A Brief Introduction to PDE-Constrained Optimization

Harbir Antil and Dmitriy Leykekhman

Abstract In this chapter we give a brief overview of optimization problems with partial differential equation (PDE) constraints, i.e., PDE-constrained optimization (PDECO). We start with three potentially different formulations of a general PDECO problem and focus on the so-called reduced form. We present a derivation of the optimality conditions. Later we discuss the linear and the semilinear quadratic PDECO problems. We conclude with the discretization and the convergence rates for these problems. For illustration, we make a MATLAB code available at

https://bitbucket.org/harbirantil/pde_constrained_opt

that solves the semilinear PDECO problem with control constraints.

1 Introduction

This volume is aimed at early-career graduate students or whoever is interested in entering into the field of PDE-constrained optimization (PDECO), i.e., optimization problems with PDE constraints. The purpose of this particular chapter is to provide a brief mathematical introduction to this topic with an emphasis on optimality conditions and finite-element discretization. Most of the material presented here is well-known and can be found in several textbooks. In particular, we mention [14, 53, 56, 62, 81].

PDE-constrained optimization problems arise in many applications, for instance in flow control [42] and shape optimization [43, 80]. To name a few: controlling pollutants in a river [5]; drug delivery using magnetic fields [8, 9]; optimal

H. Antil (✉)
Department of Mathematical Sciences, George Mason University, Fairfax, VA 22030, USA
e-mail: hantil@gmu.edu

D. Leykekhman
Department of Mathematics, University of Connecticut, Storrs, CT 06269, USA
e-mail: dmitriy.leykekhman@uconn.edu

placement of droplets on electrowetting on dielectric (EWOD) devices [7]; shape optimization of microfluidic biochips [3, 4]. In the first two examples, the control is distributed (confined to either the entire domain or a subdomain). In the third example, the control acts on the boundary, and in the final example, the control is the shape of the domain. These problems range from macroscale to microscale and contain varying levels of difficulties due to the underlying physics.

A PDE-constrained optimization problem has four or five major components: i) a control variable z; ii) state variable u; iii) a state equation (i.e., PDE or system of PDEs) which associates a state u with every control z; iv) a cost functional J to be minimized which depends on z and u; and, possibly, v) constraints imposed on z (control constraints) or/and on u (state constraints). In the most abstract form, this amounts to

$$\text{minimize} \quad J(u, z) \tag{1}$$

subject to $u \in U_{ad}$ and $z \in Z_{ad}$ satisfying

$$e(u, z) = 0. \tag{2}$$

Thus we want to minimize the cost J which is a function of the state u and the control z. The state equation is given by (2) that represents a PDE or system of PDEs (usually in the weak form) with u in an admissible set U_{ad} and the control z in an admissible set Z_{ad}.

To fully understand PDECO problems, the following topics must be considered: (i) functional analysis of the problem; (ii) solver development; (iii) discretization and software development. We shall briefly elaborate on each next, a more detailed discussion will be given in the subsequent chapters. By *analysis*, we mean

- Existence, uniqueness, and regularity of a solution to (2).
- Existence and uniqueness of a solution to the PDECO problem (1)–(2).
- First-order necessary optimality conditions using adjoint equations.
- If possible the second-order necessary/sufficient conditions.

We can carry out the *solver development*, typically gradient based, either at the continuous (infinite dimensional) level [49, 50, 82] or at the discrete (finite dimensional) level [57, 66]. The choice depends on the two options (a) "First discretize then optimize approach" or (b) "First optimize then discretize approach." As the names suggest in case (a) one first discretizes the optimization problem and then writes the optimality conditions at the finite dimensional level. On the other hand, (b) requires first writing the optimality conditions at the continuous level and then discretizing them. These two approaches in general are different. For instance the discretization of the adjoint variables is tied to the state variables if one proceeds with (a). On the other hand, (b) provides more flexibility with respect to this particular aspect. There is no universal approach to select either of the approaches [53, Section 3.2.4]. However, a choice should be made so that the discrete system preserves the structure of the continuous problem. It is worth

noting that the optimization methods can be derived directly at the continuous level. In any event, one should avoid using off-the-shelf finite-dimensional algorithms to solve PDECO problems. This may lead to *mesh dependence,*, i.e., when the number of optimization iterations drastically increases with mesh refinement. In order to overcome such issues, we must adapt such algorithms (if we are using them) to incorporate functional-analytic concepts, for instance, suitable inner products. Nevertheless, *discretization and software development* amounts to

- Either finite-difference, finite-volume, or finite-element discretizations of the PDE and the control [19, 46, 70]. Other discretizations such as wavelets are possible as well, see [30, 31, 78] and references therein.
- Analysis of discrete PDE and optimization solvers and check for mesh independence [51, 82].
- Make the solver efficient, for instance, by using

    - Adaptive finite-element methods (AFEM) [59, 67].
    - Preconditioning techniques [15, 76, 84], time-parallel approaches [16, 32, 33, 83].
    - Model reduction: Proper orthogonal decomposition [6, 41, 54, 55], reduced-basis method [17, 45, 69].

- Software development, for instance, Rapid Optimization Library (ROL) [60], Dolfin-Adjoint [36].

Our goal for this chapter is to collect the necessary ingredients to understand basic PDECO problems. The remaining chapters of this volume provide comprehensive treatments of variational inequalities, algorithm development, optimization under uncertainty, inverse problems, shape optimization, and several applications. We introduce notation, relevant function spaces, and notions of derivatives. We provide an abstract treatment to solve (1)–(2). Moreover, we discuss the linear quadratic PDECO problem in detail (quadratic cost functional with linear elliptic PDEs and control constraints), provide a flavor of the semilinear quadratic PDECO problem, and discuss the necessary components for the numerical analysis of the linear and semilinear quadratic PDECO problems. We also provide a MATLAB implementation of the semilinear quadratic PDECO problem where the space discretization is carried out using the finite-element method (FEM). We solve the resulting optimization problem using the Newton-CG method, or LBFGS method in the absence of the control constraints ($Z_{ad} = Z$) and the semismooth Newton method in the presence of the control constraints. The code can be downloaded using the link

<center>https://bitbucket.org/harbirantil/pde_constrained_opt</center>

The first part of this chapter is organized as follows: in Section 2, we first consider a general optimization problem. We discuss existence of solutions to this problem in Section 2.1 and provide notions of derivatives in function spaces in Section 2.2. We conclude this section with the first-order necessary optimality conditions (cf. Section 2.3). We apply the approach discussed in Section 2 to the general PDECO

problem (1)–(2) in Section 3 and discuss the full and the reduced forms. Derivation of optimality conditions is performed using both full and reduced forms.

In the second part of this chapter, we focus on the linear and semilinear quadratic PDECO problems. We introduce basic Sobolev spaces in Section 4. In Section 5, we mention the results relating to the well-posedness (existence and uniqueness whenever possible) of linear elliptic PDEs and their regularity. Next in Section 6, we formulate the linear quadratic PDECO problem and study its well-posedness in the reduced form. Section 7 discusses the semilinear quadratic PDECO problem. We conclude by introducing a finite-element discretization for the linear and semilinear quadratic PDECO problems in Section 8.

2 Abstract Optimization Problem

The purpose of this section is to consider an abstract optimization problem and study its well-posedness (cf. Section 2.1). We derive the first-order necessary optimality conditions in Section 2.3. This requires the notions of derivatives in function spaces (cf. Section 2.2). The results presented in this section will be applied to the PDECO problems in the subsequent sections.

Let Z be a real reflexive Banach space and Z_{ad} is a closed convex subset. We consider the minimization problem

$$\min_{z \in Z_{ad}} f(z). \tag{3}$$

2.1 Existence

We first show existence of solution to the minimization problem (3) using the direct method of calculus of variations. This is also known as the Weierstrass theorem.

Theorem 1 *Suppose $f : Z \to \mathbb{R}$ is weakly lower semicontinuous with Z a reflexive Banach space and $Z_{ad} \subset Z$ is closed convex. Let the lower γ-level set $\{z \in Z_{ad} : f(z) \leq \gamma\}$ of f is nonempty and bounded for some $\gamma \in \mathbb{R}$. Then problem (3) has an optimal solution, i.e., there exists $\bar{z} \in Z_{ad}$ such that $f(\bar{z}) \leq f(z)$ for all $z \in Z_{ad}$. If f is strictly convex then the solution is unique.*

Proof The proof is based on the direct method of calculus of variations. Following the proof of [14, Theorem 3.2.1], we can construct a minimizing sequence $\{z_n\}_{n\in\mathbb{N}}$ contained in the lower γ-level set such that $f(z_n) \to \inf f(z)$ as $n \to \infty$. Since the lower γ-level set is convex and closed, therefore it is weakly sequentially closed [81, Theorem 2.11]. In addition, since Z is reflexive and the lower γ-level set is bounded, therefore it is weakly sequentially compact [81, Theorem 2.11]. As a result, there exists a subsequence (not relabeled) such that

$$z_n \rightharpoonup \bar{z} \quad \text{in } Z$$

with $\bar{z}$ in the lower γ-level set. It then remains to show that $\bar{z}$ is the optimal solution. Due to the weak lower semicontinuity of f, we conclude that

$$f(\bar{z}) \leq \liminf f(z_n) = \inf_{z \in Z_{ad}} f(z).$$

Finally, in order to show the uniqueness let us assume that z_1 and z_2 be two optimal solutions. Using the definition of strict convexity, we have

$$f\left(\frac{z_1 + z_2}{2}\right) < \frac{1}{2}f(z_1) + \frac{1}{2}f(z_2) = \inf f(z)$$

which is a contradiction. This completes the proof.

After showing the existence of minimizers, a natural question to ask is what the first-order optimality conditions are. However, we need to understand the notions of derivatives in function spaces before we can proceed further.

2.2 Differentiation in Banach Spaces

We introduce the notions of derivatives in function spaces [18, 79]. As an example, we shall apply the ideas to a quadratic cost functional. We will also derive the first-order optimality conditions for the problem (3) based on the derivatives introduced in this section.

Let $\mathscr{L}(A, B)$ denote the space of bounded linear operators from Banach space A to B. Let $(Z, \| \cdot \|_Z)$, $(V, \| \cdot \|_V)$ be real Banach spaces, $\mathscr{Z} \subset Z$, open and $F : \mathscr{Z} \to V$. Moreover, let $z \in \mathscr{Z}$.

Definition 1 (Directional Derivative) F is said to be directionally differentiable at z if the limit $\lim_{t \downarrow 0} \frac{1}{t}(F(z + th) - F(z))$ exists in V for all $h \in Z$. If such limit exists, we denote

$$F'(z, h) := \lim_{t \downarrow 0} \frac{1}{t}(F(z + th) - F(z))$$

and say that $F'(z, h)$ is the directional derivative of F at z in the direction h.

Notice that for a given z, $h \mapsto F'(z, h)$ is not necessarily a linear mapping but it is positive homogeneous, we refer to [79, 81] for examples.

Definition 2 (Gâteaux Derivative) F is said to be Gâteaux differentiable at z if it is directionally differentiable and $F'(z, h) = F'(z)h$ for $F'(z) \in \mathscr{L}(Z, V)$. We refer to $F'(z)$ as the Gâteaux derivative at z.

We next introduce a stronger notion.

Definition 3 (Fréchet Derivative) F is said to be Fréchet differentiable at z if and only if F is Gâteaux differentiable at z and the following holds:

$$F(z + h) = F(z) + F'(z)h + r(z, h) \quad \text{with} \quad \frac{\|r(z, h)\|_V}{\|h\|_Z} \to 0 \quad \text{as} \quad \|h\|_Z \to 0.$$

We refer to $F'(z)$ as the Fréchet derivative at z.

Remark 1 (Few Facts)

(i) If the Fréchet derivative exists so does the Gâteaux derivative and they coincide. However, the converse is not true in general.

(ii) We say that F is continuously Gâteaux differentiable if $F'(\cdot)$ exists and $F'(\cdot)$ is continuous. In that case F is Fréchet differentiable [18, pp. 35–36].

(iii) Let $E = G(F(z))$ where F is Gâteaux differentiable at z and G is Fréchet differentiable at $F(z)$, then E is Gâteaux differentiable.[1]

Notice that when $V = \mathbb{R}$ then $\mathscr{L}(Z, V) = Z^*$. In addition if F is Gâteaux differentiable at z then we have

$$F'(z)h = \langle F'(z), h \rangle_{Z^*, Z},$$

where Z^* is the dual space of Z and $\langle \cdot, \cdot \rangle_{Z^*, Z}$ denotes the duality pairing.

Before we conclude this subsection, we apply the above introduced definitions to two prototypical quadratic functionals. The derivatives in both these examples are Fréchet derivatives.

Example 1 Let $(H, (\cdot, \cdot)_H)$ be a real Hilbert space and $F : H \to \mathbb{R}$ defined as $F(z) := \|z\|_H^2 = (z, z)_H$, then for all $z, h \in H$ we have

$$F(z + h) - F(z) = 2(z, h)_H + \|h\|_H^2.$$

Thus,

$$F'(z)h = (2z, h)_H.$$

Using the Riesz Representation Theorem (identify H with its dual H^*), we can write

$$(\nabla F(z), h)_H = \langle F'(z), h \rangle_{H^*, H},$$

where $\nabla F(z) \in H$ is the representative of $F'(z) \in H^*$. We refer to $\nabla F(z) \in H$ as the gradient of F at z. In the above case, we have $\nabla F(z) = 2z$.

---

[1] The chain rule only requires the outer function to be Hadamard directionally differentiable and the inner function to be Hadamard (Gâteaux) directionally differentiable [79, Proposition 3.6].

Remark 2 (Gradient) As can be seen from the above example, the expression that we obtain by identifying $F'(z) \in H^*$ with an element of H is called the gradient of F. We will use the notation $\nabla F(z)$ to denote the gradient. We further notice that the definition of the gradient depends on the underlying inner product.

Example 2 Let $(Z, (\cdot, \cdot)_Z)$, $(H, (\cdot, \cdot)_H)$ be real Hilbert spaces and $u_d \in H$ be fixed. Let $S \in \mathscr{L}(Z, H)$. Consider $E : Z \to \mathbb{R}$,

$$E(z) = \|Sz - u_d\|_H^2.$$

Then $E(z) = G(F(z))$, where $G(v) = \|v\|_H^2$ and $F(z) = Sz - u_d$.

Next using the chain rule, we obtain that

$$\langle E'(z), h \rangle_{Z^*, Z} = \langle G'(F(z)), F'(z)h \rangle_{H^*, H} = (2v, F'(z)h)_H$$
$$= 2(Sz - u_d, Sh)_H = 2\langle S^*(Sz - u_d), h \rangle_{Z^*, Z},$$

where $S^* \in \mathscr{L}(H^*, Z^*)$ is the adjoint of S. Here we have assumed that S^*S and S^*u_d are well defined. Thus $E'(z) = S^*(Sz - u_d) \in Z^*$. Since Z is a Hilbert space, similar to the previous example, we can again apply the Riesz representation theorem to get the representative $\nabla E(z) \in Z$ of $E'(z)$.

2.3 First-Order Necessary Optimality Conditions

We conclude this section with the following result on the first-order necessary optimality conditions.

Theorem 2 *Let Z be a real Banach space (not necessarily reflexive). Let $f : \mathscr{Z} \to \mathbb{R}$ be Gâteaux differentiable in $\mathscr{Z}$, where $Z_{ad} \subset \mathscr{Z} \subset Z$, $\mathscr{Z}$ open. If $\bar{z} \in Z_{ad}$ is a solution to (3) then the first-order necessary optimality conditions are*

$$\langle f'(\bar{z}), z - \bar{z} \rangle_{Z^*, Z} \geq 0 \quad \forall z \in Z_{ad}. \tag{4}$$

In addition, if f is convex and $\bar{z} \in Z_{ad}$ solves (4) then $\bar{z}$ is a solution to (3), i.e., (4) is necessary and sufficient.

Proof The proof of the first part is a direct consequence of the definition of Gâteaux derivative. Let $z \in Z_{ad}$ be arbitrary. By the convexity of Z_{ad} we have that $\bar{z} + t(z - \bar{z}) \in Z_{ad}$ for all $t \in [0, 1]$. From the optimality of $\bar{z}$ it follows that

$$f(\bar{z} + t(z - \bar{z})) - f(\bar{z}) \geq 0 \quad \forall t \in [0, 1].$$

Dividing both sides by t and taking the limit as t approaches 0^+, we arrive at (4).

Next we use the convexity of f, i.e., for all $t \in (0, 1]$ $f(\bar{z} + t(z - \bar{z})) \leq (1 - t) f(\bar{z}) + t f(z)$. By rearranging terms and taking the limit as t approaches 0^+, we arrive at

$$f(z) - f(\bar{z}) \geq \langle f'(\bar{z}), z - \bar{z} \rangle_{Z^*, Z} \quad \forall z \in Z_{ad}.$$

We then obtain the desired sufficient condition by using (4).

Remark 3 We make the following observations:

i. Theorem 2 only requires f to be directionally differentiable and it holds after replacing $\langle f'(\bar{z}), z - \bar{z} \rangle_{Z^*, Z}$ by $f'(z, z - \bar{z})$.
ii. Notice that the existence of a $\bar{z} \in Z_{ad}$ in Theorem 2 that solves (3) can be satisfied under the assumptions of Theorem 1.
iii. In general, for a nonconvex f, we cannot expect to achieve a global minimum but only a local minimum. We call $\bar{z} \in Z_{ad}$ a local minimum to (3) if there exists an $\varepsilon > 0$ such that

$$f(\bar{z}) \leq f(z) \quad \forall z \in Z_{ad} \cap B_\varepsilon(\bar{z})$$

where $B_\varepsilon(\bar{z}) \subset Z$ is a ball of radius ε centered at $\bar{z}$.
iv. Equation (4) is known as a variational inequality.

3 Application to PDE-Constrained Optimization Problems

Let Z be a real reflexive Banach space and U, Y be real Banach spaces. We begin by recalling the abstract problem (1)–(2)

$$\min_{(u,z) \in U \times Z} J(u, z) \quad \text{subject to} \quad e(u, z) = 0, \quad z \in Z_{ad}, \ u \in U_{ad}, \tag{5}$$

where $J : U \times Z \to \mathbb{R}$ and $e : U \times Z_{ad} \to Y$ where $Z_{ad} \subset Z$ is closed convex. We refer to (5) as the *full-space* form. Before we proceed further, we remark that often the cost functional J has two components

$$J(u, z) = J_1(u) + J_2(z),$$

where $J_1 : U \to \mathbb{R}$ is the objective (to be attained) and $J_2 : Z \to \mathbb{R}$ is the control penalty.

Another way to write (5) is by letting $X := U \times Z$, $X_{ad} := U_{ad} \times Z_{ad}$. Then we seek $x \in X_{ad}$ such that

$$\min_{x \in X_{ad}} J(x) \quad \text{subject to} \quad e(x) = 0. \tag{6}$$

Notice that (6) does not assume splitting of the control and state variables and is a generalization of (5).

By eliminating the state variables, we obtain a third form of (5) and we call it the *reduced form*. Specifically, this requires existence of a solution operator: $S : Z \rightarrow U$, which assigns each control to a unique state

$$z \mapsto S(z) = u(z) \quad \text{where } u(z) \text{ satisfies} \quad e(u(z), z) = 0.$$

Thus we define the reduced cost functional as $\mathscr{J} : Z \rightarrow \mathbb{R}$

$$\mathscr{J}(z) := J(S(z), z).$$

Instead of (5) we then solve

$$\min_{z \in Z_{ad}} \mathscr{J}(z)$$

subject to   (7)

$$S(z) \in U_{ad}.$$

We remark that the formulations (5) and (7) are not equivalent in general. For instance, there are applications where the solution operator S does not exist, i.e., the problem (7) may not have a solution but (5) is still solvable.

We next state existence result for (7). For simplicity of presentation from here on we will assume that $U_{ad} = U$, i.e., no state constraints. However the discussion can be easily adapted to this more general situation.

Corollary 1 *Let the following assumptions hold*

(i) $Z_{ad} \subseteq Z$ *is closed and convex.*
(ii) *For each* $z \in Z_{ad}$, *there exists a unique mapping* $S(z)$ *that solves* $e(S(z), z) = 0$.
(iii) S *is weakly continuous, i.e., if* $z_n \rightharpoonup z$ *in* Z_{ad} *then* $S(z_n) \rightharpoonup S(z)$ *in* U.
(iv) *The lower* γ-*level set* $\{z \in Z_{ad} : \mathscr{J}(z) \leq \gamma\}$ *of* $\mathscr{J}$ *is nonempty and bounded for some* $\gamma \in \mathbb{R}$.
(v) J_1 *is continuous and convex and* J_2 *is weakly lower semicontinuous.*

Then there exists a solution to (7).

Proof We notice that since J_1 is continuous and convex, therefore it is weakly lower semicontinuous [81, Theorem 2.12]. The weak continuity of S combined with the weak lower semicontinuity of J_2 implies that $\mathscr{J}$ is weakly lower semicontinuous. The proof then follows using Theorem 1.

We remark that in many applications we replace the lower γ-level set by either boundedness of Z_{ad} or the coercivity of J_2. More details will be provided in the subsequent sections.

At times it is more suitable to directly work with the full-space (5) form as the reduced form (7) may not even exist. This requires us to use the Lagrangian functional; we will discuss this in Section 3.2. Another advantage of using Lagrangian

formulation is the ease with which it allows us to derive the first- and second-order derivatives. This will be discussed in Section 3.2. We first consider the derivation of first-order optimality conditions for the reduced form in Section 3.1.

3.1 Reduced Form: First-Order Necessary Optimality Conditions

Corollary 2 *Let all the assumptions of Corollary 1 hold except Z being reflexive. Let $\mathscr{Z}$ be an open set in Z such that $Z_{ad} \subset \mathscr{Z}$ such that $z \mapsto S(z) : \mathscr{Z} \to U$ is Gâteaux differentiable with derivative*

$$S'(z) \in \mathscr{L}(Z, U),$$

$(u, z) \mapsto J(u, z) : U \times Z \to \mathbb{R}$ is Fréchet differentiable with

$$J'(u, z) \in \mathscr{L}(U \times Z, \mathbb{R}).$$

If $\bar{z}$ is minimizer of (5) over Z_{ad} then the first-order necessary optimality conditions are given by

$$\langle S'(\bar{z})^* J_u(S(\bar{z}), \bar{z}) + J_z(S(\bar{z}), \bar{z}), z - \bar{z} \rangle_{Z^*, Z} \geq 0 \quad \forall z \in Z_{ad}, \tag{8}$$

where J_u and J_z are the partial derivatives of J. If $\mathscr{J}$ is convex then the condition (8) is sufficient.

Proof The proof is a consequence of Theorem 2. Let $\bar{z}$ be a solution of (5) then from Theorem 2 we have that $\langle \mathscr{J}'(\bar{z}), z - \bar{z} \rangle_{Z^*, Z} \geq 0$ for all $z \in Z_{ad}$. Combining this with the directional derivative and setting $h := z - \bar{z}$, we obtain that

$$\langle \mathscr{J}'(\bar{z}), h \rangle_{Z^*, Z} = J'(S(z), z)h = \langle J_u(S(z), z), S'(z)h \rangle_{U^*, U} + \langle J_z(S(z), z), h \rangle_{Z^*, Z}$$

where we have used the chain rule for derivatives which holds under the stated differentiability assumptions on S and J. Since $S'(z)^*$ is well defined, we conclude that

$$\langle \mathscr{J}'(\bar{z}), h \rangle_{Z^*, Z} = \langle S'(z)^* J_u(S(z), z), h \rangle_{Z^*, Z} + \langle J_z(S(z), z), h \rangle_{Z^*, Z}$$

This completes the proof.

To further understand the structure of $S'(\bar{z})$, we assume that the PDE constraint function e is sufficiently smooth and the conditions of the implicit function theorem hold. Upon differentiating the state equation, we obtain that

$$e_u(S(\bar{z}), \bar{z})S'(\bar{z})h = -e_z(S(\bar{z}), \bar{z})h.$$

Subsequently, we arrive at

$$S'(\bar{z})h = -e_u(S(\bar{z}), \bar{z})^{-1} (e_z(S(\bar{z}), \bar{z})h) . \tag{9}$$

Substituting this in (8), we obtain that

$$-\left\langle e_z(S(\bar{z}), \bar{z})^* \left((e_u(S(\bar{z}), \bar{z})^{-1})^* J_u(S(\bar{z}), \bar{z})\right), z - \bar{z}\right\rangle_{Z^*, Z}$$
$$+\langle J_z(S(\bar{z}), \bar{z}), z - \bar{z}\rangle_{Z^*, Z} \geq 0.$$

Introducing the adjoint variable $\bar{p}$ and solving the adjoint equation,

$$e_u(S(\bar{z}), \bar{z})^* \bar{p} = J_u(S(\bar{z}), \bar{z}), \tag{10}$$

we arrive at the following reformulation of (8)

$$-\left\langle e_z(S(\bar{z}), \bar{z})^* \bar{p}, z - \bar{z}\right\rangle_{Z^*, Z} + \langle J_z(S(\bar{z}), \bar{z}), z - \bar{z}\rangle_{Z^*, Z} \geq 0. \tag{11}$$

Notice that

$$\mathscr{J}'(z) = -e_z(S(z), z)^* p + J_z(S(z), z) \in Z^*, \tag{12}$$

is the derivative of $\mathscr{J}$ at z. We summarize the computation of $\mathscr{J}'(z)$ in Algorithm 1.

Algorithm 1 requires two PDE solvers (possibly nonlinear in Step 1, linear PDE in Step 2).

In order to get the gradient, we can again apply the Riesz representation theorem (under the assumption that Z is a Hilbert space or admits a representation) to get a representative $\nabla \mathscr{J}(z)$ satisfying

$$(\nabla \mathscr{J}(z), v)_Z = \langle \mathscr{J}'(z), v\rangle_{Z^*, Z} \quad \forall v \in Z.$$

Having the expression of the gradient in hand, we can develop a gradient-based optimization solver. We will provide another derivation of the first-order optimality conditions in Section 3.2 using the Lagrangian approach.

In order to design Newton-based methods, it is desirable to have the second-order derivative information of the reduced functional. This can be obtained by using either the reduced functional approach or the Lagrangian approach. We will provide a brief discussion in the next section as well. More details will be given in subsequent chapters, we also refer to [44, 53].

Algorithm 1 Derivative computation using adjoints

1: Given z, solve $e(u, z) = 0$ for the state u.
2: Solve the adjoint equation $e_u(u(z), z)^* p = J_u(u(z), z)$ for p.
3: Compute $\mathscr{J}'(z) = J_u(u(z), z) - e_z(u(z), z)^* p(z)$.

3.2 Lagrangian Formulation

3.2.1 First-Order Optimality Conditions

The full-space form requires us to introduce Lagrangian functional: $L : U \times Z_{ad} \times Y^* \to \mathbb{R}$,

$$L(u, z, p) = J(u, z) - \langle e(u, z), p \rangle_{Y,Y^*}. \tag{13}$$

where Y^* is the dual space of Y (recall that $e : U \times Z_{ad} \to Y$). Notice that if we set $u = S(z)$ in (13) then $e(S(z), z) = 0$ and we obtain that

$$\mathscr{J}(z) = J(S(z), z) = L(S(z), z, p) \quad \text{for any } p \in Y^*. \tag{14}$$

Now if $(\bar{u}, \bar{z}, \bar{p})$ denotes a stationary point, then the partial derivatives of $L(u, z, p)$ with respect to u, z, and p at $(\bar{u}, \bar{z}, \bar{p})$ vanish and as a result we obtain

$$L_p(\bar{u}, \bar{z}, \bar{p}) = 0,$$

which reduces to the state equation

$$e(\bar{u}, \bar{z}) = 0.$$

Also

$$L_u(\bar{u}, \bar{z}, \bar{p}) = 0, \tag{15}$$

which is just the adjoint equation (10). Indeed

$$\langle L_u(\bar{u}, \bar{z}, \bar{p}), \xi \rangle_{U^*,U} = \langle J_u(\bar{u}, \bar{z}), \xi \rangle_{U^*,U} - \langle \bar{p}, e_u(\bar{u}, \bar{z})\xi \rangle_{Y^*,Y}$$
$$= \langle J_u(\bar{u}, \bar{z}) - e_u(\bar{u}, \bar{z})^* \bar{p}, \xi \rangle_{U^*,U}.$$

In other words

$$L_u(\bar{u}, \bar{z}, \bar{p}) = J_u(\bar{u}, \bar{z}) - e_u(\bar{u}, \bar{z})^* \bar{p}.$$

Finally,

$$\langle L_z(\bar{u}, \bar{z}, \bar{p}), z - \bar{z} \rangle_{Z^*,Z} \geq 0 \quad \forall z \in Z_{ad},$$

which is equivalent to the variational inequality for the control (11). Indeed we have that the gradient of the reduced function $\mathscr{J}$ at z (12) is

$$\mathscr{J}'(z) = L_z(u, z, p),$$

where u and p solve the state and the adjoint equations, respectively.

Few comments are in order, first of all the above approach provides an elegant way to derive the first-order necessary optimality conditions and is highly recommended. As we will discuss below, the above approach also allows us to easily derive the second-order derivatives for the reduced functional. Secondly, even though the above introduced Lagrangian L is rigorous, however, we have not yet addressed the question of existence of the Lagrange multiplier p which makes the approach "formal." The existence of Lagrange multiplier p can be shown by using the Karush–Kuhn–Tucker (KKT) theory in function spaces [81, Chapter 6]. This theory requires the Robinson's regularity condition [71] or the Zowe–Kurcyusz constraint qualification [85], see [53, Chapter 1]. In certain cases, in particular, in the absence of control constraints and linear PDE constraints, one can also use the inf-sup theory for saddle point problems [39] to show existence of the Lagrange multipliers.

3.2.2 Second-Order Derivatives

Next we focus on deriving the expression of the second-order derivative of the reduced functional $\mathscr{J}$. The second-order derivative information can significantly improve the convergence rates for optimization algorithms. For instance in the absence of control constraints, the first-order necessary optimality conditions are $\mathscr{J}'(\bar{z}) = 0$ in Z^*. In order to solve for z, one can use Newton's method which is quadratically convergent (locally). Each iteration of Newton's method requires solving

$$\mathscr{J}''(z)v = -\mathscr{J}'(z) \quad \text{in } Z^* \tag{16}$$

for a direction $v \in Z$. In general (for large problems), it is often too expensive to store and factorize the Hessian. Instead it is often more practical to solve (16) using iterative methods that only require Hessian-times-vector products. We will discuss the computation of Hessian-times-vector product next. We remark that in case of bound constraints on the control one can use a superlinearly (locally) convergent semismooth Newton method [48, 56, 82].

We will proceed by using the Lagrangian approach. We operate under the assumption that J and e are twice continuously differentiable.

From (14), we recall that

$$\mathscr{J}(z) = J(u(z), z) = L(u(z), z, p)$$

$u(z) = S(z)$ solves the state equation and $p \in Y^*$ is arbitrary. After differentiating this expression in a direction h_1, we obtain that

$$\langle \mathscr{J}'(z), h_1 \rangle_{Z^*,Z} = \langle L_u(u(z), z, p), u'(z)h_1 \rangle_{U^*,U} + \langle L_z(u(z), z, p), h_1 \rangle_{Z^*,Z}.$$

Again differentiating this expression in a direction h_2 and choosing a particular p that solves the adjoint equation (15), we arrive at

$$\mathscr{J}''(z) = T(S(z), z)^* H(S(z), z, p) T(S(z), z),$$

where

$$T(u, z) = \begin{pmatrix} -e_u(u, z)^{-1} e_z(u, z) \\ I_Z \end{pmatrix}$$

with I_Z denoting the identity map on Z and $H(u, z, p)$ is given by

$$H(u, z, p) = \begin{pmatrix} L_{uu}(u, z, p) \ L_{uz}(u, z, p) \\ L_{zu}(u, z, p) \ L_{zz}(u, z, p) \end{pmatrix}.$$

Then one can compute the Hessian vector product by using Algorithm 2. Notice that Algorithm 2 requires two linear PDE solvers in Steps 3 and 4. We refer to [53, Chapter 1] and [44] for further details.

So far our approach has been general. For the chapter remainder, we will focus on two particular examples where the cost functional is quadratic and the PDE constraints are linear and semilinear elliptic PDEs. In order to develop the notion of solutions to these PDEs, we first introduce Sobolev spaces.

4   Sobolev Spaces

In this section, we introduce the necessary function spaces to be used throughout the chapter remainder. Let $\Omega \subset \mathbb{R}^n$ be an open, bounded domain with Lipschitz boundary $\partial \Omega$. For $1 \le p < \infty$, we denote by $L^p(\Omega)$ the Banach space

$$L^p(\Omega) := \left\{ v : \Omega \to \mathbb{R} \ : \ v \text{ is measurable and } \int_\Omega |v(x)|^p \, dx < \infty \right\}$$

Algorithm 2 Hessian-Times-Vector Computation

1: Given z, solve $e(u, z) = 0$ for u (if not done already).
2: Solve adjoint equation: $e_u(u, z)^* p = J_u(u, z)$ for p (if not done already).
3: Solve $e_u(u, z)w = -e_z(u, z)v$.
4: Solve $e_u(u, z)^* q = L_{uu}(u, z, p)w + L_{uz}(u, z, p)v$.
5: Compute

$$\mathscr{J}''(z)v = -e_z(u, z)^* q + L_{uz}(u, z, p)w + L_{zz}(u, z, p)v.$$

---

with the norm $\|v\|_{L^p(\Omega)} := \left(\int_\Omega |v(x)|^p \, dx \right)^{\frac{1}{p}}$. These spaces are equivalence classes of functions equal up to a set of measure zero [37]. In particular when $p = 2$, we obtain $L^2(\Omega)$ which is a Hilbert space with inner product $(u, v)_{L^2(\Omega)} = \int_\Omega u(x)v(x) \, dx$. When $p = \infty$, we obtain $L^\infty(\Omega)$, a Banach space with norm $\|v\|_{L^\infty(\Omega)} := \text{ess sup}_\Omega |v|$.

Moving forward, we use multi-index notation to define partial derivatives. For a multi-index $\gamma = (\gamma_1, \ldots, \gamma_n) \in \mathbb{N}_0^n := (\mathbb{N} \cup \{0\})^n$, we let its order to be $|\gamma| := \sum_{i=1}^n \gamma_i$. The associated $|\gamma|$-th order partial derivative of a function u at x is

$$D^\gamma u(x) := \frac{\partial^{|\gamma|} u}{\partial x_1^{\gamma_1} \ldots \partial x_n^{\gamma_n}}(x).$$

Then we denote by $W^{k,p}(\Omega)$ the Sobolev spaces with the norm

$$\|u\|_{W^{k,p}(\Omega)} := \begin{cases} \left(\sum_{|\gamma| \le k} \int_\Omega |D^\gamma u|^p \, dx \right)^{1/p} & 1 \le p < \infty \\ \sum_{|\gamma| \le k} \text{ess sup}_\Omega |D^\gamma u| & p = \infty. \end{cases}$$

If $p = 2$, we write

$$H^k(\Omega) = W^{k,2}(\Omega), \quad k = 0, 1, \ldots$$

which is a Hilbert space with inner product

$$(u, v)_{H^k(\Omega)} = \sum_{|\gamma| \le k} (D^\gamma u, D^\gamma v)_{L^2(\Omega)}.$$

Notice that $H^0(\Omega) = L^2(\Omega)$.

We denote by $W_0^{k,p}(\Omega)$ the closure of $C_0^\infty(\Omega)$ with respect to $W^{k,p}(\Omega)$-norm. Thus $u \in W_0^{k,p}(\Omega)$ if and only if there exist functions $u_m \in C_0^\infty(\Omega)$ such that $u_m \to u$ in $W^{k,p}(\Omega)$. The space $H_0^1(\Omega)$ consists of functions $u \in H^1(\Omega)$ such that

$$u = 0 \quad \text{on } \partial\Omega,$$

in the trace sense. Using the Poincaré inequality $\|u\|_{L^2(\Omega)} \le C\|\nabla u\|_{L^2(\Omega)}$, where $C = C(\Omega)$, we have

$$\|u\|_{H^1(\Omega)} \le C\|\nabla u\|_{L^2(\Omega)} \quad \forall u \in H_0^1(\Omega).$$

Finally, we denote the dual of $H_0^1(\Omega)$ by $H^{-1}(\Omega)$. It is easy to see that $L^2(\Omega)$ is continuously embedded in $H^{-1}(\Omega)$. For more details, we refer to [1].

5   Second-Order Linear Elliptic PDEs

In this section, we study the second-order elliptic PDEs. In general, we cannot expect classical solutions to these PDEs, therefore we first introduce the notion of weak solutions in Section 5.1. We will also study the notion of 'strong' solutions for this problem in Section 5.2. This higher regularity (strong solutions) will help us to establish the approximability of the continuous solution using the finite-element method (cf. Section 8). Strong solutions, in addition, also play a role in other situations, for instance, in studying the regularity of multipliers in state-constrained problems [20] and PDECO problems with variational inequalities [49].

5.1   Existence and Uniqueness

We begin this section by making certain uniform ellipticity assumptions.

Assumption 3 (Coefficient Matrix) *Let A be an $n \times n$ matrix with entries a_{ij} for $1 \leq i, j \leq n$. We assume that a_{ij} are measurable, belong to $L^\infty(\Omega)$, and are symmetric, that is, $a_{ij}(x) = a_{ji}(x)$ for all $1 \leq i, j \leq n$ and for a.e. $x \in \Omega$. We further assume that A is positive definite and satisfy the uniform ellipticity condition*

$$\exists \, \beta \geq \alpha > 0 \quad such \; that \quad \alpha |\xi|^2 \leq A(x)\xi \cdot \xi \leq \beta |\xi|^2 \quad \forall \xi \in \mathbb{R}^n, \; a.e. \; x \; in \; \Omega. \tag{17}$$

Given f, we consider the following linear elliptic PDE

$$-\mathrm{div}\,(A\nabla u) = f \quad \text{in } \Omega$$
$$u = 0 \quad \text{on } \partial\Omega. \tag{18}$$

We understand (18) in a *weak sense*, i.e., given $f \in H^{-1}(\Omega)$, we seek a solution $u \in H_0^1(\Omega)$ that satisfies

$$\int_\Omega A\nabla u \cdot \nabla v \, dx = \langle f, v \rangle_{-1,1}, \quad \forall v \in H_0^1(\Omega), \tag{19}$$

where $\langle \cdot, \cdot \rangle_{-1,1}$ denotes the duality pairing between $H^{-1}(\Omega)$ and $H_0^1(\Omega)$.

Theorem 4 *For every $f \in H^{-1}(\Omega)$, there exists a unique weak solution $u \in H_0^1(\Omega)$ that fulfills*

$$\|u\|_{H^1(\Omega)} \leq C \|f\|_{H^{-1}(\Omega)}, \tag{20}$$

where the constant C only depends on Ω and α.

Proof The existence and uniqueness is due to the Lax–Milgram Lemma. Moreover, the bound (20) immediately follows by using the fact that A is uniformly positive definite and the Poincaré inequality [38].

Remark 4 In general, for Sobolev spaces $W^{1,p}(\Omega)$ with $p \neq 2$, we need the inf-sup conditions to prove Theorem 4. The Banach–Nečas theorem then guarantees existence and uniqueness of u [34, Theorem 2.6]. The latter is a necessary and sufficient condition. See also [67].

5.2 Regularity

It is imperative to understand the regularity of solution u to (19). For instance, such an understanding allows us to develop numerical schemes with optimal rate of convergence (see Section 8). It can assist us with proving rates of convergence for the optimization algorithms [47].

Theorem 5 *Let $u \in H_0^1(\Omega)$ be a weak solution of (19) and let the coefficient matrix A satisfy Assumption 3.*

- *If $f \in L^2(\Omega)$ and Ω is a convex polytope or $C^{1,1}$ domain in $\mathbb{R}^n$, then $u \in H^2(\Omega) \cap H_0^1(\Omega)$ and there exists a constant $C = C(\alpha, \beta, \Omega)$ such that*

$$\|u\|_{H^2(\Omega)} \leq C \|f\|_{L^2(\Omega)}.$$

- *If $f \in L^p(\Omega)$ for $1 < p < \infty$ and Ω is $C^{1,1}$, then $u \in W^{2,p}(\Omega) \cap W_0^{1,p}(\Omega)$ and there exists a constant $C = C(\alpha, \beta, \Omega, p)$ such that*

$$\|u\|_{W^{2,p}(\Omega)} \leq C \|f\|_{L^p(\Omega)}.$$

If $p > n$, then $u \in C^{1,\alpha}(\bar{\Omega})$ with $\alpha = 1 - n/p$.

Proof If Ω is a convex polygonal/polyhedral domain, then H^2-regularity is in [40, 3.2.1.2]. When $\partial\Omega$ is $C^{1,1}$ and $f \in L^p(\Omega)$ for any $1 < p < \infty$, the result is due to [38, Theorem 9.15]. In the case $p > n$, the $C^{1,\alpha}$ regularity follows from $W^{2,p}$ regularity and the Sobolev embedding.

6 Linear Quadratic PDE-Constrained Optimization Problem

Having some basic understanding of elliptic PDEs in hand, we next apply the results of Section 3 to a linear quadratic PDECO problem (cf. Section 6.1). In Section 6.2, we formulate it as a reduced PDECO problem only in terms of the control variable z. This allows us to use the direct method of calculus of variations from Theorem 1 to show the existence of solution.

6.1 Problem Formulation

Let $u_d \in L^2(\Omega)$ and $z_a, z_b \in L^2(\Omega)$ with $z_a < z_b$ a.e. in Ω being given. Moreover, let $\lambda \geq 0$ denotes the penalty parameter. Then we are interested in minimizing

$$J(u, z) = \frac{1}{2}\|u - u_d\|_{L^2(\Omega)}^2 + \frac{\lambda}{2}\|z\|_{L^2(\Omega)}^2 \tag{21}$$

subject to (in the weak form)

$$\begin{aligned}
-\text{div}\,(A\nabla u) = z &\quad \text{in } \Omega, \\
u = 0 &\quad \text{on } \partial\Omega,
\end{aligned} \tag{22}$$

and the pointwise control constraints

$$z \in Z_{ad} := \{v \in L^2(\Omega) : z_a(x) \leq v(x) \leq z_b(x)\,,\ a.e.\ x \in \Omega\}. \tag{23}$$

Notice that in our formulation above we also allow $z_a = -\infty$ and $z_b = \infty$. We call this case as an unconstrained case, i.e., $Z_{ad} = L^2(\Omega)$.

For the above problem, we have (cf. Section 3)

$$U = H_0^1(\Omega), \quad Y = H^{-1}(\Omega), \quad Z = L^2(\Omega).$$

In order to understand the problem (21)–(23) similar to (7), first we condense the variables and write J only in terms of z. We again call this as the reduced form and the resulting cost functional as the reduced functional. We discuss this next.

6.2 Reduced PDECO Problem

For every $z \in Y$, there exists a unique solution $u = u(z) \in U$ to (22). As a result, we can define the solution operator to (22) as

$$\mathscr{S} : Y \to U, \quad z \mapsto u(z) = \mathscr{S}z,$$

which is linear and continuous. In view of the continuous embedding, $H_0^1(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow H^{-1}(\Omega)$ we may also consider $\mathscr{S}$ as a map from $L^2(\Omega)$ to $L^2(\Omega)$. In other words instead of $\mathscr{S}$, we consider the operator $S := \mathscr{E}_u\mathscr{S}\mathscr{E}_z$, where $\mathscr{E}_z : L^2(\Omega) \to H^{-1}(\Omega)$ and $\mathscr{E}_u : H_0^1(\Omega) \to L^2(\Omega)$ denote the embedding operators that assign to each $z \in L^2(\Omega)$ and $u \in H_0^1(\Omega)$ the functions $z \in H^{-1}(\Omega)$ and $u \in L^2(\Omega)$ so that when these new functions are restricted to $L^2(\Omega)$ and $H_0^1(\Omega)$ the operator yields the original functions, respectively. Notice that $S : L^2(\Omega) \to L^2(\Omega)$. One of the main advantages of using S is the fact that the adjoint operator

S^* also acts on $L^2(\Omega)$ (cf. Section 6.3). Using the solution map S, we arrive at the so-called reduced cost $\mathscr{J} : L^2(\Omega) \to \mathbb{R}$ which is defined as

$$\mathscr{J}(z) := J(Sz, z),$$

and the minimization problem (21)–(23) is equivalent to the reduced problem:

$$\min_{z \in Z_{ad}} \mathscr{J}(z). \tag{24}$$

We notice that it is more convenient to analyze (24) in comparison to (21)–(23). In fact, we have the following well-posedness result.

Corollary 3 *Let either $\lambda > 0$ or Z_{ad} be bounded. Then there exists a solution to the minimization problem* (24). *If in addition $\lambda > 0$ or S is injective then the solution is unique.*

Proof The proof is a consequence of Theorem 1. We will first show that $\mathscr{J}$ is weakly lower semicontinuous. Notice that $\mathscr{J}(z) = \mathscr{J}_1(Sz) + \mathscr{J}_2(z)$ where $\mathscr{J}_1(Sz) := \frac{1}{2}\|Sz - u_d\|_{L^2(\Omega)}^2$ and $\mathscr{J}_2(z) := \frac{\lambda}{2}\|z\|_{L^2(\Omega)}^2$. Clearly $\mathscr{J}_2$ is weakly lower semicontinuous (convexity and continuity imply weak lower semicontinuity [18, Theorem 2.23]). On the other hand, due to the compact embedding of $H_0^1(\Omega)$ in $L^2(\Omega)$, we have that $S : L^2(\Omega) \to L^2(\Omega)$ is completely continuous, i.e., if $z_n \rightharpoonup z$ in $L^2(\Omega)$ then $Sz_n \to Sz$ in $L^2(\Omega)$. Thus, owing to the continuity and convexity of $\mathscr{J}_1$, we conclude that $\mathscr{J}_1$ is weakly lower semicontinuous. Whence $\mathscr{J}$ is weakly lower semicontinuous.

It then remains to characterize the lower γ-level set. Here we replace the lower γ-level set condition by the coercivity of $\mathscr{J}$ (if $\lambda > 0$) or by the closed convex bounded set Z_{ad}.

Finally uniqueness is due to strict convexity of $\mathscr{J}$.

For the remainder of this section, we will consider $\lambda > 0$.

6.3 First-Order Optimality Conditions

We are now ready to derive the first-order optimality conditions by following Section 3.1 and the expression of the gradient of the reduced objective function. In Section 6.4, we follow Section 3.2 and consider an alternate strategy to derive the optimality conditions using the Lagrangian formulation.

We recall that the reduced functional is

$$\mathscr{J}(z) = \frac{1}{2}\|Sz - u_d\|_{L^2(\Omega)}^2 + \frac{\lambda}{2}\|z\|_{L^2(\Omega)}^2.$$

Using Examples 1 and 2, the gradient of $\mathscr{J}$ is given by

$$\nabla \mathscr{J}(z) = S^*(Sz - u_d) + \lambda z \in L^2(\Omega).$$

Here we have $S : L^2(\Omega) \to L^2(\Omega)$. The first-order necessary and sufficient (due to the convexity of $\mathcal{J}$) optimality condition (4) then becomes

$$(\nabla \mathcal{J}(\bar{z}), z - \bar{z})_{L^2(\Omega)} \geq 0 \quad \forall z \in Z_{ad}. \tag{25}$$

In order to efficiently evaluate $\nabla \mathcal{J}(z)$, we introduce the so-called adjoint variable $p \in H_0^1(\Omega)$ solving

$$-\text{div}\,(A\nabla p) = u - u_d \quad \text{in } \Omega$$
$$p = 0 \quad \text{on } \partial\Omega. \tag{26}$$

We will next show that the adjoint operator $S^* : L^2(\Omega) \to L^2(\Omega)$ can be defined by $S^*\zeta := p$ where p solves (26) with right-hand side given by ζ. Here $\zeta \in L^2(\Omega)$ is arbitrary. Let $z \in L^2(\Omega)$ be an arbitrary right-hand side of the state equation and the resulting state variable is $u = Sz \in H_0^1(\Omega)$. By testing the equation for u with p and vice versa, we obtain that $(z, p)_{L^2(\Omega)} = (\zeta, u)_{L^2(\Omega)}$. Since $u = Sz$, we deduce that $S^*\zeta = p$. As a result, $S^*(u - u_d) = p$ where p solves (26).

Thus the gradient computation reduces to evaluation of the following expression:

$$\nabla \mathcal{J}(z) = p + \lambda z \in L^2(\Omega).$$

Finally, we gather the first-order necessary and sufficient optimality system:

$$\bar{u} \in H_0^1(\Omega) : \quad \int_{\Omega} A\nabla\bar{u} \cdot \nabla v \, dx = \int_{\Omega} \bar{z}v \, dx \quad \forall v \in H_0^1(\Omega) \tag{27a}$$

$$\bar{p} \in H_0^1(\Omega) : \quad \int_{\Omega} A\nabla\bar{p} \cdot \nabla v \, dx = \int_{\Omega} (\bar{u} - u_d)v \, dx \quad \forall v \in H_0^1(\Omega) \tag{27b}$$

$$\bar{z} \in Z_{ad} : \quad (\bar{p} + \lambda\bar{z}, z - \bar{z})_{L^2(\Omega)} \geq 0, \quad \forall z \in Z_{ad}. \tag{27c}$$

Notice that (27) is a coupled system, namely $\bar{u}$ in (27a) depends on the unknown optimal control $\bar{z}$ which fulfills the inequality (27c). The latter depends on the adjoint variable $\bar{p}$ that solves the adjoint equation (27b). This in turn depends on $\bar{u}$. We further remark the variational inequality (27c) for the control is equivalent to the following projection formula (see [81]):

$$\bar{z}(x) = \mathscr{P}_{Z_{ad}}\left\{-\frac{1}{\lambda}\bar{p}(x)\right\} \quad \text{a.e. } x \in \Omega. \tag{28}$$

Here $\mathscr{P}_{Z_{ad}}(v)$ denotes projection of v onto Z_{ad}. In our case, we can further write this as $\mathscr{P}_{Z_{ad}}(v) := \min\{b(x), \max\{a(x), v(x)\}\}$, x a.e. in Ω.

6.4 Lagrange Method

An alternative approach to derive the first-order optimality system is using the Lagrangian as described in Section 3.2. We again emphasize that even though this approach is used formally in the following, it can be made rigorous. It provides a systematic way of deriving the optimality system, especially for tedious problems and this step is strongly recommended.

Introduce $L : H_0^1(\Omega) \times Z_{ad} \times H^1(\Omega) \to \mathbb{R}$, defined as

$$L(u, z, p) := J(u, z) - \int_{\Omega} (A\nabla u \cdot \nabla p - zp) \ dx.$$

If $(\bar{u}, \bar{z}, \bar{p})$ is a stationary point then

$$\langle L_p(\bar{u}, \bar{z}, \bar{p}), h \rangle_{H^{-1}(\Omega), H_0^1(\Omega)} = 0 \quad \forall h \in H_0^1(\Omega),$$

$$\langle L_u(\bar{u}, \bar{z}, \bar{p}), h \rangle_{H^{-1}(\Omega), H_0^1(\Omega)} = 0 \quad \forall h \in H_0^1(\Omega), \tag{29}$$

$$(L_z(\bar{u}, \bar{z}, \bar{p}), (z - \bar{z}))_{L^2(\Omega)} \geq 0 \quad \forall z \in Z_{ad}.$$

It is not hard to see that (29) leads to the same optimality system as in (27).

7 Semilinear Quadratic PDE-Constrained Optimization Problem

The focus of the previous section was on the linear quadratic PDECO problem. However, things are more delicate when we replace the linear PDE constraint by a semilinear one. In this section, we provide a brief discussion of a PDECO problem governed by a semilinear PDE.

Let $\Omega \subset \mathbb{R}^n$, with $n \geq 2$, be a Lipschitz domain and the Assumption 3 holds. Moreover, let $u_d \in L^2(\Omega)$ and $z_a, z_b \in L^\infty(\Omega)$ with $z_a(x) < z_b(x)$ a.e. $x \in \Omega$ and $\lambda \geq 0$ be given. We then consider the following semilinear optimal control problem:

$$\min J(u, z) := \frac{1}{2}\|u - u_d\|_{L^2(\Omega)}^2 + \frac{\lambda}{2}\|z\|_{L^2(\Omega)}^2 \tag{30}$$

subject to $u \in L^\infty(\Omega) \cap H_0^1(\Omega)$ solving the weak form of

$$-\text{div} (A\nabla u) + f(\cdot, u) = z \quad \text{in } \Omega$$
$$u = 0 \quad \text{on } \partial\Omega \tag{31}$$

and

$$z \in Z_{ad} := \{v \in L^\infty(\Omega) : z_a(x) \leq v(x) \leq z_b(x) , \ a.e. \ x \in \Omega\}. \tag{32}$$

Notice the difference between the semilinear state equation (31) and the linear state equation (22). The key difficulty in the above problem is due to the nonlinearity introduced by f. The control bounds fulfill $z_a, z_b \in L^\infty(\Omega)$. This is different than the linear case where we assumed the bounds to be in $L^2(\Omega)$. This choice enforces the control $z \in L^\infty(\Omega)$ which in turn provides additional regularity for the state u as we discuss next.

In order to establish existence of solution to the state equation (31), we make certain assumptions on the nonlinearity f.

Assumption 6 *For a function $f : \Omega \times \mathbb{R} \to \mathbb{R}$ we consider the following assumption:*

$$
\begin{cases}
f(x, \cdot) \text{ is strictly increasing} & \text{for a.e. } x \in \Omega, \\
f(x, 0) = 0 & \text{for a.e. } x \in \Omega, \\
f(x, \cdot) \text{ is continuous} & \text{for a.e. } x \in \Omega, \\
f(\cdot, t) \text{ is measurable} & \text{for all } t \in \mathbb{R}, \\
\lim_{t \to \infty} f(x, t) = \infty & \text{for a.e. } x \in \Omega.
\end{cases}
$$

Remark 5 The condition $f(x, 0) = 0$ in Assumption 6 is not a restriction. If this condition on f cannot be verified, then it is enough to rewrite Equation (31) in Ω as

$$-\text{div} (A\nabla u) + f(\cdot, u) - f(\cdot, 0) = z - f(\cdot, 0) \quad \text{in } \Omega.$$

A typical example of f that fulfills Assumption 6 is given next (cf. [10, 11]).

Example 3 Let $q \in [1, \infty)$ and let $b : \Omega \to (0, \infty)$ be a function in $L^\infty(\Omega)$, that is, $b(x) > 0$ for a.e. $x \in \Omega$. Define the function $f : \Omega \times \mathbb{R} \to \mathbb{R}$ by $f(x, t) = b(x)|t|^{q-1}t$.

Theorem 7 (Existence and Uniqueness for Semilinear PDE) *Let Assumptions 3 and 6 hold. Then for every $z \in L^p(\Omega)$ with $p > \frac{n}{2}$, there exists a unique weak solution $u \in L^\infty(\Omega) \cap H_0^1(\Omega)$ to the state equation (31) and there exists a constant $C = C(\alpha, \beta, \Omega) > 0$ such that*

$$\|u\|_{H^1(\Omega)} + \|u\|_{L^\infty(\Omega)} \leq C\|z\|_{L^p(\Omega)}. \tag{33}$$

Proof The existence of solution is using the Browder–Minty theorem (cf [11, proposition 3.2] after setting $s = 1$). On the other hand, the $L^\infty(\Omega)$ regularity is by using a technique of Stampacchia [11, Theorem 3.5], see also [2, 21].

For the above minimization problem, we have (cf. Section 3)

$$U = H_0^1(\Omega) \cap L^\infty(\Omega), \quad Y = H^{-1}(\Omega), \quad Z = L^p(\Omega) \quad \text{with } p > n/2.$$

Notice that

$$Z_{ad} \subset L^\infty(\Omega) \subset Z.$$

As a result and owing to Theorem 7, the control to state map is well defined

$$S : L^\infty(\Omega) \to L^\infty(\Omega) \cap H_0^1(\Omega).$$

We notice that S is also well defined as a map from Z to $L^\infty(\Omega) \cap H_0^1(\Omega)$.[2] Due to S we can write the reduced problem as

$$\min_{z \in Z_{ad}} \mathscr{J}(z) := J(S(z), z). \tag{34}$$

In order to show the existence of solution to (34), typically f is assumed to be locally Lipschitz in the second argument to fulfill the assumptions on S in Corollary 1. This assumption was recently weakened in [11] and was replaced by the following growth condition on f: There exists a constant $c \in (0, 1]$ such that

$$c|f(x, \xi - \eta)| \le |f(x, \xi) - f(x, \eta)| \tag{35}$$

for a.e. $x \in \Omega$ and for all $\xi, \eta \in \mathbb{R}$. Such a growth condition is fulfilled by Example 3 (cf. [11]). Under this condition, we have the following existence result for (34).

Corollary 4 *Let the Assumptions of Theorem 7 hold. In addition, let f fulfills the growth condition* (35) *and that*

$$f(\cdot, w(\cdot)) \in L^2(\Omega) \text{ for every } w \in L^\infty(\Omega). \tag{36}$$

Then there exists a solution to (34).

Proof Similar to Corollary 3, the proof is again a consequence of Theorem 1. We interpret Z_{ad} as a subset of Z which is a reflexive Banach space. However, care must be taken to show the weak lower semicontinuity of $\mathscr{J}$. One has to carefully study the convergence of the state sequence $\{S(z_n)\}_{n\in\mathbb{N}}$. See for instance [11, Theorem 4.2].

Notice that the condition (36) is also fulfilled by Example 3.

Remark 6 We mention that all the results given in Corollary 4 remain true if one replaces the growth condition (35) and (36) with the following local Lipschitz continuity condition: For all $M > 0$ there exists a constant $L_M > 0$ such that f satisfies

$$|f(x, \xi) - f(x, \eta)| \le L_M|\xi - \eta| \tag{37}$$

for a.e. $x \in \Omega$ and $\xi, \eta \in \mathbb{R}$ with $|\eta|, |\xi| \le M$.

---

[2]Note both the choices of spaces for S are motivated by the theory of Nemytskii or superposition operators. Care must to taken to ensure their differentiability [81, Section 4.3].

For the remainder of this section, we will assume that $\lambda > 0$. Before we proceed further to derive the optimality conditions, we need some additional assumptions on f. Notice that the second-order derivatives are needed if one is interested in studying the second-order sufficient conditions.

Assumption 8 *We assume the following:*

(i) *The function $f(x, \cdot)$ is k-times, with $k = 1, 2$, continuously differentiable for a.e. $x \in \Omega$.*

(ii) *For all $M > 0$, there exists a constant $L_M > 0$ such that f satisfies (37) and*

$$\left| D_u^k f(x, \xi) - D_u^k f(x, \eta) \right| \leq L_M |\xi - \eta|, \quad k = 1, 2,$$

for a.e. $x \in \Omega$ and $\xi, \eta \in \mathbb{R}$ with $|\xi|, |\eta| \leq M$. Here D_u denotes the partial derivatives with respect to the second component.

(iii) *$D_u f(\cdot, 0) \in L^\infty(\Omega)$.*

(iv) *$D_{uu} f(\cdot, u(\cdot)) \in L^\infty(\Omega)$ whenever $u(\cdot) \in L^\infty(\Omega)$.*

Assumptions 8 help us prove that $S : L^\infty(\Omega) \to L^\infty(\Omega) \cap H_0^1(\Omega)$ is not only twice Fréchet differentiable (using the Implicit Function Theorem) but also twice continuous Fréchet differentiability of $\mathscr{J}$.

By invoking Corollary 2 the *first-order necessary optimality conditions* are given as follows: For every solution $\bar{z}$ of the problem (34), there exists a unique optimal state $\bar{u} = S(\bar{z})$ and an optimal adjoint state $\bar{p}$ such that

$$\bar{u} \in L^\infty(\Omega) \cap H_0^1(\Omega) : \int_\Omega A\nabla \bar{u} \cdot \nabla v \, dx + \int_\Omega f(x, \bar{u})v \, dx$$
$$= \int_\Omega \bar{z}v \, dx \quad \forall v \in H_0^1(\Omega)$$
$$\bar{p} \in H_0^1(\Omega) : \int_\Omega A\nabla \bar{p} \cdot \nabla v \, dx + \int_\Omega D_u f(x, \bar{u})\bar{p} \, dx$$
$$= \int_\Omega (\bar{u} - u_d)v \, dx \quad \forall v \in H_0^1(\Omega)$$
$$\bar{z} \in Z_{ad} : (\bar{p} + \lambda \bar{z}, z - \bar{z})_{L^2(\Omega)} \geq 0, \quad \forall z \in Z_{ad}.$$
$$(38)$$

Alternatively, one can use the Lagrangian approach of Section 6.4 to derive (38).

Notice that the variational inequality in (38) again can be written using the Projection formula as

$$\bar{z}(x) = \mathscr{P}_{Z_{ad}} \left\{ -\frac{1}{\lambda} \bar{p}(x) \right\} \quad \text{a.e. } x \in \Omega. \tag{39}$$

We further remark that since $\mathscr{J}$ is non-convex, in general due to the semilinear state equation, we cannot expect a global unique solution to the PDECO problem but only a local one. This local uniqueness can be shown by studying second order sufficient conditions. Nevertheless, care must be taken to prove such a result. This

is due to the fact that the penalty term on the control in the cost functional is in $L^2(\Omega)$. However, the constraints in Z_{ad} are in $L^\infty(\Omega)$. This leads to the so-called $L^2(\Omega) - L^\infty(\Omega)$ norm discrepancy and should be taken into account before considering second-order sufficient conditions. We refer to [81, Theorem 4.29] for details. We further remark that the second-order sufficient conditions are a useful tool to derive the discretization error estimates [13]. A further discussion is provided in Theorem 12.

8 Discrete Optimal Control Problem

We will illustrate the main ideas of the finite-element approximation of the PDE-constrained optimization problems in the case of linear elliptic problem with control constraints (21)–(23). First, we briefly review the basics of the finite-element discretization just for the state equation. For what follows, it is sufficient to take $f \in L^2(\Omega)$. We consider the weak form of Equation (19),

$$(A\nabla u, \nabla v)_{L^2(\Omega)} = (f, v)_{L^2(\Omega)} \quad \forall v \in H_0^1(\Omega).$$

We partition the domain Ω into elements. For simplicity, we only discuss the case when elements are simplices. For $h \in (0, h_0]; h_0 > 0$, let $\mathscr{T}_h$ denote a quasi-uniform triangulation of Ω with mesh size h, i.e., $\mathscr{T}_h = \{\tau\}$ is a partition of Ω into triangles or tetrahedrons τ of diameter h_τ such that for $h = \max_\tau h_\tau$,

$$\text{diam}(\tau) \leq h \leq C|\tau|^{\frac{1}{n}}, \quad \forall \tau \in \mathscr{T}_h,$$

where the constant $C > 0$ independent of h. For simplicity, we assume $\cup \tau = \Omega$. Let V_h be the set of all functions in $H_0^1(\Omega)$ that are continuous on Ω and linear on each τ. V_h is usually called the space of conforming Lagrange piecewise linear elements.

Now we define the finite-element Galerkin approximation $u_h \in V_h$ of (8), as the unique solution of

$$(A\nabla u_h, \nabla v_h)_{L^2(\Omega)} = (f, v_h)_{L^2(\Omega)} \quad \forall v_h \in V_h. \tag{40}$$

Expanding u_h in terms of basis functions, it is easy to see that (40) is equivalent to a system of linear equations and since $V_h \subset H_0^1(\Omega)$ the resulting matrix is nonsingular. Notice that by construction

$$(A\nabla(u - u_h), \nabla v_h)_{L^2(\Omega)} = 0 \quad \forall v_h \in V_h. \tag{41}$$

Thus, the Galerkin solution u_h is the orthogonal projection of u onto V_h with respect to the inner-product $(A\nabla \cdot, \nabla \cdot)_{L^2(\Omega)}$. Almost immediately we obtain the following key result.

Lemma 1 (Céa Lemma) *Let u and u_h satisfy* (41). *Then the following estimate holds*

$$\|u - u_h\|_{H^1(\Omega)} \leq C \min_{\chi \in V_h} \|u - \chi\|_{H^1(\Omega)}.$$

The constant C depends only on ellipticity, boundedness of the matrix A, and the domain Ω.

The above result says that the Galerkin solution is the best approximation to u from V_h in $H^1(\Omega)$-norm up to a constant. We can use Cea's lemma to derive a priori error estimates. Let $I_h : H^1(\Omega) \to V_h$ be a projection with the approximation properties

$$\|u - I_h u\|_{H^s(\Omega)} \leq Ch^{2-s}\|u\|_{H^2(\Omega)}, \quad s = 0, 1, \tag{42}$$

then from Cea's lemma immediately follows

$$\|u - u_h\|_{H^1(\Omega)} \leq Ch\|u\|_{H^2(\Omega)}.$$

Notice that the constant $C > 0$ above is independent of h. The error estimates in $L^2(\Omega)$-norm are not immediate, since the Galerkin solution does not have a property of being best approximation in $L^2(\Omega)$-norm; nevertheless, one can still establish optimal error estimates in $L^2(\Omega)$ via a duality argument, also known as Nitsche's trick. This result requires $H^2(\Omega)$-regularity.

Lemma 2 *Let Ω be convex or $C^{1,1}$ and u and u_h satisfy* (41). *Then there exists a constant C independent of h such that*

$$\|u - u_h\|_{L^2(\Omega)} \leq Ch \min_{\chi \in V_h} \|u - \chi\|_{H^1(\Omega)}.$$

Proof Let $e = u - u_h$ and consider a dual problem

$$(A\nabla w, \nabla v)_{L^2(\Omega)} = (e, v)_{L^2(\Omega)} \quad \forall v \in H_0^1(\Omega).$$

By setting, $v = e$, we obtain

$$\|e\|_{L^2(\Omega)}^2 = (A\nabla w, \nabla e)_{L^2(\Omega)} = (A\nabla(w - w_h), \nabla e)_{L^2(\Omega)},$$

where the last equality is due to (41). Next using the Cauchy–Schwarz inequality and the fact that, under given regularity of the domain $\|w\|_{H^2(\Omega)} \leq C\|e\|_{L^2(\Omega)}$, we obtain the required result.

Combining Cea's lemma and Lemma 2, we immediately establish the optimal a priori error estimate

$$\|u - u_h\|_{L^2(\Omega)} \leq Ch^2\|u\|_{H^2(\Omega)}.$$

Notice that the above estimate does require the convexity of Ω.

Corollary 5 *H^2 regularity also allows us to express the error in terms of data. Thus from Lemmas 1 and 2 it follows*

$$\|u - u_h\|_{L^2(\Omega)} + h\|u - u_h\|_{H^1(\Omega)} \leq Ch^2\|f\|_{L^2(\Omega)}.$$

8.1 Discrete Linear Quadratic PDE-Constrained Optimization Problem

For the remainder chapter, we will assume that $\lambda > 0$. To discretize the problem, we replace the state space $H_0^1(\Omega)$ with $V_h \subset H_0^1(\Omega)$ and the control space of admissible functions Z_{ad} with $Z_{ad,h} \subset Z_{ad}$. In case of unconstrained control, we can choose $Z_{ad,h} = V_h$. Theoretically, the mesh for the discretization of the state variable and the mesh for the discretization of the control can be different. However having two different meshes adds more technical difficulties for implementation. For this reason, it is more convenient to work with the same mesh which is what we assume from now on. Thus, the discretized problem (21)–(23) becomes

$$\min_{u_h \in V_h, \, z_h \in Z_{ad,h}} J_h(u_h, z_h) = \frac{1}{2}\|u_h - u_d\|_{L^2(\Omega)}^2 + \frac{\lambda}{2}\|z_h\|_{L^2(\Omega)}^2 \tag{43}$$

subject to

$$(A\nabla u_h, \nabla v_h) = (z_h, v_h) \quad \forall v_h \in V_h. \tag{44}$$

Similar to the infinite dimensional case, we define a discrete solution operator $S_h : Z_{ad,h} \to V_h$ to (44) and the reduced discrete problem becomes

$$\min_{z_h \in Z_{ad,h}} \mathscr{J}_h(z_h) := \min_{z_h \in Z_{ad,h}} J_h(S_h z_h, z_h). \tag{45}$$

Similar to the continuous problem, one can show that problem (45) has a unique solution $\bar{z}_h \in Z_{ad,h}$, the corresponding discrete optimal state is $\bar{u}_h = S_h(\bar{z}_h)$, and similar to Theorem 2 the first-order necessary and sufficient optimality condition is

$$\mathscr{J}_h'(\bar{z}_h)(z_h - \bar{z}_h) \geq 0, \quad \forall z_h \in Z_{ad,h}. \tag{46}$$

8.2 Optimization Problem Without Control Constraints

In this situation, $Z_{ad} = L^2(\Omega)$ and $Z_{ad,h} = V_h$ and as a result (25) and (46) reduce to equalities

$$\bar{z} = -\frac{1}{\lambda}\bar{p}, \quad \bar{z}_h = -\frac{1}{\lambda}\bar{p}_h, \tag{47}$$

correspondingly, and as a result the continuous and discrete PDECO problems are equivalent to the systems of equations

$$\bar{u} = S(-\frac{1}{\lambda}\bar{p})$$

$$\bar{p} = S^*(\bar{u} - u_d)$$

and

$$\bar{u}_h = S_h(-\frac{1}{\lambda}\bar{p}_h)$$

$$\bar{p}_h = S_h^*(\bar{u}_h - u_d).$$

As a result $\bar{z}, \bar{u}, \bar{p} \in H^2(\Omega)$ and we can expect second-order convergence for the optimal control in L^2 norm. Indeed, one can establish the following result.

Theorem 9 *Let $\bar{z}$ and $\bar{z}_h$ be optimal solutions to continuous and discrete PDECO problems* (21) *and* (43)*, respectively, without control constraints. Assume in addition that Ω is convex or $C^{1,1}$. Then there exists a constant C independent of h such that*

$$\|\bar{z} - \bar{z}_h\|_{L^2(\Omega)} \leq Ch^2 \left(\|\bar{z}\|_{L^2(\Omega)} + \|u_d\|_{L^2(\Omega)} \right).$$

Proof We begin by recalling the unconstrained optimality conditions $\lambda\bar{z} + \bar{p} = 0 = \lambda\bar{z}_h + \bar{p}_h$, which yields $(\lambda\bar{z} + \bar{p}, \bar{z}_h - \bar{z}) = 0 = (\lambda\bar{z}_h + \bar{p}_h, \bar{z} - \bar{z}_h)$. Adding these last two equalities, we arrive at

$$\lambda\|\bar{z}-\bar{z}_h\|_{L^2(\Omega)}^2 = (\bar{p}-\bar{p}_h, \bar{z}_h-\bar{z})_{L^2(\Omega)} = (S^*(S\bar{z}-u_d)-S_h^*(S_h\bar{z}_h-u_d), \bar{z}_h-\bar{z})_{L^2(\Omega)}, \tag{48}$$

where in the last equality we have used the representation of $\bar{p}$ and $\bar{p}_h$. Up on rewriting (48), we obtain

$$\lambda\|\bar{z}-\bar{z}_h\|_{L^2(\Omega)}^2 = (S^*S\bar{z}-S_h^*S_h\bar{z}_h, \bar{z}_h-\bar{z})_{L^2(\Omega)}+((S^*-S_h^*)u_d, \bar{z}_h-\bar{z})_{L^2(\Omega)} = I+II. \tag{49}$$

It follows from Corollary 5 that

$$|II| \leq Ch^2\|u_d\|_{L^2(\Omega)}\|\bar{z} - \bar{z}_h\|_{L^2(\Omega)}. \tag{50}$$

It then remains to estimate I in (49). We add and subtract $(S_h^*S_h\bar{z}, \bar{z}_h - \bar{z})_{L^2(\Omega)}$ to I and arrive at

$$I = ((S^*S - S_h^*S_h)\bar{z}, \bar{z}_h - \bar{z})_{L^2(\Omega)} + (S_h(\bar{z} - \bar{z}_h), S_h(\bar{z}_h - \bar{z}))_{L^2(\Omega)}$$

$$\leq ((S^*S - S_h^*S_h)\bar{z}, \bar{z}_h - \bar{z})_{L^2(\Omega)}, \tag{51}$$

where we have used the fact that $(S_h(\bar{z}-\bar{z}_h), S_h(\bar{z}_h-\bar{z}))_{L^2(\Omega)} = -\|S_h(\bar{z}-\bar{z}_h)\|^2 \le 0$. Again adding and subtracting $(S_h^* S\bar{z}, , \bar{z}_h - \bar{z})_{L^2(\Omega)}$ to (51), we arrive at

$$
\begin{aligned}
|I| &\le |((S^* - S_h^*)S\bar{z}, \bar{z}_h - \bar{z})_{L^2(\Omega)} + (S_h^*(S - S_h)\bar{z}, \bar{z}_h - \bar{z})_{L^2(\Omega)}| \\
&\le Ch^2 \|\bar{z}\|_{L^2(\Omega)} \|\bar{z} - \bar{z}_h\|_{L^2(\Omega)},
\end{aligned}
\tag{52}
$$

where we have first used the triangle inequality and have estimated the first term using Corollary 5 and continuity of $S : L^2(\Omega) \to L^2(\Omega)$: $|((S^* - S_h^*)S\bar{z}, \bar{z}_h - \bar{z})_{L^2(\Omega)}| \le Ch^2 \|S\bar{z}\|_{L^2(\Omega)} \|\bar{z} - \bar{z}_h\|_{L^2(\Omega)} \le Ch^2 \|\bar{z}\|_{L^2(\Omega)} \|\bar{z} - \bar{z}_h\|_{L^2(\Omega)}$. The estimate of the remaining term follows again using Corollary 5 and the continuity of $S_h^* : L^2(\Omega) \to L^2(\Omega)$. Finally, substituting the estimates of I and II from (52) and (50) in (49), we arrive at the asserted result.

8.3 Optimization Problem with Control Constraints

For the rest of this section, we assume constant box constraints, i.e., $z_a, z_b \in \mathbb{R}$, with $z_a < z_b$. We remind the reader that in this situation the optimal control is given by a projection formula:

$$
\bar{z}(x) = \mathscr{P}_{Z_{ad}}\left\{ -\frac{1}{\lambda}\bar{p}(x) \right\}.
\tag{53}
$$

If the constraints are active, $\bar{z} \notin H^2(\Omega)$. However, we can still conclude $\bar{z} \in H^1(\Omega)$ and even $\bar{z} \in W^1_\infty(\Omega)$ by using [58, Theorem A.1]. In light of this, the second-order convergence cannot in general be expected. There are several approaches to treat the problem.

8.3.1 Cell-Wise Constant Control Discretization

One idea is to consider a cellwise constant discretization of the control variable, i.e., we choose $Z_{ad,h} = Z_{ad} \cap Z_h^0$, where Z_h^0 is a space of piecewise constant functions on the partition $\mathscr{T}_h$. This idea goes back to Falk [35]. Since we consider piecewise constant discretization, only first-order convergence for the control can be expected. Indeed for such discretization, one can establish the following convergence result.

Theorem 10 *Let $\bar{z}$ and $\bar{z}_h$ be optimal solutions to continuous and discrete PDECO problems* (21) *and* (43)*, respectively, with control constraints* (23)*. Let $Z_{ad,h} = Z_{ad} \cap Z_h^0$. Assume in addition that Ω is convex or $C^{1,1}$. Then there exists a constant C independent of h such that*

$$
\|\bar{z} - \bar{z}_h\|_{L^2(\Omega)} \le Ch\left(\|\bar{z}\|_{H^1(\Omega)} + \|u_d\|_{L^2(\Omega)} \right).
$$

Proof First we define a projection $\pi_h : Z_{ad} \rightarrow Z_{ad} \cap Z_h^0$ by

$$\pi_h v \mid_\tau = \frac{1}{|\tau|} \int_\tau v \, dx, \quad \forall \tau \in \mathscr{T}_h. \tag{54}$$

Thus the projection π_h is the orthogonal projection onto Z_h^0 with respect to L^2-inner-product, i.e.

$$(v - \pi_h v, w)_{L^2(\Omega)} = 0, \quad w \in Z_h^0 \tag{55}$$

and has the following approximation property

$$\|v - \pi_h v\|_{L^2(\Omega)} \leq Ch\|\nabla v\|_{L^2(\Omega)}, \quad v \in H^1(\Omega). \tag{56}$$

Then replacing z by $\bar{z}_h$ in (25) and z_h by $\pi_h \bar{z}$ in (46), we arrive at

$$(\lambda \bar{z} + \bar{p}, \bar{z}_h - \bar{z})_{L^2(\Omega)} \geq 0, \quad (\lambda \bar{z}_h + \bar{p}_h, \pi_h \bar{z} - \bar{z}_h)_{L^2(\Omega)} \geq 0.$$

Adding these inequalities, we obtain that

$$\lambda \|\bar{z} - \bar{z}_h\|_{L^2(\Omega)}^2 \leq (\bar{p} - \bar{p}_h, \bar{z}_h - \bar{z})_{L^2(\Omega)} + (\lambda \bar{z}_h + \bar{p}_h, \pi_h \bar{z} - \bar{z})_{L^2(\Omega)} = I + II. \tag{57}$$

The estimate of I is exactly the same as in Theorem 9, i.e.,

$$|I| \leq Ch^2(\|\bar{z}\|_{L^2(\Omega)} + \|u_d\|_{L^2(\Omega)})\|\bar{z} - \bar{z}_h\|_{L^2(\Omega)}$$

$$\leq C_\lambda h^4 (\|\bar{z}\|_{L^2(\Omega)} + \|u_d\|_{L^2(\Omega)})^2 + \frac{\lambda}{4}\|\bar{z} - \bar{z}_h\|_{L^2(\Omega)}^2, \tag{58}$$

where we have used the Young's inequality in addition. Next we provide estimate for II. Using the characterization of $\bar{p}_h$, followed by, adding and subtracting $(S_h^*(S_h\bar{z} - u_d), \pi_h\bar{z} - \bar{z})_{L^2(\Omega)}$, and using the continuity of S_h^*, we obtain

$$II = (S_h^*(S_h\bar{z}_h - u_d), \pi_h\bar{z} - \bar{z})_{L^2(\Omega)}$$

$$= (S_h^* S_h(\bar{z}_h - \bar{z}), \pi_h\bar{z} - \bar{z})_{L^2(\Omega)} + (S_h^*(S_h\bar{z} - u_d), \pi_h\bar{z} - \bar{z})_{L^2(\Omega)} =: II_1 + II_2. \tag{59}$$

To estimate II_1, we use (56) and Young's inequality to arrive at

$$|II_1| \leq C_\lambda h^2 \|\bar{z}\|_{H^1(\Omega)}^2 + \frac{\lambda}{4}\|\bar{z} - \bar{z}_h\|_{L^2(\Omega)}^2. \tag{60}$$

It then remains to estimate II_2 in (59). By adding and subtracting $(S_h^*(S\bar{z} - u_d), \pi_h\bar{z} - \bar{z})_{L^2(\Omega)}$ in II_2, we obtain that

$$|II_2| = |(S_h^*(S_h - S)\bar{z}, \pi_h\bar{z} - \bar{z})_{L^2(\Omega)} + (S_h^*(S\bar{z} - u_d), \pi_h\bar{z} - \bar{z})_{L^2(\Omega)}|$$

$$\leq Ch^3\|\bar{z}\|_{H^1(\Omega)} + |((S_h^* - S^*)(S\bar{z} - u_d), \pi_h\bar{z} - \bar{z})_{L^2(\Omega)}|$$

$$+ |(S^*(S\bar{z} - u_d), \pi_h\bar{z} - \bar{z})_{L^2(\Omega)}| =: II_{2,1} + II_{2,2} + II_{2,3}, \tag{61}$$

where we have used Corollary 5 and (56) to estimate the first term. Moreover, we have added and subtracted $(S^*(S\bar{z} - u_d), \pi_h\bar{z} - \bar{z})_{L^2(\Omega)}$ to the second term. It then remains to estimate $II_{2,2}$ and $II_{2,3}$. Again using Corollary 5 and (56), we obtain that $II_{2,2} \leq Ch^3(\|\bar{z}\|_{L^2(\Omega)} + \|u_d\|_{L^2(\Omega)})\|\bar{z}\|_{H^1(\Omega)}$. Finally, to estimate $II_{2,3}$ we first recall that $S^*(S\bar{z} - u_d) = \bar{p}$. Since π_h is L^2-orthogonal projection, we obtain

$$II_{2,3} = (\bar{p} - \pi_h\bar{p}, \pi_h\bar{z} - \bar{z})_{L^2(\Omega)} \leq Ch^3(\|\bar{z}\|_{L^2(\Omega)} + \|u_d\|_{L^2(\Omega)})\|\bar{z}\|_{H^1(\Omega)}. \tag{62}$$

Collecting all the estimates, we arrive at the asserted result.

Comparing this result with the unconstrained case, we have only first-order convergence. This is mainly due to the choice of the discrete control space which does not take the full advantage of the regularity of the optimal control, namely $\bar{z} \in W_\infty^1(\Omega)$. Moreover, away from the active constraints $\bar{z}$ is still in H^2. Taking this in consideration, there are some alternatives to increase the order of the convergence.

8.3.2 Cell-Wise Linear Control Discretization

To improve the convergence rate of the above result, we consider $Z_{ad,h} = Z_{ad} \cap V_h$, i.e., the control space consists of piecewise linear functions satisfying constraints (22). The approximation properties in this setting were investigated in a number of papers, for example [72, 74]. We will not provide all the details, only highlight the main ideas. To take advantage of the regularity for $\bar{z}$ discussed above, we consider the following sets:

$$\mathcal{T}_h^1 = \{\tau \in \mathcal{T}_h \mid \bar{z}\mid_\tau = z_a \text{ or } \bar{z}\mid_\tau = z_b\},$$

which is the set of active cells,

$$\mathcal{T}_h^2 = \{\tau \in \mathcal{T}_h \mid z_a < \bar{z}\mid_\tau < z_b\},$$

the set of not active cells, and the rest

$$\mathcal{T}_h^3 = \mathcal{T}_h \backslash (\mathcal{T}_h^1 \cup \mathcal{T}_h^2).$$

Then under the assumption that

$$meas(\mathcal{T}_h^3) \leq Ch, \tag{63}$$

which is valid, for example, if the boundary of the active set consists of a finite number of rectifiable curves, one can establish the following result.

Theorem 11 *Let $\bar{z}$ and $\bar{z}_h$ be optimal solutions to continuous and discrete PDECO problems* (21) *and* (43)*, respectively, with control constraints. Assume in addition that Ω is convex or $C^{1,1}$, $u_d \in L^p(\Omega)$ for $p > n$ and assumptions* (63) *hold. Then there exists a constant C independent of h such that*

$$\|\bar{z} - \bar{z}_h\|_{L^2(\Omega)} \leq Ch^{\frac{3}{2}} \left(\|\bar{p}\|_{H^2(\Omega)} + \|\nabla\bar{z}\|_{L^\infty(\Omega)} \right).$$

Proof The proof of this result can be found in [73].

Remark 7 (Variational Discretization) The idea of the variational discretization approach introduced by Hinze [52] is not to discretize the control variable, i.e., to choose $Z_{ad,h} = Z_{ad}$. This approach does give second-order convergence for the control but requires a nonstandard implementation, especially for $n > 1$.

8.4 Semilinear Equations

Similar to the linear case, we discretize the problem with finite elements. Thus, we replace the state space $H_0^1(\Omega)$ with $V_h \subset H_0^1(\Omega)$ and the control space of admissible functions Z_{ad} with $Z_{ad,h} \subset Z_{ad}$. We additionally assume that $z_a, z_b \in \mathbb{R} \cup \{\pm\infty\}$, with $z_a < z_b$. In case of unconstrained control, we take $z_a = -\infty$ and $z_b = \infty$, i.e., $Z_{ad,h} = V_h$. The discretized problem (30)–(32) becomes

$$\min_{u_h \in V_h, \, z_h \in Z_{ad,h}} J_h(u_h, z_h) = \frac{1}{2}\|u_h - u_d\|_{L^2(\Omega)}^2 + \frac{\lambda}{2}\|z_h\|_{L^2(\Omega)}^2 \tag{64}$$

subject to

$$(A\nabla u_h, \nabla v_h) + (f(\cdot, u_h), v_h) = (z_h, v_h) \quad \forall v_h \in V_h. \tag{65}$$

All of the above strategies for choosing $Z_{ad,h}$ can be applied for the semilinear problem as well and similar error estimates can be obtained. Of course the arguments are more technical and we refer to [13, 22, 29] for details.

Theorem 12 *Assume that Ω is convex domain with $C^{1,1}$ boundary and let Assumptions 6 and 8 be satisfied. Let $\bar{z}$ be a strict local solution to* (30)–(32) *that fulfills the second-order optimality condition: There exist $\delta > 0$ and $\tau > 0$ such that*

$$\mathcal{J}''(\bar{z})(z, z) \geq \delta\|z\|_{L^2(\Omega)}^2$$

holds for all $z \in L^\infty(\Omega)$ satisfying

$$z(x) \begin{cases} \geq 0 & \text{if } \bar{z}(x) = z_a, \\ \leq 0 & \text{if } \bar{z}(x) = z_b, \\ = 0 & \text{if } |\lambda\bar{z} + \bar{p}| \geq \tau > 0. \end{cases} \tag{66}$$

Then:

- *[13, Thm. 5.1] (Cell-wise constant control) Let $Z_{ad,h} = Z_{ad} \cap Z_h^0$ and $\{\bar{z}_h\}$ be a sequence of locally optimal piecewise constant solutions to (64)–(65) that converges strongly in $L^2(\Omega)$ to $\bar{z}$. Then there exists a constant C independent of h such that*

$$\|\bar{z} - \bar{z}_h\|_{L^2(\Omega)} \leq Ch.$$

- *[29, Thm. 4.5](Cell-wise linear control) Let $Z_{ad,h} = Z_{ad} \cap V_h$ and $\{\bar{z}_h\}$ be a sequence of locally optimal piecewise linear solutions to (64)–(65) that converges strongly in $L^2(\Omega)$ to $\bar{z}$. If in addition (63) holds, then there exists a constant C independent of h such that*

$$\|\bar{z} - \bar{z}_h\|_{L^2(\Omega)} \leq Ch^{\frac{3}{2}}.$$

9 Conclusion and the Current State of the Art

In this introductory chapter, we reviewed the main ideas behind the study of PDECO. We briefly mentioned numerical approximation of such problems by the finite-element method and showed some convergence results in the case of control constraints. However, the subject is vast with many active research directions. In this final section, let us mention some topics that we skipped and some active areas of research that we did not touch.

- In the above discussion, we only touched problems with control constraints. However, problems with state constraints are equally important. In contrast to the control constraints, which basically amount to projection onto the feasible set, the state constraints require much more care since the Lagrange multipliers are not functions, only measures. We refer to [81, Chapter 6] for a nice introduction on the subject. For such problems, the error analysis is also more subtle and one often has to reserve to more technical pointwise error estimates to derive optimal-order convergence. We refer to [28, 64, 65] for some recent developments of the subject.
- All above error estimates were a priori error estimates. However, a large fraction of the literature on finite elements is devoted to a posteriori error estimates, i.e., estimates where the error between the true solution and the discrete approximated solution is expressed in terms of computable quantities. We refer to [75, 77] for more recent development of the subject.

- One can consider more complicated state equations or even systems, which can be linear or nonlinear, time dependent, variational inequalities, and so on. Currently, the theory is well developed for problems constrained by linear and semilinear elliptic problems, but the research is very much active for nonlinear, time dependent, and variational inequalities [61, 68].
- We only consider a quadratic cost functional in our error analysis. However, other choices may be desired. For example, it was observed numerically that seeking the control from the space of regular Borel measures forces the sparsity of optimal solution, meaning that the support of solution is small. This phenomenon was analyzed for elliptic and parabolic problems in a number of papers [23–25]; however, there are still some remaining open questions.
- In all our examples, we considered the distributed control, i.e., the control z was acting in the interior of the domain. However, problems where the control acts on the boundary are important in many applications. The control can enter as Dirichlet, Neumann, or Robin boundary conditions. Because of the variational structure of the problems, the Neumann and Robin boundary conditions naturally enter the variational form of the state equation and as a result Neumann boundary controls can be naturally analyzed, see [81, Chapter 2], we refer to [26, 27] for the Robin case. Dirichlet boundary conditions do not have this property and one has to use more sophisticated machinery to overcome technical difficulties and to derive optimal error estimate for the optimal solution [12]. Alternatively, one can also use the penalized Robin boundary conditions to study the Dirichlet boundary control problems [63].

References

1. R.A. Adams and J.J.F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003.
2. J.-J. Alibert and J.-P. Raymond. Boundary control of semilinear elliptic equations with discontinuous leading coefficients and unbounded controls. *Numer. Funct. Anal. Optim.*, 18(3–4):235–250, 1997.
3. H. Antil, M. Heinkenschloss, and R. H.W. Hoppe. Domain decomposition and balanced truncation model reduction for shape optimization of the Stokes system. *Optimization Methods and Software*, 26(4–5):643–669, 2011.
4. H. Antil, M. Heinkenschloss, R.H.W. Hoppe, C. Linsenmann, and A. Wixforth. Reduced order modeling based shape optimization of surface acoustic wave driven microfluidic biochips. *Math. Comput. Simul.*, 82(10):1986–2003, June 2012.
5. H. Antil, M. Heinkenschloss, R.H.W. Hoppe, and D. C. Sorensen. Domain decomposition and model reduction for the numerical solution of PDE constrained optimization problems with localized optimization variables. *Comput. Vis. Sci.*, 13(6):249–264, 2010.
6. H. Antil, M. Heinkenschloss, and D. C. Sorensen. Application of the discrete empirical interpolation method to reduced order modeling of nonlinear and parametric systems. volume 8 of *Springer MS&A series: Reduced Order Methods for modeling and computational G. Rozza, Eds*. Springer-Verlag Italia, Milano, 2013.

7. H. Antil, M. Hintermüller, R. Nochetto, T. Surowiec, and D. Wegner. Finite horizon model predictive control of electrowetting on dielectric with pinning. *Interfaces Free Bound.*, 19(1): 1–30, 2017.

8. H. Antil, R.H. Nochetto, and P. Venegas. Controlling the Kelvin force: Basic strategies and applications to magnetic drug targeting. *arXiv preprint arXiv:1704.06872*, 2017.

9. H. Antil, R.H. Nochetto, and P. Venegas. Optimizing the Kelvin force in a moving target subdomain. *Math. Models Methods Appl. Sci.*, 28(1):95–130, 2018.

10. H. Antil, J. Pfefferer, and M. Warma. A note on semilinear fractional elliptic equation: analysis and discretization. *Math. Model. Numer. Anal. (ESAIM: M2AN)*, 51:2049–2067, 2017.

11. H. Antil and M. Warma. Optimal control of fractional semilinear PDEs. *arXiv preprint arXiv:1712.04336*, 2017.

12. T. Apel, M. Mateos, J. Pfefferer, and A. Rösch. On the regularity of the solutions of Dirichlet optimal control problems in polygonal domains. *SIAM J. Control Optim.*, 53(6):3620–3641, 2015.

13. N. Arada, E. Casas, and F. Tröltzsch. Error estimates for the numerical approximation of a semilinear elliptic control problem. *Comput. Optim. Appl.*, 23(2):201–229, 2002.

14. H. Attouch, G. Buttazzo, and G. Michaille. *Variational analysis in Sobolev and BV spaces*, volume 17 of *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, second edition, 2014. Applications to PDEs and optimization.

15. M. Benzi, G.H. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta Numer.*, 14:1–137, 2005.

16. M. Berggren and M. Heinkenschloss. Parallel solution of optimal-control problems by time-domain decomposition. In *Computational science for the 21st century. Symposium*, pages 102–112, 1997.

17. P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence rates for greedy algorithms in reduced basis methods. *SIAM J. Math. Anal.*, 43(3):1457–1472, 2011.

18. J.F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer Series in Operations Research. Springer-Verlag, New York, 2000.

19. S.C. Brenner and L.R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

20. E. Casas. Control of an elliptic problem with pointwise state constraints. *SIAM J. Control Optim.*, 24(6):1309–1318, 1986.

21. E. Casas. Boundary control of semilinear elliptic equations with pointwise state constraints. *SIAM J. Control Optim.*, 31(4):993–1006, 1993.

22. E. Casas. Using piecewise linear functions in the numerical approximation of semilinear elliptic control problems. *Adv. Comput. Math.*, 26(1–3):137–153, 2007.

23. E. Casas, C. Clason, and K. Kunisch. Approximation of elliptic control problems in measure spaces with sparse solutions. *SIAM J. Control Optim.*, 50(4):1735–1752, 2012.

24. E. Casas, C. Clason, and K. Kunisch. Parabolic control problems in measure spaces with sparse solutions. *SIAM J. Control Optim.*, 51(1):28–63, 2013.

25. E. Casas and K. Kunisch. Optimal control of semilinear elliptic equations in measure spaces. *SIAM J. Control Optim.*, 52(1):339–364, 2014.

26. E. Casas, M. Mateos, and J.-P. Raymond. Penalization of Dirichlet optimal control problems. *ESAIM Control Optim. Calc. Var.*, 15(4):782–809, 2009.

27. E. Casas, M. Mateos, and F. Tröltzsch. Error estimates for the numerical approximation of boundary semilinear elliptic control problems. *Comput. Optim. Appl.*, 31(2):193–219, 2005.

28. E. Casas, M. Mateos, and B. Vexler. New regularity results and improved error estimates for optimal control problems with state constraints. *ESAIM Control Optim. Calc. Var.*, 20(3):803–822, 2014.

29. E. Casas and F. Tröltzsch. A general theorem on error estimates with application to a quasilinear elliptic optimal control problem. *Comput. Optim. Appl.*, 53(1):173–206, 2012.

30. A. Cohen, M. Hoffmann, and M. Reiß. Adaptive wavelet Galerkin methods for linear inverse problems. *SIAM J. Numer. Anal.*, 42(4):1479–1501, 2004.

31. W. Dahmen, A. Kunoth, and K. Urban. A wavelet Galerkin method for the Stokes equations. *Computing*, 56(3):259–301, 1996. International GAMM-Workshop on Multi-level Methods (Meisdorf, 1994).

32. X. Deng, X.-C. Cai, and J. Zou. Two-level space-time domain decomposition methods for three-dimensional unsteady inverse source problems. *J. Sci. Comput.*, 67(3):860–882, 2016.

33. X. Deng and M. Heinkenschloss. A parallel-in-time gradient-type method for discrete time optimal control problems. 2016.

34. A. Ern and J.L. Guermond. *Theory and practice of finite elements*, volume 159 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2004.

35. R.S. Falk. Approximation of a class of optimal control problems with order of convergence estimates. *J. Math. Anal. Appl.*, 44:28–47, 1973.

36. P.E. Farrell, D.A. Ham, S.W. Funke, and M.E. Rognes. Automated derivation of the adjoint of high-level transient finite element programs. *SIAM J. Sci. Comput.*, 35(4):C369–C393, 2013.

37. G.B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

38. D. Gilbarg and N.S. Trudinger. *Elliptic partial differential equations of second order*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.

39. V. Girault and P.-A. Raviart. *Finite element methods for Navier-Stokes equations*, volume 5 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1986. Theory and algorithms.

40. P. Grisvard. *Singularities in boundary value problems*, volume 22 of *Recherches en Mathématiques Appliquées [Research in Applied Mathematics]*. Masson, Paris, 1992.

41. M. Gubisch and S. Volkwein. Proper orthogonal decomposition for linear-quadratic optimal control. In *Model reduction and approximation*, volume 15 of *Comput. Sci. Eng.*, pages 3–63. SIAM, Philadelphia, PA, 2017.

42. M.D. Gunzburger. *Perspectives in flow control and optimization*, volume 5 of *Advances in Design and Control*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2003.

43. J. Haslinger and R. A. E. Mäkinen. *Introduction to shape optimization*, volume 7 of *Advances in Design and Control*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2003. Theory, approximation, and computation.

44. M. Heinkenschloss. Numerical solution of implicitly constrained optimization problems. 2008.

45. J.S. Hesthaven, G. Rozza, and B. Stamm. *Certified reduced basis methods for parametrized partial differential equations*. SpringerBriefs in Mathematics. Springer, Cham; BCAM Basque Center for Applied Mathematics, Bilbao, 2016. BCAM SpringerBriefs.

46. J.S. Hesthaven and T. Warburton. *Nodal discontinuous Galerkin methods*, volume 54 of *Texts in Applied Mathematics*. Springer, New York, 2008. Algorithms, analysis, and applications.

47. M. Hintermüller and M. Hinze. Moreau-Yosida regularization in state-constrained elliptic control problems: error estimates and parameter adjustment. *SIAM J. Numer. Anal.*, 47(3):1666–1683, 2009.

48. M. Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set strategy as a semismooth Newton method. *SIAM J. Optim.*, 13(3):865–888 (2003), 2002.

49. M. Hintermüller and I. Kopacka. Mathematical programs with complementarity constraints in function space: C- and strong stationarity and a path-following algorithm. *SIAM J. Optim.*, 20(2):868–902, 2009.

50. M. Hintermüller and I. Kopacka. A smooth penalty approach and a nonlinear multigrid algorithm for elliptic MPECs. *Comput. Optim. Appl.*, 50(1):111–145, 2011.

51. M. Hintermüller and M. Ulbrich. A mesh-independence result for semismooth Newton methods. *Math. Program.*, 101(1, Ser. B):151–184, 2004.

52. M. Hinze. A variational discretization concept in control constrained optimization: the linear-quadratic case. *Comput. Optim. Appl.*, 30(1):45–61, 2005.

53. M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE constraints*, volume 23 of *Mathematical Modelling: Theory and Applications*. Springer, New York, 2009.

54. M. Hinze and S. Volkwein. Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: error estimates and suboptimal control. In *Dimension reduction of large-scale systems*, volume 45 of *Lect. Notes Comput. Sci. Eng.*, pages 261–306. Springer, Berlin, 2005.

55. M. Hinze and S. Volkwein. Error estimates for abstract linear-quadratic optimal control problems using proper orthogonal decomposition. *Comput. Optim. Appl.*, 39(3):319–345, 2008.

56. K. Ito and K. Kunisch. *Lagrange multiplier approach to variational problems and applications*, volume 15 of *Advances in Design and Control*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008.

57. C.T. Kelley. *Iterative methods for optimization*, volume 18 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999.

58. D. Kinderlehrer and G. Stampacchia. *An introduction to variational inequalities and their applications*, volume 31 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000. Reprint of the 1980 original.

59. K. Kohls, A. Rösch, and K.G. Siebert. A posteriori error analysis of optimal control problems with control constraints. *SIAM J. Control Optim.*, 52(3):1832–1861, 2014.

60. D. Kouri, D. Ridzal, and G.von Winckel. Webpage. https://trilinos.org/packages/rol/.

61. G. Leugering, S. Engell, A. Griewank, M. Hinze, R. Rannacher, V. Schulz, M. Ulbrich, and S. Ulbrich, editors. *Constrained optimization and optimal control for partial differential equations*, volume 160 of *International Series of Numerical Mathematics*. Birkhäuser/Springer Basel AG, Basel, 2012.

62. J.-L. Lions. *Optimal control of systems governed by partial differential equations.* Translated from the French by S. K. Mitter. Die Grundlehren der mathematischen Wissenschaften, Band 170. Springer-Verlag, New York-Berlin, 1971.

63. S. May, R. Rannacher, and B. Vexler. Error analysis for a finite element approximation of elliptic Dirichlet boundary control problems. *SIAM J. Control Optim.*, 51(3):2585–2611, 2013.

64. I. Neitzel, J. Pfefferer, and A. Rösch. Finite element discretization of state-constrained elliptic optimal control problems with semilinear state equation. *SIAM J. Control Optim.*, 53(2):874–904, 2015.

65. I. Neitzel and F. Tröltzsch. Numerical analysis of state-constrained optimal control problems for PDEs. In *Constrained optimization and optimal control for partial differential equations*, volume 160 of *Internat. Ser. Numer. Math.*, pages 467–482. Birkhäuser/Springer Basel AG, Basel, 2012.

66. J. Nocedal and S.J. Wright. *Numerical optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2006.

67. R.H. Nochetto, K.G. Siebert, and A. Veeser. Theory of adaptive finite element methods: an introduction. In *Multiscale, nonlinear and adaptive approximation*, pages 409–542. Springer, Berlin, 2009.

68. E. Rocca G. Schimperna J. Sprekels P. Colli, A. Favini, editor. *Solvability, Regularity, and Optimal Control of Boundary Value Problems for PDEs*, volume 22 of *Springer INdAM Series*. Springer, Cham, 2017. In Honour of Prof. Gianni Gilardi.

69. A. Quarteroni, A. Manzoni, and F. Negri. *Reduced basis methods for partial differential equations*, volume 92 of *Unitext*. Springer, Cham, 2016. An introduction, La Matematica per il 3+2.

70. B. Rivière. *Discontinuous Galerkin methods for solving elliptic and parabolic equations*, volume 35 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008. Theory and implementation.

71. S.M. Robinson. Stability theory for systems of inequalities. II. Differentiable nonlinear systems. *SIAM J. Numer. Anal.*, 13(4):497–513, 1976.

72. A. Rösch. Error estimates for linear-quadratic control problems with control constraints. *Optim. Methods Softw.*, 21(1):121–134, 2006.

73. A. Rösch and R. Simon. Linear and discontinuous approximations for optimal control problems. *Numer. Funct. Anal. Optim.*, 26(3):427–448, 2005.
74. A. Rösch and R. Simon. Superconvergence properties for optimal control problems discretized by piecewise linear and discontinuous functions. *Numer. Funct. Anal. Optim.*, 28(3–4): 425–443, 2007.
75. A. Rösch and D. Wachsmuth. A-posteriori error estimates for optimal control problems with state and control constraints. *Numer. Math.*, 120(4):733–762, 2012.
76. A. Schiela and S. Ulbrich. Operator preconditioning for a class of inequality constrained optimal control problems. *SIAM J. Optim.*, 24(1):435–466, 2014.
77. R. Schneider and G. Wachsmuth. A posteriori error estimation for control-constrained, linear-quadratic optimal control problems. *SIAM J. Numer. Anal.*, 54(2):1169–1192, 2016.
78. C. Schwab and C.J. Gittelson. Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numer.*, 20:291–467, 2011.
79. A. Shapiro. On concepts of directional differentiability. *J. Optim. Theory Appl.*, 66(3):477–487, 1990.
80. J. Sokołowski and J.-P. Zolésio. *Introduction to shape optimization*, volume 16 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1992. Shape sensitivity analysis.
81. F. Tröltzsch. *Optimal control of partial differential equations*, volume 112 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010. Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels.
82. M. Ulbrich. *Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces*, volume 11 of *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, 2011.
83. S. Ulbrich. Preconditioners based on "parareal" time-domain decomposition for time-dependent PDE-constrained optimization. In *Multiple Shooting and Time Domain Decomposition Methods*, pages 203–232. Springer, 2015.
84. A.J. Wathen. Preconditioning. *Acta Numer.*, 24:329–376, 2015.
85. J. Zowe and S. Kurcyusz. Regularity and stability for the mathematical programming problem in Banach spaces. *Appl. Math. Optim.*, 5(1):49–62, 1979.