OPTIMAL CONTROL OF ELLIPTIC PARTIAL DIFFERENTIAL EQUATIONS

S. VOLKWEIN

Abstract. This lecture is an introduction to the theory of optimal control problems governed by elliptic partial differential equations. The main focus is on existence results for optimal controls as well as on optimality conditions. Linear-quadratic and semilinear problems are considered. It is basically based on the books [2, 3].

1. Optimal control in finite dimension

Some basic concepts in optimal control theory can be illustrated very well in the context of finite-dimensional optimization. In particular, we do not have to deal with partial differential equations and several aspects from functional analysis.

1.1. Finite-dimensional optimal control problem. Let us consider the minimization problem

$$\min J(y,u) \quad \text{subject to (s.t.)} \quad Ay = Bu \quad \text{and} \quad u \in U_{ad} \tag{1.1}$$

where $J : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ denotes the cost functional, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ and $\emptyset \neq U_{ad} \subset \mathbb{R}^m$ is the set of admissible controls.

We look for vectors $y \in \mathbb{R}^n$ and $u \in \mathbb{R}^m$ which solve (1.1).

Example 1.1. Often the cost functional is quadratic, e.g.,

$$J(y,u) = |y - y_d|^2 + \lambda |u|^2,$$

where $|\cdot|$ stands for the Euclidean norm and $y_d \in \mathbb{R}^n$, $\lambda \geq 0$ hold. $\Diamond$

Problem (1.1) has the form of an optimization problem. Now we assume that A is an invertible matrix. Then we have

$$y = A^{-1}Bu. \tag{1.2}$$

In this case there exists a unique vector $y \in \mathbb{R}^n$ for any $u \in \mathbb{R}^m$. Hence, y is a dependent variable. We call u the control and y the state. In this way, (1.1) becomes a finite-dimensional optimal control problem.

We define the matrix $S \in \mathbb{R}^{m\times n}$ by $S = A^{-1}B$. Then, S is the solution matrix of our control system: $y = Su$. Utilizing the matrix S we introduce the so-called reduced cost functional

$$\hat{J}(u) = J(Su, u).$$

Date: June 30, 2016.

Key words and phrases. Optimal control, elliptic partial differential equations, optimality conditions.

Thanks to Roberta Mancini who compiled this LaTeX version of the scriptum.

This leads to the reduced problem

$$\min \hat{J}(u) \quad \text{s.t.} \quad u \in U_{ad}. \tag{1.3}$$

In (1.3) the state variable is eliminated.

1.2. Existence of optimal controls.

Definition 1.2. *The vector $u^* \in U_{ad}$ is called an* optimal control *for* (1.1) *provided*

$$\hat{J}(u^*) \leq \hat{J}(u) \qquad \text{for all } u \in U_{ad}.$$

The vector $y^* = Su^*$ is the associated optimal state.

Theorem 1.3. *Suppose that J is continuous on $\mathbb{R}^n \times U_{ad}$, that U_{ad} is nonempty, bounded, closed and that A is invertible. Then, there exists at least one optimal control for* (1.1).

Proof. Since the cost functional J is continuous on $\mathbb{R}^n \times U_{ad}$, the reduced cost $\hat{J}$ is continuous on U_{ad}. Furthermore, $U_{ad} \subset \mathbb{R}^m$ is bounded and closed. This implies that U_{ad} is compact. Due to the theorem of Weierstrass $\hat{J}$ has a minimum $u^* \in U_{ad} \neq \emptyset$, i.e., $\hat{J}(u^*) = \min_{u \in U_{ad}} \hat{J}(u)$. □

In the context of partial differential equations the proof for the existence of optimal controls is more complicated. The reason for this fact is that bounded and closed sets in infinite-dimensional function spaces need not to be compact.

1.3. First-order necessary optimality conditions. To compute solutions to optimal control problems we make use of optimality conditions. For that purpose we study first-order conditions for optimality.

We use the following notation for a function $\hat{J} : \mathbb{R}^m \to \mathbb{R}$:

$$\begin{aligned}
D_i &= \frac{\partial}{\partial x_i}, \quad D_x = \frac{\partial}{\partial x}, \quad D_{xx} = \frac{\partial^2}{\partial x^2} && \text{(partial derivatives)},\\
\hat{J}'(x) &= (D_1 \hat{J}(x), \dots, D_m \hat{J}(x)) \in \mathbb{R}^{1 \times m} && \text{(derivative)},\\
\nabla \hat{J}(x) &= \hat{J}'(x)^\top && \text{(gradient)}.
\end{aligned}$$

For functions $J : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ we denote by $D_y J(y,u) \in \mathbb{R}^{1 \times n}$ the derivative with respect to $y \in \mathbb{R}^n$, i.e., $D_y J(y,u) = (D_{y_1} J(y,u), \dots, D_{y_n} J(y,u))$. The vector $\nabla_y J(y,u) = D_y J(y,u)^\top \in \mathbb{R}^{n \times 1}$ is the gradient of J with respect to y. Analogously, $D_u J(y,u)$ and $\nabla_u J(y,u)$ are defined.
The Euclidean inner product is denoted by

$$\langle u, v \rangle_{\mathbb{R}^m} = u \cdot v = \sum_{i=1}^m u_i v_i \quad \text{for} \quad u = (u_1, \dots, u_m)^\top,\, v = (v_1, \dots, v_m)^\top.$$

For the directional derivative in direction $h \in \mathbb{R}^m$ we have

$$\hat{J}'(u)h = \langle \nabla \hat{J}(u), h \rangle_{\mathbb{R}^m} = \nabla \hat{J}(u) \cdot h.$$

Throughout we assume that all partial derivatives of J exist and are continuous. From the chain rule it follows that $\hat{J}(u) = J(Su, u)$ is continuously differentiable.

Example 1.4. Let us consider the cost functional

$$\hat{J}(u) = \frac{1}{2}|Su - y_d|^2 + \frac{\lambda}{2}|u|^2,$$

see Example 1.1. We obtain

$$\begin{aligned} \nabla\hat{J}(u) &= S^\top(Su - y_d) + \lambda u, \\ \hat{J}'(u) &= (S^\top(Su - y_d) + \lambda u)^\top, \\ \hat{J}'(u)h &= \langle S^\top(Su - y_d) + \lambda u, h\rangle_{\mathbb{R}^m} \end{aligned}$$

at $u \in \mathbb{R}^m$ and for $h \in \mathbb{R}^m$. ◊

Theorem 1.5. *Suppose that u^* is an optimal control for (1.1), U_{ad} is convex and $\hat{J}$ is differentiable. Then the variational inequality*

$$\hat{J}'(u^*)(u - u^*) \geq 0 \qquad \text{for all } u \in U_{ad} \tag{1.4}$$

holds.

It follows from Theorem 1.5 that at u^* the cost functional $\hat{J}$ can not decrease in any feasible direction. The proof follows from a more general result (see [3, pag. 63]).

From the chain rule we derive

$$\begin{aligned} \hat{J}'(u^*)h &= D_yJ(Su^*, u^*)Sh + D_uJ(Su^*, u^*)h \\ &= \langle \nabla_y J(y^*, u^*), A^{-1}Bh\rangle_{\mathbb{R}^n} + \langle \nabla_u J(y^*, u^*), h\rangle_{\mathbb{R}^m} \\ &= \langle B^\top A^{-\top}\nabla_y J(y^*, u^*) + \nabla_u J(y^*, u^*), h\rangle_{\mathbb{R}^m}, \end{aligned} \tag{1.5}$$

where $(A^\top)^{-1} = (A^{-1})^\top := A^{-\top}$ holds. Thus, we derive from (1.4)

$$\langle B^\top A^{-\top}\nabla_y J(y^*, u^*) + \nabla_u J(y^*, u^*), u - u^*\rangle_{\mathbb{R}^m} \geq 0 \tag{1.6}$$

for all $u \in U_{ad}$. In the following subsection we will introduce the so-called adjoint or dual variable. Then, we can express (1.6) in a simpler way.

1.4. Adjoint variable and reduced gradient. In a numerical realization the computation of A^{-1} is avoided. The same holds for the matrix $A^{-\top}$. Thus, we replace the term $A^{-\top}\nabla_y J(y^*, u^*)$ by $p^* := -A^{-\top}\nabla_y J(y^*, u^*)$, which is equivalent with

$$A^\top p^* = -\nabla_y J(y^*, u^*). \tag{1.7}$$

Definition 1.6. *Equation* (1.7) *is called the* adjoint *or* dual equation*. Its solution p^* is the* adjoint *or* dual variable *associated with (y^*, u^*).*

Example 1.7. For the quadratic cost functional $J(y, u) = \frac{1}{2}|y - y_d|^2 + \frac{1}{2}\lambda|u|^2$ with $y, y_d \in \mathbb{R}^m$ and $\lambda \geq 0$ we derive the adjoint equation

$$A^\top p^* = y_d - y^*.$$

Here we have used $\nabla_y J(y, u) = y - y_d$. ◊

The introduction of the dual variable yields two advantages:

1) We obtain an expression for (1.6) without the matrix $A^{-\top}$.
2) The expression (1.6) can be written in a more readable form.

Utilizing $y^* = Su^*$ in (1.5) we find that

$$\nabla \hat{J}(u^*) = -B^\top p^* + \nabla_u J(y^*, u^*).$$

The vector $\nabla \hat{J}(u^*)$ is called the *reduced gradient*. The directional derivative of the reduced cost functional $\hat{J}$ at an arbitrary $u \in U_{ad}$ in direction h is given by

$$\hat{J}'(u)h = \langle -B^\top p + \nabla_u J(y, u), h\rangle_{\mathbb{R}^m},$$

where $y = Su$ and $p = -A^\top \nabla_y J(y, u)$ hold. From Theorem 1.5 and (1.6) we derive directly the following theorem.

Theorem 1.8. *Suppose that A is invertible, u^* is an optimal control for* (1.1), *$y^* = Su^*$ the associated optimal state and J is differentiable. Then, there exists a unique dual variable p^* satisfying* (1.7). *Moreover, the variational inequality*

$$\langle -B^\top p^* + \nabla_u J(y^*, u^*), u - u^*\rangle_{\mathbb{R}^m} \geq 0 \quad \textit{for all } u \in U_{ad} \tag{1.8}$$

holds true.

We have derived an optimality system for the unknown variables y^*, u^* and p^*:

$$\begin{aligned} &Ay^* = Bu^*, \qquad u^* \in U_{ad} \\ &A^\top p^* = -\nabla_y J(y^*, u^*) \\ &\langle -B^\top p^* + \nabla_u J(y^*, u^*), v - u^*\rangle_{\mathbb{R}^m} \geq 0 \quad \text{for all } v \in U_{ad}. \end{aligned} \tag{1.9}$$

Every solution (y^*, u^*) to (1.1) must satisfy, together with the dual variable p^*, the necessary conditions (1.9).

If $U_{ad} = \mathbb{R}^m$ holds, then the term $u - u^*$ can attain any value $h \in \mathbb{R}^m$. Therefore, the variational inequality (1.8) implies the equation

$$-B^\top p^* + \nabla_u J(y^*, u^*) = 0.$$

Example 1.9. We consider the cost functional

$$J(y, u) = \frac{1}{2}|Cy - y_d|^2 + \frac{\lambda}{2}|u|^2$$

with $C \in \mathbb{R}^{n\times n}$, y, $y_d \in \mathbb{R}^n$, $\lambda \geq 0$ and $u \in \mathbb{R}^m$. Then,

$$\nabla_y J(y, u) = C^\top (Cy - y_d), \qquad \nabla_u J(y, u) = \lambda u.$$

Thus, we obtain the optimality system

$$\begin{aligned} &Ay^* = Bu^*, \quad u^* \in U_{ad} \\ &A^\top p^* = C^\top (y_d - Cy^*) \\ &\langle -B^\top p^* + \lambda u^*, v - u^*\rangle_{\mathbb{R}^m} \geq 0 \quad \text{for all } v \in U_{ad} \end{aligned}$$

If $U_{ad} = \mathbb{R}^m$ holds, we find $-B^\top p^* + \lambda u^* = 0$. For $\lambda > 0$ we have

$$u^* = \frac{1}{\lambda} B^\top p^*. \tag{1.10}$$

Inserting (1.10) into the state equation, we obtain a linear system in the state and dual variables:

$$\begin{aligned} Ay^* &= \frac{1}{\lambda} BB^\top p^* \\ A^\top p^* &= C^\top (y_d - Cy^*). \end{aligned}$$

If (y^*, p^*) is computed, u^* is given by (1.10). $\Diamond$

1.5. The Lagrange function. The optimality condition can be expressed by utilizing the Lagrange function.

Definition 1.10. *The function* $\mathcal{L} : \mathbb{R}^{2n+m} \to \mathbb{R}$ *defined by*

$$\mathcal{L}(y, u, p) = J(y, u) + \langle Ay - Bu, p\rangle_{\mathbb{R}^n}, \qquad (y, u, p) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n,$$

is called the Lagrange function *for* (1.1).

It follows that the second and third conditions of (1.9) can be expressed as

$$\begin{aligned} &\nabla_y \mathcal{L}(y^*, u^*, p^*) = 0 \\ &\langle \nabla_u \mathcal{L}(y^*, u^*, p^*), u - u^*\rangle_{\mathbb{R}^m} \geq 0 \quad \text{for all } u \in U_{ad}. \end{aligned}$$

Remark 1.11. The adjoint equation (1.7) is equivalent to $\nabla_y \mathcal{L}(y^*, u^*, p^*) = 0$. Thus, (1.7) can be derived from the derivative of the Lagrange functional with respect to the state variable y. Analogously, the variational inequality follows from the gradient $\nabla_u \mathcal{L}(y^*, u^*, p^*)$. ♢

It follows from Remark 1.11 that (y^*, u^*) satisfies the necessary optimality conditions of the minimization problem

$$\min \mathcal{L}(y, u, p^*) \quad \text{s.t. } (y, u) \in \mathbb{R}^n \times U_{ad}. \tag{1.11}$$

Notice that (1.11) has no equality constraints (in contrast to (1.1)). In most applications p^* is not known a-priori. Thus, (y^*, u^*) can not be computed from (1.11).

1.6. Discussion of the variational inequality. In many applications the set of admissible controls has the form

$$U_{ad} = \{u \in \mathbb{R}^m \,|\, u_a \leq u \leq u_b\}, \tag{1.12}$$

where $u_a \leq u_b$ are given vectors in $\mathbb{R}^m$ and "$\leq$" means less or equal in each component: $u_{a,i} \leq u_i \leq u_{b,i}$ for $i = 1, \ldots, m$. From (1.8) it follows that

$$\langle -B^\top p^* + \nabla_u J(y^*, u^*), u^*\rangle_{\mathbb{R}^m} \leq \langle -B^\top p^* + \nabla_u J(y^*, u^*), u\rangle_{\mathbb{R}^m}$$

for all $u \in U_{ad}$. This implies that u^* solves the minimization problem

$$\min_{u \in U_{ad}} \langle -B^\top p^* + \nabla_u J(y^*, u^*), u\rangle_{\mathbb{R}^m} = \min_{u \in U_{ad}} \sum_{i=1}^m (-B^\top p^* + \nabla_u J(y^*, u^*))_i u_i.$$

If U_{ad} is of the form (1.12), then the minimization of a component u_i is independent of u_j, $i \neq j$:

$$(-B^\top p^* + \nabla_u J(y^*, u^*))_i u_i^* = \min_{u_{a,i} \leq u_i \leq u_{b,i}} (-B^\top p^* + \nabla_u J(y^*, u^*))_i u_i,$$

$1 \leq i \leq m$. Thus,

$$u_i^* = \begin{cases} u_{b,i} & \text{if } (-B^\top p^* + \nabla_u J(y^*, u^*))_i < 0 \\ u_{a,i} & \text{if } (-B^\top p^* + \nabla_u J(y^*, u^*))_i > 0. \end{cases} \tag{1.13}$$

If $(-B^\top p^* + \nabla_u J(y^*, u^*))_i = 0$ holds, we have no information from the variational inequality. In many cases we can use the equation $(-B^\top p^* + \nabla_u J(y^*, u^*))_i = 0$ to obtain an explicit equation for one of the components of u^*.

1.7. The Karush–Kuhn–Tucker system. Define the vectors

$$\begin{aligned} \mu_a &:= (-B^\top p^* + \nabla_u J(y^*, u^*))_+ \\ \mu_b &:= (-B^\top p^* + \nabla_u J(y^*, u^*))_-. \end{aligned} \tag{1.14}$$

where $\mu_{a,i} = (-B^\top p^* + \nabla_u J(y^*, u^*))_i$ if the right-hand side is positive and $\mu_{a,i} = 0$ otherwise. Analogously, $\mu_{b,i} = |(-B^\top p^* + \nabla_u J(y^*, u^*))_i|$ if the right-hand side is negative and $\mu_{b,i} = 0$ otherwise. Utilizing (1.13) we have

$$\begin{aligned} &\mu_a \geq 0, \quad u_a - u^* \leq 0, \quad \langle u_a - u^*, \mu_a \rangle_{\mathbb{R}^m} = 0 \\ &\mu_b \geq 0, \quad u^* - u_b \leq 0, \quad \langle u^* - u_b, \mu_b \rangle_{\mathbb{R}^m} = 0 \end{aligned}$$

These conditions are called *complementarity conditions*. The inequalities are clear. We prove $\langle u_a - u^*, \mu_a \rangle_{\mathbb{R}^m} = 0$. Suppose that $u_{a,i} < u_i^*$ holds. Due to (1.13) we have $(-B^\top p^* + \nabla_u J(y^*, u^*))_i \leq 0$. Thus, $\mu_{a,i} = 0$ which gives $(u_{a,i} - u_i^*)\mu_{a,i} = 0$. Now we assume $\mu_{a,i} > 0$. Using (1.14) we derive $(-B^\top p^* + \nabla_u J(y^*, u^*))_i > 0$. It follows from (1.13) that $u_{a,i} = u_i^*$ holds. Again, we have $(u_{a,i} - u_i^*)\mu_{a,i} = 0$. Summation over $i = 1, \ldots, m$ yields $\langle u_a - u^*, \mu_a \rangle_{\mathbb{R}^m} = 0$.
Notice that

$$\mu_a - \mu_b = -B^\top p^* + \nabla_u J(y^*, u^*).$$

Hence,

$$\nabla_u J(y^*, u^*) - B^\top p^* + \mu_b - \mu_a = 0. \tag{1.15}$$

Let us consider an augmented Lagrange functional

$$\tilde{\mathcal{L}}(y, u, p, \mu_a, \mu_b) = J(y, u) + \langle Ay - Bu, p \rangle_{\mathbb{R}^n} + \langle u_a - u, \mu_a \rangle_{\mathbb{R}^m} + \langle u - u_b, \mu_b \rangle_{\mathbb{R}^m}$$

Then, (1.15) can be written as

$$\nabla_u \tilde{\mathcal{L}}(y^*, u^*, p^*, \mu_a, \mu_b) = 0.$$

Moreover, the adjoint equation is equivalent with

$$\nabla_y \tilde{\mathcal{L}}(y^*, u^*, p^*, \mu_a, \mu_b) = 0.$$

Here, we have used that $\nabla_y \mathcal{L} = \nabla_y \tilde{\mathcal{L}}$. The vectors μ_a and μ_b are the Lagrange multipliers for the inequality constraints $u_a - u^* \leq 0$ and $u^* - u_b \leq 0$.

Theorem 1.12. *Suppose that u^* is an optimal control for* (1.1)*, A is invertible and U_{ad} has the form* (1.12)*. Then, there exist Lagrange multipliers $p^* \in \mathbb{R}^n$ and $\mu_a, \mu_b \in \mathbb{R}^m$ satisfying*

$$\begin{aligned} &\nabla_y \tilde{\mathcal{L}}(y^*, u^*, p^*, \mu_a, \mu_b) = 0 \\ &\nabla_u \tilde{\mathcal{L}}(y^*, u^*, p^*, \mu_a, \mu_b) = 0 \\ &\mu_a \geq 0, \quad \mu_b \geq 0 \\ &\langle u_a - u^*, \mu_a \rangle_{\mathbb{R}^m} = \langle u^* - u_b, \mu_b \rangle_{\mathbb{R}^m} = 0 \\ &Ay^* = Bu^*, \quad u_a \leq u \leq u_b. \end{aligned} \tag{1.16}$$

The optimality system (1.16) is called the *Karush-Kuhn-Tucker (KKT) system.*

2. Existence of optimal controls for linear quadratic optimal control problems

In this section we present strategies to prove existence of optimal controls for linear quadratic problems. The cost functional is quadratic and the constraints are linear elliptic equations together with linear inequality constraints.

Assumption 1. *Let $\Omega \subset \mathbb{R}^N$ be a bounded domain with Lipschitz-continuous boundary and suppose that $\lambda \geq 0$, $y_\Omega \in L^2(\Omega)$, $y_\Gamma \in L^2(\Gamma)$, $\beta \in L^\infty(\Omega)$, $\alpha \in L^\infty(\Gamma)$ with $\alpha(x) \geq 0$ for almost all (f.a.a) $x \in \Gamma$ and u_a, u_b, v_a, $v_b \in L^2(E)$ with $u_a(x) \leq u_b(x)$, $v_a(x) \leq v_b(x)$ f.a.a. $x \in E$. Here, $E = \Omega$ or $E = \Gamma$.*

2.1. Optimal stationary heat source. We consider the problem

$$\min J(y,u) := \frac{1}{2}\|y - y_\Omega\|^2_{L^2(\Omega)} + \frac{\lambda}{2}\|u\|^2_{L^2(\Omega)} \tag{2.1a}$$

subject to

$$-\Delta y = \beta u \text{ in } \Omega, \quad y = 0 \text{ on } \Gamma, \tag{2.1b}$$

$$u_a(x) \leq u(x) \leq u_b(x) \quad \text{f.a.a. } x \in \Omega. \tag{2.1c}$$

Let us introduce the set of admissible control by

$$U_{ad} = \left\{ u \in L^2(\Omega) : u_a(x) \leq u(x) \leq u_b(x) \text{ f.a.a. } x \in \Omega \right\}.$$

Note that U_{ad} is non-empty, convex and bounded in $L^2(\Omega)$.

The following proposition follows from the Lax-Milgram lemma. For a proof we refer to [3, pag. 33], for instance.

Proposition 2.1. *With Assumption 1 holding there exists a unique weak solution $y \in H^1_0(\Omega)$ to (2.1b), for every $u \in L^2(\Omega)$, i.e.*

$$\int_\Omega \nabla y \cdot \nabla \varphi \,\mathrm{d}x = \int_\Omega \beta u \varphi \,\mathrm{d}x \quad \textit{for all } \varphi \in H^1_0(\Omega).$$

Furthermore,

$$\|y\|_{H^1(\Omega)} \leq C\,\|u\|_{L^2(\Omega)} \tag{2.2}$$

for a constant C depending on $\beta \in L^\infty(\Omega)$.

Remark 2.2. Let us introduce the linear operator $e : H^1_0(\Omega) \times L^2(\Omega) \to H^{-1}(\Omega)$ by

$$\langle e(y,u), \varphi \rangle_{H^{-1}(\Omega), H^1_0(\Omega)} = \int_\Omega \nabla y \cdot \nabla \varphi - \beta u \varphi \,\mathrm{d}x \quad \text{for } \varphi \in H^1_0(\Omega)$$

with $(y,u) \in H^1_0(\Omega) \times L^2(\Omega)$. Then, (2.1) can be expressed equivalently as

$$\min J(x) \quad \text{subject to (s.t.)} \quad x = (y,u) \in X_{ad} \text{ and } e(x) = 0 \text{ in } H^{-1}(\Omega)$$

with $X_{ad} = H^1_0(\Omega) \times U_{ad}$. ◊

The unique solution y to (2.1b) is called *the state associated with u*. We define *the state space*

$$Y = H^1_0(\Omega)$$

and we write $y = y(u)$ to emphasize the dependence on u.

Definition 2.3. *An element $u^* \in U_{ad}$ is called* optimal control *and $y^* = y(u^*)$ the associated* optimal state *provided*

$$J(y^*, u^*) \leq J(y(u), u) \quad \textit{for all } u \in U_{ad}.$$

By Proposition 2.1 the solution operator $G : L^2(\Omega) \to H_0^1(\Omega)$, $u \mapsto y(u)$ is well-defined. We call G the control-to-state mapping. Notice that G is linear and continuous. The continuity follows from (2.2).

Remark 2.4. The space $H^1(\Omega)$ (and therefore also $H_0^1(\Omega) \subset H^1(\Omega)$) is continuously embedded into $L^2(\Omega)$. In particular,

$$\|y\|_{L^2(\Omega)} \le \|y\|_{H^1(\Omega)} \le C\|u\|_{L^2(\Omega)}$$

for $y = Gu$ and $u \in L^2(\Omega)$. Hence, we consider G as a mapping from $L^2(\Omega)$ to $L^2(\Omega)$. More precisely, we define the solution operator

$$S = E_Y G : L^2(\Omega) \to L^2(\Omega),$$

where $E_Y : H^1(\Omega) \to L^2(\Omega)$ denotes the canonical embedding operator. The advantage of the operator S is that its adjoint S^* is also defined on $L^2(\Omega)$ and we have

$$\langle Su, \varphi\rangle_{L^2(\Omega)} = \langle u, S^*\varphi\rangle_{L^2(\Omega)} \quad \text{for all } u, \varphi \in L^2(\Omega).$$

This will be used in Section 3.3. $\Diamond$

We introduce the so-called *reduced cost functional* $\hat{J}$ by

$$\hat{J}(u) = J(Su, u) = \frac{1}{2}\,\|Su - y_\Omega\|^2_{L^2(\Omega)} + \frac{\lambda}{2}\,\|u\|^2_{L(\Omega)^2} \quad \text{for all } u \in U_{ad}.$$

To prove the existence and uniqueness of an optimal control for (2.1) we make use of some facts from functional analysis.

Definition 2.5. *A subset M of a real Banach space U is called* weakly sequentially closed *if $u_n \in M$, $u_n \rightharpoonup u \in U$ $(n \to \infty)$ imply $u \in M$. The set M is called* weakly sequentially compact *if every sequence $\{u_n\}_{n\in\mathbb{N}}$ in M has a weakly convergent subsequence in U and if M is weakly sequentially closed.*

Definition 2.6. *A subset C of a real Banach space U is called* convex *if for all $u, v \in C$ and for all $\lambda \in [0,1]$ we have*

$$\lambda u + (1-\lambda)v \in C.$$

The mapping $f : C \to \mathbb{R}$ is said to be convex *if for all $u, v \in C$ and for all $\lambda \in [0,1]$ it follows that*

$$f(\lambda u + (1-\lambda)v) \le \lambda f(u) + (1-\lambda)f(v).$$

We call f strictly convex *if for all $u, v \in U$ with $u \ne v$ and for all $\lambda \in (0,1)$ we have*

$$f(\lambda u + (1-\lambda)v) < \lambda f(u) + (1-\lambda)f(v).$$

To prove the existence of optimal controls we will make use of the following result.

Theorem 2.7. *A convex and closed subset of a Banach space is weakly sequentially closed. If the space is reflexive (e.g. a Hilbert space) and if the subset is also bounded, then the subset is weakly sequentially compact.*

Theorem 2.8. *Every convex and continuous functional f defined on a Banach space U is weakly lower semicontinuous, i.e., for any sequence $\{u_n\}_{n\in\mathbb{N}}$ in U with $u_n \rightharpoonup u$ for $n \to \infty$ we have*

$$\liminf_{n\to\infty} f(u_n) \ge f(u).$$

Example 2.9. The norm is weakly lower semicontinuous. From

$$\|\lambda u + (1-\lambda)v\| \le \lambda \|u\| + (1-\lambda) \|v\| \quad \text{for all } \lambda \in [0,1]$$

it follows that the norm is convex. Moreover, $\|\cdot\|$ is continuous. Thus, the claim follows from Theorem 2.8. ♢

Theorem 2.10. *Suppose that U and H are given Hilbert spaces with norms $\|\cdot\|_U$ and $\|\cdot\|_H$, respectively. Furthermore, let $U_{ad} \subset U$ be non-empty, bounded, closed, convex and $y_d \in H$, $\lambda \ge 0$. The mapping $S : U \to H$ is assumed to be a linear and continuous operator. Then there exists an optimal control u^* solving*

$$\min_{u \in U_{ad}} \hat{J}(u) := \frac{1}{2} \|Su - y_d\|_H^2 + \frac{\lambda}{2} \|u\|_U^2. \tag{2.3}$$

If $\lambda > 0$ holds or if S is injective, then u^* is uniquely determined.

Proof. Since $\hat{J}(u) \ge 0$ holds, the infimum

$$j = \inf_{u \in U_{ad}} \hat{J}(u)$$

exists. By assumption, $U_{ad} \neq \emptyset$. Thus, there is a minimizing sequence $\{u_n\}_{n\in\mathbb{N}}$ satisfying $\lim_{n\to\infty} \hat{J}(u_n) = j$. The set U_{ad} is bounded and closed (but in general not compact). From the convexity of U_{ad} and Theorem 2.7, we infer that U_{ad} is weakly sequentially compact. Thus, there exists a subsequence $\{u_{n_k}\}_{k\in\mathbb{N}}$ of $\{u_n\}_{n\in\mathbb{N}}$ and an element $u^* \in U_{ad}$ satisfying

$$u_{n_k} \rightharpoonup u^* \quad \text{for } k \to \infty.$$

Since S is continuous, $\hat{J}$ is continuous. From the convexity of $\hat{J}$ and Theorem 2.8 we infer

$$\hat{J}(u^*) \le \lim_{k\to\infty} \hat{J}(u_{n_k}) = j.$$

Recall that j is the infimum of all function values $\hat{J}(u)$, $u \in U_{ad}$. From $u^* \in U_{ad}$ we have $\hat{J}(u^*) \ge j$. Thus, $\hat{J}(u^*) = j$ and u^* is an optimal control for (2.3).

Note that $\hat{J}''(u) = S^*S + \lambda I : U \to U$ is the hessian of $\hat{J}$, where $S^* : H \to U$ is the Hilbert space adjoint of $S : U \to H$ satisfying

$$\langle Su, h\rangle_H = \langle u, S^*h\rangle_U \quad \text{for all } (u,h) \in U \times H.$$

Notice that

$$\langle \hat{J}''(u)v, v\rangle_U = \langle S^*Sv + \lambda v, v\rangle_U = \|Sv\|_H^2 + \lambda \|v\|_H^2.$$

If $\lambda > 0$ then $\langle \hat{J}''(u)v, v\rangle > 0$ for all $v \in U \setminus \{0\}$. On the other hand we have that S is injective. Then $\|Sv\|_H^2 > 0$ for all $v \in U \setminus \{0\}$. Thus, we have in both cases that $\hat{J}''(u)$ is a positive operator. This implies that $\hat{J}$ is strictly convex and there exists a unique optimal control. □

Remark 2.11. In the proof of Theorem 2.10 we have only used that $\hat{J}$ is continuous and convex. Therefore, the existence of an optimal control follows for general convex and continuous cost functionals $\hat{J} : U \to \mathbb{R}$ with a Hilbert space U. ♢

Next we can use Theorem 2.10 to obtain an existence result for the optimal control problem (2.1).

Theorem 2.12. *Let Assumption* 1 *be satisfied. Then* (2.1) *has a local solution* (y^*, u^*) *with* $y^* = Su^*$. *For* $\lambda > 0$ *or* $\beta(x) \neq 0$ *f.a.a.* $x \in \Omega$ *the optimal solution is unique.*

Proof. In the context of Theorem 2.10 we choose $U = H = L^2(\Omega)$, $y_d = y_\Omega$ and $S = E_Y G$. The set $U_{ad} = \{u \in L^2(\Omega) : u_a \leq u \leq u_b\}$ is bounded, convex and closed. From Theorem 2.10 we derive the existence of an optimal solution (y^*, u^*), $y^* = Su^*$. For $\beta \neq 0$ the operator S is injective: from $Su = 0$ we have $y = 0$. This implies $\beta u = Ay = 0$ and, hence, $u = 0$ follows. $\square$

Remark 2.13. In the proof of Theorem 2.10 the optimal control u^* is the limit of the weakly convergent sequence $\{u_{n_k}\}_{k\in\mathbb{N}}$. Since $G : L^2(\Omega) \to H_0^1(\Omega)$ is linear and continuous, the sequence $\{y_{n_k}\}_{k\in\mathbb{N}}$, $y_{n_k} = Su_{n_k}$, converges also weakly (in $H_0^1(\Omega)$) to $y^* = Su^*$. $\Diamond$

Next we consider the case that $u_a = -\infty$ or/and $u_b = +\infty$. In this case U_{ad} is not bounded. Hence, U_{ad} is not weakly sequentially compact.

Theorem 2.14. *If* $\lambda > 0$ *holds and* U_{ad} *is nonempty, convex, closed, problem* (2.3) *admits a unique solution.*

Proof. By assumption there exists an element $u_0 \in U_{ad}$. For $u \in U$ with $\|u\|_U^2 > 2\lambda^{-1}\ \hat{J}(u_0)$ we have

$$\hat{J}(u) = \frac{1}{2}\,\|Su - y_d\|_H^2 + \frac{\lambda}{2}\,\|u\|_U^2 \geq \frac{\lambda}{2}\,\|u\|_U^2 > \hat{J}(u_0).$$

Thus, the minimization of $\hat{J}$ over U_{ad} is equivalent with the minimization of $\hat{J}$ over the bounded, convex and closed set

$$U_{ad} \cap \left\{u \in U \,:\, \|u\|_U^2 \leq 2\lambda^{-1}\hat{J}(u_0)\right\}.$$

Now the claim follows as in the proof of Theorem 2.10. $\square$

We directly obtain the next result.

Theorem 2.15. *Let* $u_a = -\infty$ *or/and* $u_b = +\infty$. *Moreover,* $\lambda > 0$ *holds. Then,* (2.1) *has a unique solution provided Assumption* 1 *is satisfied.*

Next we modify our state equation by changing the boundary conditions. For given control variable u the state variable is given by

$$-\Delta y = \beta u \text{ in } \Omega, \quad \frac{\partial y}{\partial n} = \alpha(y_a - y) \text{ on } \Gamma. \tag{2.4}$$

Here, $y_a \in L^2(\Gamma)$ is a given temperature and the coefficient $\alpha \in L^\infty(\Gamma)$ satisfies $\alpha(x) \geq 0$ f.a.a. $x \in \Omega$ and $\int_\Gamma \alpha^2 ds > 0$. The analysis for the state equation (2.4) is similar to the one for (2.1b). We recall the following theorem, which ensures unique solvability of (2.4).

Theorem 2.16. *Suppose that* $\Omega \subset \mathbb{R}^n$ *is a bounded domain with Lipschitz continuous boundary. For given coefficients* $c_0 \in L^\infty(\Omega)$ *and* $\alpha \in L^\infty(\Gamma)$ *with* $c_0(\lambda) \geq 0$ *f.a.a.* $x \in \Omega$ *and* $\alpha(x) \geq 0$ *f.a.a.* $x \in \Gamma$ *let* y *be given by*

$$-\Delta y + c_0 y = f \text{ in } \Omega, \quad \frac{\partial y}{\partial n} + \alpha y = g \text{ on } \Gamma. \tag{2.5}$$

Suppose that

$$\int_\Omega c_0(x)\,\mathrm{d}x + \int_\Gamma \alpha(x)^2\,\mathrm{d}s(x) > 0.$$

Then, (2.5) *has a unique solution* $y \in H^1(\Omega)$ *for any* $f \in L^2(\Omega)$ *and* $g \in L^2(\Gamma)$. *Moreover, there exists a constant* $C > 0$ *(independent of* f *and* g*) so that*

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

holds.

A proof of Theorem 2.16 is given in [3, p. 36]

In contrast to the Dirichlet problem, we choose $Y = H^1(\Omega)$. Theorem 2.16 ensures for any $u \in L^2(\Gamma)$ and $y_a \in L^2(\Gamma)$ that exists a unique solution $y \in Y$. Then, each solution to (2.4) can be expressed as

$$y = y(u) + y_0,$$

where $y(u)$ solves (2.4) for the pair $(u, y_a = 0)$ and y_0 solves (2.4) for the pair $(u = 0, y_a)$. The operator $G : u \to y(u)$ is linear and continuous from $L^2(\Omega)$ to $H^1(\Omega)$. We consider G as a mapping from $L^2(\Omega)$ to $L^2(\Omega)$:

$$S = E_Y G, \quad S : L^2(\Omega) \to L^2(\Omega).$$

Then, the state variable y is given as $y = Su + y_0$. The optimal control problem can be expressed as

$$\min_{u\in U_{ad}} \hat{J}(u) := \frac{1}{2}\|Su - (y_\Omega - y_0)\|^2_{L^2(\Omega)} + \frac{\lambda}{2}\|u\|^2_{L^2(\Omega)}. \tag{2.6}$$

From Theorems 2.10 and 2.14 the existence of an optimal control follows. If $\lambda > 0$ or $\beta \neq 0$ in Ω hold, the optimal control is unique.

2.2. Optimal stationary boundary temperature. We consider the boundary control problem

$$\min J(y, u) := \frac{1}{2}\|y - y_\Omega\|^2_{L^2(\Omega)} + \frac{\lambda}{2}\|u\|^2_{L^2(\Gamma)} \tag{2.7a}$$

subject to

$$-\Delta y = 0 \text{ in } \Omega, \quad \frac{\partial y}{\partial n} = \alpha(u - y) \text{ on } \Gamma, \tag{2.7b}$$

and

$$u_a(x) \le u(x) \le u_b(x) \quad \text{f.a.a. } x \in \Gamma. \tag{2.7c}$$

For the existence of a unique solution to the elliptic problem (2.7) we suppose

$$\int_\Gamma \alpha(x)^2\,\mathrm{d}s(x) > 0. \tag{2.8}$$

The control space is $L^2(\Gamma)$, the state space $Y = H^1(\Omega)$. We define

$$U_{ad} = \left\{u \in L^2(\Gamma) \,:\, u_a(x) \le u(x) \le u_b(x) \text{ f.a.a. } x \in \Gamma\right\}.$$

It follows from Theorem 2.16 that (2.7b) has a unique weak solution $y = y(u) \in H^1(\Omega)$ for any $u \in L^2(\Gamma)$. The operator $G : u \to y(u)$ is continuous from $L^2(\Gamma)$ to $H^1(\Omega) \subset L^2(\Omega)$. We use $S = E_Y G : L^2(\Gamma) \to L^2(\Omega)$ and obtain the next result from Theorem 2.10.

Theorem 2.17. *Let Assumption* 1 *and* (2.8) *be satisfied. Then,* (2.7) *possesses an optimal control, which is unique for* $\lambda > 0$.

Remark 2.18. With Theorem 2.14 we obtain also an existence result for unbounded sets U_{ad}. $\Diamond$

3. First-order necessary optimality conditions

Numerical methods for optimal control problem are often based on optimality conditions, which leads to gradient-type algorithms.

3.1. Differentiability in Banach spaces.

In this subsection we recall the notion of Gâteaux and Fréchet derivatives. These derivatives are needed to derive optimality conditions for PDE constrained optimization problems.

Suppose that U and V are real Banach spaces. $\mathcal{U} \subset U$ an open subset and $F : U \supset \mathcal{U} \to V$ a given mapping.

Definition 3.1. *Let $u \in \mathcal{U}$ and $h \in U$. If the limit*

$$\delta F(u,h) := \lim_{t \searrow 0} \frac{1}{t} \left(F(u+th) - F(u) \right)$$

exists in V, we call $\delta F(u,h)$ the directional derivative *of F at u in direction h. If $\delta F(u,h)$ exists for all $h \in U$, the mapping $h \mapsto \delta F(u,h)$ is called the* first variation *of F at u.*

Example 3.2. We consider the mapping $f : \mathbb{R}^2 \to \mathbb{R}$, $x = re^{i\varphi} \mapsto r\cos\varphi$. At $x = 0$ we have the first variation

$$\delta f(0,h) = \lim_{t \searrow 0} \frac{1}{t} \left(f(th) - f(0) \right) = \lim_{t \searrow 0} \frac{1}{t} \left(tr\cos\varphi - 0 \right) = r\cos\varphi = f(h),$$

where we have used that $h = re^{i\varphi}$ and $th = t\,re^{i\varphi}$. Thus, $f(th) = t\,r\cos\varphi$ and $f(h) = r\cos\varphi$. Notice that $h \mapsto f(h) = \delta f(0,h)$ is a nonlinear mapping. $\Diamond$

Definition 3.3. *Let $u \in \mathcal{U}$. Suppose that there exist the first variation $\delta F(u,h)$ and a linear, continuous operator $A : U \to V$ satisfying*

$$\delta F(u,h) = Ah \quad \textit{for all } h \in U.$$

Then, F is Gâteaux-differentiable *at u and A is the* Gâteaux derivative *of F at u. We write $A = F'(u)$.*

Remark 3.4. The Gâteaux derivative can be derived from the directional derivative. If $F : U \supset \mathcal{U} \to \mathbb{R}$ is Gâteaux-differentiable at u, then $F'(u)$ belongs to the dual space U^*. $\Diamond$

Example 3.5. 1) A nonlinear point functional: Let $U = \mathcal{U} = C([0,1])$ and $f : U \to \mathbb{R}$ be given by $f(u(\cdot)) = \sin u(1)$. Then, f is well-defined. Suppose that $h \in C([0,1])$. We compute the directional derivative of f at u in direction h. It follows that

$$\begin{aligned}
\lim_{t \searrow 0} \frac{1}{t} \left(f(u+th) - f(u) \right) &= \lim_{t \searrow 0} \frac{1}{t} \left(\sin(u(1) + th(1)) - \sin u(1) \right) \\
&= \frac{\mathrm{d}}{\mathrm{d}t} \sin(u(1) + th(1)) \Big|_{t=0} \\
&= \left(\cos(u(1) + th(1)) h(1) \right) \Big|_{t=0} = \cos u(1) h(1).
\end{aligned}$$

Thus, $\delta f(u,h) = \cos u(1) h(1)$. The mapping $h \mapsto \delta f(u,h)$ is linear and continuous in $C([0,1])$. Thus, f is Gâteaux-differentiable and $f'(u)h = \cos(u(1))h(1)$.

2) Quadratic function in a Hilbert space: Let H be a Hilbert space with inner product $\langle \cdot\,, \cdot\rangle_H$ and induced norm $\|\cdot\|_H$. We set $f(u) = \|u\|_H^2$. Then,

$$\begin{aligned}\lim_{t\searrow 0} \frac{1}{t}\left(f(u+th) - f(u)\right) &= \lim_{t\searrow 0} \frac{1}{t}\left(\|u+th\|_H^2 - \|u\|_H^2\right) \\ &= \lim_{t\searrow 0} \frac{1}{t}\left(2t\,\langle u, h\rangle_H + t^2\,\|h\|_H^2\right) = 2\,\langle u, h\rangle_H.\end{aligned}$$

Hence, $u \mapsto f(u)$ is Gâteaux-differentiable and $f'(u)h = \langle 2u, h\rangle_H$. By Riesz theorem, the dual space H^* can be identified with H. The Riesz representation of $f'(u) = \langle 2u, \cdot\rangle_H$ is $2u \in H$. Often we write $\nabla f(u) = 2u \in H$ and call $\nabla f(u)$ in this case the gradient of f at u.

3) Application to the Hilbert space $L^2(\Omega)$: Consider

$$f(u) = \|u\|_{L^2(\Omega)}^2 = \int_\Omega u(x)^2\,\mathrm{d}x.$$

From part 2) we obtain the derivative

$$f'(u)h = \int_\Omega 2u(x)h(x)\,\mathrm{d}x$$

and – by identifying $L^2(\Omega)^*$ with $L^2(\Omega)$ – the gradient $(f'(u))(x) = 2u(x)$ f.a.a. $x \in \Omega$. ♢

Definition 3.6. *Let $u \in \mathcal{U}$ and $F : U \supset \mathcal{U} \to V$ be given. The mapping F is called* Fréchet differentiable at $u \in \mathcal{U}$ *if there exists an operator $A \in \mathcal{L}(U, V)$ and a mapping $r(u, \cdot) : U \to V$ satisfying*

$$F(u + h) = F(u) + Ah + r(u, h) \quad \text{for all } h \in U \text{ with } u + h \in \mathcal{U},$$

where

$$\frac{\|r(u, h)\|_V}{\|h\|_U} \to 0 \qquad \text{for } \|h\|_U \to 0.$$

The operator A is said to be the Fréchet derivative *of F at u. We write $A = F'(u)$.*

Remark 3.7. To prove Fréchet differentiability, one often considers

$$\frac{\|F(u + h) - F(u) - Ah\|_V}{\|h\|_U} \to 0 \quad \text{for } \|h\|_U \to 0$$

with a candidate A for the derivative. ♢

Example 3.8. 1) Define the mapping $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x, y) = \begin{cases} 1 & \text{if } y = x^2 \text{ and } x \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then, f is Gâteaux-differentiable in $(0, 0)$, but not continuous. Thus, f is not Fréchet-differentiable in $(0, 0)$.

2) The mapping $f(u) = \sin u(1)$ is Fréchet-differentiable in $C([0, 1])$.
3) The function $f(u) = \|u\|_H^2$ is Fréchet-differentiable in a Hilbert space H.
4) Any linear and continuous operator A is Fréchet-differentiable. From

$$A(u + h) = Au + Ah + 0$$

we observe that $r \equiv 0$ holds. In this case, the operator A is the Fréchet derivative of itself. ♢

If F is Fréchet-differentiable, then F is Gâteaux-differentiable. Thus, a Fréchet derivative can be computed utilizing the directional derivative.

Theorem 3.9 (Chain rule)**.** *Suppose that U, V, Z are Banach spaces and $\mathcal{U} \subset U$, $\mathcal{V} \subset V$ are open sets. Let $F : \mathcal{U} \to \mathcal{V}$, $G : \mathcal{V} \to Z$ be Fréchet-differentiable at $u \in \mathcal{U}$ and $F(u) \in \mathcal{V}$, respectively. Then $E = G \circ F : \mathcal{U} \to Z$, $u \mapsto G(F(u))$ is Fréchet-differentiable at u satisfying*

$$E'(u) = G'(F(u))F'(u).$$

Example 3.10. Let $(U, \langle \cdot\,, \cdot \rangle_U)$, $(H, \langle \cdot\,, \cdot \rangle_H)$ be real Hilbert spaces, $z \in H$ be fixed, $S \in \mathcal{L}(U, H)$ and

$$E(u) = \|Su - z\|_H^2 \quad \text{for } u \in U.$$

Then, $E(u) = G(F(u))$ with $G(v) = \|v\|_H^2$ and $F(u) = Su - z$. From Example 3.5-2) and Example 3.8-4) we find

$$G'(v)h = \langle 2v, h \rangle_H, \quad F'(u)h = Sh.$$

Using Theorem 3.9 it follows that

$$\begin{aligned} E'(u)h &= G'(F(u))F'(u)h = \langle 2F(u), F'(u)h \rangle_H \\ &= 2\,\langle Su - z, Sh \rangle_H = 2\,\langle S^*(Su - z), h \rangle_U \end{aligned} \tag{3.1}$$

for any $h \in U$. In (3.1) the operator $S^* \in \mathcal{L}(H, U)$ is the adjoint operator of S and will be defined in the next subsection. ♢

3.2. Adjoint operators.

If $A \in \mathbb{R}^{m \times n}$, then,

$$\langle Au, v \rangle_{\mathbb{R}^m} = \langle u, A^\top v \rangle_{\mathbb{R}^n} \quad \text{for all } u \in \mathbb{R}^n \text{ and } v \in \mathbb{R}^m.$$

Here, $A^\top$ is the transpose of A. Analogously, we can define an adjoint operator A^* for $A \in \mathcal{L}(U, V)$ with real Hilbert spaces U, V:

$$\langle Au, v \rangle_V = \langle u, A^* v \rangle_U \quad \text{for all } u \in U \text{ and } v \in V.$$

More generally, we can define a dual operator in real Banach spaces U and V. Suppose that $A \in \mathcal{L}(U, V)$ and $f \in V^* \in \{\tilde{f} : V \to \mathbb{R} \,|\, \tilde{f} \text{ is linear and continuous}\} = \mathcal{L}(V, \mathbb{R})$. We define $g : U \to \mathbb{R}$ by

$$g(u) = f(Au).$$

Since f and A are linear, g is linear. Moreover, we have

$$|g(u)| \le \|f\|_{V^*} \|Au\|_V \le \|f\|_{V^*} \|A\|_{\mathcal{L}(U,V)} \|u\|_U.$$

Consequently, g is bounded and thus $g \in U^*$. We obtain

$$\begin{aligned} \|g\|_{U^*} &= \sup_{\|u\|_U = 1} g(u) = \sup_{\|u\|_U = 1} f(Au) = \sup_{\|u\|_U = 1} \langle f, Au \rangle_{V^*, V} \\ &= \sup_{\|u\|_U = 1} \langle A^* f, u \rangle_{U^*, U} = \|A^* f\|_{U^*} \le \|A\|_{\mathcal{L}(U,V)} \|f\|_{V^*}. \end{aligned}$$

Here, we have used the notation of the duality pairing

$$f(v) = \langle f, v \rangle_{V^*, V} \quad \text{for } v \in V.$$

The mapping $V^* \ni f \mapsto g \in U^*$ is called the *adjoint* or *dual operator* associated with $A \in \mathcal{L}(U, V)$. We denote this dual mapping by A^*. Note that

$$\langle A^* f, u \rangle_{U^*, U} = g(u) = f(Au) = \langle f, Au \rangle_{V^*, V} \quad \text{for all } u \in U.$$

Thus, for the dual operator $A^* \in \mathcal{L}(V', U')$ we have

$$\langle f, Au \rangle_{V^*,V} = \langle A^* f, u \rangle_{U^*,U} \quad \text{for all } f \in V^* \text{ and } u \in U.$$

Concerning the notation we do not distinguish between the dual and the (Hilbert space) adjoint operator.

3.3. Optimality conditions. In this subsection we derive first-order necessary optimality conditions.

3.3.1. Quadratic programming in Hilbert spaces. In section 2 we have considered the quadratic programming problem

$$\min_{u \in U_{ad}} \hat{J}(u) = \frac{1}{2} \|Su - y_d\|_H^2 + \frac{\lambda}{2} \|u\|_U^2. \tag{3.2}$$

Thus we can apply the following lemma.

Lemma 3.11. *Let U be a real Banach space, $\mathcal{U} \subset U$ be open, $C \subset \mathcal{U}$ be convex and $\hat{J} : \mathcal{U} \to \mathbb{R}$ a function, which is Gâteaux-differentiable in $\mathcal{U}$. Suppose that $u^* \in C$ is a solution to*

$$\min_{u \in C} \hat{J}(u). \tag{3.3}$$

Then the following variational inequality holds

$$\hat{J}'(u^*)(u - u^*) \geq 0 \quad \text{for all } u \in C. \tag{3.4}$$

If $u^* \in C$ solves (3.4) *and $\hat{J}$ is convex, then u^* is a solution to* (3.3).

Proof. Let $u \in C$ be chosen arbitrarily. We consider the convex linear combination

$$u(t) = u^* + t(u - u^*) \quad \text{for any } t \in [0, 1].$$

Since C is convex, $u(t) \in C$ for all $t \in [0, 1]$. From the optimality of u^* we infer that there exists a $t^* \in [0, 1]$

$$\hat{J}(u(t)) \geq \hat{J}(u^*) \quad \text{for } t \in [0, t^*].$$

Thus

$$\frac{1}{t} \Big(\hat{J}(u^* + t(u - u^*)) - \hat{J}(u^*) \Big) \geq 0$$

for all $t \in (0, t^*]$. Since $\hat{J}$ is Gâteaux-differentiable on $\mathcal{U}$, we obtain (3.4) by taking the limit $t \to 0$.

Let $u \in C$ be arbitrary and $u^* \in C$ a solution to (3.4). Since $\hat{J}$ is convex, we have

$$\hat{J}(u) - \hat{J}(u^*) \geq \hat{J}'(u^*)(u - u^*).$$

By (3.4) we obtain $\hat{J}(u) \geq \hat{J}(u^*)$, so that u^* is a solution to (3.3). $\square$

In Lemma 3.11 a first-order necessary optimality condition is formulated. If $\hat{J}$ is convex, then (3.4) is also a sufficient condition. Next we apply Lemma 3.11 to (3.2).

Theorem 3.12. *Let U, H be real Hilbert spaces, $U_{ad} \subset U$ nonempty, convex and $y_d \in H$, $\lambda \geq 0$ be given. Furthermore, assume that $S \in \mathcal{L}(U, H)$. Then, $u^* \in U_{ad}$ solves* (3.2) *if the variational inequality*

$$\langle S^*(Su^* - y_d) + \lambda u^*, u - u^* \rangle_U \geq 0 \quad \text{for all } u \in U_{ad} \tag{3.5}$$

holds.

Proof. The gradient of $\hat{J}$ is given by

$$\hat{\nabla} J(u^*) = S^*(Su^* - y_d) + \lambda u^*.$$

Thus, the claim follows directly from Lemma 3.11. $\square$

The variational inequality (3.5) can be expressed as

$$\langle Su^* - y_d, Su - Su^* \rangle_H + \lambda \langle u^*, u - u^* \rangle_U \geq 0 \quad \text{for all } u \in U_{ad},$$

where the use of S^* is avoided.

3.3.2. Optimal stationary heat source. We consider the problem (compare Section 2.1)

$$\min J(y,u) := \frac{1}{2} \|y - y_\Omega\|^2_{L^2(\Omega)} + \frac{\lambda}{2} \|u\|^2_{L^2(\Omega)} \tag{3.6a}$$

subject to

$$-\Delta y = \beta u \text{ in } \Omega, \quad y = 0 \text{ on } \Gamma, \tag{3.6b}$$

and

$$u_a \leq u \leq u_b \quad \text{in } \Omega \text{ a.e.}, \tag{3.6c}$$

where "a.e." stands for "almost everywhere". We have introduced the solution operator $S : L^2(\Omega) \to L^2(\Omega)$. From (3.5) we derive that an optimal solution to (3.6) satisfies the variational inequality

$$\langle S^*(Su^* - y_\Omega) + \lambda u^*, u - u^* \rangle_{L^2(\Omega)} \geq 0 \quad \text{for all } u \in U_{ad}. \tag{3.7}$$

To determine the adjoint operator S^* we make use of the following lemmas.

Lemma 3.13. *Suppose that $z, u \in L^2(\Omega)$, $c_0, \beta \in L^\infty(\Omega)$ with $c_0 \geq 0$ in Ω a.e. and $y, p \in H^1_0(\Omega)$ are the weak solutions to*

$$\begin{aligned} -\Delta y + c_0 y &= \beta u \quad \text{in } \Omega, & -\Delta p + c_0 p &= z \quad \text{in } \Omega, \\ y &= 0 \quad \text{on } \Gamma, & p &= 0 \quad \text{on } \Gamma. \end{aligned}$$

Then,

$$\int_\Omega zy \, \mathrm{d}x = \int_\Omega \beta p u \, \mathrm{d}x. \tag{3.8}$$

Proof. The weak formulations and $p, y \in H^1_0(\Omega)$ imply

$$\int_\Omega \nabla y \cdot \nabla p + c_0 yp \, \mathrm{d}x = \int_\Omega \beta up \, \mathrm{d}x$$

and

$$\int_\Omega \nabla p \cdot \nabla y + c_0 py \, \mathrm{d}x = \int_\Omega zy \, \mathrm{d}x.$$

Since the left-hand sides are equal, we obtain (3.8). $\square$

Lemma 3.14. *The adjoint operator $S^* : L^2(\Omega) \to L^2(\Omega)$ is given by*

$$S^* z = \beta p,$$

where $p \in H^1_0(\Omega)$ is the weak solution to $-\Delta p = z$ in Ω and $p = 0$ on Γ.

Proof. The adjoint operator S^* is defined by

$$\langle z, Su\rangle_{L^2(\Omega)} = \langle S^* z, u\rangle_{L^2(\Omega)} \quad \text{for all } z \in L^2(\Omega) \text{ and } u \in L^2(\Omega).$$

The claim follows from Lemma 3.13 with $c_0 = 0$ and $y = Su$:

$$\langle z, Su\rangle_{L^2(\Omega)} = \langle z, y\rangle_{L^2(\Omega)} = \langle \beta p, u\rangle_{L^2(\Omega)}.$$

The mapping $z \mapsto \beta p$ is linear and continuous from $L^2(\Omega)$ to $L^2(\Omega)$ (by the Lax-Milgram lemma). Since z and u are arbitrarily chosen and S^* is uniquely determined, we have $S^* z = \beta p$. □

Remark 3.15. The derivation of S^* is based on Lemma 3.13. However the construction is not obvious. Later we will discuss a technique which is based on a Lagrangian framework. In this case it is straightforward how the operator S^* can be computed. ♢

If S^* is known, (3.7) can be simplified.

Definition 3.16. *The weak solution $p \in H_0^1(\Omega)$ of the adjoint or dual equation*

$$-\Delta p = y_\Omega - y^* \text{ in } \Omega, \quad p = 0 \text{ on } \Gamma \tag{3.9}$$

with $y^* = Su^*$ is called the associated adjoint *or* dual state*.*

Recall that $y_\Omega \in L^2(\Omega)$. Furthermore $y^* \in H_0^1(\Omega) \hookrightarrow L^2(\Omega)$. Thus, $y_\Omega - y^*$ belongs to $L^2(\Omega)$. By Lax-Milgram there exists a unique adjoint state $p \in H_0^1(\Omega)$ satisfying (3.9). Choosing $z = y_\Omega - y^*$ in Lemma 3.14 we find

$$S^*(Su^* - y_\Omega) = S^*(y^* - y_\Omega) = -\beta p.$$

Hence, (3.7) implies

$$\langle \lambda u^* - \beta p, u - u^*\rangle_{L^2(\Omega)} \geq 0 \quad \text{for all } u \in U_{ad}.$$

Using (3.4) we derive the following result.

Theorem 3.17. *Suppose that u^* is an optimal solution to* (3.6) *and y^* the associated optimal state. Then there exists a unique solution $p \in H_0^1(\Omega)$ to* (3.9) *satisfying the variational inequality*

$$\int_\Omega (\lambda u^*(x) - \beta(x)p(x))\,(u(x) - u^*(x))\,\mathrm{d}x \geq 0 \quad \textit{for all } u \in U_{ad}. \tag{3.10}$$

On the contrary, if $u^* \in U_{ad}$ solves together with $y^* = Su^*$ and the solution p to (3.9) *the variational inequality* (3.10)*, then u^* is an optimal solution to* (3.6)*.*

Summarizing, a control u is optimal for (3.6) if and only if u satisfies together with y and p the following first-order necessary optimality system

$$\begin{aligned} -\Delta y &= \beta u, & -\Delta p &= y_\Omega - y,\\ y|_\Gamma &= 0, & p|_\Gamma &= 0,\\ u &\in U_{ad}, & &\\ \langle \lambda u - \beta p, v - u\rangle_{L^2(\Omega)} &\geq 0 & &\text{for all } v \in U_{ad}. \end{aligned} \tag{3.11}$$

Next we turn to a pointwise discussion of the optimality conditions. From (3.11) we derive

$$\int_\Omega (\lambda u^* - \beta p) u^*\,\mathrm{d}x \leq \int_\Omega (\lambda u^* - \beta p) u\,\mathrm{d}x \quad \text{for all } u \in U_{ad}.$$

This implies

$$\int_\Omega (\lambda u^* - \beta p) u^* \,\mathrm{d}x = \min_{u \in U_{ad}} \int_\Omega (\lambda u^* - \beta p) u \,\mathrm{d}x. \tag{3.12}$$

If $\lambda u^* - \beta p$ is known, (3.12) is a linear programming problem.

Lemma 3.18. *The variational inequality* (3.10) *holds if and only if for almost all $x \in \Omega$ we have*

$$u^*(x) \begin{cases} = u_a(x), & \text{if } \lambda u^*(x) - \beta(x)p(x) > 0, \\ \in [u_a(x), u_b(x)], & \text{if } \lambda u^*(x) - \beta(x)p(x) = 0, \\ = u_b(x), & \text{if } \lambda u^*(x) - \beta(x)p(x) < 0. \end{cases} \tag{3.13}$$

The following pointwise variational inequality is equivalent to (3.13)*:*

$$(\lambda u^*(x) - \beta(x)p(x))(v - u^*(x)) \geq 0 \text{ for all } v \in [u_a(x), u_b(x)], \ \text{f.a.a. } x \in \Omega. \tag{3.14}$$

Proof. 1) (3.10) $\Rightarrow$ (3.13): We suppose that (3.13) does not hold and define the measurable sets

$$A_+(u^*) = \{x \in \Omega : \lambda u^*(x) - \beta(x)p(x) > 0\},$$
$$A_-(u^*) = \{x \in \Omega : \lambda u^*(x) - \beta(x)p(x) < 0\},$$

where u^* is an arbitrary representant of the equivalence class for u^*. Analogously, u_a and u_b stand for arbitrary, but fixed representants. If the claim follows for the chosen representants, the claim holds also for any chosen representants. By assumption, (3.13) is not satisfied. Thus, there exists a set $E_+ \subset A_+(u^*)$ with positive measure and $u^*(x) > u_a(x)$ for all $x \in E_+$ or a set $E_- \subset A_-(u^*)$ with positive measure and $u^*(x) < u_b(x)$ for all $x \in E_-$. Let

$$u(x) = u_a(x) \text{ for } x \in E_+ \text{ and } u(x) = u^*(x) \text{ for } x \in \Omega \setminus E_+.$$

Then,

$$\begin{aligned} &\int_\Omega (\lambda u^*(x) - \beta(x)p(x))(u(x) - u^*(x)) \,\mathrm{d}x \\ &= \int_{E_+} (\lambda u^*(x) - \beta(x)p(x))(u_a(x) - u^*(x)) \,\mathrm{d}x < 0, \end{aligned}$$

because $\lambda u^* - \beta p > 0$ and $u_a < u^*$ on $E_+ \subset A_+(u^*)$. Since $u \in U_{ad}$ holds, we have a contradiction to (3.10). In the second case we proceed analogously and define

$$u = u_b \text{ on } E_- \text{ and } u = u^* \text{ on } \Omega \setminus E_-.$$

2) (3.13) $\Rightarrow$ (3.14): We have $u^* = u_a$ on $A_+(u^*)$ a.e. Thus, $v - u^*(x) \geq 0$ for any real number $v \in [u_a(x), u_b(x)]$ for $x \in A_+(u^*)$ a.e. Utilizing $\lambda u^*(x) - \beta(x)p(x) > 0$ in $A_+(u^*)$ we find

$$(\lambda u^*(x) - \beta(x)p(x))(v - u^*(x)) \geq 0 \quad \text{in } A_+(u^*) \text{ a.e.}$$

Analogously, we derive

$$(\lambda u^*(x) - \beta(x)p(x))(v - u^*(x)) \geq 0 \quad \text{in } A_-(u^*) \text{ a.e.}$$

Clearly, (3.14) holds on $\Omega \setminus (A_+(u^*) \cup A_-(u^*))$ a.e.

3) (3.14) $\Rightarrow$ (3.10): Let $u \in U_{ad}$ be chosen arbitrarily. We have $u(x) \in [u_a(x), u_b(x)]$ f.a.a. $x \in \Omega$. Using (3.14) with $v := u(x)$ we have

$$(\lambda u^*(x) - \beta(x)p(x))(u(x) - u^*(x)) \geq 0 \quad \text{f.a.a. } x \in \Omega.$$

By integrating (3.10) follows immediately.

□

From (3.14) we deduce

$$(3.15)\quad (\lambda u^*(x) - \beta(x)p(x))u^*(x) \leq (\lambda u^*(x) - \beta(x)p(x))v \text{ for all } v \in [u_a(x), u_b(x)].$$

Theorem 3.19. *The control $u^* \in U_{ad}$ is optimal for* (3.6) *if and only if one of the following conditions holds f.a.a. $x \in \Omega$: the weak minimum principle*

$$\min_{v \in [u_a(x), u_b(x)]} (\lambda u^*(x) - \beta(x)p(x))v = (\lambda u^*(x) - \beta(x)p(x))u^*(x)$$

or the minimum principle

$$\min_{v \in [u_a(x), u_b(x)]} \left(\frac{\lambda}{2} v^2 - \beta(x)p(x)v \right) = \frac{\lambda}{2} u^*(x)^2 - \beta(x)p(x)u^*(x),$$

where p is the dual variable solving (3.9) *with $y^* = Su^*$.*

Proof. The weak minimum principle follows directly from (3.15). We turn to the minimum principle and consider the (convex) quadratic optimization problem in $\mathbb{R}$:

$$(3.16)\qquad \min_{v \in [u_a(x), u_b(x)]} g(v) = \frac{\lambda}{2} v^2 - \beta(x)p(x)v.$$

The real number v^* solves (3.16) for any fixed $x \in \Omega$ if and only if v^* satisfies the variational inequality

$$g'(v^*)(v - v^*) \geq 0 \quad \text{for all } v \in [u_a(x), u_b(x)],$$

Consequently,

$$(\lambda v^* - \beta(x)p(x))(v - v^*) \geq 0 \quad \text{for all } v \in [u_a(x), u_b(x)].$$

which holds with $v^* = u^*(x)$. □

From the choice for the regularization parameter λ we can deduce further consequences.

Case $\lambda = 0$. Using (3.13) we find

$$u^*(x) = \begin{cases} u_a(x) & \text{if } \beta(x)p(x) < 0, \\ u_b(x) & \text{if } \beta(x)p(x) > 0. \end{cases}$$

If $\beta(x)p(x) = 0$ holds, we do not get any information for $u^*(x)$. In the case $\beta(x)p(x) \neq 0$ f.a.a. $x \in \Omega$ we have $u^*(x) = u_a(x)$ or $u^*(x) = u_b(x)$ f.a.a. $x \in \Omega$. In this case we have a so-called bang-bang control.

Case $\lambda > 0$. We derive from (3.13) that $u^*(x) = \frac{1}{\lambda}\beta(x)p(x)$ holds if $\lambda u^*(x) - \beta(x)p(x) = 0$. This leads to the following theorem.

Theorem 3.20. *Let $\lambda > 0$. Then u^* is a solution to* (3.6) *if and only if*

$$u^*(x) = \mathbb{P}_{[u_a(x),u_b(x)]}\left(\frac{1}{\lambda}\beta(x)p(x)\right) \quad \text{f.a.a. } x \in \Omega, \tag{3.17}$$

where $\mathbb{P}_{[a,b]}$, $a < b$, is the projection of $\mathbb{R}$ on $[a,b]$ given by

$$\mathbb{P}_{[a,b]}(u) := \min(b, \max(a,u)).$$

Proof. Theorem 3.20 follows directly from Theorem 3.19: the solution to

$$\min_{v\in[u_a(x),u_b(x)]} \left(\frac{\lambda}{2}v^2 - \beta(x)p(x)v\right)$$

is $v = \mathbb{P}_{[u_a(x),u_b(x)]}(\frac{1}{\lambda}\beta(x)p(x))$, where we have used that

$$\frac{\lambda}{2}v^2 - \beta(x)p(x)v = \frac{\lambda}{2}\left(v - \frac{1}{\lambda}\beta(x)p(x)\right)^2 + \frac{1}{2\lambda}\beta(x)^2p(x)^2$$

holds. □

Case $\lambda > 0$ *and* $U_{ad} = L^2(\Omega)$ *(no control constraints).* From (3.14) or (3.17) it follows directly

$$u^* = \frac{1}{\lambda}\beta p. \tag{3.18}$$

Thus, we obtain the following optimality system

$$\begin{aligned} -\Delta y &= \frac{1}{\lambda}\beta^2 p, & \qquad -\Delta p &= y_\Omega - y, \\ y|_\Gamma &= 0, & p|_\Gamma &= 0 \end{aligned}$$

which is a coupled system of two elliptic equations. If p is computed, u^* is given by (3.18).

By introducing a Lagrange multiplier the variational inequality (3.10) can be formulated as an additional equation, compare Section 1.7.

Theorem 3.21. *The variational inequality* (3.10) *is equivalent to the existence of two functions $\mu_a, \mu_b \in L^2(\Omega)$ satisfying $\mu_a, \mu_b \geq 0$ in Ω a.e.,*

$$\lambda u^* - \beta p + \mu_b - \mu_a = 0 \tag{3.19}$$

and the complementarity condition

$$\mu_a(x)(u_a(x) - u^*(x)) = \mu_b(x)(u^*(x) - u_b(x)) = 0 \quad \text{f.a.a. } x \in \Omega. \tag{3.20}$$

Proof. a) First we show that (3.19) and (3.20) follow from (3.10). Analogously to Section 1.7 we define

$$\begin{aligned} \mu_a(x) &= (\lambda u^*(x) - \beta(x)p(x))_+ \quad \text{f.a.a. } x \in \Omega, \\ \mu_b(x) &= (\lambda u^*(x) - \beta(x)p(x))_- \quad \text{f.a.a. } x \in \Omega, \end{aligned} \tag{3.21}$$

where $s_+ = \frac{1}{2}(s + |s|)$ and $s_- = \frac{1}{2}(|s| - s)$ for $s \in \mathbb{R}$. Then, we have $\mu_a, \mu_b \geq 0$ f.a.a. $x \in \Omega$. Moreover,

$$\lambda u^*(x) - \beta(x)p(x) = \mu_a - \mu_b,$$

which is (3.19). From (3.13) it follows that

$$\begin{aligned}(\lambda u^* - \beta p)(x) > 0 \quad &\Rightarrow \quad u^*(x) = u_a(x),\\ (\lambda u^* - \beta p)(x) < 0 \quad &\Rightarrow \quad u^*(x) = u_b(x),\\ u_a(x) < u^*(x) < u_b(x) \quad &\Rightarrow \quad (\lambda u^* - \beta p)(x) = 0.\end{aligned}$$

Thus, (3.20) holds because one of the factors is zero. For instance, if $\mu_a(x) > 0$ is satisfied, then $\mu_b(x) = 0$. Hence, $\lambda u^*(x) - \beta(x)p(x) = \mu_a(x) > 0$. This implies $u^*(x) - u_a(x) = 0$. For almost all $x \in \Omega$ with $u^*(x) > u_a(x)$ we have $(\lambda u^* - \beta p)(x) \le 0$. Due to (3.21) we have $\mu_a(x) = 0$. Summarizing, $\mu_a(x)(u_a(x) - u^*(x)) = 0$. Analogously, the second identity in (3.20) follows.

b) Suppose that (3.19)-(3.20) and $u^* \in U_{ad}$ hold. Let $u \in U_{ad}$ and $x \in \Omega$ be chosen arbitrarily. Then we have three cases.
$u^*(x) \in (u_a(x), u_b(x))$: Using (3.20) we find $\mu_a(x) = \mu_b(x) = 0$. By (3.19) it follows that

$$(\lambda u^* - \beta p)(x) = 0,$$

so that

$$(\lambda u^*(x) - \beta(x)p(x))(u(x) - u^*(x)) \ge 0. \tag{3.22}$$

$u_a(x) = u^*(x)$: Now $u(x) - u^*(x) \ge 0$ for $u \in U_{ad}$. In addition, we have (3.20), so that $\mu_b(x) = 0$ holds. Using (3.19) we conclude

$$\lambda u^*(x) - \beta(x)p(x) = \mu_a(x) \ge 0,$$

which implies (3.22).
$u_b(x) = u^*(x)$: Utilizing $u(x) - u^*(x) \le 0$ for $u \in U_{ad}$, (3.19) and (3.20) we derive (3.22), because $\mu_a(x) = 0$ and

$$\lambda u^*(x) - \beta(x)p(x) = -\mu_b(x) \le 0.$$

Summarizing, by integration, (3.22) implies (3.10).

□

Theorem 3.21 directly leads to the following Karush-Kuhn-Tucker (KKT) System (compare (3.11))

$$\begin{aligned}&-\Delta y = \beta u, && -\Delta p = y_\Omega - y,\\ &y|_\Gamma = 0, && p|_\Gamma = 0,\\ &\lambda u - \beta p - \mu_a + \mu_b = 0, && \\ &u_a \le u \le u_b, \quad \mu_a,\, \mu_b \ge 0, && \mu_a(u_a - u) = \mu_b(u - u_b) = 0.\end{aligned}$$

Definition 3.22. *The functions μ_a and μ_b defined by* (3.21) *are called* Lagrange multipliers *associated with the inequality constraints $u_a \le u$ and $u \le u_b$, respectively.*

Using the adjoint state p one can express the gradient of the reduced cost functional $\hat{J}(u) = J(y(u), u)$.

Lemma 3.23. *The gradient of*

$$\hat{J}(u) = J(y(u), u) = \frac{1}{2}\|y - y_\Omega\|^2_{L^2(\Omega)} + \frac{\lambda}{2}\|u\|^2_{L^2(\Omega)}$$

is given by

$$\nabla \hat{J}(u) = \lambda u - \beta p,$$

where $p \in H_0^1(\Omega)$ *is the weak solution of the adjoint equation*

$$\Delta p = y_\Omega - y \text{ in } \Omega, \quad p = 0 \text{ on } \Gamma \tag{3.23}$$

and $y = y(u)$ *is the associated state variable.*

Proof. In the proof of Theorem 3.12 we have derived

$$\hat{J}'(u)h = \langle S^*(Su - y_\Omega) + \lambda u, h \rangle_{L^2(\Omega)} \quad \text{for any } h \in L^2(\Omega).$$

By Lemma 3.14 we have $S^*(y - y_\Omega) = -\beta p$, where p solves (3.23). Thus,

$$\hat{J}'(u)h = \langle \lambda u - \beta p, h \rangle_{L^2(\Omega)} \quad \text{for any } h \in L^2(\Omega).$$

Utilizing the Riesz theorem, we identify $\hat{J}'(u)$ with $\nabla \hat{J}(u) = \lambda u - \beta p$. ☐

3.3.3. The formal Lagrange principle. In this section we derive the optimality conditions by utilizing the Lagrange functional. We treat all Lagrange multipliers as functions in $L^2(\Omega)$ without any proof. Therefore, we call this procedure “formal”. But the main goal of this subsection is to explain the used strategy which can be also applied to much more complex problems.

We consider the following optimal control problem

$$\min J(y,u) := \frac{1}{2} \|y - y_\Omega\|^2_{L^2(\Omega)} + \frac{\lambda}{2} \|u\|^2_{L^2(\Gamma)} \tag{3.24a}$$

subject to

$$-\Delta y = 0 \text{ in } \Omega, \quad \frac{\partial y}{\partial n} + \alpha y = \alpha u \text{ on } \Gamma, \tag{3.24b}$$

$$u_a(x) \le u(x) \le u_b(x) \quad \text{on } \Gamma \text{ a.e.} \tag{3.24c}$$

To ensures existence and uniqueness of a solution to (3.24b) we suppose that

$$\int_\Gamma \alpha(x)^2 \,\mathrm{d}x > 0$$

holds. We define formally the Lagrange function for (3.24)

$$\mathcal{L}(y,u,p) = J(y,u) + \int_\Omega -\Delta y \, p_1 \,\mathrm{d}x + \int_\Gamma \Big(\frac{\partial y}{\partial n} - \alpha(u - y) \Big) p_2 \,\mathrm{d}s,$$

where $p_1 : \Omega \mapsto \mathbb{R}$ and $p_1 : \Gamma \mapsto \mathbb{R}$ are the Lagrange multipliers associated with the partial differential equation and the boundary condition, respectively. We set $p = (p_1, p_2)$.

The definition of $\mathcal{L}$ is not completely clear. From $y \in H^1(\Omega)$ we conclude that Δy and $\frac{\partial y}{\partial n}$ need not to be functions. Without any further regularity conditions we only have $\Delta y \in H^1(\Omega)^*$ and $\frac{\partial y}{\partial n} \in H^{-1/2}(\Gamma)$. Thus, the integrals are not well-defined. Moreover, we have to specify the regularities of p_1 and p_2.

However, we continue by applying Green's formula twice

$$\mathcal{L}(y,u,p) = J(y,u) + \int_\Gamma y \frac{\partial p_1}{\partial n} - p_1 \frac{\partial y}{\partial n} \,\mathrm{d}s - \int_\Omega y \Delta p_1 \,\mathrm{d}x + \int_\Gamma \left(\frac{\partial y}{\partial n} - \alpha(u - y) \right) p_2 \,\mathrm{d}s.$$

From the Lagrange principle we conclude that (y^*, u^*) together with the Lagrange multipliers p_1, p_2 satisfies the first-order necessary optimality conditions for

$$\min_{(y,u)} \mathcal{L}(y,u,p) \quad \text{s.t. } u \in U_{ad}.$$

Since there are no restrictions for y we have

$$D_y\mathcal{L}(y^*,u^*,p)h = 0 \quad \text{for all } h \in Y = H^1(\Omega).$$

Utilizing the box constraints we find that

$$D_u\mathcal{L}(y^*,u^*,p)(u-u^*) \geq 0 \quad \text{for all } u \in U_{ad}.$$

Notice that

$$\begin{aligned} D_y\mathcal{L}(y^*,u^*,p)h &= \int_\Omega ((y^*-y_\Omega)-\Delta p_1)h\,\mathrm{d}x + \int_\Gamma (p_2-p_1)\frac{\partial h}{\partial n}\,\mathrm{d}s \\ &\quad + \int_\Gamma \left(\frac{\partial p_1}{\partial n} + \alpha p_2\right) h\,\mathrm{d}s = 0 \quad \text{for all } h \in Y, \end{aligned} \tag{3.25}$$

where we have used that the derivative of a linear mapping is the linear mapping itself. We choose specific direction $h \in Y$. More precisely, let $h \in C_0^\infty(\Omega) \subset Y$ satisfying $h = \frac{\partial h}{\partial n} = 0$ on Γ. From (3.25) we find

$$\int_\Omega (y^* - y_\Omega - \Delta p_1)\,h\,\mathrm{d}x = 0 \quad \text{for all } h \in C_0^\infty(\Omega).$$

Recall that $C_0^\infty(\Omega)$ is dense in $L^2(\Omega)$. Therefore,

$$-\Delta p_1 = y_\Omega - \bar{y} \quad \text{in } \Omega.$$

This is the first adjoint equation. In contrast to the approach in Section 3.3.2 we derive the adjoint equation utilizing the Lagrange principle. Next we assume that $h \in Y$ satisfies $h|_\Gamma = 0$. Then,

$$\int_\Gamma (p_2-p_1)\frac{\partial h}{\partial n}\,\mathrm{d}s = 0 \quad \text{for all } h \in C^\infty(\Omega) \text{ with } h|_\Gamma = 0.$$

Thus, $p_2 = p_1$ on Γ a.e. Finally, let $h \in C^\infty(\Omega)$ satisfy $\frac{\partial h}{\partial n} = 0$. Then, (3.25) implies

$$0 = \int_\Gamma \left(\frac{\partial p_1}{\partial n} + \alpha p_2\right) h\,\mathrm{d}s = \int_\Gamma \left(\frac{\partial p_1}{\partial n} + \alpha p_1\right) h\,\mathrm{d}s.$$

Consequently,

$$\frac{\partial p_1}{\partial n} + \alpha p_1 = 0 \quad \text{on } \Gamma.$$

Summarizing, p_1 and p_2 satisfy

$$\begin{aligned} -\Delta p_1 = y_\Omega - y^* \text{ in } \Omega, \quad & \frac{\partial p_1}{\partial n} + \alpha p_1 = 0 \text{ on } \Gamma, \\ p_1 = p_2 \text{ on } \Gamma. \quad & \end{aligned} \tag{3.26}$$

Analogously, we find

$$D_u\mathcal{L}(y^*,u^*,p)(u-u^*) = \int_\Gamma (\lambda u^* - \alpha p)(u-u^*)\,\mathrm{d}s \geq 0.$$

Now we introduce the Lagrange functional so that all integrals are well-defined.

Definition 3.24. *The Lagrange function* $\mathcal{L} : H^1(\Omega) \times L^2(\Gamma) \times H^1(\Omega) \to \mathbb{R}$ *for* (3.24) *is defined by*

$$\mathcal{L}(y,u,p) = J(y,u) + \int_\Omega \nabla y \cdot \nabla p\,\mathrm{d}x + \int_\Gamma \alpha(u-y)p\,\mathrm{d}s.$$

Notice that

1) $D_y\mathcal{L}(y^*, u^*, p)h = 0$ for all $h \in H^1(\Omega)$ is equivalent with the weak form of (3.26),
2) $D_u\mathcal{L}(y^*, u^*, p)(u - u^*) \geq 0$ for all $u \in U_{ad}$ is equivalent with the variational inequality $\hat{J}'(u^*)(u - u^*) \geq 0$ for all $u \in U_{ad}$.

Remark 3.25. 1) Let us mention that we have ensured the existence of the Lagrange multiplier $p \in H^1(\Omega)$ by the solvability of the adjoint equation $-\Delta p = y_\Omega - y^*$ in Ω, $\frac{\partial p}{\partial n} + \alpha p = 0$ on Γ. The existence can also be guaranteed by the Karush-Kuhn-Tucker theory in Banach spaces.

2) The gradient of the reduced cost functional $\hat{J}(u) = J(y(u), u)$ is given by

$$\hat{J}'(u) = D_u\mathcal{L}(y(u), u, p(u)).$$

3) Let us mention that the mappings

$$\tau_1 : H^2(\Omega) \to H^{3/2}(\Gamma) \times H^{1/2}(\Gamma), \quad h \mapsto \left(h|_\Gamma, \frac{\partial h}{\partial n}\Big|_\Gamma \right),$$
$$\tau_2 : H^1(\Omega) \to H^{1/2}(\Gamma), \quad h \to h|_\Gamma$$

are surjective. Therefore, we can choose h satisfying $\frac{\partial h}{\partial n}|_\Gamma = 0$ or $h|_\Gamma = 0$.$\Diamond$

We can extend the Lagrange function by including the inequality constraints. In this case we define

$$\mathcal{L}(y, u, p, \mu_a, \mu_b) = J(y, u) + \int_\Omega \nabla y \cdot \nabla p \,\mathrm{d}x + \int_\Gamma \alpha(y - u)p \,\mathrm{d}s + \int_\Omega \mu_a(u_a - u) + \mu_b(u - u_b) \,\mathrm{d}x.$$

4. Existence of optimal controls for semilinear optimal control problems

In this subsection we consider optimal control problems governed by semilinear elliptic differential equations. An example is given by the following problem

$$\min J(y, u) = \frac{1}{2}\int_\Omega |y - y_\Omega|^2 \,\mathrm{d}x + \int_\Omega u^2 \,\mathrm{d}x \tag{4.1a}$$

$$\text{s.t. } -\Delta y + y + y^3 = u \text{ in } \Omega, \quad \frac{\partial y}{\partial n} = 0 \text{ on } \Gamma \tag{4.1b}$$

$$u_a \leq u \leq u_b \quad \text{for a.e. } x \in \Omega. \tag{4.1c}$$

We call the elliptic equation (4.1b) *semilinear*. It will be shown that the adjoint equation for (4.1) has the form

$$-\Delta p + p + 3(y^*)^2 p = y_\Omega - y^* \text{ in } \Omega, \quad \frac{\partial p}{\partial n} = 0 \text{ on } \Gamma, \tag{4.2}$$

where y^* is an optimal state for (4.1), i.e.,

$$-\Delta y^* + y^* + (y^*)^3 = u^* \text{ in } \Omega, \quad \frac{\partial y^*}{\partial n} = 0 \text{ on } \Gamma, \tag{4.3}$$

with an optimal control u^*. Moreover, we have

$$u^* = \mathbb{P}_{[u_a(x), u_b(x)]}\left(\frac{1}{\lambda}p(x)\right) \tag{4.4}$$

as in the linear-quadratic case; see (3.17). Summarizing, the first-order necessary optimality conditions consist in (4.2), (4.3), (4.4) and $u^* \in U_{ad} = \{v \in$

$L^2(\Omega) \,|\, u_a(x) \le v(x) \le u_b(x)$ in Ω a.e.$\}$. To justify our formal derivation we have to prove the differentiability of the non-linear mapping $y(\cdot) \mapsto y(\cdot)^3$. Here, the choice of the right function space is not trivial. Moreover, (4.1) is a non-convex optimal control problem. Although J is convex, the non-linear term in (4.1b) makes the problem non-convex. Therefore, the first-order necessary optimality conditions are not sufficient, so that the second-order conditions come into play. As a consequence, (4.1) may posses several solutions.

4.1. Semilinear elliptic equations. Let us consider the following problem

$$\mathcal{A}y + c_0(x)y + d(x,y) = f \text{ in } \Omega \tag{4.5a}$$

$$\partial_{\nu_{\mathcal{A}}} y + \alpha(x)y + b(x,y) = g \text{ on } \Gamma. \tag{4.5b}$$

First we discuss existence and uniqueness of solutions to (4.5) in $H^1(\Omega)$. The proofs are based on the theory of monotone operators. The basic idea is the following. Suppose that $a : \mathbb{R} \mapsto \mathbb{R}$ is continuous and strictly monotonously increasing with $\lim_{x\to\pm\infty} a(x) = \pm\infty$. Then there exists a unique solution y to $a(y) = f$ for any $f \in \mathbb{R}$. This argument can also be applied to more complex equations in Banach spaces. Suppose that V is a real and separable Banach space, e.g., $V = H^1(\Omega)$ or $V = H_0^1(\Omega)$. Recall that a Banach space is called *separable* if there exists a countable and dense subset $\mathcal{V} \subset V$.

Definition 4.1. *The operator $A : V \mapsto V^*$ is called* monotone *if*

$$\langle Ay_1 - Ay_2, y_1 - y_2\rangle_{V^*,V} \ge 0 \quad \text{for all } y_1, y_2 \in V$$

holds. Furthermore, A is said to be strictly monotone *if*

$$\langle Ay_1 - Ay_2, y_1 - y_2\rangle_{V^*,V} > 0 \quad \text{for all } y_1, y_2 \in V \text{ with } y_1 \ne y_2.$$

If

$$\frac{\langle Ay, y\rangle_{V^*,V}}{\|y\|_V} \to \infty \quad \text{for } \|y\|_V \to \infty,$$

the operator A is called coercive. *The operator A is said to be* hemicontinuous *if the function*

$$\varphi : [0,1] \to \mathbb{R},\ t \mapsto (A(y+tv), w)_{V^*,V}$$

is continuous for any fixed $y, v, w \in V$. If there exist a constant $\beta_0 > 0$ with

$$\langle Ay_1 - Ay_2, y_1 - y_2\rangle_{V^*,V} \ge \beta_0 \|y_1 - y_2\|_V^2 \quad \text{for all } y_1, y_2 \in V$$

then we say that A is strong monotone.

Theorem 4.2. *Suppose that V is a separable Hilbert space and $A : V \mapsto V^*$ a monotone, coercive, hemicontinuous operator. Then there exists a solution to the equation $Ay = f$ for every $f \in V^*$. The set of solutions is bounded, convex and closed. If A is strictly monotone, then y is uniquely determined. If A is strong monotone, then $A^{-1} : V^* \to V$ is Lipschitz continuous.*

A proof of Theorem 4.2 can be found in [4]. We apply Theorem 4.2 to problem (4.5) with $V = H^1(\Omega)$. First we introduce a weak solution to (4.5). For that purpose we multiply (4.5a) by a test function in $H^1(\Omega)$ and integrate over Ω. After applying Green's formula and using (4.5b) we find

$$\begin{aligned}\int_\Omega (\mathcal{A}\nabla y)\cdot\nabla\varphi + c_0 y\varphi + d(\cdot,y)\varphi\,\mathrm{d}x + \int_\Gamma \alpha y\varphi + b(\cdot,y)\varphi\,\mathrm{d}s(x)\\ = \int_\Omega f\varphi\,\mathrm{d}x + \int_\Gamma g\varphi\,\mathrm{d}s \quad \text{for all } \varphi\in H^1(\Omega).\end{aligned} \tag{4.6}$$

In (4.6) we have the problem that the integrals

$$\int_\Omega d(\cdot,y)\varphi\,\mathrm{d}x \quad\text{and}\quad \int_\Gamma b(\cdot,y)\varphi\,\mathrm{d}s(x)$$

are not well-defined if d and b are unbounded, respectively. For instance, the functions e^y and y^k may not be bounded for functions $y\in H^1(\Omega)$. Therefore, we make use of the following assumptions.

Assumption 2. *Let $\Omega\subset\mathbb{R}^N$, $N\geq 2$, be a bounded Lipschitz domain with boundary Γ. We suppose that f is of the form*

$$(fy)(x) = -\sum_{i,j=1}^N D_i(a_{ij}(x)D_j y(x))$$

with bounded and measurable coefficients a_{ij} satisfying

$$a_{ij}=a_{ji},\ 1\leq i,j\leq N, \quad\textit{and}\quad \sum_{i,j=1}^N a_{ij}(x)\xi_i\xi_j \geq \gamma_0|\xi|^2 \quad \textit{for all } \xi\in\mathbb{R}^N,$$

where $\gamma_0>0$ is a constant. Let $c_0\in L^\infty(\Omega)$ and $\alpha_0\in L^\infty(\Gamma)$ with $c_0\geq 0$ in Ω a.e. and $\alpha_0\geq 0$ on Γ a.e., satisfying

$$\int_\Omega c_0^2(x)\,\mathrm{d}x + \int_\Gamma \alpha^2(x)\,\mathrm{d}s(x) > 0.$$

The functions $d=d(x,y):\Omega\times\mathbb{R}\mapsto\mathbb{R}$ and $b=b(x,y):\Gamma\times\mathbb{R}\mapsto\mathbb{R}$ are bounded and measurable with respect to x for any fixed y and are continuous and monotonously increasing with respect to y for almost all x. (In particular, $d(x,0)$ and $b(x,0)$ are bounded and measurable).

To handle the problem of unboundedness we also utilize the following assumption.

Assumption 3. *For almost all $x\in\Gamma$ or $x\in\Omega$ we have $b(x,0)=0$ or $d(x,0)=0$, respectively. Moreover, there exists a constant $M>0$ so that*

$$|b(x,y)|\leq M \quad\textit{and}\quad |d(x,y)|\leq M$$

for almost all $x\in\Gamma$ or $x\in\Omega$ and for all $y\in\mathbb{R}$.

Let us define the bilinear form

$$a(y,v)=\int_\Omega\sum_{i,j=1}^N a_{ij}D_i y(x)D_j v(x)\,\mathrm{d}x + \int_\Omega c_0(x)y(x)v(x)\,\mathrm{d}x + \int_\Gamma \alpha(x)y(x)v(x)\,\mathrm{d}s.$$

Definition 4.3. *With Assumptions* 2 *and* 3 *holding, we call $y\in H^1(\Omega)$ a* weak solution to (4.5) *provided*

$$a(y,v)+\int_\Omega d(\cdot,y)v\,\mathrm{d}x + \int_\Gamma b(\cdot,y)v\,\mathrm{d}s(x) = \int_\Omega fv\,\mathrm{d}x + \int_\Gamma gv\,\mathrm{d}s(x).$$

Theorem 4.4. *With Assumptions* 2 *and* 3 *holding there exists a unique weak solution to* (4.5) *for every* $(f,g) \in L^2(\Omega) \times L^2(\Gamma)$*. In particular, there exists a constant* $c_M > 0$ *independent of* d, b, f *and* g *so that*

$$\|y\|_{H^1(\Omega)} \leq c_M \big(\|f\|_{L^2(\Omega)} + \|g\|_{L^2(\Gamma)} \big).$$

Remark 4.5. a) Notice that in (4.1b) we have $d(x,y) = y^3$, which does not satisfy Assumption 3. In the proof of Theorem 4.4 the boundedness of d is utilized to prove that d belongs to $L^2(\Omega)$. From $y \in H^1(\Omega)$ we derive that $y \in L^6(\Omega)$ for $\Omega \subset \mathbb{R}^3$. Hence, $y^3 \in L^2(\Omega)$ holds.

b) Theorem 4.4 remains valid provided $f \in L^r(\Omega)$ and $g \in L^s(\Gamma)$ with $r > \frac{N}{2}$ and $s > N-1$; compare Theorem 4.6 below. For $N \in \{2,3\}$ we observe that both r and s can be smaller than 2. ♢

Following [1] we can prove continuity of the weak solution.

Theorem 4.6. *Suppose that Assumptions* 2 *and* 3 *hold. Moreover, let* $r > N/2$, $s > N-1$*. Then there exists a unique weak solution* $y \in H^1(\Omega) \cap L^\infty(\Omega)$ *to* (4.5) *for any* $f \in L^r(\Omega)$ *and* $g \in L^s(\Gamma)$*. Furthermore, the following estimate hold*

$$\|y\|_{L^\infty(\Omega)} \leq c_\infty \big(\|f\|_{L^r(\Omega)} + \|g\|_{L^s(\Gamma)} \big) \tag{4.7}$$

for a constant $c_\infty > 0$*, which is independent of* d, b, f, g*.*

Remark 4.7. Let us mention that

$$\|y\|_{L^\infty(\Gamma)} \leq \|y\|_{L^\infty(\Omega)}$$

Thus, estimate (4.7) also holds for $\|y\|_{L^\infty(\Gamma)}$. ♢

The boundedness of d and b is not essential for the previous existence results. The important property is the monotonicity. In [1], this fact is utilized to prove the next theorem.

Theorem 4.8. *Suppose that Assumption* 2 *holds,* $\Omega \subset \mathbb{R}^N$ *is a bounded Lipschitz domain and* $r > N/2$, $s > N-1$*. In addition,* $b(x,0) = 0$ *and* $d(x,0) = 0$ *for almost all* $x \in \Gamma$ *and* $x \in \Omega$*, respectively. Then,* (4.5) *possesses a unique solution* $y \in H^1(\Omega) \cap L^\infty(\Omega)$ *for any* $(f,y) \in L^r(\Omega) \times L^s(\Gamma)$*. Furthermore, there exists a constant – independent of* d, b, f *and* y *– satisfying*

$$\|y\|_{H^1(\Omega)} + \|y\|_{C(\overline{\Omega})} \leq c_\infty \big(\|f\|_{L^r(\Omega)} + \|g\|_{L^s(\Gamma)} \big) \tag{4.8}$$

Remark 4.9. 1) If the assumptions $b(x,0) = 0$ and $d(x,0) = 0$ do not hold, we obtain

$$\|y\|_{H^1(\Omega)} + \|y\|_{C(\overline{\Omega})} \leq c_\infty \Big(\|f - d(\cdot,0)\|_{L^r(\Omega)} + \|g - b(\cdot,0)\|_{L^s(\Gamma)} \Big).$$

2) In [3] weaker assumptions are presented, so that the existence of a unique solution $y \in H^1(\Omega) \cap C(\overline{\Omega})$ to

$$-\Delta y + e^y = f \text{ in } \Omega, \quad \frac{\partial y}{\partial n} = 0 \text{ on } \Gamma$$

and

$$-\Delta y + y^3 = f \text{ in } \Omega, \quad \frac{\partial y}{\partial n} = 0 \text{ on } \Gamma$$

follows. ♢

4.2. Existence of optimal controls. We make use of the following assumption.

Assumption 4. a) *$\Omega \subset \mathbb{R}^N$ is a bounded Lipschitz domain.*

b) *The functions $d = d(x,y)$, $\varphi = \varphi(x,y) : \Omega\times\mathbb{R} \to \mathbb{R}$, $b = b(x,y) : \Gamma\times\mathbb{R} \to \mathbb{R}$ and $\psi = \psi(x,y) : E \times \mathbb{R} \to \mathbb{R}$, $E = \Omega$ or $E = \Gamma$, are measurable in x for all $y \in \mathbb{R}$ or $u \in \mathbb{R}$ and two times differentiable with respect to y or u for almost all $x \in \Omega$ or $x \in \Gamma$. Moreover there exist constant $R \geq 0$ and $L(M) \geq 0$, such that the following conditions hold*

$$\begin{aligned} |d(x,0)| + |d_y(x,0)| + |d_{yy}(x,0)| &\le R && \text{f.a.a. } x \in \Omega,\\ |\varphi(x,0)| + |\varphi_y(x,0)| + |\varphi_{yy}(x,0)| &\le R && \text{f.a.a. } x \in \Omega,\\ |b(x,0)| + |b_y(x,0)| + |b_{yy}(x,0)| &\le R && \text{f.a.a. } x \in \Gamma,\\ |\psi(x,0)| + |\psi_y(x,0)| + |\psi_{yy}(x,0)| &\le R && \text{f.a.a. } x \in E \end{aligned}$$

and

$$\begin{aligned} |d_{yy}(x,y_1)| + |d_{yy}(x,y_2)| &\le L(M)\,|y_1 - y_2| && \text{f.a.a. } x \in \Omega,\\ |\varphi_{yy}(x,y_1)| + |\varphi_{yy}(x,y_2)| &\le L(M)\,|y_1 - y_2| && \text{f.a.a. } x \in \Omega,\\ |b_{yy}(x,y_1)| + |b_{yy}(x,y_2)| &\le L(M)\,|y_1 - y_2| && \text{f.a.a. } x \in \Gamma,\\ |\psi_{yy}(x,u_1)| + |\psi_{yy}(x,u_2)| &\le L(M)\,|u_1 - u_2| && \text{f.a.a. } x \in E, \end{aligned}$$

and for all $y_1, y_2 \in [-M,M]$ and all $u_1, u_2 \in [-M,M]$. Here, $L = L(M)$ depends on $M > 0$.

c) *For almost all $x \in \Omega$ or $x \in \Gamma$ and for all $y \in \mathbb{R}$ we have*

$$d_y(x,y) \geq 0 \quad \text{and} \quad b_y(x,y) \geq 0.$$

In addition, there are subsets $E_d \subset \Omega$ and $E_b \subset \Gamma$ with positive measures satisfying

$$d_y(x,y) \geq \lambda_d \text{ for all } (x,y) \in E_d \times \mathbb{R},\ b_y(x,y) \geq \lambda_b \text{ for all } (x,y) \in E_b \times \mathbb{R}$$

for positive constants λ_d and λ_b.

d) *The lower and upper bounds u_a, v_a, u_b, $v_b : E \to \mathbb{R}$ belong to $L^\infty(E)$ for $E = \Omega$ or $E = \Gamma$. Moreover, $u_a(x) \leq u_b(x)$ and $v_a(x) \leq v_b(x)$ for almost all $x \in E$.*

Remark 4.10. To prove existence of optimal controls we only need part b) for the functions φ and ψ, but not for the their derivatives. For the first-order optimality conditions, the conditions for the second-order derivatives are not needed. ◊

Example 4.11. The following functions satisfy Assumption 4:

$$\begin{aligned} &\varphi(x,y) = a(x)y + \beta(x)(y - y_\Omega(x))^2 \quad \text{with } a, \beta, y_\Omega \in L^\infty(\Omega),\\ &d(x,y) = c_0(x)y + y^k \quad \text{if } k \in \mathbb{N} \text{ is odd and } c_0(x) \geq 0 \text{ in } \Omega, \|c_0\|_{L^\infty(\Omega)} > 0,\\ &d(x,y) = c_0(x)y + \exp(a(x)y) \quad \text{with } 0 \leq a \in L^\infty(\Omega) \text{ and } c_0 \text{ as above.} \end{aligned}$$

◊

With Assumption 4 holding, we can apply Theorem 4.6. We write

$$\begin{aligned} d(x,y) &= c_0(x)y + (d(x,y) - c_0(x)y) = c_0(x)y + \tilde{d}(x,y)\\ b(x,y) &= \alpha_0(x)y + (b(x,y) - \alpha_0(x)y) = \alpha_0(x)y + \tilde{b}(x,y) \end{aligned} \tag{4.9}$$

with $c_0 = \chi_{E_d}\lambda_d$ and $\alpha_0 = \chi_{E_b}\lambda_b$. Then, $\tilde{d}$ and $\tilde{b}$ satisfy Assumption 2-2). Analogously, we proceed for b and define $\alpha = \chi_{E_b}\lambda_b$.

Let us consider the problem

$$\min J(y,u) = \int_\Omega \varphi(x, y(x))\,dx + \int_\Omega \psi(x, u(x))\,\mathrm{d}x \tag{4.10a}$$

subject to

$$-\Delta y + d(x,y) = u \qquad \text{in } \Omega, \tag{4.10b}$$

$$\frac{\partial y}{\partial n} = 0 \qquad \text{on } \Gamma, \tag{4.10c}$$

$$u_a(x) \le u(x) \le u_b(x) \qquad \text{f.a.a. } x \in \Omega. \tag{4.10d}$$

Since u occurs as a source term on the right-hand side of (4.10b), we call (4.10) a distributed control problem. We define

$$U_{ad} = \{u \in L^\infty(\Omega) \,:\, u_a \le u \le u_b \text{ in } \Omega \text{ a.e.}\}.$$

Definition 4.12. *We call $u^* \in U_{ad}$ an* optimal control *for* (4.10) *if*

$$J(y(u^*), u^*) \le J(y(u), u) \quad \text{for all } u \in U_{ad}.$$

The function $y = y(u^*)$ is said to be the (associated) optimal state*. We say that $u^* \in U_{ad}$ is a local optimal solution for* (4.10) *in $L^r(\Omega)$, if there exists an $\varepsilon > 0$ so that*

$$J(y(u^*), u^*) \le J(y(u), u) \quad \text{for all } u \in U_{ad} \text{ with } \|u - u^*\|_{L^r(\Omega)} \le \varepsilon.$$

We define

$$F(y) = \int_\Omega \varphi(x, y(x))\,dx. \tag{4.11a}$$

Then the following lemma holds; see [3].

Lemma 4.13. *Let Assumption* 4 *hold. Then F is continuous on $L^\infty(\Omega)$. Moreover, for all $r \in (1, \infty]$ we have*

$$\|F(y) - F(z)\|_{L^r(\Omega)} \le L(M)\,\|y - z\|_{L^r(\Omega)}$$

for all $y, z \in L^\infty(\Omega)$ satisfying $\|y\|_{L^\infty(\Omega)} \le M$, $\|z\|_{L^\infty(\Omega)} \le M$ for a constant $M \ge 0$.

Remark 4.14. 1) It follows from Lemma 4.13 that F is Lipschitz-continuous on the set $\{y \in L^2(\Omega) \,:\, \|y\|_{L^\infty(\Omega)} \le M\}$ with an arbitrary $M > 0$.

2) We set

$$Q(u) = \int_\Omega \psi(x, u(x))\,\mathrm{d}x. \tag{4.11b}$$

Then Q is also Lipschitz-continuous on U_{ad}, since U_{ad} is bounded in $L^\infty(\Omega)$. Moreover, Q is convex on U_{ad} if ψ is convex in u for almost all $x \in E$, i.e.,

$$\psi(x, \lambda u + (1-\lambda)v) \le \lambda\psi(x,u) + (1-\lambda)\psi(x,v) \tag{4.12}$$

holds for almost all $x \in \Omega$, for all $u, v \in \mathbb{R}$ and for all $\lambda \in [0,1]$. $\Diamond$

Theorem 4.15. *We suppose that Assumption* 4 *holds and ψ satisfies* (4.12)*. Then,* (4.10) *possesses at least one optimal control with associated optimal state $y^* = y(u^*)$.*

Proof. Using (4.9) we can express (4.10b)-(4.10c) as

$$-\Delta y + c_0(x)y + \tilde{d}(x,y) = u \quad \text{in } \Omega$$
$$\frac{\partial y}{\partial n} = 0 \quad \text{on } \Gamma,$$

compare (4.5). Now we apply Theorem 4.8. It follows that for every $u \in U_{ad}$ there exists a unique $y = y(u) \in H^1(\Omega) \cap C(\overline{\Omega})$. Since U_{ad} is bounded in $L^\infty(\Omega)$, U_{ad} is bounded in $L^r(\Omega)$ for $r > N/2$. We suppose $r \geq 2$. Then, we conclude from (4.8) that there exists a constant $M > 0$ satisfying

$$\|y(u)\|_{C(\overline{\Omega})} \leq M \tag{4.13}$$

for all states $y(u)$ with $u \in U_{ad}$. By Assumption 4 the cost functional J is bounded from below. For example, from Assumption 4 and (4.13) we conclude that

$$\big|\varphi(x, y(x))\big| \leq \big|\varphi(x,0)\big| + \big|\varphi(x,y(x)) - \varphi(x,0)\big| \leq R + L(M)M \quad \text{f.a.a. } x \in \Omega.$$

Since Ω is bounded, we infer that $F(y)$ is bounded. Hence, there exists $j \in \mathbb{R}$ so that

$$j = \inf_{u \in U_{ad}} J(y(u), u).$$

Let $\{(y_n, u_n)\}_{n\in\mathbb{N}}$ be a minimizing sequence, i.e., $u_n \in U_{ad}$, $y_n = y(u_n)$ and $J(y_n, u_n) \to j$ for $n \to \infty$.
Notice that U_{ad} is a subset of $L^r(\Omega)$. Then, U_{ad} is nonempty, closed, convex and bounded in $L^r(\Omega)$. Since $L^r(\Omega)$ is a Banach space, U_{ad} is weakly sequentially compact. Thus, there exists subsequence $\{u_{n_k}\}_{k\in\mathbb{N}}$ that converges weakly to an element $u^k \in U_{ad}$. To simplify the notation we denote this subsequence again by $\{u_n\}_{n\in\mathbb{N}}$.

$$u_n \rightharpoonup u^k \text{ as } n \to \infty.$$

We proof that u^k is an optimal control. For that purpose we have to show the convergence of the associated state sequence $\{y_n\}_{n\in\mathbb{N}}$.
Consider the sequence

$$z_n = d(\cdot, y_n(\cdot)) - y_n.$$

All states y_n are bounded in $L^\infty(\Omega)$ by M. In particular, the states are bounded in $L^r(\Omega)$. Therefore, there is a subsequence $\{z_{n_l}\}_{l\in\mathbb{N}}$ which converges weakly to an element $z \in L^r(\Omega)$. As above, we suppose that already z_n converges weakly to z. Then,

$$-\Delta y_n + y_n = \underbrace{-d(x,y_n) + y_n}_{\rightharpoonup -z} + \underbrace{u_n}_{\rightharpoonup u^*} =: R_n$$

where $R_n \rightharpoonup -z + u^*$. Note that y_n solves

$$-\Delta y_n + y_n = R_n \qquad \text{in } \Omega$$
$$\frac{\partial y_n}{\partial n} = 0 \qquad \text{on } \Gamma.$$

The mapping $R_n \to y_n$ is linear and continuous from $L^2(\Omega)$ to $H^1(\Omega)$, in particular from $L^r(\Omega)$ to $H^1(\Omega)$. Since any linear and continuous mapping is weakly continuous, the existence of an element $y^* \in H^1(\Omega)$ follows with

$$y_n \rightharpoonup y^* \text{ as } n \to \infty \text{ in } H^1(\Omega).$$

Recall that $H^1(\Omega)$ is compactly embedded into $L^2(\Omega)$. Thus, $y_n \to y^*$ as $n \to \infty$ in $L^2(\Omega)$. From (4.13) we derive that $|y_n(x)| \le M$ for all $x \in \bar{\Omega}$. Utilizing similar arguments as for the proof of Lemma 4.13, we see that

$$\|d(\cdot, y_n) - d(\cdot, y^*)\|_{L^2(\Omega)} \le L(M)\,\|y_n - y^*\|_{L^2(\Omega)}.$$

Thus, $d(\cdot, y_n) \to d(\cdot, y^*)$ in $L^2(\Omega)$. Next we prove $y^* = y(u^*)$.
We have for an arbitrary $\varphi \in H^1(\Omega)$

$$\int_\Omega \nabla y_n \cdot \nabla\varphi + d(\cdot, y_n)\varphi \,\mathrm{d}x = \int_\Omega u_n \varphi \,\mathrm{d}x.$$

From $y_n \rightharpoonup y^*$ in $H^1(\Omega)$ we derive

$$\int_\Omega \nabla y_n \cdot \nabla\varphi \,\mathrm{d}x \to \int_\Omega \nabla y^* \cdot \nabla\varphi \,\mathrm{d}x \text{ as } n \to \infty.$$

Utilizing $y_n \to y^*$ in $L^2(\Omega)$ and $\|y_n\|_{L^\infty(\Omega)} \le M$ we have

$$\int_\Omega d(\cdot, y_n)\varphi \,\mathrm{d}x \to \int_\Omega d(\cdot, y^*)\varphi \,\mathrm{d}x \text{ as } n \to \infty.$$

Finally, $u_n \rightharpoonup u^*$ in $L^r(\Omega)$ implies that

$$\int_\Omega u_n \varphi \,\mathrm{d}x \to \int_\Omega u^* \varphi \,\mathrm{d}x \text{ as } n \to \infty.$$

Consequently, $y^* = y(u^*)$. It remains to prove the optimality of u^*. By Remark 4.14-2) the functional Q is convex and continuous. It follows form Theorem 2.8 that Q is weakly lower semicontinuous:

$$u_n \rightharpoonup u^* \quad \Rightarrow \quad \liminf_{n\to\infty} Q(u_n) \ge Q(u^*).$$

Therefore,

$$\begin{aligned} j &= \lim_{n\to\infty} J(y_n, u_n) = \lim_{n\to\infty} F(y_n) + \lim_{n\to\infty} Q(u_n) \\ &= F(y^*) + \liminf_{n\to\infty} Q(u_n) \ge F(y^*) + Q(u^*) = J(y^*, u^*), \end{aligned}$$

where we have used that $y_n \to y^*$ in $L^2(\Omega)$. We conclude that u^* is optimal for (4.10). □

Remark 4.16. 1) The proof for homogeneous Dirichlet boundary conditions is analogous.
2) Notice that $y_n \rightharpoonup y$ does not imply $d(\cdot, y_n) \to d(\cdot, y)$ in general. Nonlinear mapping need not to be weakly continuous.
3) The assumptions on the boundedness and Lipschitz-continuity for φ and ψ are only required for the functions itself, but not for their derivatives.
4) Due to the nonlinear mapping φ, ψ, d and b, problem (4.10) is not convex. Thus, uniqueness of an optimal control can not be proved without further assumptions. ♢

Example 4.17. Consider the problem

$$\min \hat{J}(u) = -\int_0^1 \cos(u(x))\,\mathrm{d}x \quad \text{s.t. } 0 \le u(x) \le 2\pi,\ u \in L^\infty(0,1).$$

Obviously, the optimal value is -1, choosing $u^*(x) = 1$. However, every $u \in L^\infty(0,1)$ attaining either the value 0 or 2π at $x \in (0,1)$ gives the same optimal value. Thus, there are infinitely many optimal solutions. ♢

4.3. The control-to-state mapping. We consider the problem

$$-\Delta y + d(x,y) = u \quad \text{in } \Omega, \quad \frac{\partial y}{\partial n} = 0 \quad \text{on } \Gamma; \tag{4.14}$$

see (4.10b) and (4.10c). Due to Theorem 4.8 and Remark 4.9-1) there exists a unique state $y \in Y = H^1(\Omega) \cap C(\bar{\Omega})$ for every control $u \in U = L^r(\Omega)$ with $r > N/2$ (provided Assumption 4 holds). We define by $G : U \to Y$, $u \mapsto y = G(u)$ the associated solution operator.

Theorem 4.18. *With Assumption 4 holding, the operator G is Lipschitz-continuous from $L^r(\Omega)$ to $Y = H^1(\Omega) \cap C(\overline{\Omega})$, $r > N/2$: there exists a constant $L > 0$ so that*

$$\|y_1 - y_2\|_{H^1(\Omega)} + \|y_1 - y_2\|_{C(\overline{\Omega})} \le L\,\|u_1 - u_2\|_{L^r(\Omega)}$$

for all $u_i \in L^r(\Omega)$, $i = 1,2$ and associated $y_i = G(u_i)$.

Proof. The continuity follows from Theorem 4.8. Using (4.14) with u_1 and u_2 we obtain

$$-\Delta(y_1 - y_2) + d(\cdot, y_1) - d(\cdot, y_2) = u_1 - u_2 \text{ in } \Omega, \quad \frac{\partial(y_1 - y_2)}{\partial n} = 0 \text{ on } \Gamma.$$

Notice that

$$\begin{aligned} d(x, y_1(x)) - d(x, y_2(x)) &= -\int_0^1 \frac{\mathrm{d}}{\mathrm{d}s} d(x, y_1(x) + s(y_2(x) - y_1(x)))\,\mathrm{d}s \\ &= \int_0^1 d_y(x, y_1(x) + s(y_2(x) - y_1(x)))\,\mathrm{d}s\,(y_1(x) - y_2(x)). \end{aligned}$$

Since $d_y \ge 0$ holds, the integral is a non-negative function $k_0 = k_0(x)$ in $L^\infty(\Omega)$ (recall that d_y, y_1 and y_2 are continuous). On E_d we have a positive integrand. Setting $y = y_1 - y_2$ and $u = u_1 - u_2$ we obtain

$$-\Delta y + k_0(x) y = u \text{ in } \Omega, \quad \frac{\partial y}{\partial n} = 0 \text{ on } \Gamma.$$

From $k_0 \ge 0$ in Ω we infer that $k_0(x)y$ is monotone increasing in y. Applying Theorem 4.8 (with $c_0 = k_0$, $d = 0$, $f = u$, $g = b = \alpha = 0$) we have

$$\|y\|_{H^1(\Omega)} + \|y\|_{C(\overline{\Omega})} \le L\,\|u\|_{L^r(\Omega)}$$

for all u and associated y, in particular for $u = u_1 - u_2$ and $y = y_1 - y_2$. $\square$

Remark 4.19. For the next results we only need the assumptions on the nonlinear mappings and their derivatives. $\Diamond$

Lemma 4.20. *Suppose that $\varphi = \varphi(x,y)$ is measurable in $x \in \Omega$ for all $y \in \mathbb{R}$ and differentiable with respect to y for almost all $x \in \Omega$. Furthermore, there exist constants K and $L(M)$ satisfying*

$|D_y^l \varphi(x,0)| \le K$ *for almost all $x \in \Omega$ and $l = 0,1$*

$|D_y^l \varphi(x,y_1) - D_y^l \varphi(x,y_a)| \le L(M)|y_1 - y_2|$ *for all $y_i \in \mathbb{R}$ with $|y_i| \le M$, $i = 1,2$.*

Then, the mapping $\Phi(y) = \varphi(x,y)$ is Fréchet differentiable in $L^\infty(\Omega)$. For all $h \in L^\infty(\Omega)$ we have

$$(\Phi'(y)h)(x) = \varphi_y(x, y(x))h(x) \text{ for almost all } x \in \Omega.$$

Proof. For arbitrary $y, h \in L^\infty(\Omega)$ satisfying $\|y(x)\| \le M$ and $|h(x)| \le M$ for almost all $x \in \Omega$ we obtain

$$\varphi(x, y(x) + h(x)) - \varphi(x - y(x)) = \varphi_y(x, y(x))h(x) + r(y, h)(x),$$

where

$$r(y, h)(x) = \int_0^1 \Big(\varphi_y(x, y(x) + sh(x)) - \varphi_y(x, y(x)) \Big) \,\mathrm{d}s\, h(x)$$

holds. Since φ_y is Lipschitz-continuous and $|y(x) + h(x)| \le 2M$, we have

$$\begin{aligned} |r(y, h)(x)| &\le L(2M) \int_0^1 s\,|h(x)| \,\mathrm{d}s\, |h(x)| \\ &\le \frac{L(2M)}{2}\, |h(x)|^2 \le \frac{L(2M)}{2}\, \|h\|^2_{L^\infty(\Omega)} \end{aligned}$$

for almost all $x \in \Omega$. Consequently, $\|r(y, h)\|_{L^\infty(\Omega)} \le c\|h\|^2_{L^\infty(\Omega)}$ and

$$\frac{1}{\|h\|_{L^\infty(\Omega)}}\, \|r(y, h)\|_{L^\infty(\Omega)} \to 0 \quad \text{if } \|h\|_{L^\infty(\Omega)} \to 0.$$

The multiplication operator $h \mapsto \varphi_y(\cdot, y(\cdot))h$ is linear and continuous in $L^\infty(\Omega)$. Here we use that

$$\begin{aligned} |\varphi_y(x, y(x))| &\le |\varphi_y(x, y(x)) - \varphi_y(x, 0)| + |\varphi_y(x, 0)| \\ &\le L(M)|y(x)| + K \le L(M)M + K = C \end{aligned}$$

for almost all $x \in \Omega$. $\square$

Using Lemma 4.20 we can prove the differentiability of the control-to-state mapping G.

Theorem 4.21. *Let Assumption* 4 *hold and* $r > N/2$*. Then,* G *is Fréchet differentiable from* $L^r(\Omega)$ *to* $Y = H^1(\Omega) \cap C(\overline{\Omega})$*. Its derivative at* $\bar{u} \in L^r(\Omega)$ *is given by* $G'(\bar{u})u = y$*, where* y *is the weak solution to*

$$-\Delta y + d_y(x, \bar{y})y = u \text{ in } \Omega, \quad \frac{\partial y}{\partial n} = 0 \text{ on } \Gamma. \tag{4.15}$$

Notice that (4.15) *is the linearization of* (4.14) *at* $\bar{y} = G(\bar{u})$*.*

Proof. We have to prove

$$G(\bar{u} + u) - G(\bar{u}) = Du + r(\bar{u}, u),$$

where $D : L^r(\Omega) \to H^1(\Omega) \cap C(\overline{\Omega})$ is a linear continuous operator and

$$\frac{\|r(\bar{u}, u)\|_{H^1(\Omega)\cap C(\overline{\Omega})}}{\|u\|_{L^r(\Omega)}} \to 0 \quad \text{for } \|u\|_{L^r(\Omega)} \to 0, \tag{4.16}$$

with $\|r\|_{H^1(\Omega)\cap C(\overline{\Omega})} = \|r\|_{H^1(\Omega)} + \|r\|_{C(\overline{\Omega})}$. In this case, $G'(\bar{u}) = D$.
For $\bar{y} = y(\bar{u})$ and $\tilde{y} = y(\bar{u} + u)$ we have

$$\begin{aligned} -\Delta \bar{y} + d(x, \bar{y}) &= \bar{u}, & \qquad -\Delta \tilde{y} + d(x, \tilde{y}) &= \bar{u} + u, \\ \frac{\partial \bar{y}}{\partial n} &= 0, & \frac{\partial \tilde{y}}{\partial n} &= 0. \end{aligned}$$

Thus,

$$-\Delta(\tilde{y} - \bar{y}) + \underbrace{d(x, \tilde{y}) - d(x, \bar{y})}_{= d_y(x, \bar{y})(\tilde{y} - \bar{y}) + r_d} = u, \qquad \frac{\partial(\tilde{y} - \bar{y})}{\partial n} = 0.$$

By Lemma 4.20 and Assumption 4, $\Phi(y) = d(\cdot, y)$ is Fréchet differentiable from $C(\overline{\Omega})$ to $L^\infty(\Omega)$. Therefore,

$$\Phi(\tilde{y}) - \Phi(\bar{y}) = d(\cdot, \tilde{y}(\cdot)) - d(\cdot, \bar{y}(\cdot)) = d_y(\cdot, \bar{y}(\cdot))(\tilde{y}(\cdot) - \bar{y}(\cdot)) + r_d,$$

where

$$\frac{\|r_d\|_{L^\infty(\Omega)}}{\|\bar{y} - \tilde{y}\|_{C(\overline{\Omega})}} \to 0 \quad \text{for } \|\bar{y} - \tilde{y}\|_{C(\bar{\Omega})} \to 0. \tag{4.17}$$

It follows that

$$\tilde{y} - \bar{y} = y + y_\rho,$$

where y solves (4.15) and y_ρ is the solution to

$$-\Delta y_\rho + d_y(\cdot, \bar{y}) y_\rho = -r_d \text{ in } \Omega, \quad \frac{\partial y_\rho}{\partial n} = 0 \text{ on } \Gamma. \tag{4.18}$$

Namely, we have

$$\begin{aligned} -\Delta\big(y + y_\rho\big) &= -d_y(\cdot, \bar{y})(y + y_\rho) + u - r_d = -d_y(\cdot, \bar{y})(\tilde{y} - \bar{y}) - r_d + u = -d(\cdot, \tilde{y}) + d(\cdot, \bar{y}) + u \\ &= -\Delta\big(\tilde{y} - \bar{y}\big). \end{aligned}$$

Recall that $d_y(x, \bar{y}) \geq \lambda_d > 0$ in $E_d \subset \Omega$. Hence, (4.18) is uniquely solvable. From Theorem 4.18 we conclude that

$$\|\tilde{y} - \bar{y}\|_{H^1(\Omega)} + \|\tilde{y} - \bar{y}\|_{C(\overline{\Omega})} \leq L\|u\|_{L^r(\Omega)}.$$

Moreover, by (4.17)

$$\frac{\|r_d\|_{L^\infty(\Omega)}}{\|u\|_{L^r(\Omega)}} = \frac{\|r_d\|_{L^\infty(\Omega)}}{\|\tilde{y} - \bar{y}\|_{C(\overline{\Omega})}} \frac{\|\tilde{y} - \bar{y}\|_{C(\overline{\Omega})}}{\|u\|_{L^r(\Omega)}} \leq \frac{\|r_d\|_{L^\infty(\Omega)} L\|u\|_{L^r(\Omega)}}{\|\tilde{y} - \bar{y}\|_{C(\overline{\Omega})}\|u\|_{L^r(\Omega)}} = \frac{\|r_d\|_{L^\infty(\Omega)}}{\|\tilde{y} - \bar{y}\|_{C(\overline{\Omega})}},$$

for $\|\tilde{y} - \bar{y}\|_{C(\overline{\Omega})} \to 0$. Consequently, $\|r_d\|_{L^\infty(\Omega)} = o(\|u\|_{L^r(\Omega)})$. By (4.18) we have

$$\|y_\rho\|_{H^1(\Omega)} + \|y_\rho\|_{C(\overline{\Omega})} = o\big(\|u\|_{L^r(\Omega)}\big).$$

We denote the linear and continuous mapping $u \mapsto y$ by D. Therefore,

$$G(\bar{u} + u) - G(\bar{u}) = \tilde{y} - \bar{y} = y + y_\rho = Du + y_\rho = Du + r(\bar{u}, u),$$

where $r(\bar{u}, u) = y_\rho$ satisfies (4.16). □

Remark 4.22. Notice that Theorem 4.21 implies that G is Fréchet-differentiable from $L^\infty(\Omega)$ to $H^1(\Omega) \cap C(\overline{\Omega})$. ◊

Remark 4.23. In case of boundary control the results are similar. Let us consider

$$-\Delta y = 0 \text{ in } \Omega, \qquad \frac{\partial y}{\partial n} + b(\cdot, y) = u \text{ on } \Gamma. \tag{4.19}$$

Using Assumption 4 the mapping G is continuously Fréchet-differentiable from $U = L^s(\Gamma)$ to $Y = H^1(\Omega) \cap C(\overline{\Omega})$ for $s > N - 1$. The derivative G' at $\bar{u}$ is given by $G'(\bar{u}) = y$, where y is the weak solution to

$$-\Delta y = 0 \text{ in } \Omega, \qquad \frac{\partial y}{\partial n} + b_y(\cdot, \bar{y}) y = u \text{ on } \Gamma.$$

and $\bar{y} = G(\bar{u})$ is the weak solution to (4.19) with $u = \bar{u}$. ◊

4.4. Necessary optimality conditions. We suppose that $u^* \in L^\infty(\Omega)$ is a local optimal solution to (4.10), i.e.,

$$J(y(u^*), u^*) \le J(y(u), u) \quad \text{for all } u \in U_{ad} \text{ with } \|u - u^*\|_{L^\infty(\Omega)} \le \varepsilon.$$

Let $G : L^\infty(\Omega) \to H^1(\Omega) \cap C(\bar{\Omega})$ denote the control-to-state mapping. Then we can introduce the reduced cost functional

$$\hat{J}(u) := J(G(u), u) = F(G(u)) + Q(u),$$

where the mappings F and G have been defined in (4.11a)-(4.11b). With Assumption 4 holding, $\hat{J}$ is Fréchet differentiable in $L^\infty(\Omega)$. Here, we use that F, Q and G are Fréchet differentiable (see Lemma 4.13 and Theorem 4.21).

If u^* is optimal, U_{ad} is convex and $u \in U_{ad}$ arbitrary, then $v = u^* + \lambda(u - u^*) \in U_{ad}$ for all sufficiently small $\lambda > 0$. Moreover, v lies in an ε-neighborhood of u^* so that $\hat{J}(u^*) \le \hat{J}(v)$. Thus,

$$\hat{J}(u^* + \lambda(u - u^*)) \ge \hat{J}(u^*) \quad \text{for all } 0 < \lambda \le \lambda_0.$$

Dividing by λ and taking the limit $\lambda > 0$ we get

$$\hat{J}'(u^*)(u - u^*) \ge 0 \quad \text{for all } u \in U_{ad}.$$

The derivative $\hat{J}'$ can be computed by using the derivative:

$$\begin{aligned}\hat{J}'(u^*)h &= F'(G(u))G'(u)h + Q'(u)h = F'(y^*)v + Q'(u^*)h \\ &= \int_\Omega \varphi_y(x, y^*)v \,\mathrm{d}x + \int_\Omega \psi_y(x, u^*)h \,\mathrm{d}x,\end{aligned}$$

where $v = G'(u^*)h$ solves the linearized problem

$$-\Delta v + d_y(\cdot, y^*)v = h \text{ in } \Omega, \quad \frac{\partial v}{\partial n} = 0 \text{ on } \Gamma \tag{4.20}$$

(compare Theorem 4.21). Let us define the adjoint state $p \in H^1(\Omega) \cap C(\overline{\Omega})$ as the (weak) solution to

$$-\Delta p + d_y(\cdot, y^*)p = -\varphi_y(\cdot, y^*) \text{ in } \Omega, \quad \frac{\partial p}{\partial n} = 0 \text{ on } \Gamma. \tag{4.21}$$

Lemma 4.24. *Suppose that v is the (weak) solution to* (4.20) *for an arbitrary given $h \in L^2(\Omega)$. Moreover, p solves* (4.21)*. Then, we have*

$$-\int_\Omega \varphi(\cdot, y^*)v \,\mathrm{d}x = \int_\Omega ph \,\mathrm{d}x.$$

Proof. Since v and p solve (4.20) and (4.21), respectively, we have

$$\begin{aligned}\int_\Omega \nabla v \cdot \nabla \varphi + d_y(\cdot, y^*)v\varphi \,\mathrm{d}x &= \int_\Omega h\varphi \,\mathrm{d}x && \text{for all } \varphi \in H^1(\Omega), \\ \int_\Omega \nabla p \cdot \nabla \psi + d_y(\cdot, y^*)p\psi \,\mathrm{d}x &= -\int_\Omega \varphi_y(\cdot, y^*)\varphi \,\mathrm{d}x && \text{for all } \psi \in H^1(\Omega).\end{aligned}$$

Choosing $\varphi = p$ and $\psi = v$ we obtain the claim. □

Using Lemma 4.24 we derive a formula for the gradient

$$\hat{J}'(u^*)h = \int_\Omega (\psi_u(\cdot, u^*) - p)h \,\mathrm{d}x.$$

This implies the next result.

Theorem 4.25. *Let Assumption 4 hold. Suppose that $u^* \in U_{ad}$ is a local solution to (4.10) and y^* denotes the associated state. Then,*

$$\int_\Omega (\psi_u(\cdot, u^*) - p)(u - u^*)\,dx \geq 0 \quad \text{for all } u \in U_{ad}, \tag{4.22}$$

where $p \in H^1(\Omega) \cap C(\overline{\Omega})$ is the solution to (4.21).

Proceeding as in Section 3 we can formulate (4.22) as a minimization principle.

Corollary 4.26. *Suppose that Assumption 4 hold, $u^* \in U_{ad}$ is a local solution to (4.10) and $p \in H^1(\Omega) \cap C(\overline{\Omega})$ denotes the solution to (4.21). Then, the solution to*

$$\min_{u_a(x) \leq v \leq u_b(x)} (\psi_u(x, u^*(x)) - p(x))v \tag{4.23}$$

is given by $v = u^*(x)$ for almost all $x \in \Omega$.

If $\psi(x, u) = \frac{\lambda}{2}u^2$, $\lambda > 0$, holds, then ψ satisfies Assumption 4. Moreover, (4.23) leads to

$$\min_{u_a(x) \leq v \leq u_b(x)} (\lambda u^*(x) - p(x))v$$

for almost all $x \in \Omega$. As in Section 3 we obtain

$$u^*(x) = \mathbb{P}_{[u_a(x), u_b(x)]}\left(\frac{1}{\lambda}p(x)\right)$$

(for $\lambda > 0$). If u_a and u_b are continuous, then u^* is continuous. In fact, p is continuous and $\mathbb{P}_{[u_a(x), u_b(x)]}$ maps continuous functions to continuous functions. For $u_a, u_b \in H^1(\Omega)$ we even have $u^* \in H^1(\Omega) \cap C(\overline{\Omega})$.

Example 4.27. Let us consider

$$\min J(y, u) = \frac{1}{2}\left\|y - y_\Omega\right\|^2_{L^2(\Omega)} + \frac{\lambda}{2}\left\|u\right\|^2_{L^2(\Omega)}$$

subject to

$$\begin{aligned} &-\Delta y + y + y^3 = u \text{ in } \Omega, \quad \frac{\partial y}{\partial n} = 0 \text{ on } \Gamma, \\ &-2 \leq u \leq 2 \quad \text{in } \Omega. \end{aligned}$$

Setting

$$\varphi(x, y) = \frac{1}{2}(y - y_\Omega(x))^2, \quad \psi(x, u) = \frac{\lambda}{2}u^2, \quad d(x, y) = y + y^3$$

with $y_\Omega \in L^\infty(\Omega)$, we can verify that Assumption 4 holds. The existence of at least one solution u^* follows from Theorem 4.15. The adjoint equation reads

$$-\Delta p + p + 3(y^*)^2 p = y_\Omega - y \quad \text{in } \Omega, \quad \frac{\partial p}{\partial n} = 0 \quad \text{on } \Gamma.$$

We derive the variational inequality

$$\int_\Omega (\lambda u^* - p)(u - u^*)\,\mathrm{d}x \geq 0 \quad \text{for all } u \in U_{ad}.$$

For $\lambda > 0$ we obtain

$$u^*(x) = \mathbb{P}_{[-2,2]}\left(\frac{1}{\lambda}p(x)\right) \in H^1(\Omega) \cap C(\overline{\Omega}).$$

If $\lambda = 0$ holds, we have (compare Lemma 3.18)

$$u^*(x) = \begin{cases} -2 & \text{if } p(x) < 0, \\ 2 & \text{if } p(x) > 0. \end{cases}$$

Thus, $u^*(x) = 2 sign\, p(x)$. $\Diamond$

Analogously, we obtain first-order optimality conditions for boundary control problem. We cite a result from [2].

Theorem 4.28. *Let Assumption* 4 *hold. suppose that* u^* *is a (local) optimal solution to*

$$\min J(y,u) := \int_\Omega \varphi(x, y(x))\,\mathrm{d}x + \int_\Gamma \psi(x, u(x))\,\mathrm{d}s(x)$$

subject to

$$\begin{aligned} &-\Delta y + y = 0 \text{ in } \Omega, \quad \frac{\partial y}{\partial n} + b(\cdot, y) = u \text{ on } \Gamma, \\ &u \in U_{ad} := \{\tilde{u} \in L^\infty(\Gamma) \,:\, u_a \le u \le u_b \text{ on } \Gamma \text{ a.e.}\} \end{aligned}$$

and $y^* = y(u^*)$ *is the associated state. Then,*

$$\int_\Gamma (\psi_u(\cdot, u^*) - p)(u - u^*)\,\mathrm{d}s(x) \ge 0 \quad \text{for all } u \in U_{ad}$$

where $p \in H^1(\Omega) \cap C(\overline{\Omega})$ *is the (weak) solution to*

$$-\Delta p + p = -\varphi_y(\cdot, y^*) \text{ in } \Omega, \quad \frac{\partial p}{\partial n} + b_y(\cdot, y^*)p = 0 \text{ on } \Gamma.$$

4.5. Application of the formal Lagrange principle. Let us consider the problem

$$\min J(y,v,u) = \int_\Omega \varphi(x, y(x), v(x))\,\mathrm{d}x + \int_\Gamma \psi(x, y(x), u(x)\,\mathrm{d}s(x) \tag{4.24a}$$

subject to

$$-\Delta y + d(\cdot, y, v) = 0 \text{ in } \Omega, \quad \frac{\partial y}{\partial n} + b(\cdot, y, u) = 0 \text{ on } \Gamma \tag{4.24b}$$

$$u \in U_{ad} := \{\tilde{u} \in L^\infty(\Gamma) \,:\, u_a \le u \le u_b \text{ on } \Gamma \text{ a.e.}\} \tag{4.24c}$$

$$v \in V_{ad} := \{\tilde{v} \in L^\infty(\Omega) \,:\, v_a \le v \le v_b \text{ in } \Omega \text{ a.e.}\}. \tag{4.24d}$$

Notice that problem (4.24) is more general as the problems considered in the previous sections. We suppose that φ, ψ, d, b are measurable in x and differentiable in y, v, u. Moreover d_y and b_y are non-negative.

We suppose that (4.24) has a local optimal solution pair $(v^*, u^*) \in V_{ad} \times U_{ad}$ and $y^* = y(u^*, v^*)$ is the associated optimal state. To derive first-order necessary optimality conditions we introduce the Lagrange function by

$$\mathcal{L}(y,v,u,p) = J(y,v,u) + \int_\Omega \nabla y \cdot \nabla p + d(\cdot, y, v)p\,\mathrm{d}x + \int_\Gamma b(\cdot, y, u)p\,\mathrm{d}s(x).$$

We expect the existence of a function $p \in H^1(\Omega) \cap C(\overline{\Omega})$ satisfying

$$\begin{aligned} &D_y\mathcal{L}(y^*, v^*, u^*, p)y = 0 && \text{for all } y \in H^1(\Omega), \\ &D_v\mathcal{L}(y^*, v^*, u^*, p)(v - v^*) \ge 0 && \text{for all } v \in V_{ad}, \\ &D_u\mathcal{L}(y^*, v^*, u^*, p)(u - u^*) \ge 0 && \text{for all } u \in U_{ad}. \end{aligned}$$

We find

$$D_y\mathcal{L}(y^*,v^*,u^*,p)y = \int_\Omega \varphi_y(\cdot,y^*,v^*)y\,\mathrm{d}x + \int_\Gamma \psi_y(\cdot,y^*,u^*)y\,\mathrm{d}s(x)$$
$$+ \int_\Omega \nabla y\cdot\nabla p + d_y(\cdot,y^*,v^*)p\,\mathrm{d}x + \int_\Gamma b_y(\cdot,y^*,u^*)y\,\mathrm{d}s(x) = 0$$

for all $y \in H^1(\Omega)$. This leads to the weak formulation for the solution p to

$$\begin{aligned} -\Delta p + d_y(\cdot,y^*,v^*)p &= -\varphi_y(\cdot,y^*,v^*) \text{ in } \Omega \\ \frac{\partial p}{\partial n} + b_y(\cdot,y^*,u^*)p &= 0 - \psi_y(\cdot,y^*,u^*) \text{ on } \Gamma. \end{aligned} \tag{4.25}$$

We call (4.25) the adjoint equation, and p is the adjoint state. The solution exists if b_y, d_y, φ_y, ψ_y are bounded and measurable, b_y, d_y are non negative and

$$\int_\Omega d_y(\cdot,y^*,v^*)^2\,\mathrm{d}x + \int_\Gamma b_y(\cdot,y^*,u^*)^2\,\mathrm{d}s(x) > 0$$

holds.

Furthermore, we find the variational inequalities

$$\int_\Omega \left(\varphi_v(\cdot,y^*,v^*) + p\,d_v(\cdot,y^*,v^*)\right)(v-v^*)\,\mathrm{d}x \ge 0 \qquad \text{for all } v \in V_{ad},$$
$$\int_\Gamma \left(\psi_u(\cdot,y^*,u^*) + p\,b_u(\cdot,y^*,u^*)\right)(u-u^*)\,\mathrm{d}s(x) \ge 0 \qquad \text{for all } u \in U_{ad}.$$

Example 4.29. Let us consider the following problem

$$\min J(y,u,v) = \int_\Omega y^2 + y_\Omega y + \lambda_1 v^2 + v_\Omega v\,\mathrm{d}x + \int_\Gamma \lambda_2 u^8\,\mathrm{d}s(x)$$

subject to

$$\begin{aligned} &-\Delta y + y + e^y = v \text{ in } \Omega, \quad \frac{\partial y}{\partial n} + |y|y^3 = u^4 \text{ on } \Gamma, \\ &-1 \le v \le 1 \text{ in } \Omega \text{ a.e.}, \quad 0 \le u \le 1 \text{ on } \Gamma \text{ a.e.}, \end{aligned}$$

where y_Ω, v_Ω belong to $L^\infty(\Omega)$. The adjoint equation is given by

$$-\Delta p + p + e^{y^*}p = -2y^* - y_\Omega \text{ in } \Omega, \quad \frac{\partial p}{\partial n} + 4(y^*)^2|y^*|p = 0 \text{ on } \Gamma.$$

Furthermore, we have the variational inequalities

$$\int_\Omega (2\lambda_1 v^* + v_\Omega - p)(v-v^*)\,\mathrm{d}x \ge 0 \qquad \text{for all } v \in V_{ad},$$
$$\int_\Gamma \left(8\lambda_2(u^*)^7 - 4(u^*)^3 p\right)(u-u^*)\,\mathrm{d}s(x) \ge 0 \qquad \text{for all } u \in U_{ad}.$$

$\Diamond$

4.6. Second-order optimality conditions. First we introduce the notion of the second Fréchet derivative.

Definition 4.30. *Let U, V be Banach spaces, $\mathcal{U} \subset U$ be an open set and $F : U \supset \mathcal{U} \to V$ be a Fréchet differentiable mapping. If $u \mapsto F'(u)$ is Fréchet differentiable at $u \in \mathcal{U}$, then F is called* twice Fréchet differentiable *at u. The second derivative is denoted by $(F')'(u) := F''(u)$.*

By definition $F''(u)$ is a linear and continuous operator from U to $Z = \mathcal{L}(U, V)$, i.e., $F''(u) \in \mathcal{L}(U, \mathcal{L}(U, V))$. Here we do not need the operator $F''(u)$ itself, but its application in given directions.

For $u_1 \in U$ we have $F''(u)u_1 \in \mathcal{L}(U, V)$. Hence, $(F''(u)u_1)u_2 \in V$. We make use of the notation

$$F''(u)(u_1, u_2) := (F''(u)u_1)u_2.$$

Notice that for given u the mapping $(u_1, u_2) \mapsto F''(u)(u_1, u_2)$ is a symmetric, continuous bilinear form. Using Taylor expansion we find for twice differentiable functions $F : U \to V$:

$$F(u + h) = F(u) + F'(u)h + \frac{1}{2} F''(u)(h, h) + r_2^F(u, h),$$

where

$$\frac{\|r_2^F(u, h)\|_V}{\|h\|_U^2} \to 0 \quad \text{for } h \to 0.$$

We call F twice continuously Fréchet differentiable if $u \mapsto F''(u)$ is a continuous mapping, i.e.,

$$\|u - \bar{u}\|_U \to 0 \text{ implies } \|F''(u) - F''(\bar{u})\|_{\mathcal{L}(U,\mathcal{L}(U,V))} \to 0.$$

Here,

$$\begin{aligned}\|F''(u)\|_{\mathcal{L}(U,\mathcal{L}(U,V))} &= \sup_{\|u_1\|_U = 1} \|F''(u)u_1\|_{\mathcal{L}(U,V)} \\ &= \sup_{\|u_1\|_U = 1} \left(\sup_{\|u_2\|_U = 1} \|(F''(u)u_1)u_2\|_V \right),\end{aligned}$$

which can be expressed as

$$\|F''(u)\|_{\mathcal{L}(U,\mathcal{L}(U,V))} = \sup_{\|u_2\|_U = 1, \|u_2\|_U = 1} \|F''(u)(u_1, u_2)\|_V.$$

To compute the bilinear form $F''(u)(u_1, u_2)$ we determine the directional derivatives. Let $\tilde{F}(u) = F'(u)u_1$. Then, $\tilde{F}$ is a mapping from U to V. We find

$$\begin{aligned}\tilde{F}'(u)u_2 &= \frac{\mathrm{d}}{\mathrm{d}t} \tilde{F}(u + tu_2)|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t} \left(F'(u + tu_2)u_1\right)|_{t=0} \\ &= (F''(u + tu_2)u_1)u_2|_{t=0} = (F''(u)u_1)u_2 = F''(u)(u_1, u_2),\end{aligned}$$

which is the bilinear form. Now we turn to second-order optimality conditions.

Theorem 4.31. *Let U be a Banach space, $C \subset U$ a convex set and $\hat{J} : U \to \mathbb{R}$ twice continuously Fréchet differentiable in a neighborhood of $u^* \in C$. Furthermore, u^* satisfies the first-order necessary optimality condition*

$$\hat{J}'(u)(u - u^*) \geq 0 \quad \textit{for all } u \in C.$$

Moreover, there exists a $\delta > 0$ so that

$$\hat{J}''(u^*)(h, h) \geq \delta \|h\|_U^2 \quad \textit{for all } h \in U. \tag{4.26}$$

Then there exist constants $\varepsilon > 0$ and $\sigma > 0$ so that

$$\hat{J}(u) \geq \hat{J}(u^*) + \sigma \|u - u^*\|_U^2 \quad \textit{for all } u \in C \textit{ satisfying } \|u - u^*\|_U \leq \varepsilon.$$

Thus, u^* is a local minimum for $\hat{J}$ on C.

Proof. Notice that the proof is the same as in the finite-dimensional case. Let $F : [0,1] \to \mathbb{R}$ be given by $F(s) = \hat{J}(u^* + s(u-u^*))$. Then, $F(1) = \hat{J}(u)$ and $F(0) = \hat{J}(u^*)$. Using Taylor expansion

$$F(1) = F(0) + F'(0) + \frac{1}{2}\, F''(\theta), \quad \theta \in (0,1) \tag{4.27}$$

we obtain

$$\begin{aligned}
\hat{J}(u) &= \hat{J}(u^*) + \hat{J}'(u^*)(u-u^*) + \frac{1}{2}\, \hat{J}''(u^* + \theta(u-u^*))(u-u^*, u-u^*) \\
&\geq \hat{J}(u^*) + \frac{1}{2}\hat{J}''(u^* + \theta(u-u^*))(u-u^*, u-u^*) \\
&= \hat{J}(u^*) + \frac{1}{2}\hat{J}''(u^*)(u-u^*, u-u^*) \\
&\quad + \frac{1}{2}\left(\hat{J}''(u^* + \theta(u-u^*)) - \hat{J}''(u^*)\right)(u-u^*, u-u^*).
\end{aligned}$$

Since $u \mapsto \hat{J}''(u)$ is continuous (in a neighborhood of u^*), we conclude from (4.26)

$$\hat{J}(u) \geq \hat{J}(u^*) + \frac{\delta}{2}\, \|u-u^*\|_U^2 - \frac{\delta}{4}\, \|u-u^*\|_U^2$$

provided $\|u-u^*\|_U \leq \varepsilon$. Thus,

$$\hat{J}(u) \geq \hat{J}(u^*) + \frac{\delta}{4}\, \|u-u^*\|_U^2$$

which gives the claim for $\sigma = \delta/4$. □

Theorem 4.31 can be used for optimal control problems governed by semilinear equations, if the control-to-state mapping G is twice continuously Fréchet differentiable and the control is at most quadratic in the cost functional. An essential equation is given by

$$\hat{J}''(u^*)(h_1,h_2) = \mathcal{L}_{(y,u),(y,u)}(y^*,u^*,p^*)((y_1,h_1),(y_2,h_2)),$$

where $y^* = G(u^*)$ and p are the associated state and dual variables and y_i, $i = 1,2$ solves the linerized state equation for h_i, i.e., $y_i = G'(u^*)h_i$.

References

[1] E. Casas. *Boundary control of semilinear elliptic equations with pointwise state constraints.* SIAM J. Control Optim. 31 (1993), 993-1006.

[2] F. Tröltzsch. *Optimale Steuerung partieller Differentialgleichungen. Theorie, Verfahren und Anwendungen.* Vieweg Verlag, 2nd edition, 2009.

[3] F. Tröltzsch. *Optimal Control of Partial Differential Equations: Theory, Methods and Applications.* American Mathematical Society, Graduate Studies in Mathematics, vol. 112, 2010.

[4] E. Zeidler. *Nonliner Functional Analysis and its Applications II/B: Nonlinear Monotone Operators.* Springer, New York, 1990.

S. Volkwein, Fachbereich Mathematik und Statistik, Universität Konstanz, Universitätsstrasse 10, D-78457 Konstanz, Deutschland

E-mail address: Stefan.Volkwein@uni-konstanz.de