropriate scaling function.

Let us assume we have constructed $\alpha_{ij}, \mu_i, \theta$, such that (2.1) holds. Then, the above reasoning shows that if $y \in H^2(\Omega)$ is a solution of (1.1), we obtain by using integration by parts

$$a(y, v) = \int_{\Omega} f v \, dx + \int_{\Gamma} \theta g v \, ds \quad \text{for all } v \in H^1(\Omega), \quad (2.2)$$

where the bounded bilinear form $a : H^1(\Omega) \times H^1(\Omega) \rightarrow \mathbb{R}$ is given by

$$\begin{aligned} a(y, v) = & \int_{\Omega} [(a_{ij} + \alpha_{ij} - \alpha_{ji}) \partial_i y + \mu_j y] \partial_j v \\ & + (a_i + \partial_j(\alpha_{ij} - \alpha_{ji}) + \mu_i) \partial_i y v + (a_0 + \partial_j \mu_j) y v \, dx. \end{aligned} \quad (2.3)$$

Conversely, if $y \in H^2(\Omega)$ solves (2.2), y is also a strong solution of (1.1), see [Troianiello, 1987, Lem. 2.6]. Moreover, this shows that for all $f \in L^2(\Omega)$ and $g \in H^{1/2}(\Omega)$ the solution of (2.2) is independent of α_{ij} and μ_i (as long as (2.1) is satisfied), since the solution of (2.2) coincides with the strong solution of (1.1) and the strong solution is unique by Theorem 2.1.

It remains to construct $\alpha_{ij}, \mu_i \in C^{0,1}(\bar{\Omega})$ and $\theta \in C^{0,1}(\Gamma)$ such that (2.1) is satisfied. Multiplying the first equation of (2.1) by ν_i (and consequently summing over i) yields

$$\theta = \frac{a_{ij} \nu_i \nu_j}{b_i \nu_i} \quad \text{on } \Gamma. \quad (2.4)$$

Due to (1.2) and (1.3), $\theta \in C^{0,1}(\Gamma)$ is well defined and uniformly positive. Owing to the second equation of (2.1), we could choose μ_j such that $\mu_j =$

$\theta b_0 \nu_j$ on Γ . By extension, we find a function $\mu_j \in C^{0,1}(\bar{\Omega})$ such that $\mu_j = \theta b_0 \nu_j$ on Γ , see [Troianiello, 1987, Thm. 1.2].

It remains to choose the parameter α_{ij} . Note that the first equation of (2.1) is equivalent to

$$\nu_j (\alpha_{ij} - \alpha_{ji}) = \theta b_i - \nu_j a_{ij}.$$

Now, we define $\tau_i = \theta b_i - \nu_j a_{ij}$ and find $\tau_i \nu_i = 0$ by definition of θ , see (2.4). It remains to choose α_{ij} such that $\nu_j (\alpha_{ij} - \alpha_{ji}) = \tau_i$. This can be accomplished by choosing $\alpha_{ij} \in C^{0,1}(\bar{\Omega})$ such that $\alpha_{ij} = \nu_j \tau_i$ on Γ . This implies

$$\nu_j (\alpha_{ij} - \alpha_{ji}) = \tau_i = \theta b_i - \nu_j a_{ij} \quad \text{on } \Gamma. \quad (2.5)$$

Hence, (2.1) is satisfied by this choice of $\alpha_{ij}, \mu_i \in C^{0,1}(\bar{\Omega})$ and $\theta \in C^{0,1}(\Gamma)$.

Now, we define the notion of *weak solutions* of (1.1). The solution of the variational formulation (2.2) can be analogously defined for less regular functions. Let $f \in (H^1(\Omega))'$ and $g \in (H^{1/2}(\Gamma))'$ be given. We call $y \in H^1(\Omega)$ a weak solution of (1.1) if and only if

$$a(y, v) = \langle f, v \rangle_{(H^1(\Omega))', H^1(\Omega)} + \langle g, \theta v \rangle_{(H^{1/2}(\Gamma))', H^{1/2}(\Gamma)} \quad \text{for all } v \in H^1(\Omega) \quad (2.6)$$

holds. Note that multiplication with $\theta \in C^{0,1}(\Gamma)$ is a bounded, linear operator in $H^{1/2}(\Gamma)$, see Lemma 1.1. The above reasoning shows that every strong solution $y \in H^2(\Omega)$ is also a weak solution.

Theorem 2.2. *For every $f \in (H^1(\Omega))'$ and $g \in (H^{1/2}(\Gamma))'$, there exists a unique weak solution $y = y(f, g)$ of (1.1). Moreover, there exists $C > 0$ independent of f and g such that*

$$\|y(f, g)\|_{H^1(\Omega)} \leq C (\|f\|_{(H^1(\Omega))'} + \|g\|_{(H^{1/2}(\Gamma))'}).$$

Proof. We have

$$a(1, v) = \int_{\Omega} a_0 v \, dx + \int_{\Gamma} \theta b_0 v \, ds.$$

Hence, $a(1, v) \geq 0$ for all $v \in H^1(\Omega)$, $v \geq 0$ and there exists $v \in H^1(\Omega)$, $v \geq 0$ such that $a(1, v) > 0$, see (1.4). By classical arguments based on the weak maximum principle and the Fredholm alternative one finds, see e.g. [Troianiello, 1987, Cor. on p. 99] [Trudinger, 1973, Thm. 3.2],

$$a(y, v) = \langle F, v \rangle_{(H^1(\Omega))', H^1(\Omega)} \quad \text{for all } v \in H^1(\Omega)$$

possesses a unique solution $y = y(F) \in H^1(\Omega)$ for all $F \in (H^1(\Omega))'$. Moreover, the open mapping theorem implies the existence of $C > 0$ such that

$$\|y(F)\|_{H^1(\Omega)} \leq C \|F\|_{(H^1(\Omega))'}.$$

Choosing

$$\langle F, v \rangle_{(H^1(\Omega))', H^1(\Omega)} = \langle f, v \rangle_{(H^1(\Omega))', H^1(\Omega)} + \langle g, \theta v \rangle_{(H^{1/2}(\Gamma))', H^{1/2}(\Omega)}$$

yields the claim. \square

It remains to discuss the dependency of the weak solution of (1.1) on the (more or less arbitrarily chosen) functions α_{ij} and μ_i .

Lemma 2.3. *The bilinear form $a : H^1(\Omega) \times H^1(\Omega) \rightarrow \mathbb{R}$ does not depend on α_{ij}, μ_i . In particular, the weak solution of (1.1) is independent of those functions.*

We give two different proofs of this lemma. In the first one, we show directly that $a(u, v)$ for $u \in H^2(\Omega)$ is independent of α_{ij}, μ_i , whereas in the second one, we use the independence of the weak solutions in the regular case.

First proof of Lemma 2.3. We will show that $a(y, v)$ is independent of α_{ij} and μ_i for $y \in H^2(\Omega)$ and $v \in H^1(\Omega)$. The density of $H^2(\Omega)$ in $H^1(\Omega)$ yields the claim. We consider the terms involving α_{ij} and μ_i separately. We have

$$\begin{aligned} & \int_{\Omega} (\alpha_{ij} - \alpha_{ji}) \partial_i y \partial_j v + \partial_j (\alpha_{ij} - \alpha_{ji}) \partial_i y v \, dx \\ &= \int_{\Omega} \partial_j [(\alpha_{ij} - \alpha_{ji}) v] \partial_i y \, dx \\ &= - \int_{\Omega} [(\alpha_{ij} - \alpha_{ji}) v] \partial_j \partial_i y \, dx + \int_{\Gamma} (\alpha_{ij} - \alpha_{ji}) v \partial_i y \nu_j \, ds \\ &= 0 + \int_{\Gamma} \tau_i v \partial_i y \, ds \end{aligned}$$

In the last line, we used symmetry of the Hessian and (2.5). The last expression is independent of α_{ij} .

Now, we consider the terms in $a(y, v)$ depending on μ_i . We have

$$\begin{aligned} \int_{\Omega} \mu_j y \partial_j v + \mu_i \partial_i y v + \partial_j \mu_j y v \, dx &= \int_{\Omega} \partial_j (\mu_j y v) \, dx \\ &= \int_{\Gamma} \mu_j \nu_j y v \, ds = \int_{\Gamma} \theta b_0 y v \, ds. \end{aligned}$$

This expression is independent of μ_i .

This shows that $a(u, v)$ is independent of α_{ij} and μ_i . \square

Second proof of Lemma 2.3. We already know that if $f \in L^2(\Omega)$ and $g \in H^{1/2}(\Gamma)$, the weak solution $y \in H^1(\Omega)$ belongs even to $H^2(\Omega)$ and is therefore independent of α_{ij}, μ_i by Theorem 2.1. Since the mapping $(f, g) \mapsto y(f, g)$ is continuous by Theorem 2.2 and since $L^2(\Omega)$ and $H^{1/2}(\Gamma)$ are dense in $(H^1(\Omega))'$ and $(H^{1/2}(\Gamma))'$, the weak solution of (1.1) is independent of the chosen functions α_{ij}, μ_i . Hence, also the bilinear form a is independent of α_{ij} and μ_i . \square

Remark 2.4. *We remark that the requirement $a_{ij} = a_{ji}$ can be dropped. The bilinear form would then take the form*

$$\begin{aligned} a(y, v) = \int_{\Omega} & \left[\left(\frac{1}{2}(a_{ij} + a_{ji}) + \alpha_{ij} - \alpha_{ji} \right) \partial_i y + \mu_j y \right] \partial_j v \\ & + \left(a_i + \partial_j \left(\frac{1}{2}(a_{ij} - a_{ji}) + \alpha_{ij} - \alpha_{ji} \right) + \mu_i \right) \partial_i y v \\ & + (a_0 + \partial_j \mu_j) y v \, dx. \end{aligned}$$

The proof of the existence theorem 2.2 in Troianiello [1987] does not rely on the symmetry of the bilinear form. The H^2 -regularity of solutions, Theorem 2.1, as proven in Troianiello [1987] needs to be modified to accommodate for unsymmetric coefficients.

3 Coercivity of the bilinear form

In this section, we study the coercivity of the bilinear form a , which was introduced in (2.3).

It is known from Gårding's inequality, see also [Troianiello, 1987, Section 2.2.1], that

$$a(v, v) \geq C_1 \|v\|_{H^1(\Omega)}^2 - C_2 \|v\|_{L^2(\Omega)}^2 \quad \text{for all } v \in H^1(\Omega),$$

with $C_1 > 0$ and $C_2 \in \mathbb{R}$ is satisfied.

In this section, we will estimate the constant C_2 . In particular, we will study which terms in the bilinear form a contribute to C_2 . As a by-product, we give conditions which allow the choice $C_2 = 0$, i.e., under which a is coercive in $H^1(\Omega)$.

In order to use an integration by parts formula on the boundary, we assume that Ω possesses a C^2 boundary.

By definition of a , see (2.3), we have

$$\begin{aligned}
a(v, v) &= \int_{\Omega} a_{ij} \partial_i y \partial_j y + a_0 y^2 \, dx \\
&\quad + \int_{\Omega} (\alpha_{ij} - \alpha_{ji}) \partial_i y \partial_j y + \partial_j (\alpha_{ij} - \alpha_{ji}) \partial_i y y \, dx \\
&\quad + \int_{\Omega} \mu_j y \partial_j y + \mu_i \partial_i y y + \partial_j \mu_j y^2 \, dx \\
&\quad + \int_{\Omega} a_i \partial_i y y \, dx.
\end{aligned} \tag{3.1}$$

Let us rewrite the second and third line of the right-hand side of (3.1). By symmetry, the first term on the second line is zero. Let us assume $y \in C^\infty(\bar{\Omega})$ in order to rewrite

$$\begin{aligned}
&\int_{\Omega} \partial_j (\alpha_{ij} - \alpha_{ji}) \partial_i y y \, dx \\
&= - \int_{\Omega} (\alpha_{ij} - \alpha_{ji}) \partial_j (\partial_i y y) \, dx + \int_{\Gamma} (\alpha_{ij} - \alpha_{ji}) \nu_j \partial_i y y \, ds \\
&= 0 + \frac{1}{2} \int_{\Gamma} \tau_i \partial_i (y^2) \, ds = \frac{1}{2} \int_{\Gamma} \tau \nabla_{\Gamma} (y^2) \, ds \\
&= -\frac{1}{2} \int_{\Gamma} \operatorname{div}_{\Gamma}(\tau) y^2 \, ds,
\end{aligned}$$

where ∇_{Γ} , $\operatorname{div}_{\Gamma}(\tau)$ are the surface gradient and divergence of τ , see [Delfour and Zolésio, 2001, Def. 9.5.1, (9.5.6)]. Here, we used the integration-by-parts formula [Delfour and Zolésio, 2001, (9.5.27)] (and, therein, $\tau_i \nu_i = 0$). Note that this formula actually requires $\tau \in C^1(\Gamma)$, but this can be relaxed by a density argument. Using the density of $C^\infty(\bar{\Omega})$ in $H^1(\Omega)$, see [Delfour and Zolésio, 2001, Thm. 2.6.3] or [Attouch et al., 2006, Prop. 5.4.1], and using $\operatorname{div}_{\Gamma}(\tau) \in L^\infty(\Gamma)$, we find that

$$\int_{\Omega} \partial_j (\alpha_{ij} - \alpha_{ji}) \partial_i y y \, dx = -\frac{1}{2} \int_{\Gamma} \operatorname{div}_{\Gamma}(\tau) y^2 \, ds$$

holds for all $y \in H^1(\Omega)$.

It remains to study the third line in (3.1). We have

$$\begin{aligned}
\int_{\Omega} \mu_j y \partial_j y + \mu_i \partial_i y y + \partial_j \mu_j y^2 \, dx &= \int_{\Omega} \partial_j (\mu_j y^2) \, dx \\
&= \int_{\Gamma} \nu_j \mu_j y^2 \, ds = \int_{\Gamma} \theta b_0 y^2 \, ds.
\end{aligned}$$

Altogether, we obtain

$$\begin{aligned} a(v, v) &= \int_{\Omega} (a_{ij} \partial_i y \partial_j y + a_0 y^2) dx \\ &\quad + \int_{\Gamma} (\theta b_0 - \frac{1}{2} \operatorname{div}_{\Gamma}(\tau)) y^2 ds + \int_{\Omega} a_i \partial_i y y dx. \end{aligned} \quad (3.2)$$

Note that the last term comes from the convection term $a_i \partial_i y$ in the PDE (1.1). If we neglect this term then the bilinear form a can only be not coercive if $\theta b_0 - \frac{1}{2} \operatorname{div}_{\Gamma} \tau < 0$ holds. This is only possible if τ is not constant, i.e., the angle between the normal vector ν_i and the oblique vector b_i is not constant!

Note that the condition

$$\theta b_0 - \frac{1}{2} \operatorname{div}_{\Gamma}(\tau) \geq \kappa > 0 \quad \text{on } \Gamma$$

is used sometimes in the literature to prove existence of weak solutions, see, e.g., [Raskop and Grothaus, 2006, Thm. 3.7]. However, this condition is not necessary for existence and uniqueness, see Theorem 2.1, Theorem 2.2 and the example in Section 5.3.

4 The optimal control problem

Let us now turn to analyzing the optimal control problem. It is given as: minimize the functional

$$J(y, u) := j(y) + \frac{\alpha}{2} \|u\|_{L^2(\Gamma)}^2 \quad (4.1)$$

over all pairs $(y, u) \in H^1(\Omega) \times L^2(\Gamma)$ satisfying the weak formulation

$$a(y, v) = \int_{\Gamma} u \theta v ds \quad \text{for all } v \in H^1(\Omega) \quad (4.2)$$

of the PDE (1.1) and the control constraint

$$u \in U_{ad} := \{v \in L^2(\Gamma) : u_a(x) \leq v(x) \leq u_b(x) \text{ f.a.a. } x \in \Gamma\}. \quad (4.3)$$

Here, $j : H^1(\Omega) \rightarrow \mathbb{R}$ is a given Fréchet differentiable function, $\alpha > 0$, and $u_a, u_b \in L^2(\Gamma)$ satisfy $u_a(x) \leq u_b(x)$ for almost all $x \in \Gamma$.

Theorem 4.1. *The optimal control problem (4.1)–(4.3) admits solutions.*

Proof. Let us denote the feasible set for the problem (4.1)–(4.3) by F , i.e.

$$F := \{(y, u) \in H^1(\Omega) \times L^2(\Gamma) : u \in U_{ad}, (y, u) \text{ satisfy (4.2)}\}.$$

By assumption, the set U_{ad} is non-empty. Moreover, for each control $u \in L^2(\Gamma)$ the weak formulation (2.6) is uniquely solvable for $y \in H^1(\Omega)$. Hence, the set of feasible points F of the optimal control problem is not empty.

In addition, the set U_{ad} is compact with respect to the weak topology of $L^2(\Gamma)$. Let us argue that the set of associated states y is compact in the norm topology of $H^1(\Omega)$. The linear mapping $u \mapsto y$, where y solves (4.2), is linear and continuous from $H^{-1/2}(\Gamma)$ to $H^1(\Omega)$, hence compact from $L^2(\Gamma)$ to $H^1(\Omega)$, cf. [Troianiello, 1987, Lemma 1.51]. This proves that the set of states solving (4.2) with $u \in U_{ad}$ is compact in $H^1(\Omega)$. Thus, the feasible set F is compact in $H^1(\Omega) \times L^2(\Gamma)$ with the norm topology and weak topology, respectively.

The function J is continuous with respect to the first argument, lower semicontinuous with respect to the second argument in the mentioned topologies. Now the existence of optimal controls and states follows from the Weierstraß theorem. \square

Let us now turn to necessary optimality conditions.

Theorem 4.2. *Let (\bar{y}, \bar{u}) be a local solution of (4.1)–(4.3). Then there exists $\bar{p} \in H^1(\Omega)$ such that*

$$a(v, \bar{p}) = j'(\bar{y})v \quad \forall v \in H^1(\Omega) \tag{4.4}$$

and

$$(\alpha \bar{u} + \theta \bar{p}, u - \bar{u})_{L^2(\Gamma)} \geq 0 \quad \forall u \in U_{ad},$$

where θ is given by (2.4).

Proof. Let us denote by $S : (H^1(\Omega))' \rightarrow H^1(\Omega)$ the linear mapping $F \mapsto S(F)$, where $S(F)$ solves $a(S(f), v) = \langle f, v \rangle_{(H^1(\Omega))', H^1(\Omega)}$. According to Theorem 2.2, S is continuous. Let us denote by $S^* : (H^1(\Omega))' \rightarrow H^1(\Omega)$ its adjoint operator. Let now $\phi, F \in (H^1(\Omega))'$ be given, and set $p := S^*\phi \in H^1(\Omega)$. Then it holds

$$\begin{aligned} a(SF, p) &= \langle F, p \rangle_{(H^1(\Omega))', H^1(\Omega)} = \langle F, S^*\phi \rangle_{(H^1(\Omega))', H^1(\Omega)} \\ &= \langle SF, \phi \rangle_{(H^1(\Omega))', H^1(\Omega)}. \end{aligned}$$

Since $F \in (H^1(\Omega))'$ is arbitrary, and S is surjective, it follows that

$$a(v, p) = \langle v, \phi \rangle_{(H^1(\Omega))', H^1(\Omega)} \quad \text{for all } v \in H^1(\Omega).$$

Defining $\bar{p} := S^* j'(\bar{y})$, the proof follows from standard arguments, see, e.g., [Tröltzsch, 2010, Sect. 2.8]. \square

Let us now investigate the adjoint equation (4.4). As the bilinear form a is not symmetric, the strong formulation of (4.4) will differ in general from (1.1), which is the strong formulation of the state equation (4.2).

In order to establish the strong formulation, we first prove $H^2(\Omega)$ -regularity of the adjoint state \bar{p} . We cannot conclude this regularity of \bar{p} without further assumptions on the coefficients of the differential operator, which is due to the fact that the role of test function and solution is switched when compared to the state equation.

Theorem 4.3. *Let us assume that Ω has $C^{2,1}$ -boundary, and the coefficient functions satisfy $a_{ij} \in C^{1,1}(\bar{\Omega})$, $a_i \in C^{0,1}(\Omega)$ and $b_i \in C^{1,1}(\Gamma)$.*

Let $j'(\bar{y}) \in L^2(\Omega)$ and $\bar{p} \in H^1(\Omega)$ solve (4.4). Then $\bar{p} \in H^2(\Omega)$.

Proof. The weak formulation of the adjoint equation (4.4) reads

$$\begin{aligned} a(v, p) = \int_{\Omega} [(a_{ij} + \alpha_{ij} - \alpha_{ji}) \partial_i v + \mu_j v] \partial_j p + (a_i + \partial_j(\alpha_{ij} - \alpha_{ji}) + \mu_i) \partial_i v p \\ + (a_0 + \partial_j \mu_j) v p \, dx = j'(\bar{y}) v \quad \forall v \in H^1(\Omega). \end{aligned} \quad (4.5)$$

Due to the increased smoothness of the coefficients, the coefficients in the weak formulation can be constructed to satisfy $\alpha_{ij} \in C^{1,1}(\bar{\Omega})$: The function θ defined in (2.4) satisfies $\theta \in C^{1,1}(\Gamma)$, which implies $\tau_i \in C^{1,1}(\Gamma)$. Then $\alpha_{ij} \in C^{1,1}(\bar{\Omega})$ can be chosen as an extension of $\nu_j \tau_i \in C^{1,1}(\Gamma)$, see [Troianiello, 1987, Thm. 1.3].

Hence, the coefficients in the weak formulation (4.5) satisfy the assumptions of [Troianiello, 1987, Theorem 3.17 (ii)], in particular $a_i + \partial_j(\alpha_{ij} - \alpha_{ji}) + \mu_i \in C^{0,1}(\bar{\Omega})$, which gives the regularity $\bar{p} \in H^2(\Omega)$. \square

With the help of this regularity result, we can prove that the adjoint state \bar{p} is the strong solution of a boundary value problem with non-variational boundary conditions. Here again, the regularity of coefficients of the differential operator is essential.

Theorem 4.4. *Let the assumptions of Theorem 4.3 be satisfied. Then $\bar{p} \in H^2(\Omega)$ satisfies*

$$-\partial_i(a_{ij} \partial_j p) - \partial_i(a_i p) + a_0 p = j'(\bar{y}) \quad \text{in } \Omega, \quad (4.6a)$$

$$(2 \nu_i a_{ij} - \theta b_j) \partial_j p + (a_i \nu_i + \theta b_0 + \operatorname{div}_{\Gamma}(\theta b - a \nu)) p = 0 \quad \text{on } \Gamma. \quad (4.6b)$$

Here, $\theta b - a \nu$ refers to the vector field with components $\theta b_j - a_{ij} \nu_j$.

Proof. By assumption, it holds $a(v, \bar{p}) = j'(\bar{y})v$ for all $v \in H^1(\Omega)$. Using integrating by parts in (4.5) in the terms involving derivatives of the test function v , we obtain

$$\begin{aligned} a(v, p) = & \int_{\Omega} -\partial_i [(a_{ij} + \alpha_{ij} - \alpha_{ji}) \partial_j p] v - \partial_i [(a_i + \partial_j(\alpha_{ij} - \alpha_{ji})) p] v + a_0 p v \, dx \\ & + \int_{\Gamma} (a_{ij} + \alpha_{ij} - \alpha_{ji}) \nu_i \partial_j p v + (a_i + \partial_j(\alpha_{ij} - \alpha_{ji}) + \mu_i) \nu_i p v \, ds, \end{aligned} \quad (4.7)$$

where we used $\mu_j \partial_j p - \partial_i(\mu_i p) + \partial_j \mu_j p = 0$. Differentiating the terms involving α_{ij} we find

$$\partial_i [(\alpha_{ij} - \alpha_{ji}) \partial_j p] = [\partial_i(\alpha_{ij} - \alpha_{ji}) \partial_j p] + (\alpha_{ij} - \alpha_{ji}) \partial_i \partial_j p = [\partial_i(\alpha_{ij} - \alpha_{ji}) \partial_j p]$$

and

$$\begin{aligned} \partial_i [(\partial_j(\alpha_{ij} - \alpha_{ji})) p] &= (\partial_i \partial_j(\alpha_{ij} - \alpha_{ji})) p + (\partial_j(\alpha_{ij} - \alpha_{ji})) \partial_i p \\ &= -[\partial_i(\alpha_{ij} - \alpha_{ji}) \partial_j p]. \end{aligned}$$

Hence, the domain integral in (4.7) becomes

$$\begin{aligned} \int_{\Omega} -\partial_i [(a_{ij} + \alpha_{ij} - \alpha_{ji}) \partial_j p] v - \partial_i [(a_i + \partial_j(\alpha_{ij} - \alpha_{ji})) p] v + a_0 p v \, dx \\ = \int_{\Omega} -\partial_i [a_{ij} \partial_j p] v - \partial_i [a_i p] v + a_0 p v \, dx. \end{aligned}$$

Since $v \in H^1(\Omega)$ was arbitrary, this shows (4.6a).

Employing the relations (2.1) we find

$$\nu_i (\alpha_{ij} - \alpha_{ji}) = \nu_i a_{ji} - \theta b_j = \nu_i a_{ij} - \theta b_j$$

and we can transform the boundary integral in (4.7) to

$$\begin{aligned} \int_{\Gamma} (a_{ij} + \alpha_{ij} - \alpha_{ji}) \nu_i \partial_j p v + (a_i + \partial_j(\alpha_{ij} - \alpha_{ji}) + \mu_i) \nu_i p v \, ds \\ = \int_{\Gamma} (2\nu_i a_{ij} - \theta b_j) \partial_j p v + ([a_i + \partial_j(\alpha_{ij} - \alpha_{ji})] \nu_i + \theta b_0) p v \, ds. \end{aligned}$$

Here, only the term $\partial_j(\alpha_{ij} - \alpha_{ji}) \nu_i$ depends on the parameterization. Reverting to vector notation, we obtain

$$\begin{aligned} \nu_i \partial_j(\alpha_{ij} - \alpha_{ji}) &= \nu^\top \operatorname{div}(\alpha - \alpha^\top) \\ &= \nu^\top \operatorname{div}_\Gamma(\alpha - \alpha^\top) + \nu_i \partial_l(\alpha_{ij} - \alpha_{ji}) \nu_l \nu_j \\ &= \nu^\top \operatorname{div}_\Gamma(\alpha - \alpha^\top). \end{aligned}$$

We continue with

$$\begin{aligned} \nu^\top \operatorname{div}_\Gamma(\alpha - \alpha^\top) &= \operatorname{div}_\Gamma(\nu^\top(\alpha - \alpha^\top)) - \nabla_\Gamma \nu : (\alpha - \alpha^\top) \\ &= \operatorname{div}_\Gamma \tau = \operatorname{div}_\Gamma(\theta b - a\nu) \end{aligned}$$

where we have used $\nabla_\Gamma \nu = (\nabla_\Gamma \nu)^\top$, see [Delfour and Zolésio, 2001, eq. (5.10)], and relation (2.1). Collecting these results we have the following transformation of the boundary integrals

$$\begin{aligned} &\int_\Gamma (a_{ij} + \alpha_{ij} - \alpha_{ji}) \nu_i \partial_j p v + (a_i + \partial_j(\alpha_{ij} - \alpha_{ji}) + \mu_i) \nu_i p v \, ds \\ &= \int_\Gamma (2\nu_i a_{ij} - \theta b_j) \partial_j p v + ((a_i + \partial_j(\alpha_{ij} - \alpha_{ji})) \nu_i + \theta b_0) p v \, ds \\ &= \int_\Gamma (2\nu_i a_{ij} - \theta b_j) \partial_j p v + (a_i \nu_i + \theta b_0 + \operatorname{div}_\Gamma(\theta b - a\nu)) p v \, ds. \end{aligned}$$

Since $v \in H^1(\Omega)$ was arbitrary, the claim follows. \square

Example 4.5. *Let us discuss the strong formulation of the adjoint equation for the following simple optimal control problem: minimize*

$$J(y, u) = \frac{1}{2} \|y - y_d\|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \|u\|_{L^2(\Gamma)}^2$$

subject to

$$\begin{aligned} -\Delta y + y &= 0 \quad \text{in } \Omega, \\ \nabla y^\top \cdot (\nu + \tau) &= u \quad \text{on } \Gamma, \end{aligned}$$

where τ is a tangential vector field, $\tau^\top \nu = 0$. In the notation as above, we have

$$b = \nu + \tau, \quad b_0 = 0, \quad \theta = 1, \quad \theta b - a\nu = \tau.$$

Hence the adjoint equation is given by

$$\begin{aligned} -\Delta p + p &= y - y_d \quad \text{in } \Omega, \\ \nabla p^\top \cdot (\nu - \tau) + \operatorname{div}_\Gamma(\tau) \cdot p &= 0 \quad \text{on } \Gamma. \end{aligned}$$

5 Numerics

In this section, we will discuss the numerical solution of the PDE (1.1) and the associated optimal control problem, see Section 4.

5.1 A specific setting

Throughout this section we will study numerical aspects for one particular instance of (1.1) and the associated control problem. Let $\Omega = B_1(0) \subset \mathbb{R}^2$ be the unit circle. The outer unit normal vector ν and the (left) unit tangential vector t are given by

$$\nu(x) = (x_1, x_2)^\top, \quad t(x) = (-x_2, x_1)^\top,$$

respectively. Note that both vectors can be extended to smooth functions on \mathbb{R}^2 . We consider the PDE

$$-\Delta y + y = f \quad \text{in } \Omega, \quad (5.1a)$$

$$\nabla y^\top \cdot [\nu + (c_1 + c_2 x_1) t] = g \quad \text{on } \Gamma, \quad (5.1b)$$

where $c_1, c_2 \in \mathbb{R}$ are constant parameters. In the notation of (1.1), we have

$$\begin{aligned} a_{ij} &= \delta_{ij}, & a_i &= 0, & a_0 &= 1, \\ b_i(x_1, x_2) &= \nu_i(x_1, x_2) + (c_1 + c_2 x_1) t_i, & b_0 &= 0. \end{aligned}$$

5.2 The discrete forward problem

The choice

$$\begin{aligned} \theta &= 1, & \mu_i &= 0, \\ \tau_i(x_1, x_2) &= (c_1 + c_2 x_1) t_i(x_1, x_2), & \alpha_{ij} &= \tau_i \nu_j, \end{aligned}$$

satisfies (2.1). In vector and matrix notation, we have

$$\left. \begin{aligned} \tau(x) &= (-(c_1 + c_2 x_1) x_2, (c_1 + c_2 x_1) x_1)^\top \\ \alpha(x) &= \tau(x) \nu(x)^\top = (c_1 + c_2 x_1) \begin{pmatrix} -x_2 x_1 & -x_2^2 \\ x_1^2 & x_1 x_2 \end{pmatrix} \end{aligned} \right\} \quad (5.2)$$

The weak formulation (2.6) is solved by linear finite elements on a triangular mesh. Note that the discrete domain Ω_h is strictly included in Ω .

The discrete solution y for the right-hand side

$$f(x) = \exp\left(- (x_1 - 1/2)^2 - x_2^2\right), \quad g(x) = 0$$

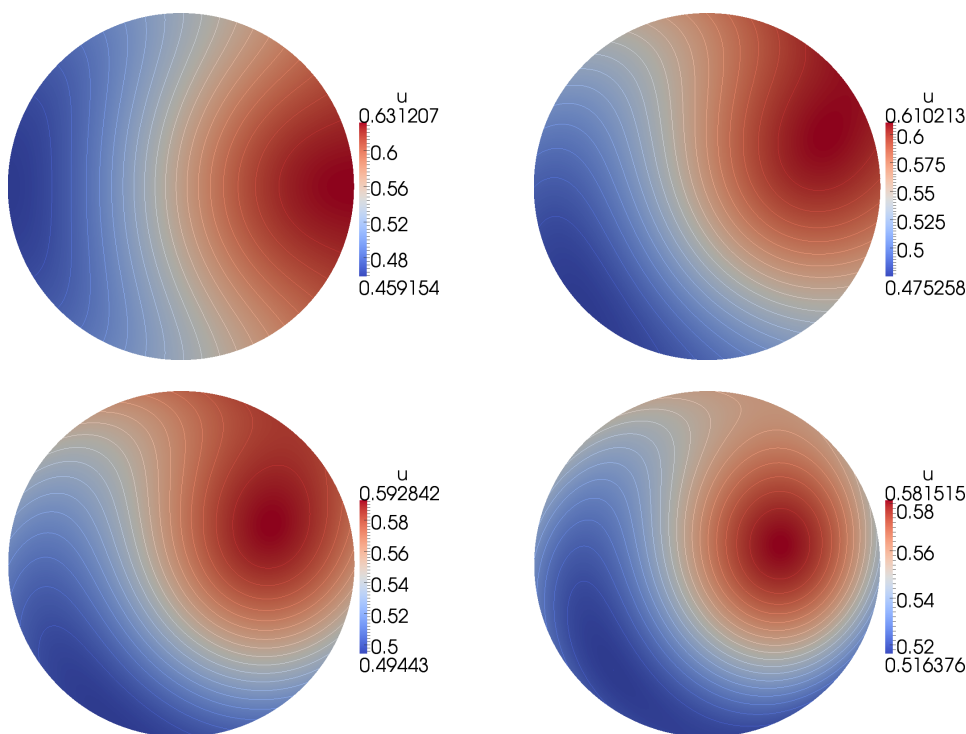


Figure 1: Solution of the problem for $c_1 \in \{0, 1, 2, 5\}$ and $c_2 = 0$ for the choice (5.2).

with parameterization (5.2) is plotted in Figure 1.

Alternatively, one may choose

$$\left. \begin{aligned} \theta &= 1, & \mu &= 0, \\ \tau(x_1, x_2) &= (c_1 + c_2 x_1) t(x_1, x_2), & \alpha(x_1, x_2) &= \begin{pmatrix} 0 & -(c_1 + c_2 x_1) \\ 0 & 0 \end{pmatrix}, \end{aligned} \right\} \quad (5.3)$$

which also satisfies (2.1). The solution with right-hand side f as above and parameterization (5.3) is plotted in Figure 2. Note that both solutions are slightly different in the discrete setting. In the continuous setting, both approaches are equivalent and their solutions coincide, see Lemma 2.3. However, the proof of Lemma 2.3 requires integration by parts and that (2.1) is satisfied on $\Gamma = \partial\Omega$. In the discrete setting, a similar proof would require that (2.1) is satisfied on $\partial\Omega_h$. This does not hold for the choices (5.2) and (5.3). Moreover, it is, in general, not possible to construct Lipschitz con-

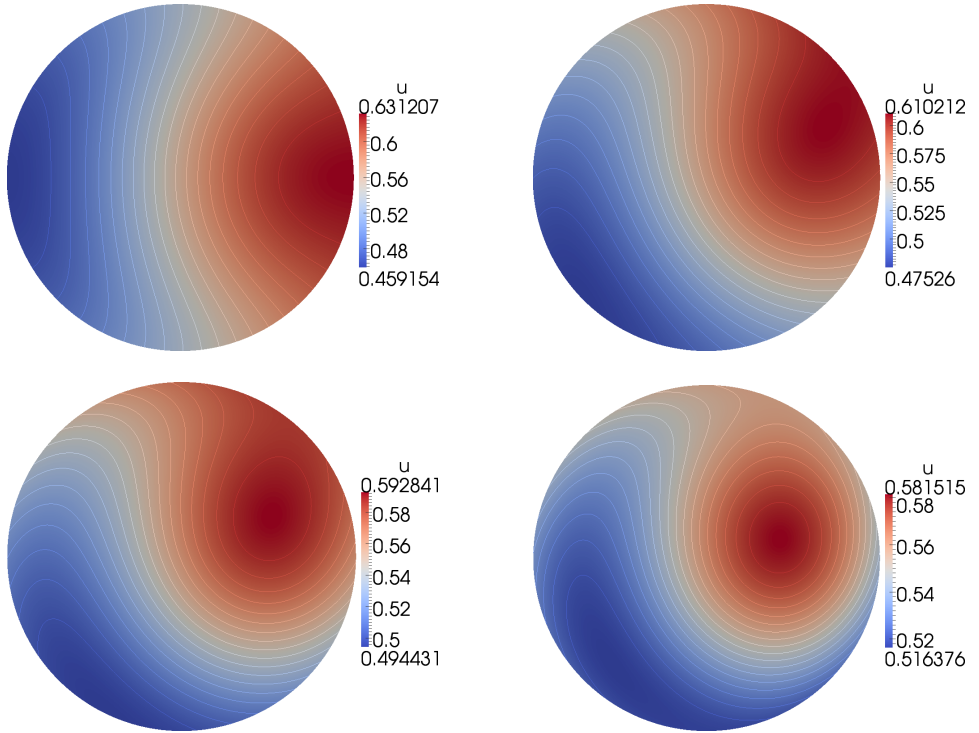


Figure 2: Solution of the problem for $c_1 \in \{0, 1, 2, 5\}$ and $c_2 = 0$ for the choice (5.3).

tinuous α_{ij} , μ_j , θ such that (2.1) holds on $\partial\Omega_h$, since $\partial\Omega_h$ is only Lipschitz continuous and its normal vector is discontinuous.

We remark that the convergence of the discretization can be proved by standard arguments, see Schatz [1974].

5.3 Coercivity of the bilinear form

In this section, we will study the coercivity of the bilinear form associated to the PDE (5.1). By (3.2) we have

$$a(y, y) = \int_{\Omega} |\nabla y|^2 + y^2 dx - \frac{1}{2} \int_{\Gamma} \operatorname{div}_{\Gamma}(\tau) y^2 ds.$$

By using [Delfour and Zolésio, 2001, (9.5.6)], we find

$$\operatorname{div}_{\Gamma}(\tau) = \operatorname{div}(\tau) - \nu^{\top} \tau' \nu = -c_2 x_2$$

where τ' is the Jacobian of τ . We have

$$\int_{\Gamma} x_2 y^2 ds = \int_{\Omega} \nabla(y^2) \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} dx = 2 \int_{\Omega} y \nabla y \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} dx \leq \int_{\Omega} |\nabla y|^2 + y^2 dx.$$

Note that this estimate is sharp since equality holds for the choice $y = \exp(x_2)$. Analogously, we have

$$\int_{\Gamma} x_2 y^2 ds \geq - \int_{\Omega} |\nabla y|^2 + y^2 dx.$$

Again, this estimate is sharp (set $y = \exp(-x_2)$). Altogether, we have

$$a(y, y) \geq \|y\|_{H^1}^2 + \frac{c_2}{2} \int_{\Gamma} x_2 y^2 ds \geq \left(1 - \frac{|c_2|}{2}\right) \|y\|_{H^1}^2.$$

This estimate is sharp (set $y = \exp(\pm x_2)$). Hence, we find that bilinear form is coercive if and only if $|c_2| < 2$. In the case $|c_2| \geq 2$, the bilinear form is no longer coercive. However, the unique solvability of the weak formulation (2.6) follows from Theorem 2.2.

Let us denote by A the stiffness matrix associated to the bilinear form a and by K the matrix associated with the inner product of $H^1(\Omega)$. Then, the bilinear form a is coercive on the discrete subspace $V_h \subset H^1(\Omega)$ if and only if the smallest eigenvalue of the symmetric part $(A + A^T)/2$ of A is positive w.r.t. K . Numerically, this smallest eigenvalue behaves like $1 - |c_2|/2$. Hence, the above analysis is confirmed by the numerical experiments.

5.4 Discrete optimal control problem

In this section, we consider the discretized optimal control problem. The state y and the adjoint state p are discretized by piecewise linear finite elements on Ω_h , whereas the boundary control u is discretized by piecewise linear finite elements on $\partial\Omega_h$. The associated spaces are denoted by V_h and U_h , respectively. We denote discrete functions and their coefficient vectors by the same symbol.

We denote by A the stiffness matrix associated with the bilinear form a and by M, M_{Γ} the mass matrices associated with the inner products of $L^2(\Omega_h), L^2(\partial\Omega_h)$, respectively. Moreover, $M_{\Omega, \Gamma}$ is the (rectangular) matrix associated with the bilinear form

$$\int_{\Gamma} \theta v u ds \quad v \in V_h, u \in U_h.$$

The discrete optimal control problem is given by

$$\begin{aligned} & \text{Minimize} && \frac{1}{2} (y - y_d)^\top M (y - y_d) + \frac{\alpha}{2} u^\top M_\Gamma u, \\ & \text{such that} && Ay = M_{\Omega, \Gamma} u \\ & && \text{and } u_a \leq u \leq u_b. \end{aligned}$$

Here, $u_a, u_b \in U_h$ are discrete variants of $u_a, u_b \in L^2(\Gamma)$, e.g., their projections. By standard calculations, the optimality system is given by

$$\begin{aligned} A^\top \bar{p} + M(\bar{y} - y_d) &= 0, \\ [\alpha M_\Gamma \bar{u} - M_{\Omega, \Gamma}^\top \bar{p}]^\top (u - \bar{u}) &\geq 0 \quad \text{for all } u_a \leq u \leq u_b, \\ A\bar{y} - M_{\Omega, \Gamma} \bar{u} &= 0. \end{aligned}$$

The variational inequality can be rewritten as

$$\bar{u} - \text{proj}_{[u_a, u_b]} \left[\bar{u} - M_\Gamma \bar{u} + \frac{1}{\alpha} M_{\Omega, \Gamma}^\top \bar{p} \right] = 0.$$

Here, the projection is to be understood coefficient-wise. Let us introduce the nonlinear function

$$F(y, u, p) = \begin{pmatrix} A^\top \bar{p} + M(\bar{y} - y_d) \\ \bar{u} - \text{proj}_{[u_a, u_b]} \left[\bar{u} - M_\Gamma \bar{u} + \frac{1}{\alpha} M_{\Omega, \Gamma}^\top \bar{p} \right] \\ A\bar{y} - M_{\Omega, \Gamma} \bar{u} \end{pmatrix}.$$

The optimality system can be written as $F(\bar{y}, \bar{u}, \bar{p}) = 0$. The mapping F is Newton differentiable. Here, a mapping $f : X \rightarrow Y$ is called Newton differentiable if there exists a mapping $f' : X \rightarrow \mathcal{L}(X, Y)$ such that

$$\lim_{\|h\|_X \rightarrow 0} \frac{1}{\|h\|_X} \|f(x+h) - f(x) - f'(x)h\|_Y = 0.$$

A generalized Jacobian (in the sense of Newton differentiability) of F is

$$F'(y, u, p) = \begin{pmatrix} M & 0 & A^\top \\ 0 & I - I_A(I - M_\Gamma) & -I_A M_{\Omega, \Gamma}^\top / \alpha \\ A & -M_{\Omega, \Gamma} & 0 \end{pmatrix}.$$

Here, the components of the diagonal matrix I_A are 1 if the components of $\bar{u} - M_\Gamma \bar{u} + \frac{1}{\alpha} M_{\Omega, \Gamma}^\top \bar{p}$ are between u_a and u_b , and 0 otherwise.

Now, we use a generalized Newton method to solve $F(\bar{y}, \bar{u}, \bar{p}) = 0$. The Newton system formulated in the next iterates $(y_{k+1}, u_{k+1}, p_{k+1})$ reads

$$\begin{pmatrix} M & 0 & A^\top \\ 0 & I - I_A(I - M_\Gamma) & -I_A M_{\Omega, \Gamma}^\top / \alpha \\ A & -M_{\Omega, \Gamma} & 0 \end{pmatrix} \begin{pmatrix} y_{k+1} \\ u_{k+1} \\ p_{k+1} \end{pmatrix} = \begin{pmatrix} M y_d \\ I_b u_b + I_a u_a \\ 0 \end{pmatrix}.$$

Here, the components of the diagonal matrix I_b (I_a) are 1 if the components of $\bar{u} - M_\Gamma \bar{u} + \frac{1}{\alpha} M_{\Omega, \Gamma}^\top \bar{p}$ are bigger than u_b (smaller than u_a) and 0 otherwise.

Unfortunately, the Newton system is not symmetric. However, it is possible to modify this system, such that it becomes symmetric.

For convenience, let us denote $\tilde{u} = I_b u_b + I_a u_a$. By the second equation, we immediately find $(I - I_A) u_{k+1} = \tilde{u}$. This can be employed in the second and third row of our system, and we obtain

$$\begin{pmatrix} M & 0 & A^\top \\ 0 & I - I_A + I_A M_\Gamma I_A & -I_A M_{\Omega, \Gamma}^\top / \alpha \\ A & -M_{\Omega, \Gamma} I_A & 0 \end{pmatrix} \begin{pmatrix} y_{k+1} \\ u_{k+1} \\ p_{k+1} \end{pmatrix} = \begin{pmatrix} M y_d \\ \tilde{u} - I_A M_\Gamma \tilde{u} \\ M_{\Omega, \Gamma} \tilde{u} \end{pmatrix}.$$

Now, we rescale the components of $I - I_A$ in the second row by the diagonal matrix M_Γ^L , which is the lumped version of M_Γ and obtain

$$\begin{pmatrix} M & 0 & A^\top \\ 0 & M_\Gamma^L (I - I_A) + I_A M_\Gamma I_A & -I_A M_{\Omega, \Gamma}^\top / \alpha \\ A & -M_{\Omega, \Gamma} I_A & 0 \end{pmatrix} \begin{pmatrix} y_{k+1} \\ u_{k+1} \\ p_{k+1} \end{pmatrix} = \begin{pmatrix} M y_d \\ M_\Gamma^L \tilde{u} - I_A M_\Gamma \tilde{u} \\ M_{\Omega, \Gamma} \tilde{u} \end{pmatrix}.$$

Finally, we scale the second row by α ,

$$\begin{pmatrix} M & 0 & A^\top \\ 0 & \alpha M_\Gamma^L (I - I_A) + \alpha I_A M_\Gamma I_A & -I_A M_{\Omega, \Gamma}^\top \\ A & -M_{\Omega, \Gamma} I_A & 0 \end{pmatrix} \begin{pmatrix} y_{k+1} \\ u_{k+1} \\ p_{k+1} \end{pmatrix} = \begin{pmatrix} M y_d \\ \alpha [M_\Gamma^L \tilde{u} - I_A M_\Gamma \tilde{u}] \\ M_{\Omega, \Gamma} \tilde{u} \end{pmatrix}.$$

Note that this matrix is symmetric. This system is solved by a preconditioned MINRES. We solve each linear system up to an absolute tolerance

$n \setminus (c_1, c_2) =$	(0, 0)	(1, 0)	(2, 0)	(3, 0)
0	2 / 15.5	4 / 21	2 / 23.5	2 / 27
1	3 / 73	3 / 84.3	3 / 92.3	3 / 101
2	3 / 138.67	4 / 179.5	4 / 211	4 / 230
3	4 / 178.75	4 / 217.5	4 / 286.75	4 / 358
4	4 / 175.75	4 / 228	5 / 307.8	5 / 393.6
5	4 / 182.5	5 / 225.4	5 / 318.8	5 / 424.4
6	5 / 179.4	5 / 224.2	5 / 324.4	5 / 431.2
7	5 / 172	5 / 223.6	5 / 322.4	5 / 428.4
8	4 / 183.75	5 / 224.8	6 / 304.83	5 / 429.8

Table 1: Number of newton iterations and the average MINRES iterations for different values of c_1 and mesh refinement levels n .

of 10^{-12} . The block-diagonal preconditioner is an approximation of the $H^1(\Omega_h) \times L^2(\partial\Omega_h) \times H^1(\Omega_h)$ -inner product. The inner products of $H^1(\Omega_h)$ are approximated by a geometric multigrid V-cycle. The inner product of $L^2(\partial\Omega_h)$ is approximated by solving with the lumped mass matrix M_Γ^L . We use the same tolerance of 10^{-12} for the outer Newton loop.

The matrices are assembled by the FE library FEniCS, Logg et al. [2012]. As a geometric multigrid implementation we use FMG, Ospald [2012]. We use the MINRES implementation from PETSc, Balay et al. [2013b,a, 1997], but with a modified convergence criterion, which uses the preconditioned norm of the residual (this should not be confused with the 2-norm of the preconditioned residual).

Let us report some iteration numbers for the choice

$$\begin{aligned} y_d(x_1, x_2) &= \exp(x_1) \sin(x_2), & \alpha &= 10^{-2}, \\ u_a &= -1.5, & u_b &= 1.5. \end{aligned}$$

We give the number of newton iterations and the average MINRES iterations for different values of (c_1, c_2) and mesh refinement levels n in Table 1 and Table 2. As it can be seen from those tables, the iteration numbers depend only slightly on the mesh refinement level n , whereas they depend heavily on the parameters c_1, c_2 . This is due to the fact that the preconditioner, which is the inner product of $H^1(\Omega_h)$, coincides with the bilinear form A only in the case $c_1 = c_2 = 0$.

The solution of the optimal control problem for the mesh refinement parameter $n = 6$ and $c_1 = c_2 = 1$ is shown in Figure 3.

$n \setminus (c_1, c_2) =$	(0, 0)	(0, 1)	(0, 2)	(0, 3)
0	2 / 15.5	2 / 34	2 / 39	3 / 36
1	3 / 73	3 / 80.67	3 / 105	4 / 107
2	3 / 138.67	3 / 183.67	4 / 226	4 / 263.25
3	4 / 178.75	4 / 217.75	4 / 286.75	4 / 365.75
4	4 / 175.75	5 / 219.4	5 / 306.6	4 / 416.75
5	4 / 182.5	5 / 226.6	5 / 323.4	5 / 432
6	5 / 179.4	4 / 236	5 / 333.6	5 / 441.4
7	5 / 172	5 / 225	6 / 325.67	5 / 470.8
8	4 / 183.75	5 / 221.6	5 / 343.4	5 / 476.4

Table 2: Number of newton iterations and the average MINRES iterations for different values of c_2 and mesh refinement levels n .

References

- H. Attouch, G. Buttazzo, and G. Michaille. *Variational analysis in Sobolev and BV spaces*, volume 6 of *MPS/SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006. Applications to PDEs and optimization.
- S. Balay, W. D. Gropp, L. C. McInnes, and B. F. Smith. Efficient management of parallelism in object oriented numerical software libraries. In E. Arge, A. M. Bruaset, and H. P. Langtangen, editors, *Modern Software Tools in Scientific Computing*, pages 163–202. Birkhäuser Press, 1997.
- S. Balay, J. Brown, K. Buschelman, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang. PETSc users manual. Technical Report ANL-95/11 - Revision 3.4, Argonne National Laboratory, 2013a.
- S. Balay, J. Brown, K. Buschelman, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang. PETSc Web page, 2013b. <http://www.mcs.anl.gov/petsc>.
- E. Casas and V. Dharmo. Error estimates for the numerical approximation of Neumann control problems governed by a class of quasilinear elliptic equations. *Comput. Optim. Appl.*, 52(3):719–756, 2012. doi: 10.1007/s10589-011-9440-0.
- E. Casas, M. Mateos, and F. Tröltzsch. Error estimates for the numerical

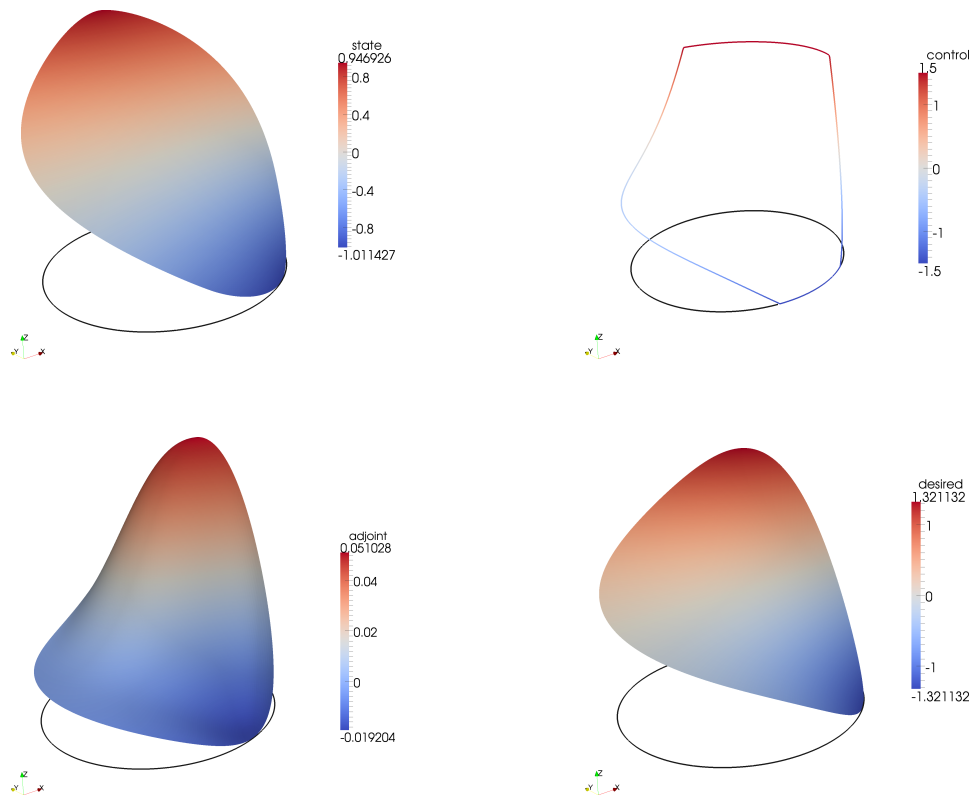


Figure 3: Solution of the optimal control problem for mesh refinement $n = 6$ and $c_1 = c_2 = 1$. The upper figures show the optimal state \bar{y} and the optimal control \bar{u} . The lower figures show the optimal adjoint state \bar{p} and the desired state y_d .

approximation of boundary semilinear elliptic control problems. *Comput. Optim. Appl.*, 31(2):193–219, 2005. doi: 10.1007/s10589-005-2180-2.

M. Delfour and J.-P. Zolésio. *Shapes and Geometries. Analysis, Differential Calculus, and Optimization*. SIAM, Philadelphia, 2001.

J. S. Dowker and K. Kirsten. Heat-kernel coefficients for oblique boundary conditions. *Classical and Quantum Gravity*, 14(9):L169–L175, 1997. ISSN 0264-9381. doi: 10.1088/0264-9381/14/9/004.

J. S. Dowker and K. Kirsten. The $a_{3/2}$ heat kernel coefficient for oblique boundary conditions. *Classical and Quantum Gravity*, 16(6):1917–1936, 1999. ISSN 0264-9381. doi: 10.1088/0264-9381/16/6/322.

- D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*, volume 224 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 1983.
- P. Grisvard. *Elliptic Problems in Nonsmooth Domains*. Pitman, Boston, 1985.
- R. Herzog, K. Kunisch, and J. Sass. Primal-dual methods for the computation of trading regions under proportional transaction costs. *Mathematical Methods of Operations Research*, 77(1):101–130, 2013. doi: 10.1007/s00186-012-0416-3.
- A. Logg, K.-A. Mardal, G. N. Wells, et al. *Automated Solution of Differential Equations by the Finite Element Method*. Springer, 2012. doi: 10.1007/978-3-642-23099-8.
- F. Ospald. Implementation of a geometric multigrid method for FEniCS and its application. Diploma thesis, Technische Universität Chemnitz, Germany, 2012.
- T. Raskop and M. Grothaus. On the oblique boundary problem with a stochastic inhomogeneity. *Stochastics. An International Journal of Probability and Stochastic Processes*, 78(4):233–257, 2006. ISSN 1744-2508. doi: 10.1080/17442500600715782.
- Y. Rozanov and F. Sansò. On the stochastic versions of Neumann and oblique derivative problems. *Stochastics and Stochastics Reports*, 74(1-2):371–391, 2002. ISSN 1045-1129. doi: 10.1080/10451120290028104.
- A. H. Schatz. An observation concerning Ritz-Galerkin methods with indefinite bilinear forms. *Mathematics of Computation*, 28:959–962, 1974.
- G. M. Troianiello. *Elliptic differential equations and obstacle problems*. The University Series in Mathematics. Plenum Press, New York, 1987.
- 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.
- N. S. Trudinger. Linear elliptic operators with measurable coefficients. *Ann. Scuola Norm. Sup. Pisa (3)*, 27:265–308, 1973.