2 Problem formulation

The goal of this chapter is to introduce the Optimal Control Problem (OCP) formulation which serves as the point of origin for all further investigations in this thesis. To this end we recapitulate elements of the theory of parabolic Partial Differential Equations (PDEs) in Section 2.1 and present a system of PDEs coupled with Ordinary Differential Equations (ODEs) in Section 2.2. The coupled system is one of the constraints among additional boundary and path constraints for the OCP which we describe in Section 2.3. We emphasize the particular aspects in which our problem setting differs and extends the setting most often found in PDE constrained optimization.

2.1 Dynamical models described by Partial Differential Equations

We treat processes which are modeled by a state u distributed in space and evolving over time. The evolution of u is deterministic and described by PDEs. The behavior of the dynamical system can further be influenced by a time and possibly space dependent control q.

Nonlinear instationary PDEs usually do not have solutions in classical function spaces. We recapitulate the required definitions for Bochner spaces and vector-valued distributions necessary for formulations which have solutions in a weak sense. The presentation here is based on Dautray and Lions [37], Gajewski et al. [57], and Wloka [167]. We omit all proofs which can be found therein. Throughout this chapter let $\Omega \in \mathbb{R}^d$ be a bounded open domain with sufficiently regular boundary $\partial\Omega$, X be a Banach space, and $\mathrm{d}\mu$ denote the Lebesgue measure in $\mathbb{R}^d$.

We assume that the reader is familiar with basic concepts of functional analysis (see, e.g., Dunford and Schwartz [50]). We denote with $L^p(\Omega), 1 \leq p \leq \infty$, the *Lebesgue space* of μ-measurable $\mathbb{R}$-valued functions whose absolute value to the p-th power has a bounded integral over Ω if $p < \infty$ or which are essentially bounded if $p = \infty$. Functions which coincide μ-almost everywhere are considered identical. With $W^{k,p}(\Omega)$, $k \geq 0, 1 \leq p < \infty$ we denote the *Sobolev space* of functions in $L^p(\Omega)$ whose distributional derivatives up to order k lie in $L^p(\Omega)$.

A. Potschka, *A Direct Method for Parabolic PDE Constrained Optimization Problems,*
Advances in Numerical Mathematics, DOI 10.1007/978-3-658-04476-3_2,
© Springer Fachmedien Wiesbaden 2014

The spaces $L^p(\Omega), W^{k,p}(\Omega)$ endowed with their usual norms are Banach spaces. The spaces $H^k(\Omega) := W^{k,2}(\Omega)$ equipped with their usual scalar product are Hilbert spaces. The construction of $L^p(\Omega)$ and $W^{k,p}(\Omega)$ can be generalized to functions with values in Banach spaces:

Definition 2.1 (Bochner spaces)**.** By $L^p(\Omega;X)$, $1 \le p < \infty$, we denote the space of all measurable functions $v : \Omega \to X$ satisfying

$$\int_\Omega \|v\|_X^p \, \mathrm{d}\mu < \infty.$$

We identify elements of $L^p(\Omega;X)$ which coincide μ-almost everywhere and equip $L^p(\Omega;X)$ with the norm

$$\|v\|_{L^p(\Omega;X)} = \left(\int_\Omega \|v\|_X^p \, \mathrm{d}\mu \right)^{1/p}.$$

Now we proceed in the following way: Generally we are interested in weak solutions $u \in W$ in an appropriate Hilbert space $W \subset L^2((t_1,t_2) \times \Omega)$ with finite $t_1, t_2 \in \mathbb{R}$. Functions in $L^2((t_1,t_2) \times \Omega)$ need not even be continuous and hence we must exercise care to give well-defined meaning to derivatives and the traces $u(t_1,.)$ and $u(t_2,.)$. This is not trivial because altering u on any set of measure zero, e.g., $\{t_1,t_2\} \times \Omega$, yields the same u in $L^2((t_1,t_2) \times \Omega)$. The traces are important for the formulation of boundary value conditions. We address these issues concerning the state space in three steps. In a first step, we write

$$L^2((t_1,t_2) \times \Omega) = L^2((t_1,t_2); L^2(\Omega)),$$

i.e., we interpret u as an L^2 function in time with values in the space of L^2 functions in space. Second, we can formulate the time derivative $\mathrm{d}u/\mathrm{d}t$ of u via the concept of vectorial distributional derivatives.

Definition 2.2. Let Y be another Banach space. We denote the space of continuous linear mappings from X to Y with $\mathscr{L}(X,Y)$.

Definition 2.3. The space of vectorial distributions of the interval $(t_1,t_2) \subset \mathbb{R}$ with values in the Banach space X is denoted by

$$\mathscr{D}'((t_1,t_2);X) := \mathscr{L}(C^\infty([t_1,t_2];\mathbb{R}), X).$$

We can identify every $u \in L^2((t_1,t_2);X) \subset L^1((t_1,t_2);X)$ with a distribution $T \in \mathscr{D}'((t_1,t_2);X)$ via the Bochner integral

$$T\varphi = \int_{t_1}^{t_2} u(t)\varphi(t)\mathrm{d}t \quad \text{for all } \varphi \in C^\infty([t_1,t_2];\mathbb{R}).$$

Definition 2.4. The k-th derivative of T is defined via

$$\frac{\mathrm{d}^k T}{\mathrm{d}t^k}\varphi = (-1)^k \int_{t_1}^{t_2} u(t)\varphi^{(k)}(t)\mathrm{d}t.$$

Thus, $\mathrm{d}T/\mathrm{d}t \in \mathscr{D}'((t_1,t_2);X)$. We assume now that $X \hookrightarrow Y$, where $\hookrightarrow$ denotes continuous embedding. Hence it holds that

$$\mathscr{D}'((t_1,t_2);X) \hookrightarrow \mathscr{D}'((t_1,t_2);Y),$$
$$L^p((t_1,t_2);X) \hookrightarrow L^p((t_1,t_2);Y).$$

Let $u \in L^2((t_1,t_2);X)$. We say that $\mathrm{d}u/\mathrm{d}t \in L^2((t_1,t_2);Y)$ if there exists $u' \in L^2((t_1,t_2),Y)$ such that

$$\int_{t_1}^{t_2} u'(t)\varphi(t)\mathrm{d}t = \frac{\mathrm{d}T}{\mathrm{d}t}\varphi = -\int_{t_1}^{t_2} u(t)\varphi^{(1)}(t)\mathrm{d}t \quad \text{for all } \varphi \in C^\infty([t_1,t_2];\mathbb{R}),$$

and we identify $\mathrm{d}u/\mathrm{d}t := u'$. We also use the abbreviation $\partial_t u := \mathrm{d}u/\mathrm{d}t$.

In the third step, let V and H be separable Hilbert spaces and let V^* denote the dual space of V. We assume throughout that (V,H,V^*) is a *Gelfand triple*

$$V \overset{\mathrm{d}}{\hookrightarrow} H \overset{\mathrm{d}}{\hookrightarrow} V^*,$$

i.e., the embeddings of V in H and $H = H^*$ in V^* are continuous and dense. Now we choose $X = V$ and $Y = V^*$ in order to define the space of L^2 functions over V with time derivatives in L^2 over the dual V^* according to

$$W(t_1,t_2) = \{u \in L^2((t_1,t_2);V) \mid \partial_t u \in L^2((t_1,t_2);V^*)\}.$$

Lemma 2.5. *The space $W(t_1,t_2)$ is a Hilbert space when endowed with the scalar product*

$$(u,v)_{W(t_1,t_2)} = \int_{t_1}^{t_2} (u(t),v(t))_V \,\mathrm{d}t + \int_{t_1}^{t_2} (\partial_t u(t), \partial_t v(t))_{V^*} \,\mathrm{d}t.$$

Proof. See Wloka [167, Satz 25.4]. □

Theorem 2.6. *We can alter every $u \in W(t_1,t_2)$ on a set of measure zero to obtain a function in $C^0([t_1,t_2];H)$. Furthermore, if we equip $C^0([t_1,t_2];H)$ with the norm of uniform convergence then*

$$W(t_1,t_2) \hookrightarrow C^0([t_1,t_2];H).$$

Proof. See Dautray and Lions [37, Chapter XVIII, Theorem 1]. □

Corollary 2.7. *For $u \in W(t_1,t_2)$ the traces $u(t_1), u(t_2)$ have a well-defined meaning in H (but not in V in general).*

For the control variables we assume $q \in L^2((t_1,t_2);Q)$ where $Q \subseteq L^2(\Omega)^{n_q}$ or $Q \subseteq L^2(\partial\Omega)^{n_q}$ for distributed or boundary control, respectively. We can then formulate the parabolic differential equation

$$\partial_t u(t) + A(q(t),u(t)) = 0, \tag{2.1}$$

with a nonlinear elliptic differential operator $A : Q \times V \to V^*$. In the numerical approaches which we present in Chapters 5 and 6 we exploit that A is an elliptic operator. We further assume that A is defined via a semilinear (i.e., linear in the last argument) form $a : (Q \times V) \times V \to \mathbb{R}$ according to

$$\langle A(q(t),u(t)),\varphi\rangle_{V^*\times V} = a(q(t),u(t),\varphi) \quad \text{for all } \varphi \in V. \tag{2.2}$$

We consider Initial Value Problems (IVPs), i.e., PDE (2.1) subject to $u(t_1) = u^0 \in H$. The question of existence, uniqueness, and continuous dependence of solutions on the problem data u^0 and q cannot be answered satisfactorily in a general setting. However, there are problem-dependent sufficient conditions (compare, e.g., Gajewski et al. [57] for the case $A(q(t),u(t)) = A_q(q(t)) + A_u(u(t))$). A thorough discussion of this question is beyond the focus of this thesis.

Example 1. For illustration we consider the linear heat equation with Robin boundary control and initial values

$$\partial_t u = \Delta u \quad \text{in } (0,1)\times\Omega, \tag{2.3a}$$

$$\partial_\nu u + \alpha u = \beta q \quad \text{on } (0,1)\times\partial\Omega, \tag{2.3b}$$

$$u\big|_{t=0} = u^0, \tag{2.3c}$$

where $\alpha, \beta \in L^\infty(\partial\Omega)$ and ∂_ν denotes the derivative in the direction of the outwards pointing normal ν on $\partial\Omega$. We choose $V = H^1(\Omega)$ and $H = L^2(\Omega)$. Multiplication with a test function $\varphi \in V$ and integration by parts transform equations (2.3a) and (2.3b) into

$$0 = \int_\Omega \partial_t u(t)\varphi - \int_\Omega (\Delta u(t))\varphi \tag{2.4a}$$

$$= \int_\Omega \partial_t u(t)\varphi + \int_\Omega \nabla u(t)^{\mathrm{T}}\nabla\varphi - \int_{\partial\Omega} \left(\nabla u(t)^{\mathrm{T}}\nu\right)\varphi \tag{2.4b}$$

$$= \int_\Omega \partial_t u(t)\varphi + \int_\Omega \nabla u(t)^{\mathrm{T}}\nabla\varphi + \int_{\partial\Omega} \alpha u(t)\varphi - \int_{\partial\Omega} \beta q(t)\varphi \tag{2.4c}$$

$$=: \int_\Omega \partial_t u(t)\varphi + a(q(t),u(t),\varphi), \tag{2.4d}$$

which serves as the definition for the semilinear form a and the corresponding operator A. We immediately observe that a is even bilinear on $(Q \times V) \times V$ in this example.

2.2 Coupled ODEs and PDEs

In some applications, e.g., in chemical engineering, the models consist of PDEs which are coupled with ODEs. We denote the ODE states, which are not distributed in space, by $v \in C^0([t_1,t_2];\mathbb{R}^{n_v})$. These states can for instance model the accumulation of mass of a chemical species at an outflow port of a chromatographic column (compare Chapter 15). We can formulate the coupled system of differential equations as

$$\partial_t u(t) = -A(q(t),u(t),v(t)), \qquad \dot{v}(t) = f^{\mathrm{ODE}}(q(t),u(t),v(t)), \tag{2.5a}$$

where $f^{\mathrm{ODE}} : Q \times H \times \mathbb{R}^{n_v}$, subject to initial or boundary value conditions in time. We restrict ourselves to an autonomous formulation because the non-autonomous case can always be formulated as system (2.5) by introduction of an extra ODE state $\dot{v}_i = 1$ with initial value $v_i(t_1) = t_1$.

The question of existence, uniqueness, and continuous dependence on the data for the solution of IVPs with the differential equations (2.5) is even more challenging than for PDE IVPs and must be investigated for restriced problem classes (e.g., when A is not dependent on the $v(t)$ argument). Again, a thorough discussion of this question exceeds the scope of this thesis.

2.3 The Optimal Control Problem

We now state the OCP which is the point of origin for all further investigations of this thesis:

$$\underset{\substack{q\in L^2((0,1);Q)\\ u\in W(0,1)\\ v\in C^0([0,1];\mathbb{R}^{n_v})}}{\text{minimize}} \quad \Phi(u(1),v(1)) \tag{2.6a}$$

$$\text{s.t.} \quad \partial_t u = -A(q(t),u(t),v(t)), \qquad t \in (0,1), \tag{2.6b}$$

$$\dot{v} = f^{\mathrm{ODE}}(q(t),u(t),v(t)), \qquad t \in (0,1), \tag{2.6c}$$

$$(u(0),v(0)) = r^{\mathrm{b}}(u(1),v(1)), \tag{2.6d}$$

$$r^{\mathrm{c}}(q(t),v(t)) \geq 0, \qquad t \in (0,1), \tag{2.6e}$$

$$r^{\mathrm{e}}(v(1)) \geq 0, \tag{2.6f}$$

with nonlinear functions

$$\Phi : H \times \mathbb{R}^{n_v} \to \mathbb{R}, \quad r^{\mathrm{b}} : H \times \mathbb{R}^{n_v} \to H \times \mathbb{R}^{n_v},$$
$$r^{\mathrm{c}} : Q \times \mathbb{R}^{n_v} \to \mathbb{R}^{n_{\mathrm{r}}^{\mathrm{c}}}, \quad r^{\mathrm{e}} : \mathbb{R}^{n_v} \to \mathbb{R}^{n_{\mathrm{r}}^{\mathrm{e}}}.$$

We now discuss each line of OCP (2.6) in detail.

The objective function Φ in line (2.6a) is different from what is typically treated in PDE constrained optimization. Often, even for nonlinear optimal control problems, the objective functions are assumed to consist of a quadratic term for the states, e.g., L^2 tracking type in space or in the space-time cylinder, plus a quadratic Tychonoff-type regularization term for the controls (see, e.g., Tröltzsch [152]) of the type

$$\frac{1}{2}\int_0^1 \left\| u(t) - u^{\mathrm{desired}}(t) \right\|_H^2 \mathrm{d}t + \frac{\gamma}{2}\int_0^1 \|q(t)\|_Q^2 \,\mathrm{d}t. \tag{2.7}$$

We remark that tracking type problems with objective (2.7) on the space-time cylinder can always be cast in the form of OCP (2.6) by introduction of an additional ODE state variable v_i subject to

$$\dot{v}_i(t) = \left\| u(t) - u^{\mathrm{desired}}(t) \right\|_H^2 + \gamma \|q(t)\|_Q^2, \quad v_i(0) = 0,$$

with the choice $\Phi(u(1), v(1)) = v_i(1)/2$. The applications we are interested in, however, can have economical objective functions which are not of tracking type.

Constraints (2.6b) and (2.6c) determine the dynamics of the considered system. We have already described them in detail in Sections 2.1 and 2.2 of this chapter.

Initial or boundary value constraints are given by equation (2.6d). Typical examples are pure initial value conditions via constant

$$r^{\mathrm{b}}(u(1), v(1)) := (u^0, v^0)$$

or periodicity conditions

$$r^{\mathrm{b}}(u(1), v(1)) := (u(1), v(1)).$$

Compared to initial value conditions the presence of boundary value conditions makes it more difficult to use reduced approaches which rely on a solution operator for the differential equations mapping a control q to a feasible state u. Instead of solving one IVP, the solution operator would have to solve one Boundary Value Problem (BVP) which is in general both theoretically and numerically more difficult. Thus we avoid this *sequential* approach in favor of a *simultaneous* approach in which the intermediate control and state iterates of the method may be infeasible for equations (2.6b) through (2.6d). Of course feasibility must be attained in the optimal solution.

Inequality (2.6e) is supposed to hold for almost all $t \in (0,1)$ and can be used to formulate constraints on the controls and ODE states. We deliberately do not include PDE state constraints in the formulation which give rise to various theoretical difficulties and are currently a very active field of research. We allow for additional inequalities on the ODE states at the end via inequality (2.6f). In the context of chemical engineering applications, the constraints (2.6e) and (2.6f) can comprise flow rate, purity, throughput constraints, etc.

Problems with free time-independent parameters can be formulated within problem class (2.6) via introduction of additional ODE states v_i with vanishing time derivative $\dot{v}_i(t) = 0$. Although the software package MUSCOP (see Chapter 11) treats time-independent parameters explicitly, we refrain from elaborating on these issues in this thesis in order to avoid notational clutter.

OCP (2.6) also includes the cases of free start and end time via a time transformation, e.g., $\tau(t) = (1-t)\tau_1 + t\tau_2 \in [\tau_1, \tau_2], t \in [0,1]$. This case plays an important role in this thesis, e.g., in periodic applications with free period duration, see Chapter 15.

Concerning regularity of the functions involved in OCP (2.6), we take a pragmatic view point: We assume that the problem can be consistently discretized (along the lines of Chapter 3) and that the resulting finite dimensional optimization problem is sufficiently smooth on each discretization level to allow for employment of fast numerical methods (see Chapter 5).

http://www.springer.com/978-3-658-04475-6

A Direct Method for Parabolic PDE Constrained Optimization Problems
Potschka, A.
2014, XIV, 216 p. 30 illus., Softcover
ISBN: 978-3-658-04475-6