Technische Universität München
Fakultät für Mathematik
Lehrstuhl für Numerische Mathematik / Steuerungstheorie

Semilinear parabolic systems with hysteresis: Hadamard differentiability of the solution operator and optimal control

Christian Münch

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Prof. Dr. Daniel Matthes
Prüfer der Dissertation:	1.	Prof. Dr. Martin Brokate
	2.	Prof. Dr. Klemens Fellner
	3.	Prof. Dr. Christian Meyer

Die Dissertation wurde am 05.12.2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 26.03.2018 angenommen.

Abstract

This thesis considers a class of semilinear parabolic evolution systems subject to a hysteresis operator and a Bochner-Lebesgue integrable source term. The underlying spatial domain is allowed to have a very general boundary.
In the first part of the work, we prove resolvent estimates for elliptic operators.
In the second part, we apply semigroup theory to prove well-posedness and boundedness of the solution operator. Rate independence in the non-linearity complicates the analysis, since locality in time is lost. We investigate in Lipschitz continuity and Hadamard differentiability of the solution operator in the initial value and the source term. Using fixed point arguments, a representation of the derivative as an evolution system is derived.
In the third part, the results are applied in the optimal control of hysteresis-reaction-diffusion systems. We study a control problem with distributed control functions, or controls which act on a part of the boundary of the domain. The state equation is given by a reaction-diffusion system with the additional challenge that the reaction term includes a scalar stop operator.
First of all, we prove first order necessary optimality conditions for either type of control functions. Under certain regularity assumptions we derive results about the continuity properties of the adjoint system. For the case of distributed controls, we improve the optimality conditions, show uniqueness of the adjoint variables and prove criteria for possible discontinuity points and upper bounds for jumps of the adjoint variable which corresponds to the hysteresis. For the general problem, we employ the optimality system to prove higher regularity of optimal solutions. Finally, we derive regularity properties of the optimal value function and the optimal set function of a perturbed control problem when the set of controls is restricted.

Zusammenfassung

Thema der Arbeit sind Systeme semilinearer parabolischer Differentialgleichungen, deren Nicht-Linearität einen Hystereseoperator enthält. Das zugrundeliegende Definitionsgebiet ist möglicherweise nicht glatt. Im ersten Teil der Arbeit werden notwendige Resolventenabschätzungen für elliptische Operatoren bewiesen. Die Resultate werden anschließend genutzt um mit Hilfe von Halbgruppentheorie die Wohlgestelltheit sowie die Beschränktheit des Lösungsoperators zu zeigen. Außerdem werden Lipschitz Stetigkeit und Hadamard Differenzierbarkeit des Lösungsoperators sowohl im Anfangswert als auch im Quellterm gezeigt. Insbesondere wird die Ableitung als Lösung einer Evolutionsgleichung dargestellt. Ein Großteil der Beweise beruht auf Fixpunktargumenten. Der Hystereseoperator bereitet Schwierigkeiten in vielen Abschätzungen, da dieser nicht lokal in der Zeit arbeitet. Im dritten Teil der Arbeit werden die Resultate in einem Optimalsteuerungsproblem umgesetzt. Zustandsgleichung ist ein Reaktions-Diffusions-System, dessen Nicht-Linearität einen skalaren Stopp-Operator enthält. Wir betrachten verteilte Steuerungen oder Kontrollfunktionen auf einem Teil des Randes. Zunächst werden notwendige Optimalitätsbedingungen erster Ordnung für Steuerungen beider Art hergeleitet. Unter einer geeigneten Regularitätsannahme werden Aussagen zur Stetigkeit des adjungierten Systems erarbeitet. Für verteilte Steuerungen können die Optimalitätsbedingungen in stärkerer Form bewiesen werden. Außerdem werden für diesen Fall Eindeutigkeit des adjungierten Systems, sowie Kriterien für mögliche Unstetigkeitspunkte und obere Schranken für Sprünge hergeleitet. Für das allgemeine Problem zeigen wir mit Hilfe des Optimalitätssystems höhere Regularität der optimalen Lösungen. Abschließend erarbeiten wir Regularitätseigenschaften der Optimalwertfunktion und der zugehörigen Mengenfunktion für ein gestörtes Kontrollproblem.

Contents

1 Introduction **1**

2 Establishment of necessary tools **6**

2.1 Sobolev spaces including homogeneous Dirichlet boundary conditions 6
2.2 General elliptic operators 8
2.2.1 Definition of $\mathcal{T}_p$ and T_p 9
2.2.2 Resolvent estimates for $\mathcal{T}_p$ 11
2.2.3 Diffusion operators 19
2.2.4 Generators of analytic semigroups of operators 21
2.2.5 Fractional powers and fractional power spaces 24
2.2.6 Maximal parabolic Sobolev regularity 27
2.3 Spaces of Banach space valued functions and embeddings 28
2.4 Hysteresis operators 28

3 Semilinear parabolic systems with hysteresis and Bochner-Lebesgue integrable non-linearity **31**

3.1 Main assumption and notation 31
3.2 Well-posedness of the evolution equation 32
3.3 Hadamard differentiability of the solution operator for the evolution equation . . 41

4 Optimal control of hysteresis-reaction-diffusion systems **50**

4.1 Main assumption and further notation 53
4.2 Existence of an optimal control 54
4.3 Regularized control problem 56
4.3.1 Regularization of (3.11) and uniform-in-ε estimates 58
4.3.2 Dynamics of the regularized states 62
4.3.3 The regularized optimal control problem 64
4.3.4 Gâteaux differentiability of the solution operator of the regularized state equation 65
4.3.5 Adjoint system for the regularized problem 66
4.3.6 Estimates for the adjoints of the regularized problem 73
4.4 Adjoint system and optimality conditions for the optimal control problem 75
4.4.1 Adjoint system for distributed or boundary controls 76
4.4.2 Optimality conditions for distributed or boundary controls 83
4.4.3 Summary: Adjoint system and optimality conditions for distributed- or boundary controls 85
4.4.4 Improved optimality conditions and uniqueness for distributed controls . 87
4.4.4.1 Improved optimality conditions 87
4.4.4.2 Uniqueness of the adjoint variables 88
4.4.5 Properties of the measures $d\mu$ and dq for distributed controls 90
4.4.5.1 Isolated times and waiting slots 91
4.4.5.2 Continuity properties of q in I_∂ with the regularity assumption . 100
4.5 Higher regularity of the solutions of the optimal control problem 102
4.6 The value function of a perturbed control problem 105

1 Introduction

A lot of research has been made in the analysis of (systems of) partial differential equations (PDEs) and in particular in the field of non-linear and non-smooth PDEs. Moreover, many optimal control problems are subject to PDEs which depend of a control function. Especially in this context, the question of differentiability of the corresponding solution operator, the control-to-state mapping, becomes interesting. The latter is essential in order to derive (necessary) optimality conditions for the control problem at hand.
The concern of this work is the following. Firstly, we study systems of semi-linear PDEs which, represented as abstract operator equations, take the form

$$\begin{aligned} \frac{d}{dt}y(t) + (T_p y)(t) = (F[y])(t) + u(t) \quad \text{in } X \text{ for } t > 0,\\ y(0) = y_0 \in X. \end{aligned} \tag{1.1}$$

In particular, (1.1) is the weak formulation of a system of PDEs. Accordingly, the solution y takes values in a Banach space $X = \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ which is decomposed as a product of dual Sobolev spaces. All boundary conditions are included in the solution space in the way that each test function satisfies homogeneous Dirichlet boundary conditions on $\Gamma_{D_j} \subset \partial\Omega$, $j \in \{1, \dots m\}$. Additionally, the domain Ω satisfies minimal smoothness assumptions. The semi-linear elliptic operator T_p is unbounded on X and y_0 is a prescribed initial value. The main difficulty comprises of the non-linearity F, which is a Nemytski operator of the form $(F[y])(t) = f(y(t), \mathcal{W}[Sy](t))$. The function f is assumed to be locally Lipschitz continuous and directionally differentiable. Moreover, $\mathcal{W}$ is a scalar stop operator or another hysteresis operator with appropriate properties. Accordingly, the vector y has to be mapped to a scalar valued function by some linear operator $S \in X^*$ in order to serve as an input for $\mathcal{W}$. Solutions $z = \mathcal{W}[v]$ can not be written in a closed form. There are several ways to represent z. The most useful formulation for this work turned out to be the following variational inequality [cf. BK13]:

$$(\dot{z}(t) - \dot{v}(t))(z(t) - \xi) \leq 0 \qquad \text{for } \xi \in [a, b] \text{ and } t \in (0, T), \tag{1.2}$$

$$z(t) \in [a, b] \text{ for } t \in [0, T], \tag{1.3}$$

$$z(0) = z_0. \tag{1.4}$$

Lastly, the forcing term $u \in \mathrm{L}^q(J_T; X)$ is a Bochner integrable function.
The first aim of this work lies in the analysis of the solution operator G which maps each pair (y_0, u) to the solution y of (1.1). Before this question can be addressed, some semigroup theory has to be developed. In particular, we prove resolvent estimates for the operator T_p, which extends an existing result from [Hal+15] from scalar valued functions to vector fields. Subsequently, those tools are applied to prove well-posedness of (1.1) and Hadamard directional differentiability of G in y_0 and u. Those results extend the findings in [Mün17a] from diffusion operators to general elliptic operators and from zero initial value to arbitrary y_0.
In the second part of the work, we apply the results from the first part to an optimal control problem where (1.1) serves as the state equation and in which the forcing term takes the role of a control function. Specifically, we focus on the subclass of diffusion operators A_p and initial value $y_0 = 0$. Moreover, the cost function at hand is of tracking type.
For $i \in \{1, 2\}$, the control problem to study takes the form

$$\min_{u \in U_i} J(y, u) := \frac{1}{2}\|y - y_d\|^2_{U_1} + \frac{\kappa}{2}\|u\|^2_{U_i} \tag{1.5}$$

subject to

$$\begin{aligned}\dot{y}(t) + (A_p y)(t) &= f(y(t), z(t)) + (B_i u)(t) \text{ in } \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega) \text{ for } t \in (0,T), && (1.6)\\ y(0) &= 0 \qquad\qquad\qquad\qquad\qquad \text{in } \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega),\\ z &= \mathcal{W}[Sy]. && (1.7)\end{aligned}$$

Hence, the state equations (1.6)–(1.7) take the form of (1.1) with $T_p = A_p$, $y_0 = 0$ and $B_i u$ instead of u.

We consider two different types of control functions. The first one corresponds to $i = 1$ and includes controls which act distributed over the domain Ω. The corresponding control space $U_1 := \mathrm{L}^2\left((0,T); [\mathrm{L}^2(\Omega)]^m\right)$ is embedded into $\mathrm{L}^2\left((0,T); \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)$ by a suitable operator B_1. The second control space $U_2 := \mathrm{L}^2\left((0,T); \prod_{j=1}^m \mathrm{L}^2(\Gamma_{N_j}, \mathcal{H}_{d-1})\right)$ consists of functions which act exclusively on the Neumann boundary parts $\Gamma_{N_j} \subset \partial\Omega$, $j \in \{1, \dots m\}$. Again, functions from U_2 are embedded into $\mathrm{L}^2\left((0,T); \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)$ by a suitable operator B_2.

We derive necessary optimality conditions for (1.5)–(1.7). In particular, we show existence of an adjoint system (p, q) and prove a maximum condition. For $i = 1$ we improve the latter condition and show uniqueness of (p, q). Moreover, we show higher regularity of the optimal solutions and study the optimal value function and the optimal set function of a perturbed control problem

$$\min_{u \in C \subset U_i} J(G(B_i(u - r, 0), u - r), \tag{1.8}$$

where $r \in U_i$ corresponds to the perturbation. Those results reflect the findings of [Mün17b], but the theory and all proofs are carried out more detailed and in several parts more accurate. The evolution of q and the optimality condition depend on an abstract measure $d\mu$ which could not be characterized in [Mün17b]. This work extends the results from [Mün17b]. In particular, for $i = 1$ we prove sign conditions and bounds for $d\mu$ and deduce sign conditions and bounds for dq. We exploit this to prove conditions for discontinuity points of q as well as upper bounds for possible jumps.

In the following, we compare this work to the literature.

A lot of the semigroup theory which we apply in this work can be found in [Paz83], [Lun95] or [Hen81]. Also results on *general non-linear abstract operator evolution equations without a forcing term* u can be found here. Typical non-linearities take the form $f(t, y(t))$, where f is (locally) Lipschitz continuous. *Semilinear parabolic problems* similar to (1.1) above have been studied in [Lun95] for example. Also here, the non-linearities $f(t, y(t))$ are Lipschitz continuous. Recent results about differentiability of the solution mapping of non-linear operator equations are due to [MS15]. So far, the non-linearity f is always defined *locally in time*, and *no hysteresis* is considered.

Good progress has also been achieved with respect to optimal control of (systems of) PDEs. Specifically, many results on *semi-linear parabolic* optimal control problems can be found in [Trö10]. Early research in this direction is due to [BC85; RZ98; Cas97]. We also call the reader's attention to [HKR13], where a parabolic control problem with rough boundary conditions is studied.

The particular subclass of optimal control problems with *reaction-diffusion systems* as state equations has already been studied in [Gri03]. Particularly, parameter sensitivity analysis is the focus of this work. Subsequently, the results have been broadened in [GV06] and several further papers. A similar problem was also analyzed in [BJT10], where optimality conditions could be shown.

So far, all optimal control problems are subject to non-linear PDEs with *smooth non-linearities*. Hence, continuous differentiability of the control-to-state operator - often up to second order - could be shown. By linearity of the derivatives, first and often second order optimality conditions could be derived.

Only few results have been established for the optimal control of *infinite-dimensional rate-independent processes*. Early studies in this direction are due to [Rin08; Rin09]. In particular, existence of optimal controls for problems subject to energetically driven processes has been studied. An application of these results in the field of shape memory materials and in particular to thermal control problems is due to [ELS13; EL14]. No optimality conditions can be found in these works. In the infinite-dimensional setting, an optimal control problem in the field of static plasticity has been studied in [HMW12; HMW13]. Subsequently, the theory was applied in [HMW14]. Specifically, a quasi-static control problem was solved numerically with a time-discretization approach. A class of time-continuous, infinite-dimensional, rate-independent control problems of quasi-static plasticity type was analyzed in [Wac12; Wac15; Wac16]. Again with a time-discretization argument, optimality conditions could be derived. Finally, in [SWW16], a time-continuous, infinite-dimensional optimal control problem of a rate-independent system is studied. The state equation is considered in its energetic formulation, and necessary optimality conditions are shown by viscous regularization.

To the best of our knowledge, the research of optimal control of *hysteresis* started with [Bro87; Bro88; Bro91]. The optimal control problem here is subject to an ODE-system with hysteresis and necessary optimality conditions could be shown. Moreover, a time discretization argument has been applied to derive an adjoint system. Closely connected, research on optimal control of sweeping processes has been done in [CMF14; Col+12; Col+16]. Comparable to this work, but for a control problem of an ODE-system with a (vectorial) stop hysteresis operator, first order optimality conditions were derived in [BK13]. Similar as (1.2)–(1.4), a variational inequality is chosen to represent the hysteresis operator. The main difficulty which occurs with the stop operator and with every other interesting hysteresis operator is its non-locality in time. As a consequence, the state $y(t)$ at each time $t \in (0,T]$ depends in general in a non-trivial way on the whole history $(0,t)$. Another problem is due to the fact that the stop operator is not differentiable in the classical sense. Hence, classical differentiability of the control-to-state operator is lost. In [BK13], regularization techniques were applied to overcome this problem, and finally an optimality system could be derived. We take advantage of several of the ideas in this work. However, since the state equation (1.6)–(1.7) is of reaction-diffusion type, we need additional arguments. In particular, the state vector $y : [0,T] \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ which solves (1.6)–(1.7) takes its values in an infinite-dimensional space. Moreover, our assumptions on the non-linearity are much weaker since we only suppose f to be locally Lipschitz continuous and directionally differentiable rather than continuously differentiable. As a consequence, we require techniques as in [MS15]. Indeed, since the domain Ω has a rough boundary, the (1.6)–(1.7) can only be written as a weak formulation and the domain of the unbounded diffusion operator A_p is contained in $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$, i.e. in a product of dual spaces.

Only few substantial results are available on the control of *hysteresis-reaction-diffusion systems*, and even less in the direction of optimal control. Particularly, some research on automatic control problems governed by reaction-diffusion systems with feedback control of relay switch and Preisach type has been done in [CC02]. In fact, global existence and uniqueness of solutions could be shown. Closed-loop control of a reaction-diffusion system coupled with ordinary differential inclusions is the subject of [DN11]. Assuming the number control devices to be finite, a feedback law was derived.

Finally, the optimal control of general *non-smooth semi-linear parabolic equations* was analyzed in [MS15]. Specifically, the non-linearity is Lipschitz continuous on bounded sets and direc-

tionally differentiable. Hence, the control-to-state operator is not differentiable in the classical sense. Nevertheless, necessary optimality conditions could be derived. The derivation of an adjoint system relies on regularization techniques. Although no hysteresis is considered in [MS15], with additional arguments from [BK13], the approach can be adapted to work for problem (1.5)–(1.7), where the hysteresis entails non-locality in time. For further research on optimal control of non-smooth parabolic equations we refer to the references in [MS15].
As described in the beginning, this work considers *non-smooth semi-linear parabolic systems with hysteresis* and the *optimal control of non-smooth reaction-diffusion systems with hysteresis*. In particular, a scalar stop operator in the non-linearity $F[y] = f(y, \mathcal{W}[Sy])$ implies *non-locality* in time. Moreover, the function f is assumed to be *locally Lipschitz continuous* and *directionally differentiable*. Finally, the domain Ω satisfies *minimal smoothness* assumptions.
The work is structured as follows:
In Chapter 2, we establish necessary tools to analyze the operator equation (1.1). We expand the framework in [Mün17a; Mün17b] to the generality of this work.
Results from the literature on Sobolev Spaces with Dirichlet boundary conditions are collected in Section 2.1. In particular, we define the space $\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$ and its dual space $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ for $p \in [2, \infty)$ and $\frac{1}{p'} + \frac{1}{p} = 1$, where the domain Ω is non-smooth.
In Section 2.2, we define a class of elliptic operators $\mathcal{T}_p$ and T_p. We prove resolvent estimates for those operators for appropriate p which are necessary for the construction of analytic semigroups $\exp(-T_p t)$, $t \geq 0$, see Theorem 2.14. This extends [Hal+15, Theorem 5.12] from scalar valued function spaces to spaces of vector valued functions. In Subsection 2.2.3, we deduce the corresponding results for the subclass of diffusion operators $\mathcal{A}_p$ and A_p as they were used in [Mün17a; Mün17b]. Subsection 2.2.4 contains the necessary background on sectorial operators and semigroups. We apply Theorem 2.14 to prove that T_p generates an analytic semigroup, see Theorem 2.22. In Theorem 2.25, we prove that the resolvent set of T_p is contained in a sector $S_{\delta,\tilde{\Phi}} = \left\{\lambda : \tilde{\Phi} \leq |\arg(\lambda - \delta)| \leq \pi,\, \lambda \neq \delta\right\}$, where $\delta > 0$ and $\tilde{\varphi} \in (0, \frac{\pi}{2})$. This yields crucial estimates for the semigroup $\exp(-T_p t)$, $t \geq 0$. In Subsection 2.2.5, we collect the necessary background on fractional powers of operators T_p^θ and the corresponding spaces $X^\theta_{T_p}$, $\theta \geq 0$. In particular, we show important embedding properties of $X^\theta_{T_p}$ for $\theta \in [0, 1]$ and that $X^1_{T_p}$ is topologically equivalent to $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$. We also deduce topological equivalence of $X^\theta_{T_p}$ to a complex interpolation space $[\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)]_\theta$ for $\theta \in (0, 1)$ if T_p satisfies additional properties. The latter hold for A_p. In Subsection 2.2.6, we introduce the concept of maximal parabolic Sobolev regularity of an operator which is required to prove higher regularity of the solution y of (1.1).
Section 2.3 contains some embedding results on Bochner spaces which we need for several weak compactness arguments.
Finally, in Section 2.4, we introduce the concept of hysteresis operators and define the scalar stop and the scalar play operator.
In Chapter 3, we study the solution operator G of (1.1).
The main assumption of the chapter as well as some notation can be found in Section 3.1.
In Section 3.2, we show that (1.1) is well defined for $u \in \mathrm{L}^q((0, T); X)$ and $y_0 \in X^\beta_{T_p}$ for appropriate $q \in (1, \infty)$ and $\beta \in (0, 1]$, see Theorem 3.2. The corresponding result for diffusion operators is stated in Corollary 3.4.
Afterwards, in Section 3.3, we define Hadamard directional differentiability and show that the solution operator G of (1.1) is Hadamard directionally differentiable in y_0 and u, see Theorem 3.11. The corresponding result for diffusion operators can be found in Corollary 3.15.
In Chapter 4, we study the optimal control problem (1.5)–(1.7). In particular, we restrict ourselves to diffusion operators A_p and zero initial value $y_0 = 0$ in (1.6).

Section 4.1 contains the main assumption of the chapter and some further notation in addition to Section 3.1.

In Section 4.2, we exploit Corollary 3.4, Corollary 3.15 and the embedding results from Chapter 2 and prove that an optimal control for problem (1.5)–(1.7) exists, see Theorem 4.6.

Towards our main objective, which is an adjoint system for problem (1.5)–(1.7), we regularize the control problem in Section 4.3. In particular, we introduce a regularization parameter $\varepsilon > 0$ and replace $F[y] = f(y, \mathcal{W}[Sy])$ in (1.1) by $F_\varepsilon[y] = f_\varepsilon(y, Z_\varepsilon(Sy))$, were f_ε and Z_ε are regular enough and approximate f and $\mathcal{W}$ in the limit $\varepsilon \to 0$. Accordingly, this yields a regularization of the state equation (1.6)–(1.7). In Subsection 4.3.1, we extend the results from [Mün17b] by proving well posedness and regularity results for Z_ε. Corollaries 3.4, 3.15 imply that the solution operators G_ε of the regularized state equations are well defined and Gâteaux-differentiable on $\mathrm{L}^q((0,T);X)$. For a fixed optimal solution $\overline{u}$ of (1.5)–(1.7), we add an additional term of the form $\frac{1}{2}\|u - \overline{u}\|_X$ to the original cost function J and define a sequence of regularized optimal control problems with state equations $y_\varepsilon = G_\varepsilon(B_i u)$, $z_\varepsilon = Z_\varepsilon(Sy_\varepsilon)$. As for problem (1.5)–(1.7), we obtain optimal solutions $\overline{u}_\varepsilon$, $\overline{y}_\varepsilon = G_\varepsilon(B_i\overline{u}_\varepsilon)$ and $\overline{z}_\varepsilon = Z_\varepsilon(S\overline{y}_\varepsilon)$. By uniform-in-$\varepsilon$ bounds and weak compactness arguments, we prove that those functions converge to the optimal solution $(\overline{u}, \overline{y}, \overline{z})$ of the original problem in the limit $\varepsilon \to 0$, see Theorem 4.16. Even though the control-to-state operators G_ε are Gâteaux-differentiable, it remains challenging to derive adjoint systems $(p_\varepsilon, q_\varepsilon)$ for the regularized problems, since G_ε is defined implicitly via Z_ε, similar as in (1.6). Nevertheless, we establish an optimality system for those problems in Theorem 4.20.

In Section 4.4, we accomplish the original aim and derive an optimality system for problem (1.5)–(1.7). In particular, we take the limit $\varepsilon \to 0$ of $(p_\varepsilon, q_\varepsilon)$ in Subsection 4.4.1 and prove necessary optimality conditions. The evolution equation of the adjoint variable p which corresponds to $\overline{y}$ follows without further effort. But the adjoint variable q which corresponds to $\overline{z}$ is of lower regularity as it occurs frequently in optimal control problems which are subject to implicit state constraints of variational inequality type. Indeed, q is only contained in $\mathrm{BV}(0,T)$, the space of functions with bounded total variation in $[0,T]$. Hence, there exists no time derivative and the evolution of q is only represented by a measure $dq \in \mathrm{C}([0,T])^*$. In order to understand the optimality system for problem (1.5)–(1.7) completely, we study q and dq in more detail. It turns out that dq depends on an abstract measure $d\mu \in \mathrm{C}([0,T])^*$. Moreover, the measure is part of the maximum condition for problem (1.5)–(1.7) which we prove in Subsection 4.4.2. This makes it even more appealing to characterize $d\mu$. For the general problem with $i \in \{1,2\}$ and without any further assumptions, we prove that $d\mu$ has its support in the subset of times $J_\partial \subset [0,T]$ where $\overline{z}$ is located at the boundary points $\{a,b\}$ of $[a,b]$. With an additional regularity assumption on $S\overline{y}$ we can further shrink the support of $d\mu$. Example 4.32 provides an example in which this assumption applies. The latter is not contained [Mün17b]. The first main results of Section 4.4 are summarized in Theorem 4.38 and Corollary 4.39. Those contain all results about the optimality system and the maximum condition for problem (1.5)–(1.7) for $i \in \{1,2\}$. In Sections 4.4.4–4.4.5 we continue to study the control problem with distributed control functions, i.e. with $i = 1$, in more detail. It turns out that the optimality conditions for this particular case can be improved, since B_1 has dense range and hence B_1^* is one-to-one for $p \geq 2$ close to two. Corollary 4.40 contains the improved maximum condition. Moreover, in Corollary 4.41 we again exploit injectivity of B_1^* and show uniqueness of p, q and $d\mu$ for $i = 1$. These together are the second main result of Section 4.4. In Subsection 4.4.5 we extend the findings of [Mün17b] by analyzing the measure $d\mu$ for the case $i = 1$ and for continuously differentiable f in more detail. In particular, we partition the interval $[0,T]$ into different categories of times. In the subset of times in which the support of $d\mu$ is located, we prove sign conditions and bounds for $d\mu$, see Lemma 4.46 and Theorem 4.47. With help of the optimality system in Corollary 4.40, we transfer those results to the measure dq, see Corollary 4.48. In the case when the regularity assumption of Subsection 4.4.1 holds, we prove conditions for discontinuity points of q and upper bounds for possible jumps, see Corollary 4.49.

In Section 4.5, we return to the general case $i \in \{1, 2\}$. We introduce an additional assumption on B_i and exploit regularity of (p, q) in time and the relation between (p, q) and $\overline{u}$ in order to establish higher regularity of the optimal control $\overline{u}$ and the optimal state $\overline{y}$, see Theorem 4.51. Example 4.52 provides an example in which Theorem 4.51 can be applied.

Finally, in Section 4.6, we study the perturbed problem (1.8). In Theorem 4.54, we prove that the optimal value function $v : r \in U_i \to \mathbb{R}$ is lower semi-continuous for C convex and closed. If C is also compact, we prove that v is continuous and that the corresponding optimal set function $V : r \in U_i \rightrightarrows C$ is upper semi-continuous.

Note that all results of Chapter 4 can be applied to control problems with general spaces of control functions of the form $U = \mathrm{L}^2\big((0,T); \tilde{U}\big)$. We only require the existence of a continuous operator $B : \tilde{U} \to X = \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. If B satisfies the properties of B_1, then also the improvements of Sections 4.4.4–4.4.5 hold. Moreover, the cost function $J(y, u)$ can be replaced by a general differentiable functional $J(y, u, z)$, as long as the corresponding reduced cost function remains coercive in $u \in U$. Of course, this results in a different optimality system. In particular, the evolution equations of the corresponding adjoint variables include the partial derivatives $J_y(\overline{y}, \overline{u}, \overline{z})$ and $J_z(\overline{y}, \overline{u}, \overline{z})$. Finally, the diffusion operator A_p can be replaced by a general semi-linear parabolic operator which satisfies maximal parabolic regularity on $X = \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$.

Notation:

Depending on the chapter, the Banach spaces in this work consist of complex or real valued functions. If Y is such a Banach space, then we denote by Y^* the corresponding (anti-)dual space. In the complex case, the duality pairing between $u \in Y^*$ and $v \in Y$ will be denoted by $\langle u, v\rangle_{Y^*,Y}$ or $\langle u, v\rangle_Y$, and the anti-duality pairing by $\langle\langle u, v\rangle\rangle_{Y^*,Y}$ or $\langle\langle u, v\rangle\rangle_Y$. Accordingly, there holds

$$\langle u, \overline{v}\rangle_{Y^*,Y} = \langle u, \overline{v}\rangle_Y = \langle\langle u, v\rangle\rangle_Y = \langle\langle u, v\rangle\rangle_{Y^*,Y}.$$

Moreover, we write $\mathcal{L}(Y, Z)$ for the space of linear operators between spaces Y and Z and $\mathcal{L}(Y)$ for the space of linear operators on Y.

2 Establishment of necessary tools

2.1 Sobolev spaces including homogeneous Dirichlet boundary conditions

The setting and the theory of this section has appeared in a similar form in [Mün17a] and [Mün17b]. Remember however, that we will not exclusively work with spaces of real valued functions, but consider spaces of complex valued functions in parts of this work. The theory in this section is strongly connected to [Hal+15]. We recall several definitions, results and assumptions. In the following, $\Omega \subset \mathbb{R}^d$ is assumed to be a bounded domain with $d \geq 2$. The boundary regularity is defined in Assumption 2.2. For some given $m \in \mathbb{N}\backslash\{0\}$, we want to define a vector space as the product of m distinct Sobolev spaces of functions, which are complex or real valued depending on the context.

For each component $j \in \{1, \ldots, m\}$ of this space of vector valued functions, see Definition 2.4, the boundary $\partial\Omega$ is the union of the corresponding Dirichlet part $\Gamma_{D_j} \subset \partial\Omega$ and the Neumann boundary $\Gamma_{N_j} := \partial\Omega\backslash\Gamma_{D_j}$, see Assumption 2.2. The cases $\Gamma_{D_j} = \emptyset$ and $\Gamma_{D_j} = \partial\Omega$ are not excluded [Hal+15, Comment after Definition 2.4] and [Aus+14, Remark 2.2 (iii)]. The definition of Sobolev spaces of functions which are zero on a part of the boundary allows us to incorporate homogeneous boundary conditions of a PDE already in the definition of the space of solutions. In many problems, $\partial\Omega$, Γ_{D_j} and Γ_{D_j} are assumed to be Lipschitz continuous manifolds. We want to admit a much broader class of possible boundary decompositions in our setting.

This leads us to the definition of an I-set, where $I \in (0, d]$ is a fixed real number.

Definition 2.1. [Hal+15, Definition 2.1] For $0 < I \le d$ and a closed set $M \subset \mathbb{R}^d$ let ρ denote the restriction of the I-dimensional Hausdorff measure $\mathcal{H}_I$ to M. Then we call M an I-set if there are constants $c_1, c_2 > 0$ such that

$$c_1 r^I \le \rho\left(B_{\mathbb{R}^d}(x, r) \cap M\right) \le c_2 r^I$$

for all x in M and $r \in (0, 1)$.

We assume throughout that the domain of existence satisfies the following:

Assumption 2.2. [Hal+15, Assumption 2.3] The domain $\Omega \subset \mathbb{R}^d$ is bounded and its closure $\overline{\Omega}$ is a d-set. For $j \in \{1, \dots, m\}$, the Neumann boundary part $\Gamma_{N_j} \subset \partial\Omega$ is relatively open and the Dirichlet boundary part $\Gamma_{D_j} = \partial\Omega \backslash \Gamma_{N_j}$ is a $(d-1)$-set.

Remark 2.3. As already mentioned in the beginning of this section, note that the cases $\Gamma_{D_j} = \emptyset$ and $\Gamma_{D_j} = \partial\Omega$ are not excluded [Hal+15, Comment after Definition 2.4] and [Aus+14, Remark 2.2 (iii)]. Assumption 2.2 allows for very general domains. For example, Ω may be a Lipschitz domain and for $j \in \{1, \dots, m\}$, Γ_{D_j} can be a $(d-1)$-dimensional manifold. But much more general cases are possible: "In particular, the Dirichlet boundary part need not be (part of) a continuous boundary in the sense of [Gri, Definition 1.2.1.1] and the domain is not required to 'lie on one side of the Dirichlet boundary part'." [DER15, 1. Introduction]

As in [Hal+15, Definition 2.4] or [Mün17a, Definition 2.4] we define Sobolev spaces which include Dirichlet boundary conditions on a part of the domain.

Definition 2.4. Let $U \subset \mathbb{R}^d$ be a domain and $p \in [1, \infty)$. All functions are either real or complex valued.

- $\mathrm{W}^{1,p}(U)$ denotes the usual Sobolev space of functions $\psi \in \mathrm{L}^p(U)$, whose weak partial derivatives exist in $\mathrm{L}^p(U)$, normed by

$$\|\psi\|_{\mathrm{W}^{1,p}(U)} = \left(\int\limits_U \left(|\psi|^2 + \sum_{j=1}^d \left| \frac{\partial \psi}{\partial x_j} \right|^2 \right)^{\frac{p}{2}} dx \right)^{\frac{1}{p}}.$$

- For a closed subset M of $\overline{U}$ we define

$$\mathrm{C}_M^\infty(U) := \{\psi|_U : \ \psi \in \mathrm{C}_0^\infty(\mathbb{R}^d), \ \mathrm{supp}(\psi) \cap \mathrm{M} = \emptyset\}$$

  and denote by $\mathrm{W}_\mathrm{M}^{1,p}(U)$ the closure of $\mathrm{C}_M^\infty(U)$ in $\mathrm{W}^{1,p}(U)$.

Remark 2.5. [Mün17a, Remark 2.5] In $\mathrm{W}_\mathrm{M}^{1,p}(U)$ in Definition 2.4, we use the same norm as in [Hal+15], which differs from the usual norm in Sobolev spaces. One reason for this choice is that it simplifies estimates concerning the duality between $\mathrm{W}_\mathrm{M}^{1,p}(U)$ and $\mathrm{W}_\mathrm{M}^{1,p'}(U)$. We may identify a function $\phi \in \mathrm{W}_\mathrm{M}^{1,p}(U)$ with an element in $\mathrm{W}_\mathrm{M}^{-1,p}(U)$ since for any $\psi \in \mathrm{W}_\mathrm{M}^{1,p'}(U)$ the Cauchy Schwarz inequality together with Hölder's inequality yields

$$\int\limits_U \left(\phi\overline{\psi} + \sum_{j=1}^d \frac{\partial \phi}{\partial x_j} \overline{\frac{\partial \psi}{\partial x_j}} \right) dx \le \int\limits_U \left(|\phi|^2 + \sum_{j=1}^d \left| \frac{\partial \phi}{\partial x_j} \right|^2 \right)^{\frac{1}{2}} \left(|\psi|^2 + \sum_{j=1}^d \left| \frac{\partial \psi}{\partial x_j} \right|^2 \right)^{\frac{1}{2}} dx$$

$$\le \left(\int\limits_U \left(|\phi|^2 + \sum_{j=1}^d \left| \frac{\partial \phi}{\partial x_j} \right|^2 \right)^{\frac{p}{2}} dx \right)^{\frac{1}{p}} \left(\int\limits_U \left(|\psi|^2 + \sum_{j=1}^d \left| \frac{\partial \psi}{\partial x_j} \right|^2 \right)^{\frac{p'}{2}} dx \right)^{\frac{1}{p'}}.$$

The results in [Hal+15, 3. Interpolation] about interpolation properties between spaces of the form $\mathrm{W}^{1,p}_{\Gamma_{D_j}}(\Omega)$ for different $p \in (1,\infty)$ are established under the assumption that a linear and continuous extension operator $\mathcal{E} : \mathrm{W}^{1,1}_{\Gamma_{D_j}}(\Omega) \to \mathrm{W}^{1,1}_{\Gamma_{D_j}}(\mathbb{R}^d)$ exists, which simultaneously defines a continuous extension operator $\mathcal{E} : \mathrm{W}^{1,p}_{\Gamma_{D_j}}(\Omega) \to \mathrm{W}^{1,p}_{\Gamma_{D_j}}(\mathbb{R}^d)$ for all $p \in (1,\infty)$ [cf. Hal+15, Assumption 3.1]. This operator is used to carry over existing interpolation results from the usual Sobolev spaces to $\mathrm{W}^{1,p}_{\Gamma_{D_j}}(\Omega)$-spaces. It is shown that such an extension operator can be constructed under the following assumption.

Assumption 2.6. [Hal+15, Assumption 4.11] In the setting of Assumption 2.2 we suppose for all $j \in \{1,\dots,m\}$ and any $x \in \overline{\Gamma_{N_j}}$ that there is an open neighborhood U_x of x and a bi-Lipschitz mapping ϕ_x from U_x onto a cube in $\mathbb{R}^d$ such that $\phi_x(\Omega \cap U_x)$ equals the lower half of the cube and such that $\partial\Omega \cap U_x$ is mapped onto the top surface of the lower half cube.

Remark 2.7. Aside from the construction of an extension operator, as described above, Assumption 2.6 is important due to the following two reasons:

1. We will need Assumption 2.6 in Section 2.2.2 to prove resolvent estimates for elliptic operators.
2. [Mün17a, Remark 2.7] Under Assumption 2.6 it can be shown that the embeddings
$$\mathrm{W}^{1,p}_{\Gamma_{D_j}}(\Omega) \hookrightarrow \mathrm{L}^q(\Omega)$$
are compact for $q \in [1, \frac{dp}{d-p})$ if $p \in (1,d)$, and for arbitrary $q \in [1,\infty)$ if $p \geq d$ [Hal+15, Remark 3.2]. The proof is almost equal to the proofs of [Eva10, Part II, 5.6.1, Theorem 2] and [Eva10, Part II, 5.7, Theorem 1].

The second remark will turn out to be very important when it comes to embedding properties of fractional power spaces and of Banach space valued functions, see Subsection 2.2.5 and Section 2.3 below.

The following definition of vector valued Sobolev spaces which include homogeneous Dirichlet conditions goes back to [Hal+15, Section 6], cf. also [Mün17a, Definition 2.8]:

Definition 2.8. With Assumption 2.2 and Assumption 2.6 and $p \in [1,\infty)$ we define a Sobolev space of vector valued functions by the product space
$$\mathbb{W}^{1,p}_{\Gamma_D}(\Omega) := \prod_{j=1}^{m} \mathrm{W}^{1,p}_{\Gamma_{D_j}}(\Omega).$$

For $p \in (1,\infty)$ we denote its (componentwise) dual by $\mathbb{W}^{-1,p'}_{\Gamma_D}(\Omega)$, or the anti-dual in the complex case respectively.

2.2 General elliptic operators

In this section, we define a class of elliptic operators $\mathcal{T}_p$ and T_p. We prove resolvent estimates for those operators. In particular, with Theorem 2.14 below we are able to extend an existing result for scalar valued function spaces to the vector valued case. As already mentioned in the introduction, this implies that the operators are sectorial and therefore generate analytic semigroups of operators, see Subsection 2.2.4. Since zero is contained in the resolvent sets $\rho(T_p)$ it can also be shown that the fractional powers T_p^θ are well defined. This leads us to the definition of fractional power spaces in Subsection 2.2.5. All spaces in this section are considered to be complex.

2.2.1 Definition of $\mathcal{T}_p$ and T_p

Before we introduce the elliptic operators $\mathcal{T}_p$ and T_p in Definition 2.11, we define two auxiliary operators which we need for the construction. The definition of $\mathcal{T}_p$ below goes back to [Hal+15, Section 6]. We extend the framework of this work in order to provide the reader a comprehensive overview about the spaces between which the individual operators act.

Definition 2.9. With Assumption 2.2 and Assumption 2.6, let $p \in [1,\infty)$. We introduce

$$\mathcal{J}_p : \mathbb{W}^{1,p}_{\Gamma_D}(\Omega) \to \mathrm{L}^p(\Omega;\mathbb{C}^m \times \mathbb{C}^{m\times d}), \quad \mathcal{J}_p(u) := (u, \nabla u) \qquad \text{and}$$
$$I_p : \mathbb{W}^{1,p}_{\Gamma_D}(\Omega) \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), \qquad \langle\langle I_p u, v\rangle\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} := \int_\Omega u \cdot \overline{v}\, dx \quad \forall v \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega).$$

In the following lemma, we study the operator I_p in more detail. Amongst others, we prove that I_p is well defined. This is necessary for the construction of T_p.

Lemma 2.10. In the framework of Definition 2.9, let $p \in (1,\infty)$. Then I_p is compact, one-to-one and has dense range, i.e. $\overline{\ker(I_p)} = \{0\}$ and $\overline{\operatorname{ran}(I_p)} = \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. Moreover, there holds

$$\|I_p\|_{\mathcal{L}\left(\mathbb{W}^{1,p}_{\Gamma_D}(\Omega),\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \leq 1.$$

Proof. Because $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ is a product of $\mathrm{W}^{1,p}_{\Gamma_{D_i}}(\Omega)$-spaces, the embedding $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega) \hookrightarrow [\mathrm{L}^p(\Omega)]^m$ is compact according to Remark 2.7. By Hölders inequality, one has

$$|\langle\langle I_p u, v\rangle\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)}| \leq \|u\|_{[\mathrm{L}^p(\Omega)]^m}\|v\|_{[\mathrm{L}^{p'}(\Omega)]^m} \leq \|u\|_{[\mathrm{L}^p(\Omega)]^m}\|v\|_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)},$$

for all $u \in \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ and $v \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$. So indeed, I_p maps every $u \in \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ to a continuous functional on $\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$. Moreover, it follows

$$\|I_p u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} \leq \|u\|_{[\mathrm{L}^p(\Omega)]^m} \leq \|u\|_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)}.$$

This already implies $\|I_p\|_{\mathcal{L}\left(\mathbb{W}^{1,p}_{\Gamma_D}(\Omega),\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \leq 1$ and that I_p is compact as the concatenation of a compact and a continuous operator. It remains to prove that I_p is one-to-one and has dense range. The embedding $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega) \hookrightarrow [\mathrm{L}^p(\Omega)]^m$ is one-to-one, since $\|u\|_{[\mathrm{L}^p(\Omega)]^m} = 0$ for $u \in \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ implies $\nabla u = 0 \in \mathrm{L}^p(\Omega;\mathbb{C}^{m\times d})$ by definition of the weak derivative, so that $u = 0$ in $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$. Note that $[\mathrm{C}^\infty_0(\Omega)]^m$ is dense in $[\mathrm{L}^p(\Omega)]^m$ [W05, Lemma V.1.10]. Moreover, $\langle\langle I_p u, v\rangle\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)}$ is equal to the anti-dual pairing between $u \in [\mathrm{L}^p(\Omega)]^m$ and $v \in [\mathrm{L}^{p'}(\Omega)]^m$. Hence, $\langle\langle I_p u, v\rangle\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} = 0$ for any $v \in [\mathrm{C}^\infty_0(\Omega)]^m$ implies that u must be zero in $[\mathrm{L}^p(\Omega)]^m$ by the Hahn-Banach theorem [W05, Korollar III.1.6]. In this case, $u = 0$ in $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ as well by the above argument. Consequently, I_p is one-to-one.
To see that I_p has dense range, note first that $\mathbb{W}^{1,p'}(\Omega)$ is reflexive for $p' \in (1,\infty)$ [AF03, 3.6 Theorem]. That is, also $\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$ is reflexive as a closed subspace of a reflexive space [W05, Satz III.3.4], so that we may identify $\left[\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right]^* = \left[\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)\right]^{**}$ with $\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$. This leads to the representation

$$I_p^* = I_{p'} : \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega) \to \mathbb{W}^{-1,p'}_{\Gamma_D}(\Omega),$$

and also the adjoint operator I_p^* of I_p is one-to-one. Now by [W05, Satz III.4.5], one has

$$\overline{\operatorname{ran}(I_p)} = (\ker(I_p^*))_\perp = (\ker(I_{p'}))_\perp = \{0\}_\perp = \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega),$$

which proves that I_p has dense range. $\square$

Building up on the results of Lemma 2.10, we introduce elliptic operators $\mathcal{T}_p$ and T_p :

Definition 2.11. For a given coefficient function

$$\mathbb{T} \in \mathrm{L}^\infty(\Omega; \mathcal{L}(\mathbb{C}^m \times \mathbb{C}^{m\times d}, \mathbb{C}^m \times \mathbb{C}^{m\times d}))$$

and fixed $p \in (1,\infty)$, the operator $\mathcal{T}_p$ is defined by

$$\mathcal{T}_p : \mathbb{W}^{1,p}_{\Gamma_D}(\Omega) \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), \quad \mathcal{T}_p := \mathcal{J}^*_{p'} \mathbb{T} \mathcal{J}_p.$$

In particular, for $u \in \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ and $v \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$, $\mathcal{T}_p u$ acts on v as

$$\langle\langle \mathcal{T}_p u, v\rangle\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} = \int_\Omega \mathbb{T}\begin{pmatrix} u \\ \nabla u\end{pmatrix} : \overline{\begin{pmatrix} v \\ \nabla v\end{pmatrix}} dx \qquad \text{and}$$

$$\langle \mathcal{T}_p u, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} = \int_\Omega \mathbb{T}\begin{pmatrix} u \\ \nabla u\end{pmatrix} : \begin{pmatrix} v \\ \nabla v\end{pmatrix} dx,$$

where we denote

$$\begin{pmatrix} e_1 \\ E_1\end{pmatrix} : \begin{pmatrix} e_2 \\ E_2\end{pmatrix} = \sum_{i=1}^m e_1^i e_2^i + \sum_{j=1}^m \sum_{k=1}^d E_1^{jk} E_2^{jk} \qquad \forall e_1, e_2 \in \mathbb{C}^m, \forall E_1, E_2 \in \mathbb{C}^{m\times d}.$$

We assume that $\mathcal{T}_2$ is elliptic, i.e. that there exists some $\omega > 0$ such that for all $v \in \mathbb{W}^{1,2}_{\Gamma_D}(\Omega)$ there holds the estimate

$$\mathrm{Re}\langle\langle \mathcal{T}_2 v, v\rangle\rangle_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)} \geq \omega \|v\|^2_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)}.$$

Since for $p \in (1,\infty)$, $\mathcal{T}_p$ corresponds to the restriction of the elliptic operator $\mathcal{T}_2$ to $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$, we call $\mathcal{T}_p$ elliptic. We define the unbounded operator

$$T_p := \mathcal{T}_p I_p^{-1} : \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega) \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$$

with $\mathrm{dom}(T_p) = \mathrm{ran}(I_p)$ as its domain of definition, and also refer to this operator as elliptic. The difference between $\mathcal{T}_p$ and T_p will become clear from the notation.

In the following remark, we introduce a representation of the matrix valued function $\mathbb{T}$ from Definition 2.11 which will be helpful in the proof of Theorem 2.14. In particular, we write $\mathbb{T}$ as a block diagonal matrix with four entries such that for $u \in \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ and $v \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$ the expression $\langle\langle \mathcal{T}_p u, v\rangle\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)}$ separates into four terms which involve the pairings $\{u, v\}$, $\{\nabla u, v\}$, $\{u, \nabla v\}$ and $\{\nabla u, \nabla v\}$.

Remark 2.12. The coefficient function $\mathbb{T}$ in Definition 2.11 can be represented in the form

$$\mathbb{T} = \begin{pmatrix} \mathbb{T}_{11} & \mathbb{T}_{12} P_{m\times d} \\ P^{-1}_{m\times d}\mathbb{T}_{21} & P^{-1}_{m\times d}\mathbb{T}_{22}P_{m\times d}\end{pmatrix},$$

with matrix valued functions $\mathbb{T}_{11} \in \mathrm{L}^\infty(\Omega; \mathbb{C}^{m\times m})$, $\mathbb{T}_{12} \in \mathrm{L}^\infty(\Omega; \mathbb{C}^{m\times md})$, $\mathbb{T}_{21} \in \mathrm{L}^\infty(\Omega; \mathbb{C}^{md\times m})$ and $\mathbb{T}_{22} \in \mathrm{L}^\infty(\Omega; \mathbb{C}^{md\times md})$. Written in this form (and omitting the dependence on x), the action of $\mathbb{T}$ on $(u, \nabla u)^\intercal$ is given by

$$\mathbb{T}\begin{pmatrix} u \\ \nabla u\end{pmatrix} = \begin{pmatrix} \mathbb{T}_{11} & \mathbb{T}_{12} P_{m\times d} \\ P^{-1}_{m\times d}\mathbb{T}_{21} & P^{-1}_{m\times d}\mathbb{T}_{22}P_{m\times d}\end{pmatrix}\begin{pmatrix} u \\ \nabla u\end{pmatrix} = \begin{pmatrix} \mathbb{T}_{11}u + \mathbb{T}_{12}P_{m\times d}\nabla u \\ P^{-1}_{m\times d}\mathbb{T}_{21}u + P^{-1}_{m\times d}\mathbb{T}_{22}P_{m\times d}\nabla u\end{pmatrix}.$$

Here, the mapping $P_{m\times d}:\mathbb{C}^{m\times d}\to\mathbb{C}^{md}$ is defined by

$$(P_{m\times d}E)_j := E_{\left(\lfloor\frac{j-1}{d}\rfloor+1\right)\left(j-d\cdot\lfloor\frac{j-1}{d}\rfloor\right)} \text{ for } 1\le j\le md, \qquad \forall E_1,E_2\in\mathbb{C}^{m\times d}.$$

Particularly, the entries of E are - row for row - written into a vector with md components. Similarly $P^{-1}_{m\times d}:\mathbb{C}^{md}\to\mathbb{C}^{m\times d}$ is defined by

$$(P^{-1}_{m\times d}e)_{kj} := e_{(k-1)\cdot d+j} \text{ for } 1\le k\le m \text{ and } 1\le j\le d, \qquad \forall e_1,e_2\in\mathbb{C}^{m}.$$

Hence, the entries of e are - row for row - written into a $\mathbb{C}^{m\times d}$-matrix, starting with the first row.

The following result is shown in [Hal+15, Theorem 6.2] and provides us the main tool to prove resolvent estimates for $\mathcal{T}_p$ and T_p in Theorem 2.14 below:

Theorem 2.13. *In the setting of Definition 2.9 let $\mathcal{M}$ be a set of coefficient functions as in Definition 2.11 with a uniform upper L^∞-bound $c^+>0$ and a common lower bound ω for the ellipticity constant of the corresponding operators. Then there exists an open interval $\mathrm{J}_\mathcal{M}$ with $2\in\mathrm{J}_\mathcal{M}$ such that for all $p\in\mathrm{J}_\mathcal{M}$ and all $\mathbb{T}\in\mathcal{M}$ the corresponding operator $\mathcal{T}_p$ is a topological isomorphism between $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ and $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. Additionally, there exists a constant $c_\mathcal{M}>0$ such that for all $p\in\mathrm{J}_\mathcal{M}$ and $f\in\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ the estimate*

$$\sup_{\mathbb{T}\in\mathcal{M}}\left\{\|\mathcal{T}_p^{-1}f\|_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)}\right\}\le c_\mathcal{M}\|f\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)}$$

holds.

2.2.2 Resolvent estimates for $\mathcal{T}_p$

In the setting of Theorem 2.13, we prove resolvent estimates for the operators $\mathcal{T}_p$. The proof mainly follows the proof of [GR89, Theorem 2], which states the result for scalar valued function spaces with regular domains. The latter is adapted in [Hal+15, Theorem 5.12] to apply for scalar valued spaces with rough boundary. However, in the vectorial framework with rough boundary we have to argue differently in several steps.

Theorem 2.14. *In the setting of Definition 2.9 let $\mathcal{M}$ be a set of coefficient functions as in Definition 2.11 with a uniform upper L^∞-bound $c^+>0$ and a common lower bound ω for the ellipticity constant of the corresponding operators. Then there exists an open interval $\mathrm{J}_\mathcal{M}$ with $2\in\mathrm{J}_\mathcal{M}$ (in general smaller than the one in Theorem 2.13) such that for all $p\in\mathrm{J}_\mathcal{M}$ and all $\mathbb{T}\in\mathcal{M}$, T_p is a densely defined and closed operator on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ which has compact resolvent. Moreover, with $\mathbb{C}_+:=\{\lambda\in\mathbb{C}:\ \mathrm{Re}(\lambda)\ge 0\}$ there holds:*

1. *For all $\lambda\in\mathbb{C}_+$ and $\mathbb{T}\in\mathcal{M}$, $\mathcal{T}_p+\lambda I_p$ is a continuous bijection from $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ onto $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$.*
2. $\displaystyle\sup_{\mathbb{T}\in\mathcal{M},\ \lambda\in\mathbb{C}_+}\|(\mathcal{T}_p+\lambda I_p)^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega),\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)\right)}<\infty.$
3. $\displaystyle\sup_{\mathbb{T}\in\mathcal{M},\ \lambda\in\mathbb{C}_+}\|\lambda I_p(\mathcal{T}_p+\lambda I_p)^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega),\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)}<\infty.$

Proof. The proof is divided into two steps. First we show the statement for $p\ge 2$ and afterwards we deduce that the theorem holds for $p<2$.
(I) $p\ge 2$:

Consider the interval $\mathrm{J}_{\mathcal{M}}$ from Theorem 2.13. Let first $p \in \mathrm{J}_{\mathcal{M}}$ be given with $p \geq 2$. Moreover, let $\mathbb{T} \in \mathcal{M}$ be arbitrary but fixed and recall the definition of the corresponding elliptic operator $\mathcal{T}_p$. W.l.o.g we assume $\frac{\omega}{c_+} < 1$. For arbitrary but fixed $\lambda \in \mathbb{C}_+$ we define $\rho := 1 - \frac{\omega}{2c_+}\mathrm{sgn}(\mathrm{Im}\lambda)i$. Note that $|\rho| < 2$. This and the definition of ρ will be important in Step I.i and Step I.ii.
The strategy of the proof is the following:
Steps I.i–I.ii show Statement 1 and that T_p is a densely defined and closed operator on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ with compact resolvent. Compactness of I_p, see Lemma 2.10, will be crucial here.
In Step I.iii we add an artificial dimension to $\mathbb{R}^d$ and extend the domain Ω to $\Omega \times (-1, 1)$. Moreover, we introduce extensions for functions from $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ by multiplication with Ψ times an exponential term which includes ρ, where Ψ is an appropriate smooth function which has compact support in $(-1, 1)$. We also define the restriction of a function of the extended domain by multiplying Ψ times the exponential term as above and integrating over $(-1, 1)$. Then we estimate the norm of those new functions against the norm of the old functions. We also extend the matrix $\mathbb{T}$ to a new matrix $\mathbb{T}_\lambda$ and obtain an elliptic operator $\mathcal{T}_{\lambda,p}$, which now acts on the extended function space. The strategy is to show that Theorem 2.13 holds for $\mathcal{T}_{\lambda,p}$ independently of λ and to use this knowledge in order to prove Statement 2 of this theorem. Partial integration and the correct choice of the extended and the restricted functions are crucial here.
In Step I.iv we conclude from Statement 2 that Statement 3 holds for $p \geq 2$.
(I.i) $\mathcal{T}_p + \lambda I_p$ is one-to-one:
We show that $\mathcal{T}_2 + \lambda I_2 : \mathbb{W}^{1,2}_{\Gamma_D}(\Omega) \to \mathbb{W}^{-1,2}_{\Gamma_D}(\Omega)$ is one-to-one. For arbitrary $u \in \mathbb{W}^{1,2}_{\Gamma_D}(\Omega)$, we first observe $|\rho| < 2$, then insert $\rho = 1 - \frac{\omega}{2c_+}\mathrm{sgn}(\mathrm{Im}\lambda)i$ and apply the estimates $\mathrm{Re}\lambda \geq 0$, $\frac{\omega}{2c_+} < 1$ and

$$\frac{\omega}{2c_+}\left|\mathrm{Im}\langle \mathcal{T}_2 u, \overline{u}\rangle_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)}\right| \leq \frac{\omega}{2}\|u\|^2_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)},$$

as well as ellipticity of $\mathcal{T}_2$ to compute:

$$\begin{aligned}
&2\|\mathcal{T}_2 u + \lambda I_2 u\|_{\mathbb{W}^{-1,2}_{\Gamma_D}(\Omega)}\|u\|_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)} \\
\geq &2\left|\langle \mathcal{T}_2 u + \lambda I_2 u, \overline{u}\rangle_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)}\right| \geq |\rho|\left|\langle \mathcal{T}_2 u + \lambda I_2 u, \overline{u}\rangle_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)}\right| \\
= &\sqrt{\left|\mathrm{Re}\Big(\rho\langle \mathcal{T}_2 u + \lambda I_2 u, \overline{u}\rangle_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)}\Big)\right|^2 + \left|\mathrm{Im}\Big(\rho\langle \mathcal{T}_2 u + \lambda I_2 u, \overline{u}\rangle_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)}\Big)\right|^2} \\
\geq &\mathrm{Re}\Big(\rho\langle \mathcal{T}_2 u + \lambda I_2 u, \overline{u}\rangle_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)}\Big) \\
\geq &\mathrm{Re}\langle \mathcal{T}_2 u, \overline{u}\rangle_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)} - \frac{\omega}{2c_+}\left|\mathrm{Im}\langle \mathcal{T}_2 u, \overline{u}\rangle_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)}\right| + \|u\|^2_{[\mathrm{L}^2(\Omega)]^m}\mathrm{Re}\lambda \\
&+ \frac{\omega}{2c_+}|\mathrm{Im}\lambda|\|u\|^2_{[\mathrm{L}^2(\Omega)]^m} \geq \frac{\omega}{2}\left(\|u\|^2_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)} + \frac{|\lambda|}{c_+}\|u\|^2_{[\mathrm{L}^2(\Omega)]^m}\right).
\end{aligned}$$

This shows

$$\|u\|_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)} \leq \frac{4}{\omega}\|\mathcal{T}_2 u + \lambda I_2 u\|_{\mathbb{W}^{-1,2}_{\Gamma_D}(\Omega)}, \tag{2.1}$$

which implies that $\mathcal{T}_2 + \lambda I_2 : \mathbb{W}^{1,2}_{\Gamma_D}(\Omega) \to \mathbb{W}^{-1,2}_{\Gamma_D}(\Omega)$ is one-to-one. Moreover, we obtain

$$\mathrm{Re}\Big(\rho\langle \mathcal{T}_2 u, \overline{u}\rangle_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)}\Big) \geq \frac{\omega}{2}\|u\|^2_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)}. \tag{2.2}$$

We will need this estimate in Step I.iii below. Since $\mathcal{T}_p + \lambda I_p$ is the restriction of $\mathcal{T}_2 + \lambda I_2$, we conclude that $\mathcal{T}_p + \lambda I_p$ is one-to-one.

(II.ii) T_p is densely defined and closed with compact resolvent and $\mathcal{T}_p + \lambda I_p$ is surjective:
We show that T_p is densely defined and closed with compact resolvent. This yields us surjectivity of $\mathcal{T}_p + \lambda I_p : \mathbb{W}^{1,p}_{\Gamma_D}(\Omega) \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$.
Theorem 2.13 entails that $\mathcal{T}_p^{-1} : \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega) \to \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ is continuous with norm less or equal than $c_{\mathcal{M}}$. Since I_p is compact by Lemma 2.10, also $T_p^{-1} = I_p \mathcal{T}_p^{-1} : \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega) \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ is compact. In particular, the resolvent set $\rho(T_p)$ contains zero. T_p is densely defined because $\mathrm{dom}(T_p) = \mathrm{ran}(I_p)$ which is a dense subset of $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$, see Lemma 2.10. To see that T_p is closed, let $(u_n)_{n\in\mathbb{N}} \subset \mathrm{dom}(T_p)$ and $v \in \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ be given such that $u_n \to v$ with $n \to \infty$. Let further $T_p u_n \to y$ with $n \to \infty$ for some $y \in \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. By continuity of $\mathcal{T}_p^{-1}$ there holds

$$I_p^{-1} u_n = \mathcal{T}_p^{-1} \mathcal{T}_p I_p^{-1} u_n = \mathcal{T}_p^{-1} T_p u_n \to \mathcal{T}_p^{-1} y \quad \text{in} \quad \mathbb{W}^{1,p}_{\Gamma_D}(\Omega) \quad \text{with} \quad n \to \infty.$$

Because I_p is continuous, we deduce $u_n \to I_p \mathcal{T}_p^{-1} y$ in $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ with $n \to \infty$ which yields us $v = I_p \mathcal{T}_p^{-1} y \in \mathrm{ran}(I_p) = \mathrm{dom}(T_p)$. We also conclude

$$T_p v = \mathcal{T}_p I_p^{-1} I_p \mathcal{T}_p^{-1} y = y,$$

which implies the closedness of T_p. Hence. T_p is densely defined, closed with $0 \in \rho(T_p)$ and T_p^{-1} is compact. [Kat80, Chp.3, Theorem 6.2.9] yields that the spectrum of T_p exclusively consists of isolated eigenvalues with finite multiplicities and that T_p has compact resolvent. Consequently, $\mathcal{T}_p + \lambda I_p : \mathbb{W}^{1,p}_{\Gamma_D}(\Omega) \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ can only fail to be surjective if λ is an eigenvalue of $-T_p$. But because T_2 is the restriction of T_p, this would imply that λ is also an eigenvalue of $-T_2$. This cannot be the case since $T_2 + \lambda = (\mathcal{T}_2 + \lambda I_2) I_2^{-1}$ and because $\mathcal{T}_2 + \lambda I_2$ is one-to-one by (2.1).
Note that Steps I.i–I.ii prove Statement 1 for the case $p \geq 2$.
(I.iii) Statement 2 holds for $p \geq 2$:
As in the proof of [Hal+15, Theorem 5.12], consider the extended domain $\tilde{\Omega} := \Omega \times (-1, 1)$. For $1 \leq j \leq m$ we define the extended open Neumann boundary part as the open set $\tilde{\Gamma}_{N_j} := \Gamma_{N_j} \times (-1, 1)$ and the corresponding Dirichlet boundary part

$$\tilde{\Gamma}_{\Gamma_{D_j}} := \partial\tilde{\Omega} \backslash \tilde{\Gamma}_{N_j} = (\overline{\Omega} \times \{-1, 1\}) \cup (\Gamma_{\Gamma_{D_j}} \times (-1, 1)) = (\overline{\Omega} \times \{-1, 1\}) \cup (\Gamma_{\Gamma_{D_j}} \times [-1, 1]).$$

Note that $\overline{\tilde{\Omega}}$ is a $(d+1)$-set and that each $\tilde{\Gamma}_{\Gamma_{D_j}}$ is a d-set which still satisfies Assumption 2.6. With a little abuse of notation we write $\mathcal{J}_p$ for the same operator as in Definition 2.9, now defined from $\mathbb{W}^{1,p}_{\tilde{\Gamma}_D}(\tilde{\Omega})$ to $\mathrm{L}^p(\tilde{\Omega}; \mathbb{C}^m \times \mathbb{C}^{m\times(d+1)})$. According to Definition 2.11 and Remark 2.12, we consider the fixed value $\lambda \in \mathbb{C}_+$ from the beginning of Step I and define the extended operator

$$\mathcal{T}_{\lambda,p} : \mathbb{W}^{1,p}_{\tilde{\Gamma}_D}(\tilde{\Omega}) \to \mathbb{W}^{-1,p}_{\tilde{\Gamma}_D}(\tilde{\Omega}), \quad \mathcal{T}_{\lambda,p} := \mathcal{J}_{p'}^* \mathbb{T}_\lambda \mathcal{J}_p,$$

where the coefficient function $\mathbb{T}_\lambda \in \mathrm{L}^\infty(\tilde{\Omega}, \mathcal{L}(\mathbb{C}^m \times \mathbb{C}^{m\times(d+1)}, \mathbb{C}^m \times \mathbb{C}^{m\times(d+1)}))$ takes the form

$$\mathbb{T}_\lambda = \begin{pmatrix} \mathbb{T}_{\lambda,11} & \mathbb{T}_{\lambda,12} P_{m\times(d+1)} \\ P^{-1}_{m\times(d+1)} \mathbb{T}_{\lambda,21} & P^{-1}_{m\times(d+1)} \mathbb{T}_{\lambda,22} P_{m\times(d+1)} \end{pmatrix}.$$

For the definition of the single components $\mathbb{T}_{\lambda,ij}$, $i, j \in \{1, 2\}$, remember the definition of the parameter $\rho = 1 - \frac{\omega}{2c_+} \mathrm{sgn}(\mathrm{Im}\lambda) i$. Then for a.e. $\tilde{x} := (x, t) \in \tilde{\Omega}$, the components of $\mathbb{T}_\lambda$ are defined by

$$
\begin{aligned}
&\mathbb{T}_{\lambda,11}(\tilde{x}) := \rho\mathbb{T}_{11}(x),\\
&(\mathbb{T}_{\lambda,12})_{kj}(\tilde{x}) := \begin{cases} \rho(\mathbb{T}_{12})_{k(j-\lfloor\frac{j}{d+1}\rfloor)}(x) & \text{if } j \bmod (d+1) \neq 0,\\ 0 & \text{else,}\end{cases}\\
&(\mathbb{T}_{\lambda,21})_{kj}(\tilde{x}) := \begin{cases} \rho(\mathbb{T}_{12})_{(k-\lfloor\frac{k}{d+1}\rfloor)j}(x) & \text{if } k \bmod (d+1) \neq 0,\\ 0 & \text{else}\end{cases} \quad \text{and}\\
&(\mathbb{T}_{\lambda,22})_{kj}(\tilde{x}) :=\\
&\quad \begin{cases} \rho(\mathbb{T}_{22})_{(k-\lfloor\frac{k}{d+1}\rfloor)(j-\lfloor\frac{j}{d+1}\rfloor)}(x) & \text{if } k \bmod (d+1), j \bmod (d+1) \neq 0,\\ 0, & \text{if } k \bmod (d+1) = 0, \text{ xor } j \bmod (d+1) = 0,\\ \frac{\lambda}{|\lambda|}(c_+ - \frac{\omega}{2}\operatorname{sgn}(\operatorname{Im}(\lambda))i) & \text{if } k \bmod (d+1) = j \bmod (d+1) = 0 \text{ and } k = j,\\ 0 & \text{else.}\end{cases}
\end{aligned}
$$

Note that $\frac{c_+\rho\lambda}{|\lambda|} = \frac{\lambda}{|\lambda|}\left(c_+ - \frac{\omega}{2}\operatorname{sgn}(\operatorname{Im}(\lambda))i\right)$. By this choice of $\mathbb{T}_\lambda$, $\mathcal{T}_{\lambda,p}$ essentially acts as $\rho\mathcal{T}_p$. The new spatial variable t is only influenced by the diffusive part

$$
\mathcal{J}_{p'}^* \begin{pmatrix} 0 & 0 \\ 0 & P_{m\times(d+1)}^{-1}\mathbb{T}_{\lambda,22}P_{m\times(d+1)} \end{pmatrix} \mathcal{J}_p
$$

of $\mathcal{T}_{\lambda,p}$ and in such a way that derivatives in t are not mixed with derivatives in the variables $x_1, \dots, x_d$. More precisely, for $\tilde{u} \in \mathbb{W}^{1,p}_{\tilde{\Gamma}_D}(\tilde{\Omega})$ and $\tilde{v} \in \mathbb{W}^{1,p'}_{\tilde{\Gamma}_D}(\tilde{\Omega})$ there holds

$$
\begin{aligned}
\langle\langle \mathcal{T}_{\lambda,p}\tilde{u}, \tilde{v}\rangle\rangle_{\mathbb{W}^{1,p'}_{\tilde{\Gamma}_D}(\tilde{\Omega})} &= \int_{-1}^{1}\int_{\Omega} \mathbb{T}_\lambda \begin{pmatrix}\tilde{u}\\ \nabla\tilde{u}\end{pmatrix} : \overline{\begin{pmatrix}\tilde{v}\\ \nabla\tilde{v}\end{pmatrix}} dxdt\\
= \int_{-1}^{1}\int_{\Omega} &\rho\left[\mathbb{T}_{11}\tilde{u} + \mathbb{T}_{12}P_{m\times d}\nabla_x\tilde{u}\right]\cdot\overline{\tilde{v}}\\
&+ \sum_{j=1}^{m}\sum_{k=1}^{d}\rho\left[\left(P_{m\times d}^{-1}\mathbb{T}_{21}\tilde{u}\right)_{jk} + \left(P_{m\times d}^{-1}\mathbb{T}_{22}P_{m\times d}\nabla_x\tilde{u}\right)_{jk}\right]\overline{(\nabla_x\tilde{v}_j)_k}\\
&+ \frac{c_+\rho\lambda}{|\lambda|}\sum_{j=1}^{m}\frac{\partial\tilde{u}_j}{\partial t}\overline{\frac{\partial\tilde{v}_j}{\partial t}}dxdt\\
= \int_{-1}^{1}\rho\int_{\Omega}&\left[\mathbb{T}\begin{pmatrix}\tilde{u}\\ \nabla_x\tilde{u}\end{pmatrix}\right] : \overline{\begin{pmatrix}\tilde{v}\\ \nabla_x\tilde{v}\end{pmatrix}}dx + \sum_{j=1}^{m}\int_{\Omega}\frac{c_+\rho\lambda}{|\lambda|}\frac{\partial\tilde{u}_j}{\partial t}\overline{\frac{\partial\tilde{v}_j}{\partial t}}dxdt\\
= \int_{-1}^{1}\rho&\langle\langle\mathcal{T}_p[\tilde{u}(\cdot,t)], \tilde{v}(\cdot,t)\rangle\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} + \sum_{j=1}^{m}\int_{\Omega}\frac{c_+\rho\lambda}{|\lambda|}\frac{\partial\tilde{u}_j}{\partial t}\overline{\frac{\partial\tilde{v}_j}{\partial t}}dxdt.
\end{aligned}
$$

Remember that the norm of $\rho\mathbb{T}$ is bounded by $|\rho|c_+ < 2c_+$. Hence, the norm of $\mathbb{T}_\lambda$ in $\mathrm{L}^\infty(\tilde{\Omega}, \mathcal{L}(\mathbb{C}^m \times \mathbb{C}^{m\times(d+1)}, \mathbb{C}^m \times \mathbb{C}^{m\times(d+1)}))$ is bounded by $2c_+$, i.e. independently of λ.
We show that $\mathcal{T}_{\lambda,2}$ is elliptic with an ellipticity constant which is independent of λ. To this aim, let $\tilde{v} \in \mathbb{W}^{1,2}_{\tilde{\Gamma}_D}(\tilde{\Omega})$ be given. There holds

$$
\operatorname{Re}\langle\langle\mathcal{T}_{\lambda,2}\tilde{v}, \tilde{v}\rangle\rangle_{\mathbb{W}^{1,2}_{\tilde{\Gamma}_D}(\tilde{\Omega})} = \int_{-1}^{1}\operatorname{Re}\left(\rho\langle\langle\mathcal{T}_p[\tilde{v}(\cdot,t)], \tilde{v}(\cdot,t)\rangle\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)}\right) + \sum_{j=1}^{m}\int_{\Omega}\operatorname{Re}\left(\frac{c_+\rho\lambda}{|\lambda|}\right)\left|\frac{\partial\tilde{v}_i}{\partial t}\right|^2 dxdt.
$$

For a.e. $t \in (-1,1)$, we apply (2.2) to the first integrand and estimate

$$\frac{\omega}{2}\|\tilde{v}(.,t)\|^2_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)} \leq \operatorname{Re}\left(\rho\langle\langle\mathcal{T}_p[\tilde{v}(\cdot,t)],\tilde{v}(\cdot,t)\rangle\rangle_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)}\right).$$

Moreover, since we assumed $\frac{\omega}{2} \leq c_+$ in the beginning of Step I, there holds

$$\frac{\omega}{2} \leq \frac{\omega}{2}\left(\frac{\operatorname{Re}(\lambda)}{|\lambda|}+\frac{|\operatorname{Im}(\lambda)|}{|\lambda|}\right) \leq \left(\frac{\operatorname{Re}(\lambda)}{|\lambda|}c_+ + \frac{\omega|\operatorname{Im}(\lambda)|}{2|\lambda|}\right) = \operatorname{Re}\left(\frac{c_+\rho\lambda}{|\lambda|}\right).$$

Hence, we deduce uniform ellipticity of $\mathcal{T}_{\lambda,2}$ from

$$\operatorname{Re}\langle\langle\mathcal{T}_{\lambda,2}\tilde{v},\tilde{v}\rangle\rangle_{\mathbb{W}^{1,2}_{\tilde{\Gamma}_D}(\tilde{\Omega})} \geq \int_{-1}^{1}\frac{\omega}{2}\|\tilde{v}(.,t)\|^2_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)} + \frac{\omega}{2}\sum_{j=1}^{m}\int_\Omega\left|\frac{\partial\tilde{v}_i}{\partial t}\right|^2 dxdt = \frac{\omega}{2}\|\tilde{v}\|^2_{\mathbb{W}^{1,2}_{\tilde{\Gamma}_D}(\tilde{\Omega})}.$$

Consequently, we can apply Theorem 2.13 to the set of coefficient matrices $\tilde{\mathcal{M}} := \{\mathbb{T}_\lambda : \mathbb{T} \in \mathcal{M},\ \lambda \in \mathbb{C}_+\}$. Note that $\mathrm{J}_{\tilde{\mathcal{M}}} \subset \mathrm{J}_{\mathcal{M}}$ because the uniform upper bound $2c_+$ of $\tilde{M}$ is larger than c_+ and the uniform ellipticity constant $\frac{\omega}{2}$ of $\mathcal{M}$ is smaller than ω. We assume w.l.o.g. that $\mathrm{J}_{\tilde{\mathcal{M}}}$ is of the form (p_0', p_0) for some $p_0 > 2$. Theorem 2.13 implies that $\mathcal{T}_{\lambda,p}$ is a topological isomorphism between $\mathbb{W}^{1,p}_{\tilde{\Gamma}_D}(\tilde{\Omega})$ and $\mathbb{W}^{-1,p}_{\tilde{\Gamma}_D}(\tilde{\Omega})$ and that the operator norm of $\mathcal{T}^{-1}_{\lambda,p}$ is bounded by a constant $c_{\tilde{\mathcal{M}}}$ which is independent of $\mathbb{T}_\lambda \in \tilde{M}$. Let $\psi \in \mathrm{C}_0^\infty(-1,1)$ be arbitrary with $0 \leq \psi \leq 1$ and $\psi(t) = 1$ for $t \in \left[-\frac{1}{2},\frac{1}{2}\right]$. We introduce the constant $\mu := \sqrt{\frac{|\lambda|}{c_+}}$. There holds

$$\frac{\rho\lambda}{\mu^2} = \frac{\lambda}{|\lambda|}\left(c_+ - \frac{\omega}{2}\operatorname{sgn}(\operatorname{Im}(\lambda))i\right).$$

For $p \in [2,\infty) \cap \mathrm{J}_{\tilde{\mathcal{M}}}$ and arbitrary $u \in \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ consider the extension

$$\tilde{u}(\tilde{x}) := u(x)\psi(t)\exp(i\mu t).$$

Note that $\tilde{u} \in \mathbb{W}^{1,p}_{\tilde{\Gamma}_D}(\tilde{\Omega})$. Moreover, we can estimate

$$\|\tilde{u}\|^p_{\mathbb{W}^{1,p}_{\tilde{\Gamma}_D}(\tilde{\Omega})} \geq \int_{-1/2}^{1/2}\int_\Omega\left[\begin{pmatrix}u(x)\\ \nabla u(x)\end{pmatrix} : \overline{\begin{pmatrix}u(x)\\ \nabla u(x)\end{pmatrix}}\right]^{\frac{p}{2}} dxdt = \|u\|^p_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)}. \tag{2.3}$$

For arbitrary $\tilde{v} \in \mathbb{W}^{1,p'}_{\tilde{\Gamma}_D}(\tilde{\Omega})$ we define the restriction $v \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$ as

$$v(x) := \int_{-1}^{1}\tilde{v}(x,t)\psi(t)\exp(i\mu t)dt.$$

Note that

$$\|v\|^{p'}_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} = \int_\Omega\left[\begin{pmatrix}v(x)\\ \nabla v(x)\end{pmatrix} : \overline{\begin{pmatrix}v(x)\\ \nabla v(x)\end{pmatrix}}\right]^{\frac{p'}{2}} dx \leq \int_\Omega\left[\left(\int_{-1}^{1}|\tilde{v}(x,t)||\psi(t)\exp(i\mu t)|dt\right)^2\right.$$

$$\left. + \sum_{j=1}^{m}\sum_{k=1}^{d}\left(\int_{-1}^{1}|(\nabla_x\tilde{v}_j(x,t))_k|\,|\psi(t)\exp(i\mu t)|dt\right)^2\right]^{\frac{p'}{2}} dx.$$

Hence, $|\psi(t)\exp(i\mu t)| \le 1$ and Hölder's inequality for the integration in t yields

$$\|v\|^{p'}_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} \le \int\limits_{\Omega} \left[2\int\limits_{-1}^{1} |\tilde{v}(x,t)|^2 + \sum_{j=1}^{m}\sum_{k=1}^{d} |(\nabla_x \tilde{v}_j(x,t))_k|^2 \, dt \right]^{\frac{p'}{2}} dx = 2^{p'/2} \|\tilde{v}\|^{p'}_{\mathbb{W}^{1,p'}_{\tilde{\Gamma}_D}(\tilde{\Omega})}.$$

This shows

$$\|v\|_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} \le 2^{\frac{p'}{2p'}} \|\tilde{v}\|_{\mathbb{W}^{1,p'}_{\tilde{\Gamma}_D}(\tilde{\Omega})} = \sqrt{2}\|\tilde{v}\|_{\mathbb{W}^{1,p'}_{\tilde{\Gamma}_D}(\tilde{\Omega})}. \tag{2.4}$$

We want to exploit ellipticity of $\mathcal{T}_{\lambda,p}$ to show Statement 2 of the theorem. To do this, we derive a relation between $\mathcal{T}_{\lambda,p}$ and $\mathcal{T}_p + \lambda I_p$. Consider the expression

$$\langle \mathcal{T}_{\lambda,p}\tilde{u}, \tilde{v}\rangle_{\mathbb{W}^{1,p'}_{\tilde{\Gamma}_D}(\tilde{\Omega})} = \rho \int_{-1}^{1} \langle \mathcal{T}_p \tilde{u}(\cdot,t), \tilde{v}(\cdot,t)\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} dt + \frac{\rho\lambda}{\mu^2}\int_{\tilde{\Omega}} \partial_t \tilde{u}\cdot \partial_t \tilde{v} d\tilde{x} =: K + L. \tag{2.5}$$

By definition of $\tilde{u}(\tilde{x}) := u(x)\psi(t)\exp(i\mu t)$ and $v(x) := \int\limits_{-1}^{1} \tilde{v}(x,t)\psi(t)\exp(i\mu t)dt$ there holds

$$K = \rho \int_{-1}^{1} \langle \mathcal{T}_p \tilde{u}(\cdot,t), \tilde{v}(\cdot,t)\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} dt = \rho \langle \mathcal{T}_p u, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)}. \tag{2.6}$$

For L we compute

$$L = \frac{\rho\lambda}{\mu^2}\int_{\tilde{\Omega}} \partial_t \tilde{u}\cdot\partial_t\tilde{v} d\tilde{x} = \frac{\rho\lambda}{\mu^2}\int_{\tilde{\Omega}} u(x)\mathrm{exp}(i\mu t)[\psi'(t) + i\mu\psi(t)]\cdot \partial_t \tilde{v}(x,t) d\tilde{x} =: L_1 + L_2. \tag{2.7}$$

Partial integration of L_2 in t yields

$$\begin{aligned} L_2 &= \frac{\rho\lambda}{\mu^2}\int_{\tilde{\Omega}} i\mu\tilde{u}\cdot\partial_t\tilde{v} d\tilde{x} = -\frac{\rho\lambda}{\mu^2}\int_{\tilde{\Omega}} [\exp(i\mu t) i\mu\psi(t)]' u(x)\cdot\tilde{v}(x,t) d\tilde{x} \\ &= -\frac{\rho\lambda}{\mu^2}\int_{\tilde{\Omega}} \exp(i\mu t) i\mu[i\mu\psi(t) + \psi'(t)]u(x)\cdot\tilde{v}(x,t)d\tilde{x} =: L_{2,1} + L_{2,2}. \end{aligned} \tag{2.8}$$

Again by definition of $\tilde{u}$ and v we obtain

$$L_{2,1} = \rho\lambda \int_{\tilde{\Omega}} \exp(i\mu t)\psi(t)u(x)\cdot\tilde{v}(x,t)d\tilde{x} = \rho\langle \lambda I_p u, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)}. \tag{2.9}$$

Another partial integration of $L_{2,2}$ in t yields

$$\begin{aligned} L_{2,2} &= \frac{\rho\lambda}{\mu^2}\int_{\tilde{\Omega}} \exp(i\mu t)u(x)\cdot\partial_t[\psi'(t)\tilde{v}(x,t)]d\tilde{x} \\ &= \frac{\rho\lambda}{\mu^2}\int_{\tilde{\Omega}} \exp(i\mu t)u(x)\cdot[\psi'(t)\partial_t\tilde{v}(x,t) + \psi''(t)\tilde{v}(x,t)]d\tilde{x} =: L_{2,2,1} + L_{2,2,2}. \end{aligned} \tag{2.10}$$

Note that $L_{2,2,1} = L_1 = \frac{\rho\lambda}{\mu^2}\int_{\tilde{\Omega}} u(x)\mathrm{exp}(i\mu t)\psi'(t)\cdot\partial_t\tilde{v}(x,t)d\tilde{x}$ by (2.7). This together with the representations (2.6)–(2.10) implies that we can rewrite (2.5) as

$$\begin{aligned} &\langle \mathcal{T}_{\lambda,p}\tilde{u},\tilde{v}\rangle_{\mathbb{W}^{1,p'}_{\tilde{\Gamma}_D}(\tilde{\Omega})} = K + L = (K + L_{2,1}) + (L_1 + L_{2,2,1}) + L_{2,2,2} \\ &= \rho\left\{ \langle \mathcal{T}_p u + \lambda I_p u, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} + \frac{\lambda}{\mu^2}\int\limits_{\Omega} u\cdot \int\limits_{-1}^{1} \exp(i\mu t)\left[2\psi'(t)\frac{\partial\tilde{v}}{\partial t} + \psi''(t)\tilde{v}\right] dt dx \right\}. \end{aligned} \tag{2.11}$$

We estimate (2.11) from above. Note that $|\exp(i\mu t)| \leq 1$, $\left|\frac{\rho\lambda}{\mu^2}\right| \leq 2c_+$, $\psi \in \mathrm{C}_0^\infty(-1,1)$ and $|\rho| \leq 2$. Moreover, remember (2.4). Hence, there exist constants $c_1, c_2 > 0$ which do not depend on λ and $\mathbb{T} \in \mathcal{M}$ such that

$$\begin{aligned}
&\left|\langle \mathcal{T}_{\lambda,p}\tilde{u}, \tilde{v}\rangle_{\mathbb{W}^{1,p'}_{\tilde{\Gamma}_D}(\tilde{\Omega})}\right| \\
&\leq \left|\rho\langle \mathcal{T}_p u + \lambda I_p u, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)}\right| + c_1\|u\|_{[\mathrm{L}^p(\Omega)]^m}\|\tilde{v}\|_{\mathbb{W}^{1,p'}_{\tilde{\Gamma}_D}(\tilde{\Omega})} \\
&\leq 2\sqrt{2}\|\mathcal{T}_p u + \lambda I_p u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)}\|v\|_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} + c_1\|u\|_{[\mathrm{L}^p(\Omega)]^m}\|\tilde{v}\|_{\mathbb{W}^{1,p'}_{\tilde{\Gamma}_D}(\tilde{\Omega})} \\
&\leq c_2\left[\|\mathcal{T}_p u + \lambda I_p u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} + \|u\|_{[\mathrm{L}^p(\Omega)]^m}\right]\|\tilde{v}\|_{\mathbb{W}^{1,p'}_{\tilde{\Gamma}_D}(\tilde{\Omega})}.
\end{aligned}$$

Since $\tilde{v} \in \mathbb{W}^{1,p'}_{\tilde{\Gamma}_D}(\tilde{\Omega})$ was arbitrary, we conclude

$$\|\mathcal{T}_{\lambda,p}\tilde{u}\|_{\mathbb{W}^{-1,p}_{\tilde{\Gamma}_D}(\tilde{\Omega})} \leq c_2\left[\|\mathcal{T}_p u + \lambda I_p u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} + \|u\|_{[\mathrm{L}^p(\Omega)]^m}\right]. \tag{2.12}$$

From (2.3) we know $\|u\|_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)} \leq \|\tilde{u}\|_{\mathbb{W}^{1,p}_{\tilde{\Gamma}_D}(\tilde{\Omega})}$ and Theorem 2.13 implies

$$\|\tilde{u}\|_{\mathbb{W}^{1,p}_{\tilde{\Gamma}_D}(\tilde{\Omega})} \leq c_{\tilde{\mathcal{M}}}\|\mathcal{T}_{\lambda,p}\tilde{u}\|_{\mathbb{W}^{-1,p}_{\tilde{\Gamma}_D}(\tilde{\Omega})} \qquad \text{for all } \mathbb{T}_\lambda \in \tilde{M}.$$

Consequently, we conclude from (2.12) that

$$\|u\|_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)} \leq c_3\left[\|\mathcal{T}_p u + \lambda I_p u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} + \|u\|_{[\mathrm{L}^p(\Omega)]^m}\right] \tag{2.13}$$

for some $c_3 > 0$ which is independent of $\lambda \in \mathbb{C}_+$ and $\mathbb{T} \in \mathcal{M}$. Remark 2.7 and $p \geq 2$ imply the embeddings $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega) \hookrightarrow\hookrightarrow [\mathrm{L}^p(\Omega)]^m \hookrightarrow [\mathrm{L}^2(\Omega)]^m$. [Nec12, Lemma 2.6.1] (see also [GR89, p. 111]) entails for each $\varepsilon > 0$ the existence of a constant $c(\varepsilon) > 0$ such that

$$\|u\|_{[\mathrm{L}^p(\Omega)]^m} \leq \varepsilon\|u\|_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)} + c(\varepsilon)\|u\|_{[\mathrm{L}^2(\Omega)]^m} \qquad \forall u \in [\mathrm{L}^p(\Omega)]^m.$$

For $\varepsilon < 1$, this estimate in (2.13) yields some $c_4 > 0$ with

$$\|u\|_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)} \leq c_4\left[\|\mathcal{T}_p u + \lambda I_p u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} + \|u\|_{[\mathrm{L}^2(\Omega)]^m}\right]. \tag{2.14}$$

Recall that $\|u\|_{[\mathrm{L}^2(\Omega)]^m} \leq \|u\|_{\mathbb{W}^{1,2}_{\Gamma_D}(\Omega)} \leq \frac{4}{\omega}\|\mathcal{T}_2 u + \lambda I_2 u\|_{\mathbb{W}^{-1,2}_{\Gamma_D}(\Omega)}$ by (2.1). Moreover, $p \geq 2$ implies $\|\mathcal{T}_2 u + \lambda I_2 u\|_{\mathbb{W}^{-1,2}_{\Gamma_D}(\Omega)} \leq \|\mathcal{T}_p u + \lambda I_p u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)}$. Both estimates applied in (2.14) finally yield us some $c_5 > 0$ with

$$\|u\|_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)} \leq c_5\|\mathcal{T}_p u + \lambda I_p u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)}. \tag{2.15}$$

Statement 1 for $p \geq 2$ has been shown in Steps I.i–I.ii. Hence, we can replace u in (2.15) by $u = (\mathcal{T}_p + \lambda I_p)^{-1}v$ for some $v \in \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$, and (2.15) shows Statement 2 for $p \geq 2$, since c_5 is independent of $\lambda \in \mathbb{C}_+$ and $\mathbb{T} \in \mathcal{M}$.
(I.iv) Statement 3 of the theorem holds for $p \geq 2$:

For any $v \in \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ and $u = (\mathcal{T}_p + \lambda I_p)^{-1} v$, (2.15) implies

$$\begin{aligned}
&\|\lambda I_p(\mathcal{T}_p + \lambda I_p)^{-1} v\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} = \|\lambda I_p u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} \\
&\leq \|\mathcal{T}_p u + \lambda I_p u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} + \|\mathcal{T}_p u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} \\
&\leq \|\mathcal{T}_p u + \lambda I_p u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} + c_+ \|u\|_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)} \\
&\leq \|v\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} + c_+ c_5 \|v\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} = (1 + c_+ c_5)\|v\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)}.
\end{aligned}$$

(II) $p < 2$:
We deduce the statements for $p < 2$ from those for $p \geq 2$. W.l.o.g. $\mathrm{J}_{\tilde{\mathcal{M}}}$ in Step I was of the form (p_0', p_0). Hence, for each $p \in \mathrm{J}_{\tilde{\mathcal{M}}} \cap (-\infty, 2)$ also $p' > 2$ is contained in $\mathrm{J}_{\tilde{\mathcal{M}}}$. Because $\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$ and $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ are reflexive, one has

$$(\mathcal{T}_p + \lambda I_p)^* = (\mathcal{T}_p^* + \lambda I_{p'}) : \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega) \to \mathbb{W}^{-1,p'}_{\Gamma_D}(\Omega) = [\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)]^*.$$

By [W05, Satz III.4.2], $\mathcal{T}_p^*$ has the same operator norm as $\mathcal{T}_p$. Hence, Theorem 2.14 can be applied to $\mathcal{T}_p^* + \lambda I_{p'}$ because $p' > 2$. Let $v \in \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$, $u \in \mathbb{W}^{-1,p'}_{\Gamma_D}(\Omega)$ and $\lambda \in \mathbb{C}_+$ be arbitrary. There holds

$$\begin{aligned}
\langle u, v\rangle_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)} &= \langle (\mathcal{T}_p^* + \lambda I_{p'})(\mathcal{T}_p^* + \lambda I_{p'})^{-1} u, v\rangle_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)} \\
&= \langle u, (\mathcal{T}_p^* + \lambda I_{p'})^{-1*}(\mathcal{T}_p + \lambda I_p) v\rangle_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)}.
\end{aligned}$$

By a corollary of the Hahn Banach Theorem [W05, Korollar III.1.6] it follows

$$v = (\mathcal{T}_p^* + \lambda I_{p'})^{-1*}(\mathcal{T}_p + \lambda I_p) v,$$

so that $\mathcal{T}_p + \lambda I_p$ is one-to-one. Consequently, λ is no eigenvalue of $-T_p = -\mathcal{T}_p I_p^{-1} : \mathrm{ran}(I_p) \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. T_p is closed and T_p^{-1} is compact by the same reasoning as in Step I.ii. We conclude that the spectrum of $-T_p$ consists only of eigenvalues so that $\lambda \in \rho(-T_p)$. This implies that $T_p + \lambda$ is surjective. Because $(\mathcal{T}_p + \lambda I_p) I_p^{-1} = (T_p + \lambda)$, also $(\mathcal{T}_p + \lambda I_p)$ is surjective and hence bijective. Since $\mathcal{T}_p + \lambda I_p$ is bounded and bijective, a corollary of the open mapping theorem, [W05, Korollar IV.3.4], yields that also $(\mathcal{T}_p + \lambda I_p)^{-1}$ is bounded which implies Statement 1 of the theorem.
For Statement 2, note that $(\mathcal{T}_p + \lambda I_p)^{-1} = (\mathcal{T}_p^* + \lambda I_{p'})^{-1*}$ and that $(\mathcal{T}_p^* + \lambda I_p)^{-1*}$ and $(\mathcal{T}_p^* + \lambda I_{p'})^{-1}$ have the same operator norm. Statement 2 of the theorem holds for $p' \geq 2$. This implies that the norm of $(\mathcal{T}_p^* + \lambda I_{p'})^{-1}$ is bounded by a constant which is independent of λ and $\mathbb{T} \in \mathcal{M}$. Consequently, Statement 2 holds also for $p < 2$. Statement 3 follows analogous to the case $p \geq 2$. □

Remark 2.15. If the coefficient matrices $\mathbb{T} \in \mathcal{M}$ of $\mathcal{T}_p$ are real-valued then Theorem 2.14 carries over to real valued spaces $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ [cf. Hal+15, 7. Applications]. To see this, we compute for arbitrary real valued $u \in \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$, $v \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$ and $\lambda \in \mathbb{C}_+$:

$$\begin{aligned}
\overline{\langle (\mathcal{T}_p + \lambda I_p) u, v\rangle}_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} &= \int_\Omega \overline{\left[\mathbb{T}(x) \begin{pmatrix} u(x) \\ \nabla u(x) \end{pmatrix}\right] : \begin{pmatrix} v(x) \\ \nabla v(x) \end{pmatrix}} + \overline{\lambda u(x) v(x)} dx \\
&= \langle (\mathcal{T}_p + \overline{\lambda} I_p) u, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)}.
\end{aligned}$$

For $\lambda \in \mathbb{R}_+$, this implies

$$\overline{\langle (\mathcal{T}_p + \lambda I_p)u, v\rangle}_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} = \langle (\mathcal{T}_p + \lambda I_p)u, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} \in \mathbb{R}.$$

Similarly, if $\mathrm{Im}(u) \neq 0$ but v is real valued and $\lambda \in \mathbb{R}_+$, then

$$\begin{aligned} \overline{\langle (\mathcal{T}_p + \lambda I_p)u, v\rangle}_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} &= \langle (\mathcal{T}_p + \lambda I_p)\overline{u}, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} && \text{and} \\ 2\mathrm{Im}\langle (\mathcal{T}_p + \lambda I_p)u, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} &= \overline{\langle (\mathcal{T}_p + \lambda I_p)u, v\rangle}_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} - \langle (\mathcal{T}_p + \lambda I_p)u, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} \\ &= \langle (\mathcal{T}_p + \lambda I_p)(\overline{u} - u), v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)}. \end{aligned}$$

But then $\mathrm{Im}\langle (\mathcal{T}_p + \lambda I_p)u, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} = 0$ does not hold for all $v \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$, since $\mathcal{T}_p + \lambda I_p$ is one-to-one and $\overline{u} - u$ is not the zero function. This implies that if $\tilde{u}$ is contained in the real valued dual space $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ of $\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$, i.e. if $\langle \tilde{u}, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} \in \mathbb{R}$ for all real valued $v \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$, then $u := (\mathcal{T}_p + \lambda I_p)^{-1}\tilde{u} \in \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ must be real valued to, since otherwise there exists some real valued $v \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$ such that

$$0 = 2\mathrm{Im}\langle \tilde{u}, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} = 2\mathrm{Im}\langle (\mathcal{T}_p + \lambda I_p)u, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} \neq 0.$$

2.2.3 Diffusion operators

In this subsection, we define an important sub-category of elliptic operators as introduced in Definition 2.11, which are diffusion operators. In particular, we specialize on diffusion operators which have a diagonal, positive definite diffusion matrix. Those are considered throughout [Mün17a] and [Mün17b]. All spaces in this subsection are assumed to be real valued. Before we introduce diffusion operators in Definition 2.17, we define an operator similar to $\mathcal{J}_p$ and recall I_p from Definition 2.9.

Definition 2.16. [Mün17a, Definition 2.8] With Assumption 2.2 and Assumption 2.6 and $p \in (1, \infty)$ we define the operators

$$\mathcal{L}_p : \mathbb{W}^{1,p}_{\Gamma_D}(\Omega) \to \mathrm{L}^p(\Omega, \mathbb{R}^{md}), \qquad \mathcal{L}_p(u) := \mathrm{vec}(\nabla u) = (\nabla u_1, \dots, \nabla u_m)^\intercal$$

and

$$I_p : \mathbb{W}^{1,p}_{\Gamma_D}(\Omega) \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), \qquad \langle I_p u, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} := \int_\Omega u \cdot v \, dx \qquad \forall v \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega).$$

We introduce diffusion operators A_p as a sub-category of general elliptic operators T_p from Definition 2.11:

Definition 2.17. [Mün17a, Definition 2.9] Let the constants $d_1, \dots, d_m > 0$ be given diffusion coefficients and

$$D = \mathrm{diag}(d_1, \dots, d_1, \dots, d_m, \dots, d_m) \in \mathbb{R}^{md \times md}.$$

For $p \in (1, \infty)$ we set

$$\mathcal{A}_p : \mathbb{W}^{1,p}_{\Gamma_D}(\Omega) \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), \ \mathcal{A}_p := \mathcal{L}^*_{p'} D \mathcal{L}_p.$$

We define the unbounded operator

$$A_p : \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega) \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), \ A_p := \mathcal{A}_p I_p^{-1}$$

with domain

$$\mathrm{dom}(A_p) = \mathrm{ran}\,(I_p) \subset \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega).$$

Remark 2.18. Note first that $\mathcal{L}_p(u)$ in Definition 2.16 is nothing else than $P_{m\times d}\nabla u$ with $P_{m\times d}$ from Remark 2.12. The relation between $\mathcal{A}_p$ in Definition 2.17 and $\mathcal{T}_p$ in Definition 2.11 is given by $\mathcal{A}_p = \mathcal{L}_{p'}^* D\mathcal{L}_p = \mathcal{J}_{p'}^*\mathbb{T}\mathcal{J}_p = \mathcal{T}_p$, where $\mathbb{T}$ is represented as in Remark 2.12 by

$$\mathbb{T} = \begin{pmatrix} \mathbb{T}_{11} & \mathbb{T}_{12}P_{m\times d} \\ P_{m\times d}^{-1}\mathbb{T}_{21} & P_{m\times d}^{-1}\mathbb{T}_{22}P_{m\times d} \end{pmatrix},$$

with $\mathbb{T}_{11} = 0 \in \mathbb{R}^{m\times m}$, $\mathbb{T}_{12} = 0 \in \mathbb{R}^{m\times md}$, $\mathbb{T}_{21} = 0 \in \mathbb{R}^{md\times m}$ and $\mathbb{T}_{22} = D \in \mathbb{R}^{md\times md}$. Note that A_p is not necessarily an elliptic operator in the sense of Definition 2.11. However, $A_p + 1 = (\mathcal{A}_p + I_p)I_p^{-1}$ is elliptic. The corresponding coefficient matrices of $\mathcal{A}_p + I_p$ are obtained by replacing $\mathbb{T}_{11} = 0 \in \mathbb{R}^{m\times m}$ by the identity matrix $\tilde{\mathbb{T}}_{11} = \mathrm{Id} \in \mathbb{R}^{m\times m}$. By Remark 2.12, for the new coefficient matrices $\tilde{\mathbb{T}}_{ij}$, $1 \le i,j \le 2$, which define $\tilde{\mathbb{T}}$, and for the corresponding elliptic operator $\tilde{\mathcal{T}}_p$ there holds $\mathcal{A}_p + I_p = \tilde{\mathcal{T}}_p$.

As a corollary of Theorem 2.14 we can prove the main statement of [Mün17a, Theorem 2.10], cf. also [Hal+15, Theorem 5.12]:

Corollary 2.19. [Mün17a, Theorem 2.10] In the setting of Definition 2.16 and Definition 2.17 there exists an open interval J around 2 such that for all $p \in \mathrm{J}$ the operator $\mathcal{A}_p + I_p$ is a topological isomorphism between $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ and $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$.
There is a constant $c > 0$ such that for all $p \in \mathrm{J}$ and $\lambda \in \mathbb{C}_+$ there holds the resolvent estimate

$$\|(A_p + 1 + \lambda)^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \le \frac{c}{1 + |\lambda|}.$$

Moreover, $-A_p$ generates an analytic semigroup of operators on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$.

Remark 2.20. In [Hal+15, Theorem 5.12], the corresponding statements are shown for the scalar case, but with a non-constant coefficient matrix $\mu \in \mathrm{L}^\infty(\Omega; \mathbb{C}^{d\times d})$ which defines an elliptic operator $\nabla\cdot\mu\nabla$. The generalization to the vectorial setting is straight forward since Theorem 2.14 holds for non-constant coefficient matrices. We maintain the diffusion matrix in Definition 2.17 constant and diagonal because the main challenge in the analysis of reaction-diffusion systems of this work is due to the hysteresis operator in the reaction term. Note however that most of the results which hold for A_p can be generalized to the case of non-constant diffusion-matrices.

Proof of Corollary 2.19. By Remark 2.18, $A_p + 1$ is bounded and elliptic with

$$A_p + 1 = (\mathcal{A}_p + I_p)I_p^{-1} = \tilde{\mathcal{T}}_p I_p^{-1},$$

where $\tilde{\mathcal{T}}_p$ is an elliptic operator. Hence, we can apply Theorem 2.14 to $A_p + 1$. We write

$$\begin{aligned}&(1 + |\lambda|)\|(A_p + 1 + \lambda)^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \\ &= \|(A_p + 1 + \lambda)^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} + |\lambda|\|(A_p + 1 + \lambda)^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)}.\end{aligned}$$

The second term is finite by Statement 3 of Theorem 2.14. For the first term, note that

$$\begin{aligned}&\|(A_p + 1 + \lambda)^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} = \|I_p(\tilde{\mathcal{T}}_p + \lambda I_p)^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \\ &\le \|I_p\|_{\mathcal{L}\left(\mathbb{W}^{1,p}_{\Gamma_D}(\Omega),\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)}\|(\tilde{\mathcal{T}}_p + \lambda I_p)^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega),\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)\right)} < \infty\end{aligned}$$

by Lemma 2.10 and Statement 2 of Theorem 2.14. This yields the required resolvent estimate. The last statement of the theorem follows from Theorem 2.22 below. In fact, by Theorem 2.22, $-T_p = -(A_p + 1)$ generates an analytic semigroup $\exp(-(A_p + 1)t)$, $t \ge 0$, of operators on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. But then $-A_p$ generates the semigroup $\exp(-(A_p + 1)t)\exp(t) = \exp(-A_p t)$, $t \ge 0$. $\square$

2.2.4 Generators of analytic semigroups of operators

In this subsection, we give a brief introduction into the theory of sectorial operators and semigroups of operators. We do not go into detail in many parts, but refer to the literature, see for example [Paz83; Lun95; Hen81] to only name a few.
The motivation for this subsection is to get some insight into the connection between resolvent estimates like in Theorem 2.14 and the property of an operator to be the generator of an analytic semigroup of operators. The latter provides a powerful tool when it comes to the theory of non-linear operator evolution equations.
We begin with defining what we mean by a semigroup of operators and its generator:

Definition 2.21. [Cf. Paz83, Definition 1.1.1, Definition 1.2.1] and [Hen81, Definition 1.3.3]. Let X be a Banach space. We call a family of bounded and linear operators $\{T(t)\}_{t\geq 0}$ on X a semigroup of bounded linear operators if

$$\begin{aligned} T(0) &= \mathrm{Id} && \text{and} \\ T(s)T(t) &= T(s+t) && \text{for } s,t \geq 0. \end{aligned}$$

We call $\{T(t)\}_{t\geq 0}$ a strongly continuous semigroup or C_0-semigroup of bounded linear operators on X if in addition $T(t)u \to u$ with $t \downarrow 0$ for all $u \in X$.
A C_0-semigroup $\{T(t)\}_{t\geq 0}$ on X is called (real) analytic, if $t \mapsto T(t)u$ is real analytic on $(0,\infty)$ for each $u \in X$.
We call A the infinitesimal generator of a semigroup of bounded linear operators $\{T(t)\}_{t\geq 0}$ if

$$Au = \lim_{t\downarrow 0} \frac{1}{t}(T(t)u - u)$$

for all $u \in X$, for which the limit exists, and if $\mathrm{dom}(A)$ is equal to the set of those u. We also write $T(t) = \exp(At)$ in this case.

Theorem 2.22. *Adopt the notation and the assumptions from Theorem 2.14 and for $p \in \mathrm{J}_{\mathcal{M}}$ let T_p be the elliptic operator which corresponds to a coefficient matrix $\mathbb{T} \in \mathcal{M}$. Then there exists some $\delta \in (0, \frac{\pi}{2})$ such that*

$$\rho(-T_p) \supset \Sigma = \left\{\lambda : |\mathrm{arg}\lambda| < \frac{\pi}{2} + \delta\right\} \cup \{0\}, \tag{2.16}$$

and there exists some $C_1 > 0$ with

$$\|(-T_p - \lambda)^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \leq \frac{C_1}{|\lambda|} \tag{2.17}$$

for all $\lambda \in \Sigma\backslash\{0\}$. Moreover, $-T_p$ is the generator of an analytic semigroup of operators $\exp(-T_p t)$, $t \geq 0$, on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ in the sense of Definition 2.21. $\exp(-T_p t)$ is uniformly bounded, i.e. for some $C_2 > 0$ and for all $t \geq 0$,

$$\|\exp(-T_p t)\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \leq C_2. \tag{2.18}$$

$\exp(-T_p t)$ is differentiable for $t > 0$ and there exists some $C_3 > 0$ with

$$\left\|\frac{d}{dt}\exp(-T_p t)\right\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} = \|(-T_p \exp(-T_p t)\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \leq \frac{C_3}{t}. \tag{2.19}$$

Proof. We want to apply [Paz83, Theorem 2.5.2], which states several properties of an unbounded, closed and densely defined operator A on a Banach space X which are equivalent to the fact that A generates an analytic semigroup $T(t)$, $t \geq 0$. We adapt the notation in [Paz83, Theorem 2.5.2] to that of Theorem 2.14, i.e. we replace A by $-T_p$ and X by $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ for some $\mathbb{T} \in \mathcal{M}$ and $p \in \mathrm{J}_M$. In [Paz83, Theorem 2.5.2] it is assumed that $-T_p$ is the generator of a uniformly bounded C_0-semigroup of operators on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ and that $0 \in \rho(-T_p)$. We have to prove this assumption in order to apply [Paz83, Theorem 2.5.2.b)]. Property [Paz83, Theorem 2.5.2.b)] is the following: There exists some $c > 0$ such that for all $\sigma > 0$ and $\tau \neq 0$ the resolvent estimate

$$\|(-T_p - (\sigma + i\tau))^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \leq \frac{c}{|\tau|} \tag{2.20}$$

holds. Property [Paz83, Theorem 2.5.2.c)] is exactly given by (2.16)–(2.17).
First of all we prove (2.20). By Theorem 2.14, T_p is densely defined and closed with $0 \in \rho(T_p)$ and there exists some $c > 0$ such that for each $\lambda \in \mathbb{C}_+\backslash\{0\}$, $T_p + \lambda$ is continuously invertible with

$$\|(T_p + \lambda)^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \leq \frac{c}{|\lambda|}.$$

This already implies (2.20). To gain access to the full statement of [Paz83, Theorem 2.5.2], we still have to prove that $-T_p$ generates a uniformly bounded C_0-semigroup of operators on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. In the proof of [Paz83, Theorem 2.5.2, *b)* $\Rightarrow$ *c)*], this property is not needed though. Therefore, (2.16)–(2.17) follows from (2.20). But by an equivalence theorem about C_0-semigroups, [Paz83, Theorem 1.7.7], (2.16)–(2.17) already implies that $-T_p$ is the generator of a uniformly bounded C_0-semigroup of operators $\exp(-T_p t)$, $t \geq 0$, on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. Hence, we conclude (2.18). Moreover, all the assumptions of [Paz83, Theorem 2.5.2] are satisfied and we gain access to the complete statement of the theorem. By [Paz83, Theorem 2.5.2.a)], $\exp(-T_p t)$, $t \geq 0$, can be extended to an analytic semigroup of operators in a sector $\Delta_\delta = \{z : |\arg z| < \delta\}$ and $\|\exp(-T_p z)\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)}$ is uniformly bounded in every closed sub-sector $\overline{\Delta}_{\delta'}$ of Δ_δ with $\delta' < \delta$. This is sufficient for $\{\exp(-T_p t)\}_{t\geq 0}$ to be an analytic semigroup of operators according to Definition 2.21. Finally, [Paz83, Theorem 2.5.2.d)] implies that $\exp(-T_p t)$ is differentiable for $t > 0$ and that (2.19) holds. ☐

We want to improve the estimates (2.18)–(2.19) in Theorem 2.22. First of all, Theorem 2.22 shows that T_p is sectorial for a sector $S_{0,\phi}$ in the following sense:

Definition 2.23. [Hen81, Definition 1.3.1] The linear operator T_p in $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ is sectorial for a sector $S_{a,\Phi}$ if it is densely defined and closed, and if for some $\phi \in \left(0, \frac{\pi}{2}\right)$, $M \geq 1$ and $a \in \mathbb{R}$,

$$\rho(T_p) \supset S_{a,\Phi} = \{\lambda : \phi \leq |\arg(\lambda - a)| \leq \pi,\, \lambda \neq a\},$$

and if for all $\lambda \in S_{a,\phi}$,

$$\|(T_p - \lambda)^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \leq \frac{M}{|\lambda - a|}.$$

Remark 2.24. Note first that each sector in Definition 2.23 contains a left half plain, but λ is now subtracted in the resolvent estimate, i.e. the resolvent set of T_p is considered. In Theorem 2.14 and Corollary 2.19, we always looked at right half plains and added λ in the corresponding estimate. Also the statement in Theorem 2.22 contains a sector which includes a right half plane, but the resolvent set of $-T_p$ is considered there. We keep the notation from [Hen81, Definition 1.3.1] in order to make it easier for the reader to compare our results to the literature.

Theorem 2.25. *Adopt the assumptions and the notation from Theorem 2.14. Let $p \in \mathrm{J}_{\mathcal{M}}$ and $\mathbb{T} \in \mathcal{M}$ be arbitrary and consider the corresponding elliptic operator T_p. Then T_p is sectorial in the sense of Definition 2.23 for a sector $S_{\delta,\tilde{\phi}}$, with $\delta > 0$ and $\tilde{\phi} \in \left(0, \frac{\pi}{2}\right)$. Let $\exp(-T_p t)$, $t \geq 0$, be the analytic semigroup generated by $-T_p$ according to Theorem 2.22. Then there exists some $C > 0$ such that for all $t > 0$ the norms of $\exp(-T_p t)$ and $T_p \exp(-T_p t)$ can be estimated by*

$$\|\exp(-T_p t)\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \leq C \exp(-\delta t), \quad \|T_p \exp(-T_p t)\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \leq \frac{C}{t} \exp(-\delta t). \tag{2.21}$$

Proof. Since T_p is closed by Theorem 2.14, $\rho(T_p)$ is open [cf. Kat80, Chp. 4, Theorem 6.7]. Moreover, $0 \in \rho(T_p)$ by Theorem 2.14. Therefore there exist constants $\varepsilon > 0$ and $C_\varepsilon > 0$ such that $\overline{B_\varepsilon} := \overline{B_{\mathbb{C}}(0, \varepsilon)} \subset \rho(T_p)$ and for all $\lambda \in \overline{B_\varepsilon}$,

$$\|(T_p - \lambda)^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \leq C_\varepsilon. \tag{2.22}$$

We choose $\varepsilon < \frac{2}{\sqrt{13}}$. Theorem 2.22 entails that T_p is sectorial for a sector $S_{0,\phi}$ with $\phi \in \left(0, \frac{\pi}{2}\right)$. We choose $\delta = \frac{\varepsilon \cos(\phi)}{2}$ Then the circle ∂B_ε intersects the angle legs of φ exactly at the points $z_1 = (2\delta, 2\delta \tan(\varphi))$, $z_2 = (2\delta, -\tan(2\delta\varphi))$ in the complex plane. We define by $\tilde{\varphi}$ the angle between the line $a = [(\delta, 0), (2\delta, 0)]$ and the line $c = [(\delta, 0), z_1]$ in the complex plane. The line $b := [(2\delta, 0), z_1]$ closes the triangle $\triangle((\delta, 0), (2\delta, 0), z_1)$. Hence, $\tilde{\varphi} = \arctan\left(\frac{|b|}{|a|}\right) = \arctan\left(\frac{2\delta \tan(\varphi)}{\delta}\right) = \arctan(2 \tan \phi) \in \left(0, \frac{\pi}{2}\right)$. Moreover, $|c| = \varepsilon$.

The goal is to prove that T_p is sectorial for the sector $S_{\delta,\tilde{\phi}}$. First of all, consider the set $V_{2\delta,\varepsilon} := \{\lambda : |\mathrm{Re}|\lambda \leq 2\delta\} \cap \overline{B_\varepsilon}$. Then $V_{2\delta,\varepsilon}$ is defined such that $V_{2\delta,\varepsilon} \subset \overline{B_\varepsilon}$ and $(S_{\delta,\tilde{\phi}} \backslash V_{2\delta,\varepsilon}) \subset S_{0,\phi}$. This way we obtain

$$S_{\delta,\tilde{\phi}} = (S_{\delta,\tilde{\phi}} \cap V_{2\delta,\varepsilon}) \cup (S_{\delta,\tilde{\phi}} \backslash V_{2\delta,\varepsilon}) \subset \rho(T_p).$$

Hence, we are left to prove the correct resolvent estimates.

(I) Consider first the case $\lambda \in (S_{\delta,\tilde{\phi}} \cap V_{2\delta,\varepsilon})$:

By definition of δ and $V_{2\delta,\varepsilon}$ there holds

$$|\mathrm{Re}(\lambda) - \delta|^2 \leq ||\mathrm{Re}(\lambda)| + \delta| \leq |3\delta|^2 \leq (9/4)|\varepsilon|^2.$$

Moreover, $|\mathrm{Im}(\lambda)|^2 \leq |\varepsilon|^2$ because $V_{2\delta,\varepsilon} \subset \overline{B}_\varepsilon$. Hence, we can estimate

$$0 < |\lambda - \delta| = \sqrt{|\mathrm{Re}(\lambda) - \delta|^2 + |\mathrm{Im}(\lambda)|^2} \leq \frac{\sqrt{13}\varepsilon}{2} < 1.$$

Because $\lambda \in V_{2\delta,\varepsilon} \subset \overline{B}_\varepsilon$, this together with (2.22) yields

$$\|(T_p - \lambda)^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \leq C_\varepsilon \leq \frac{C_\varepsilon}{|\lambda - \delta|}.$$

(II) Consider now $\lambda \in (S_{\delta,\tilde{\phi}} \backslash V_{2\delta,\varepsilon})$:

By definition of $V_{2\delta,\varepsilon}$, either $|\mathrm{Re}(\lambda)| > 2\delta$ and hence $|\lambda| > 2\delta$ or $|\lambda| > \varepsilon$. In the second case, the definition of $\delta = \frac{\varepsilon \cos(\varphi)}{2}$ yields $|\lambda| > \varepsilon = \frac{2\delta}{\cos(\phi)} > 2\delta$. Consequently, $\frac{\delta}{|\lambda|} \leq 1$ holds for any $\lambda \in (S_{\delta,\tilde{\phi}} \backslash V_{2\delta,\varepsilon})$. Furthermore, as seen above, $\lambda \in (S_{\delta,\tilde{\phi}} \backslash V_{2\delta,\varepsilon}) \subset S_{0,\phi} \subset \rho(T_p)$, which of course implies $(\lambda - \delta) \in S_{0,\phi} \subset \rho(T_p)$, since $S_{0,\phi} - \delta \subset S_{0,\phi}$. By the first resolvent equation for closed operators, [cf. Kat80, Chp. 1, (5.5)], and a remark after [Kat80, Chp. 3, Theorem 6.5], we obtain

$$\begin{aligned}(T_p - \lambda)^{-1} &= (T_p - (\lambda - \delta))^{-1} + [\lambda - (\lambda - \delta)](T_p - \lambda)^{-1}(T_p - (\lambda - \delta))^{-1} \\ &= (T_p - (\lambda - \delta))^{-1} + \delta(T_p - \lambda)^{-1}(T_p - (\lambda - \delta))^{-1}.\end{aligned}$$

Hence, Theorem 2.25 yields

$$\begin{aligned}\|(T_p-\lambda)^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \leq & \|(T_p-(\lambda-\delta))^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \\ & + \delta\|(T_p-\lambda)^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)}\|(T_p-(\lambda-\delta))^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \\ \leq & \frac{C}{|\lambda-\delta|}\left(1+\frac{\delta C}{|\lambda|}\right) \leq \frac{C(C+1)}{|\lambda-\delta|}.\end{aligned}$$

(III) Conclusion:
Since $S_{\delta,\tilde{\phi}} = (S_{\delta,\tilde{\phi}} \cap V_{2\delta,\varepsilon}) \cup (S_{\delta,\tilde{\phi}} \backslash V_{2\delta,\varepsilon})$, we conclude from Steps I–II that

$$\|(T_p-\lambda)^{-1}\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \leq \frac{\max\{C_\varepsilon, C(1+C)\}}{|\lambda-\delta|}$$

for each $\lambda \in S_{\delta,\tilde{\phi}}$. This proves that T_p is sectorial for the sector $S_{\delta,\tilde{\phi}}$. Estimate (2.21) then follows from [Hen81, Theorem 1.3.4]. □

2.2.5 Fractional powers and fractional power spaces

In this subsection, we define the fractional power of a sectorial operator and the corresponding fractional power spaces. We mostly follow the definitions in [Hen81] but also refer to [Paz83] and [Lun95]. For more results on sectorial operators we recommend the book [Haa06].

Definition 2.26. [Hen81, Definition 1.4.1] Adopt the assumptions and the notation from Theorem 2.14. Let $p \in \mathrm{J}_{\mathcal{M}}$ and $\mathbb{T} \in \mathcal{M}$ be arbitrary and consider the corresponding elliptic operator T_p. By Theorem 2.25, T_p is sectorial and the spectrum of T_p satisfies $\sigma(T_p) \subset \{\lambda \in \mathbb{C} : \mathrm{Re}(\lambda) > \delta\}$ for some $\delta > 0$. Hence, according to [Hen81, Definition 1.4.1], for $\theta > 0$ we can define the fractional power

$$T_p^{-\theta} := \frac{1}{\Gamma(\theta)} \int_0^\infty t^{\theta-1} \exp(-T_p t) dt. \tag{2.23}$$

By [Hen81, Theorem 1.4.2], $T_p^{-\theta}$ is a bounded and linear operator on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. Moreover, $T_p^{-\theta}$ is one-to-one and for $\theta, \beta > 0$ there holds $T_p^{-\theta} T_p^{-\beta} = T_p^{-(\theta+\beta)}$.
We define the densely defined and closed operator T_p^θ as the inverse of $T_p^{-\theta}$ with domain $\mathrm{dom}(T_p^{-\theta}) = \mathrm{ran}(T_p^{-\theta})$. We set T_p^0 to the identity on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$, i.e. $T_p^0 = \mathrm{Id}$.

The following corollary provides computation rules for fractional powers of operators and important norm estimates.

Corollary 2.27. Let T_p be an operator as in Definition 2.26. Then for $\gamma, \beta \in \mathbb{R}$ with $\theta \geq \beta$ there holds $\mathrm{dom}(T_p^\theta) \subset \mathrm{dom}(T_p^\beta)$. Moreover, $T_p^\theta T_p^\beta = T_p^\beta T_p^\theta = T_p^{\theta+\beta}$ as operators on $\mathrm{dom}(T_p^\gamma)$, where $\gamma = \max\{\theta, \beta, \theta+\beta\}$. For all $t > 0$, $T_p^\theta \exp(-T_p t) = \exp(-T_p t) T_p^\theta$ on $\mathrm{dom}(T_p^\theta)$.
For $t > 0$ and $\theta \geq 0$ there exists some $C_\theta \in (0, \infty)$ such that

$$\|T_p^\theta \exp(-T_p t)\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \leq C_\theta t^{-\theta} \exp(-\delta t), \tag{2.24}$$

and for $\theta \in (0, 1]$ and $u \in \mathrm{dom}(T_p^\theta)$ we can estimate

$$\|(\exp(-T_p t) - \mathrm{Id})u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} \leq \frac{1}{\theta} C_{1-\theta} t^\theta \|T_p^\theta u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)}. \tag{2.25}$$

C_θ is bounded if θ is contained in a compact subinterval of $(0, \infty)$ and for $\theta \downarrow 0$.

Proof. See the comments after [Hen81, Definition 1.4.1] and [Hen81, Theorem 1.4.3]. $\square$

Definition 2.28. [Hen81, Theorem 1.4.7] Let T_p be an operator as in Definition 2.26. For $\theta \geq 0$, we define the space

$$\begin{aligned} X^\theta_{T_p} &:= \mathrm{dom}(T^\theta_p) && \text{with the graph norm} \\ \|u\|_{X^\theta_{T_p}} &= \|T^\theta_p u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} && \forall u \in X^\theta_{T_p}. \end{aligned}$$

Remark 2.29. For $\theta \geq 0$, the space $\big(X^\theta_{T_p}, \|\cdot\|_{X^\theta_{T_p}}\big)$ in Definition 2.28 is a Banach space [Hen81, Theorem 1.4.8]. Moreover, continuity of $T^{-\theta}_p$ – see Definition 2.26 – implies equivalence of the norm $\|\cdot\|_{X^\theta_{T_p}}$ and the usual graph norm $\|u\|_{X^\theta_{T_p}} = \|T^\theta_p u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} + \|u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)}$, $u \in X^\theta_{T_p}$. By Theorem 2.14, T_p has compact resolvent since $p \in \mathrm{J}_{\mathcal{M}}$. Hence, again [Hen81, Theorem 1.4.8] entails that the embedding $X^\theta_{T_p} \hookrightarrow X^\beta_{T_p}$ is continuous and dense for $0 \leq \beta \leq \theta$. For $0 \leq \beta < \theta$ it is compact.

Remark 2.29 implies the following topological equivalences and embeddings of fractional power spaces.

Corollary 2.30. In the setting of Definition 2.28 let $0 < \beta < \theta < 1$. There holds

$$X^1_{T_p} \hookrightarrow X^\theta_{T_p} \hookrightarrow X^\beta_{T_p} \hookrightarrow X^0_{T_p} = \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega) \tag{2.26}$$

and all embeddings are dense.
Furthermore, the spaces $\big(X^1_{T_p}, \|\cdot\|_{X^1_{T_p}}\big)$ and $\big(\mathrm{dom}(T_p), \|T_p\cdot\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} + \|\cdot\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)}\big)$ are topologically equivalent. Moreover, $X^1_{T_p}$ can be identified with $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ in the following sense:
For each $u \in \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ let $\tilde{u} \in X^1_{T_p}$ be defined by $\tilde{u} := I_p u$. There exist constants $c_1, c_2 > 0$, such that

$$c_1 \|\tilde{u}\|_{X^1_{T_p}} \leq \|u\|_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)} \leq c_2 \|\tilde{u}\|_{X^1_{T_p}} \quad \forall u \in \mathbb{W}^{1,p}_{\Gamma_D}(\Omega).$$

In particular, all of the spaces $\big(\mathrm{dom}(T_p), \|T_p\cdot\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} + \|\cdot\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)}\big)$, $\big(X^1_{T_p}, \|\cdot\|_{X^1_{T_p}}\big)$ and $\big(X^1_{T_p}, \|I^{-1}_p\cdot\|_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)}\big) = \big(\mathbb{W}^{1,p}_{\Gamma_D}(\Omega), \|\cdot\|_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)}\big)$ are topologically equivalent.

Proof. Since T^0_p equals the identity on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$, $\big(X^0_{T_p}, \|\cdot\|_{X^0_{T_p}}\big)$ and $\big(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), \|\cdot\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)}\big)$ coincide. From Remark 2.29 we conclude (2.26) and topological equivalence of $\big(X^1_{T_p}, \|\cdot\|_{X^1_{T_p}}\big)$ and $\big(\mathrm{dom}(T_p), \|T_p\cdot\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} + \|\cdot\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)}\big)$. Remember the definition $T_p = \mathcal{T}_p I^{-1}_p$ and let $u \in \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ be arbitrary. By Theorem 2.14, $\mathcal{T}_p$ is a topological isomorphism since $p \in \mathrm{J}_{\mathcal{M}}$. Hence, we conclude

$$\|\tilde{u}\|_{X^1_{T_p}} = \|T_p I_p u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} = \|\mathcal{T}_p u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} \leq \|\mathcal{T}_p\|_{\mathcal{L}\left(\mathbb{W}^{1,p}_{\Gamma_D}(\Omega), \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \|u\|_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)}.$$

Furthermore, there holds

$$\begin{aligned} \|u\|_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)} &= \|\mathcal{T}^{-1}_p \mathcal{T}_p I^{-1}_p I_p u\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} \\ &\leq \|\mathcal{T}^{-1}_p\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)\right)} \|T_p \tilde{u}\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} \\ &= \|\mathcal{T}^{-1}_p\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)\right)} \|\tilde{u}\|_{X^1_{T_p}}. \end{aligned}$$

Particularly, in $X^1_{T_p}$ the norms $\|\cdot\|_{X^1_{T_p}}$ and $\|I^{-1}_p\cdot\|_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)}$ are equivalent. Hence, all three spaces are topologically equivalent under the identification of $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ with $\mathrm{dom}(T_p) = \mathrm{ran}(I_p) = X^1_{T_p}$ via the bijective function I_p. $\square$

In special cases, the fractional power spaces in Definition 2.28 can be characterized as complex interpolation spaces. We adapt [Mün17a, Remark 2.11] to our setting and add more details:

Remark 2.31. For T_p as in Definition 2.26 and for $z \in \operatorname{int}(\mathbb{C}_+) = \{\zeta \in \mathbb{C} : \operatorname{Re}(\zeta) > 0\}$, the (complex) fractional power T_p^z can be defined by the inverse of T_p^{-z} [Yag09, Chapter 7]. For $\theta \in \mathbb{R}$ and for all $u \in \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ for which the limit exists, one then defines

$$T_p^{i\theta} u := \lim_{z \in \operatorname{int}(\mathbb{C}_+):\ z \to i\theta} T_p^z u.$$

The theory of purely imaginary powers of an operator goes beyond the scope of this work. For further details we refer to [Yag09, Chapter 8]. We apply a result on bounded purely imaginary powers of an operator to characterize the fractional power spaces from Definition 2.28. If the imaginary powers of T_p are bounded, then [CA01, Theorem 11.6.1] entails that for all $\beta \in \mathbb{R}_+$ and $\theta \in (0,1)$, the complex interpolation space $\left[\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), X^{\beta}_{T_p}\right]_\theta$ is topologically equivalent to $X^{\theta\beta}_{T_p}$, i.e.

$$\left[\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), X^{\beta}_{T_p}\right]_\theta \simeq X^{\theta\beta}_{T_p}.$$

In particular, by Corollary 2.30, for $\theta \in (0,1)$ there hold the topological identities

$$\left[\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)\right]_\theta \simeq \left[\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), \operatorname{dom}(T_p)\right]_\theta \simeq \left[\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), X^{1}_{T_p}\right]_\theta \simeq X^{\theta}_{T_p}.$$

All fractional power spaces above are defined with the norm according to Definition 2.28 and $\operatorname{dom}(T_p)$ is considered with the graph norm.

For appropriate choice of p and dimension d, embedding results for some fractional power spaces of diffusion operators A_p are known:

Remark 2.32. Consider the assumptions and the notation in Corollary 2.19 and let A_p be a diffusion operator according to Definition 2.17. Let J be the interval which corresponds to A_p+1 in Corollary 2.19 and assume $p \in \mathrm{J}$. Then by Corollary 2.30, the spaces $\operatorname{dom}(A_p)$ and $X^1_{A_p+1}$ can be identified with $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$.
If $p \in \mathrm{J} \cap [2,\infty)$ then $A_p + 1$ has bounded imaginary powers according to [Aus+14, Theorem 11.5], cf. Remark 2.31. Consequently, for $\theta \in (0,1)$ we obtain

$$\left[\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)\right]_\theta \simeq \left[\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), \operatorname{dom}(A_p)\right]_\theta \simeq \left[\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), X^{1}_{A_p+1}\right]_\theta \simeq X^{\theta}_{A_p+1}.$$

Note that $p \in \mathrm{J}$ is only required to obtain the identification of $\operatorname{dom}(A_p)$ and $X^1_{A_p+1}$ with $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$.
Remember that d denotes the dimension of Ω. For $p \geq 2$ and for all $\theta > \frac{1}{2}(1+\frac{d}{p})$, [TR12, Theorem 3.3] entails that $X^{\theta}_{A_p+1}$ is a subset of $[\mathrm{L}^\infty(\Omega)]^m$. In particular, for $p > 2$ and $d = 2$ there holds $0 < \frac{1}{2}(1+\frac{d}{p}) = \frac{1}{2} + \frac{1}{p} < 1$ so that θ can be chosen in the interval $\left(\frac{1}{2}(1+\frac{d}{p}), 1\right)$.
If $p > 2$ and $d = 2$ and if in addition Ω is regular enough – for example a Lipschitz domain – then by [DER15, Theorem 4.5] there exists some $\theta \in (0,1)$ such that $X^{\theta}_{A_p+1}$ can be embedded into a Hölder space.
The embedding results above will be crucial in order to prove higher regularity of solutions of semi-linear parabolic evolution equations.

2.2.6 Maximal parabolic Sobolev regularity

In this subsection, we introduce the concept of maximal parabolic regularity of an operator, see e.g. [Ama95, Chapter III], [MS15, Definition 2.7] or [Aus+14, Definition 11.2].
Before we define maximal parabolic regularity, we explain what we mean by a solution of an operator equation.

Definition 2.33. [Lun95, Definition 7.0.2] and [Hen81, Chapter 3.2] For $p \in (1, \infty)$ let T_p be an elliptic operator in the sense of Definition 2.11. Consider a time interval $(t_0, T) \subset \mathbb{R}$, an initial state $y_0 \in \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ and a function $g : (t_0, T) \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. We say that $y : (t_0, T) \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ is a mild solution of the evolution equation

$$\frac{d}{dt} y + T_p y = g \quad \text{in } (t_0, T), \quad y(t_0) = y_0,$$

if $y \in \mathrm{L}^1\big((t_0, T); \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\big)$ and if y solves the integral equation

$$y(t) = \exp(-T_p(t - t_0))y_0 + \int_{t_0}^{t} \exp(-T_p(t - s))g(s)\, ds \quad \text{for a.e. } t \in (t_0, T),$$

provided that the semigroup and the integral are well defined. We say that a mild solution $y : [t_0, T] \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ is a (strong) solution if it is continuous, continuously differentiable on (t_0, T) with $y(t) \in \mathrm{dom}(T_p)$ for all $t \in (t_0, T)$ and if $\lim_{t \downarrow t_0} y(t) = y_0$.

In the following definition, we introduce the notion of maximal parabolic regularity of an operator. Amongst others, this concept is a powerful tool for proving that a mild solution of an operator equation is indeed a strong one. Moreover, higher regularity of the solution follows in many instances. We also introduce some notation for different function spaces which is used in [Mün17a; Mün17b].

Definition 2.34. [Aus+14, Definition 11.2] For $p \in (1, \infty)$ let T_p be an elliptic operator in the sense of Definition 2.11. For $q \in (1, \infty)$ and $(t_0, T) \subset \mathbb{R}$ we say that T_p satisfies maximal parabolic $\mathrm{L}^q((t_0, T); \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega))$-regularity if for all $g \in \mathrm{L}^q((t_0, T); \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega))$ there is a unique solution $y \in \mathrm{W}^{1,q}((t_0, T); \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)) \cap \mathrm{L}^q((t_0, T); \mathrm{dom}(T_p))$ of the operator evolution equation

$$\frac{d}{dt} y + T_p y = g \quad \text{in } (t_0, T), \quad y(t_0) = 0.$$

The time derivative is taken in the sense of distributions. For $t_0 = 0$ and $t \in [0, T]$ we introduce the following spaces:

$$\begin{aligned}
Y_{T_p,q} &:= \mathrm{W}^{1,q}((0, T); \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)) \cap \mathrm{L}^q((0, T); \mathrm{dom}(T_p)), \\
Y_{T_p,q,t} &:= \{y \in Y_{T_p,q} : \ y(t) = 0\}, \\
Y^*_{T_p,q,t} &:= \{y \in \mathrm{W}^{1,q}(0, T; [\mathrm{dom}(T_p)]^*) \cap \mathrm{L}^q((0, T); \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)) : \ y(t) = 0\}.
\end{aligned}$$

Remark 2.35. Different to other publications we decided to use a capital Y to define the spaces $Y_{T_p,q}$, $Y_{T_p,q,t}$ and $Y^*_{T_p,q,t}$ in Definition 2.34. This choice is made already with regard to applications in optimization problems in Section 4, where the state variable will always be the solution of an operator equation. The following properties go along with maximal parabolic regularity:

1. Maximal parabolic regularity is independent of $q \in (1, \infty)$ and of the interval (t_0, T) so that we just say that T_p satisfies maximal parabolic regularity on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ [Aus+14, Remark 11.3].

2. If T_p satisfies maximal parabolic regularity on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ then $(\frac{d}{dt}+T_p)^{-1}$ is bounded as an operator from $\mathrm{L}^q\big((0,T);\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\big)$ to $Y_{T_p,q,0}$ [MS15, Proof of Proposition 2.8].

3. For $p \in \mathrm{J} \cap [2,\infty)$ with J from Corollary 2.19 consider a diffusion operator A_p in the sense of Definition 2.17. Then by [Aus+14, Theorem 11.5], $A_p + 1$ satisfies maximal parabolic Sobolev regularity on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. Hence, also A_p satisfies maximal parabolic Sobolev regularity on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. Note again that $p \in \mathrm{J}$ is only needed for the identification of $\mathrm{dom}(T_p)$ and $X^1_{T_p}$ with $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$. Maximal parabolic Sobolev regularity of A_p holds for all $p \in [2,\infty)$.

2.3 Spaces of Banach space valued functions and embeddings

In this section, we collect embedding properties between spaces of functions in time which take there values in different Banach spaces. Those embeddings will play a key role in several convergence as well as regularity proofs later on in this work.
The following embedding properties are due to [Ama95, Theorem 3]:

Lemma 2.36. Consider the notation from Theorem 2.14. For a set of coefficient functions $\mathcal{M}$, let $p \in \mathrm{J}_{\mathcal{M}}$ be given. Assume that T_p is the elliptic operator which corresponds to some $\mathbb{T} \in \mathcal{M}$. Then for any $q \in (1,\infty)$ and $Y_{T_p,q}$ as in Definition 2.34 one has

$$Y_{T_p,q} \hookrightarrow \mathrm{C}^\beta((0,T);(\mathbb{W}^{-1,p}_{\Gamma_\mathrm{D}}(\Omega),\mathrm{dom}(T_p))_{\eta,1}) \hookrightarrow \mathrm{C}^\beta((0,T);[\mathbb{W}^{-1,p}_{\Gamma_\mathrm{D}}(\Omega),\mathrm{dom}(T_p)]_\theta) \text{ and}$$
$$Y_{T_p,q} \hookrightarrow \mathrm{C}([0,T];(\mathbb{W}^{-1,p}_{\Gamma_\mathrm{D}}(\Omega),\mathrm{dom}(T_p))_{\eta,q}) \hookrightarrow \mathrm{C}([0,T];[\mathbb{W}^{-1,p}_{\Gamma_\mathrm{D}}(\Omega),\mathrm{dom}(T_p)]_\theta)$$

for every $0<\theta<\eta<1-1/q$ and $0 \le \beta < 1-1/q-\eta$. $(\cdot,\cdot)_{\eta,1}$ or $(\cdot,\cdot)_{\eta,q}$ respectively denotes real interpolation here.

Proof. By Corollary 2.30, $\mathrm{dom}(T_p) \hookrightarrow \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. The rest follows from [Ama95, Theorem 3]. □

2.4 Hysteresis operators

We are interested in the analysis of non-linear, semi-linear parabolic evolution equations and there solutions y which are functions of $t \in [0,T]$ with values in a Banach space X. In this work, the non-linearity $F[y]$ which enters the right side of the evolution equation is generally non-smooth. In particular, F is usually of the form $f(y,z)$ where $z = \mathcal{W}[Sy]$ is the output of some scalar, rate-independent hysteresis operator $\mathcal{W}$. Here, $S \in X^*$ so that $(Sy)(t) := Sy(t)$, $t \in [0,T]$, is a real-valued function which serves as the input map for $\mathcal{W}$.
This section provides a short introduction to the concept of (scalar) rate-independent hysteresis, cf. [KM17, Section 2.2]. Specifically, we will mostly consider the scalar play and the scalar stop hysteresis operator. In many cases, those can be exchanged by an operator with appropriate properties.
Hysteresis effects are spread over many fields in physics such as ferromagnetism, ferroelectricity or plasticity [May03; BS96; Vis13; MR15; KP12]. Furthermore, they are used to model shape memory effects of certain materials. Viscoplastic behaviour is a particular example here. Hysteresis models are also used to describe thermostats in engineering [Vis13], and certain effects in mathematical biology follow some hysteretic law [GST13; HJ80; Kop06; Pim+12; CGT16]. Mathematically, among the most important hysteresis operators which appear in such models are the relay switch [BS96], the scalar play [BK15], the scalar stop [BR05] or the Prandtl-Ishlinskiĭ operator [Kuh03].

As mentioned, we will work with scalar hysteresis operators throughout. Given a time interval $[0, T]$, scalar hysteresis operators take an admissible time-dependent input function $y : [0, T] \to \mathbb{R}$ together with an initial value $z_0 \in \mathbb{R}$ and return a time-dependent output function $z = z(y, z_0) : [0, T] \to \mathbb{R}$. We will mostly keep z_0 fixed and write $z = z(y)$. Depending on the hysteresis operator at hand and on the smoothness of y, the output map $z : [0, T] \to \mathbb{R}$ has a certain regularity. All scalar rate-independent hysteresis operators have two properties in common [Vis13; BS96]:

Definition 2.37. [KM17]

(Vol) The output function $z(t)$ at time $t \in [0, T]$ may depend not only on the value of the input function $y(t)$ at time t, but on the whole history of y in the interval $[0, t]$. This non-locality in time is often referred to as memory effect, causality or Volterra property: for all y_1, y_2 in the domain of the operator, for all initial values z_0, and any $t \in [0, T]$ it follows that if $y_1 = y_2$ in $[0, t]$, then $[z(y_1, z_0)](t) = [z(y_2, z_0)](t)$ cf. [Vis13, Chapter III].

(RI) The output function z is invariant under time transformations. This means that for any monotone increasing and continuous function $\phi : [0, T] \to [0, T]$ with $\phi(0) = 0$ and $\phi(T) = T$ and for all admissible input functions y there holds

$$[z(y \circ \phi, z_0)](t) = z(y, z_0)(\phi(t)) \qquad \forall t \in [0, T].$$

In [Vis13, Chapter III], the function ϕ is also assumed to be bijective, i.e., the definitions differ in the literature. For our purpose one may consider either definition of admissible time transformations. Invariance under time transformations is also called rate-independence in the literature [MR15, Definition 1.2.1].

Since the most relevant hysteresis operators in this work are the scalar stop and the scalar (generalized) play, we introduce those in the following [BK13; Vis13]:

Definition 2.38. Consider a fixed initial value $z_0 \in \mathbb{R}$ together with an interval $[a, b] \subset \mathbb{R}$. Moreover, let a time interval $[0, T]$ be given. Then we denote by $\mathcal{W} := \mathcal{W}[\cdot, z_0]$ the corresponding scalar stop operator. Since z_0 is fixed we often write $\mathcal{W}[\cdot]$. If the input function $v : [0, T] \to \mathbb{R}$ has a weak derivative, then $z = \mathcal{W}[v]$ is the unique solution of the variational inequality

$$(\dot{z}(t) - \dot{v}(t))(z(t) - \xi) \leq 0 \qquad \text{for } \xi \in [a, b] \text{ and } t \in (0, T), \tag{2.27}$$

$$z(t) \in [a, b] \text{ for } t \in [0, T], \tag{2.28}$$

$$z(0) = z_0. \tag{2.29}$$

Similarly, consider an input function $v : [0, T] \to \mathbb{R}$ which has a weak derivative and any initial value $\sigma_0 \in [v(0) - b, v(0) - a]$. Then by [Vis13, Chapter III.2], the operator $\mathcal{P}[v, \sigma_0]$ which assigns to v and $\sigma_0 \in [v(0) - b, v(0) - a]$ the unique solution σ of

$$\dot{\sigma}(t)(v(t) - \sigma(t) - \xi) \geq 0 \qquad \text{for } \xi \in [a, b] \text{ and } t \in (0, T), \tag{2.30}$$

$$v(t) - \sigma(t) \in [a, b] \text{ for } t \in [0, T], \tag{2.31}$$

$$\sigma(0) = \sigma_0, \tag{2.32}$$

defines a scalar play operator. We denote by $\mathcal{P}$ the play operator defined by $\mathcal{W}$ [Vis13, Part 1 Chapter III Proposition 3.3], i.e. $\mathcal{P}$ is determined by

$$\mathcal{P} + \mathcal{W} = \mathrm{Id}. \tag{2.33}$$

More precisely, (2.33) has to be understood as

$$\mathcal{P}[v, v(0) - z_0](t) + \mathcal{W}[v, z_0](t) = v(t) \quad \text{for } t \in [0, T].$$

If $v(0) = v_0 \in \mathbb{R}$ is fixed and known from the context, we often write $\mathcal{P}[v] = \mathcal{P}[v, v_0 - z_0]$.

Remark 2.39. We make a couple of remarks with respect to Definition 2.38:

1. The conditions (2.27)–(2.29) are equivalent to the differential inclusion

$$\begin{aligned} &\dot{v}(t) \in \dot{z}(t) + \partial I_{[a,b]}(z(t)) \text{ for } t \in (0, T), \\ &z(0) = z_0, \end{aligned}$$

   where $I_{[a,b]}(v) = 0$ if $v \in [a, b]$ and $I_{[a,b]}(v) = \infty$ if $v \notin [a, b]$ [Vis13, Chapter III.3]. Similarly, conditions (2.30)–(2.32) are equivalent to the differential inclusion

$$\begin{aligned} &\dot{\sigma}(t) \in \partial I_{[a,b]}(v(t) - \sigma(t)) \text{ for } t \in (0, T), \\ &\sigma(0) = \sigma_0. \end{aligned}$$

   There are further possibilities to represent hysteresis operators. We will mostly work with variational inequalities but refer to [Mie05, Chapter 2] for further equivalent formulations such as dual variational inequalities, a subdifferential approach, an energetic formulation, and a representation in form of a sweeping-process, see also [MR15].

2. $\mathcal{W}$ is a so called linear stop operator and $\mathcal{P}$ is a linear play operator. A generalization to non-linear plays and stops is introduced e.g. in [Vis13, Chapter III].

3. In Section 4, we will analyze an optimal control problem of a hysteresis-reaction-diffusion system where the hysteresis is given by a scalar stop operator. This is the reason why we fix the initial value z_0 of $\mathcal{W}$ in Definition 2.38 and introduce the corresponding play operator $\mathcal{P}$ via (2.33) rather than defining the play and the stop operator separately.

The following regularity properties for $\mathcal{W}$ and $\mathcal{P}$ are needed. We refer to [Vis13, Part 1, Chapter III] and [BK15] for a deeper analysis, see also [Mün17a, Subsection 2.4 and Subsection 4.2] or [Mün17b, Lemma 2.9].

Theorem 2.40 (Stop and Play)**.** *The stop operator $\mathcal{W}$ and the play operator $\mathcal{P}$ from Definition 2.38 are Lipschitz continuous as mappings on $\mathrm{C}[0, T]$. In particular,*

$$|\mathcal{W}[v_1](t) - \mathcal{W}[v_2](t)| \leq 2 \sup_{0 \leq \tau \leq t} |v_1(\tau) - v_2(\tau)|, \tag{2.34}$$

$$|\mathcal{W}[v](t)| \leq 2 \sup_{0 \leq \tau \leq t} |v(\tau)| + |z_0|, \tag{2.35}$$

$$|\mathcal{P}[v_1](t) - \mathcal{P}[v_2](t)| \leq \max\left\{ \sup_{0 \leq \tau \leq t} |v_1(\tau) - v_2(\tau)|, |v_1(0) - v_2(0)| \right\} \quad \textit{and} \tag{2.36}$$

$$|\mathcal{P}[v](t)| \leq \sup_{0 \leq \tau \leq t} |v(\tau)| + |z_0| \tag{2.37}$$

for all $v, v_1, v_2 \in \mathrm{C}[0, T]$ and $t \in [0, T]$. For $q \in [1, \infty)$, $\mathcal{W}$ and $\mathcal{P}$ are bounded and weakly continuous on $\mathrm{W}^{1,q}(0, T)$. As mappings from $\mathrm{C}[0, T]$ into $\mathrm{L}^q(0, T)$ they are Hadamard directionally differentiable, see Definition 3.6 below.

Proof. The estimates (2.34)–(2.37) as well as the boundedness in $\mathrm{W}^{1,q}(0,T)$ are shown in [Vis13, Part 1, Chapter III] for example. For $r > 0$ let $(v, \sigma_0) \mapsto \mathcal{P}_r[v, \sigma_0]$ denote the symmetric play operator which is represented by (2.30)–(2.32) if the interval $[a, b]$ is of the form $[-r, r]$. For $\mathcal{P}_r$, Hadamard directional differentiability from $\mathrm{C}[0,T] \times \mathbb{R}$ to $\mathrm{L}^q(0,T)$ is shown in [BK15]. Now $\mathcal{P}$ can be constructed as follows: We set $r = \frac{b-a}{2}$ and define the affine linear transformation $\mathcal{T}_1 : [-r, r] \to [a, b]$, $\mathcal{T}_1 : x \mapsto x - \frac{b+a}{2}$, as well as the mapping $\mathcal{T}_2 : \mathrm{C}[0,T] \to \mathbb{R}$, $\mathcal{T}_1 : v \to v(0) - z_0$. Both maps are continuously differentiable. Then for $v \in \mathrm{C}[0,T]$ there holds

$$\mathcal{P}[v] = \mathcal{P}_r[\mathcal{T}_1(v), \mathcal{T}_2(v)].$$

The chain rule yields that $\mathcal{P}$ is Hadamard directionally differentiable from $\mathrm{C}[0,T] \times \mathbb{R}$ to $\mathrm{L}^q(0,T)$, see Lemma 3.7. By (2.33) and again the chain rule the same holds for $\mathcal{W}$. $\square$

Remark 2.41. Note that we have to add $|z_0|$ in (2.35) and (2.37) because we fixed the initial value of $\mathcal{W}$, i.e. by (2.28) we have $\mathcal{W}[v](0) = z_0$ for any $v \in \mathrm{C}[0,T]$.

3 Semilinear parabolic systems with hysteresis and Bochner-Lebesgue integrable non-linearity

In this chapter, we analyze a semi-linear parabolic evolution equation of the form

$$\begin{aligned} \frac{d}{dt} y(t) + (T_p y)(t) &= (F[y])(t) + u(t) \quad \text{ in } X = \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega) \text{ for } t > 0, \\ y(0) &= y_0 \in X. \end{aligned} \tag{3.1}$$

In particular, the non-linearity $F = f(y, \mathcal{W}[Sy])$ contains a scalar stop operator. The corresponding Cauchy problem with a diffusion operator A_p instead of T_p and with zero initial value has been studied in [Mün17a]. We extend the results to apply to Cauchy problems with a general elliptic operator and non-trivial initial value y_0. Section 3.1 contains the main assumption of the chapter. In Section 3.2 we show well-posedness of (3.1). Moreover, we prove that the solution operator $G : (y_0, u) \mapsto y$ is linearly bounded and locally Lipschitz continuous on appropriate function spaces. In Section 3.3 we extend the results of [Mün17a] by showing that G is Hadamard directionally differentiable in y_0 and u rather than in u only.

3.1 Main assumption and notation

This section contains the main assumption of the chapter. We also introduce some short notation for several spaces and functions.

Assumption 3.1. Cf. [Mün17a, Assumption 2.16] We always suppose that Assumption 2.2 and Assumption 2.6 hold. All spaces are supposed to consist of real-valued functions. Consider the setting and notation from Theorem 2.14. We assume:

(A0) $\Omega \subset \mathbb{R}^d$ for some $d \geq 2$.

(A1) For a set of coefficient functions $\mathcal{M} \subset \mathrm{L}^\infty(\Omega; \mathcal{L}(\mathbb{R}^m \times \mathbb{R}^{m \times d}))$ there holds $p \in \mathrm{J}_{\mathcal{M}} \cap [2, \infty)$ and $2 \geq p\left(1 - \frac{1}{d}\right)$. Moreover, T_p is the elliptic operator which corresponds to a matrix $\mathbb{T} \in \mathcal{M}$.

(A2) For some $w \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega) \simeq [\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)]^*$, $w \neq 0$, the operator $S \in [\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)]^*$ is defined by

$$Sy := \langle y, w \rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} \ \forall y \in \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega).$$

Note that S belongs to $[X^\theta_{T_p}]^*$ for all $\theta \geq 0$ because of the embedding $X^\theta_{T_p} \hookrightarrow \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$.

(A3) We will need a fractional power space with exponent strictly smaller than one. Therefore, we assume that for some $\alpha \in (0,1)$ the function $f : X^{\alpha}_{T_p} \times \mathbb{R} \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ is locally Lipschitz continuous with respect to the $X^{\alpha}_{T_p}$-norm. This means that for every $y_0 \in X^{\alpha}_{T_p}$ there is a constant $L(y_0) > 0$ and a neighborhood

$$V(y_0) = \left\{ y \in X^{\alpha}_{T_p} : \|y - y_0\|_{X^{\alpha}_{T_p}} \leq \delta \right\}$$

of y_0 such that

$$\|f(y_1, x_1) - f(y_2, x_2)\|_X \leq L(y_0) \left(\|y_1 - y_2\|_{X^{\alpha}_{T_p}} + |x_1 - x_2| \right)$$

for every $y_1, y_2 \in V(y_0)$ and all $x_1, x_2 \in \mathbb{R}$. Moreover, f is assumed to have at most linear growth along solutions, i.e.

$$\|f(y, x)\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} \leq M \left(1 + \|y\|_{X^{\alpha}_{T_p}} + |x| \right)$$

for some constant $M > 0$.

In the setting of Assumption 3.1 we collect the notation for the rest of the chapter:

(N1) For the particular p from Assumption 3.1 we set

$$X := \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$$

with $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ from Definition 2.8. We sometimes identify elements $v \in X^*$ with their representation in $\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$, i.e.

$$\langle v, y \rangle_X = \langle y, v \rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} \quad \forall y \in X.$$

(N2) The operators T_p and the spaces $X^{\theta}_{T_p} = \operatorname{dom}(T^{\theta}_p)$ are defined as in Definition 2.11 and Definition 2.28.

(N3) The spaces $Y_{T_p,q}$ and $Y_{T_p,q,t}$ are defined as in Definition 2.34.

(N4) $\mathcal{W}$ is a scalar stop operator as defined in Definition 2.38 for some prescribed initial value $z_0 \in [a, b]$.

(N5) We abbreviate $J_T = (0, T)$.

3.2 Well-posedness of the evolution equation

We recap equation (3.1) from the beginning of the chapter which is

$$\begin{aligned} \frac{d}{dt} y(t) + (T_p y)(t) &= (F[y])(t) + u(t) \quad \text{ in } X = \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega) \text{ for } t > 0, \\ y(0) &= y_0 \in X, \end{aligned}$$

where $(F[y])(t) := f(y(t), \mathcal{W}[Sy](t))$.

In this section, we show well-posedness of the problem. In particular, we prove that the solution operator $G : (y_0, u) \mapsto y$ is linearly bounded and locally Lipschitz continuous on $X^{\beta}_{T_p} \times \mathrm{L}^q(J_T; X)$ for $\beta \in [\alpha, 1)$ and $q \in \left(\frac{1}{1-\alpha}, \infty \right]$. The first aim is to show that for every $u \in \mathrm{L}^q(J_T; X)$ and for

initial values $y_0 \in X^\alpha_{T_p}$ problem (3.1) has a unique mild solution $y \in \mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ in the sense of Definition 2.33, where α is fixed by (A3). In particular, this means that $(F[y]) + u$ is contained in $\mathrm{L}^1(J_T; X)$. If T_p satisfies maximal parabolic regularity and if $\beta \in (\alpha, 1]$, we prove that the unique mild solution even belongs to $Y_{T_p,s}$ where s is arbitrary in the interval $\left(1, \frac{1}{1-\beta}\right) \cap (1, q]$ if $\beta \in (\alpha, 1)$ and in the interval $(1, q] \cap (1, \infty)$ if $\beta = 1$. Note that the latter intersection makes sense since $q = \infty$ is allowed.
The following Theorem 3.2 is a generalization of [Mün17a, Theorem 3.1] from diffusion operators A_p to general elliptic operators T_p which do not necessarily satisfy maximal parabolic Sobolev regularity. Moreover, we allow for initial values $y_0 \in X^\beta_{T_p}$ different from zero. We adapt the proof of Theorem 3.2 to apply to this generalized setting.

Theorem 3.2. *Let Assumption 3.1 hold. Then for all $y_0 \in X^\alpha_{T_p}$ and $u \in \mathrm{L}^q(J_T; X)$ with $q \in \left(\frac{1}{1-\alpha}, \infty\right]$ problem (3.1) has a unique mild solution $y = y(y_0, u)$ in $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$. Note that $X^\alpha_{T_p} \subset X = X^0_{T_p}$ since $\alpha \in (0, 1)$. The solution mapping*

$$G : (y_0, u) \mapsto y(y_0, u),\ \ X^\alpha_{T_p} \times \mathrm{L}^q(J_T; X) \to \mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$$

is locally Lipschitz continuous. G is linearly bounded with values in $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$, i.e. for some $C = C(T) > 0$ there holds

$$\|G(y_0, u)\|_{\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})} \le C(T)(1 + \|y_0\|_{X^\alpha_{T_p}} + \|u\|_{\mathrm{L}^q(J_T;X)}) \tag{3.2}$$

for all $y_0 \in X^\alpha_{T_p}$ and $u \in \mathrm{L}^q(J_T; X)$ and $C(T)$ is independent of y_0 and u. Suppose additionally that T_p satisfies maximal parabolic regularity on X and $y_0 \in X^\beta_{T_p}$ where $\beta \in [\alpha, 1]$. Then all statements concerning G remain valid with $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ replaced by $Y_{T_p,s}$ and with $\|y_0\|_{X^\alpha_{T_p}}$ replaced by $\|y_0\|_{X^\beta_{T_p}}$ in (3.2). Here, s is arbitrary in the interval $\left(1, \frac{1}{1-\beta}\right) \cap (1, q]$ if $\beta \in [\alpha, 1)$ and in the interval $(1, q] \cap (1, \infty)$ if $\beta = 1$.

Proof. The theorem is shown with help of a fixed point argument. This technique is quite common in the context of non-linear operator evolution equations, see e.g. [Lun95, Theorem 7.1.3], [Hen81, Chapter 3] or [Paz83, Section 6.3]. Several of the estimates below appeared in [MS15, Appendix A] in a similar form. We extend the standard results in [Lun95; Hen81; Paz83; MS15] by allowing for non-linearities which are only locally Lipschitz continuous and not Lipschitz continuous on bounded sets. Moreover, the hysteresis operator, which appears in the non-linearity F, acts non-local in time. This fact requires additional work in several steps. We prove the theorem directly for $u \in \mathrm{L}^q(J_T; X)$ as it is done in [Lun95, Theorem 7.1.3]. In [MS15, Appendix A] the corresponding statement is first shown for smooth right hand sides and afterwards extended by a density argument.
We denote by $c > 0$ a generic constant which is adapted during the proof. The following observation will be used several times: For $\zeta > -1$ there holds

$$\int_0^t (t-s)^\zeta \, ds = \frac{t^{1+\zeta}}{1+\zeta}. \tag{3.3}$$

The proof is divided into five steps. In Steps I–IV we assume w.l.o.g. that $\beta = \alpha$.

(I) Local existence:
First we show the existence of local mild solutions of problem (3.1) for fixed $y_0 \in X^\alpha_{T_p}$ and $u \in \mathrm{L}^q(J_T; X)$.

The function $v_u(t) := \int\limits_0^t e^{-T_p(t-s)}u(s)\,ds$ is contained in $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ for arbitrary $T > 0$.
Moreover, since $-q'\alpha > -1 \Leftrightarrow q \in \left(\frac{1}{1-\alpha}, \infty\right]$, we can apply (2.24) and (3.3) to estimate

$$\begin{aligned}\|v_u\|_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})} &\leq \left(\int_0^T \|e^{-T_p(t-s)}\|^{q'}_{\mathcal{L}(X,X^\alpha_{T_p})}\,ds\right)^{1/q'} \|u\|_{\mathrm{L}^q(J_T;X)} \\ &\leq c \max_{t\in\overline{J_T}}\{e^{-\delta t}\}T^{1/q'-\alpha}\|u\|_{\mathrm{L}^q(J_T;X)} < \infty.\end{aligned} \tag{3.4}$$

The dependence of (3.4) on $T^{1/q'-\alpha}$ will be crucial for the contraction argument below. Let $r > 0$ be small enough so that f is Lipschitz continuous in $\overline{B_{X^\alpha_{T_p}}(y_0, r)} \times \mathbb{R}$ with a constant $L(y_0) > 0$. We denote by $\overline{B_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})}(y_0, r)}$ the ball in $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ with radius $r > 0$ around the constant function y_0, i.e.

$$\overline{B_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})}(y_0, r)} = \{y \in \mathrm{C}(\overline{J_T}; X^\alpha_{T_p}) :\ \|y(t) - y_0\|_{X^\alpha_{T_p}} \leq r\ \forall t \in \overline{J_T}\}.$$

Moreover, we will identify y_0 with the constant function $y \equiv y_0$ in $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ several times.
We apply (A2) on S, (A3) on f and Lipschitz continuity of $\mathcal{W}$ from Theorem 2.40 to estimate

$$\begin{aligned}&\|(F[y_1])(t) - (F[y_2])(t)\|_X = \|(f(y_1(t), \mathcal{W}[Sy_1](t)) - f(y_2(t), \mathcal{W}[Sy_2](t))\|_X \\ &\leq L(y_0)\Big(\|y_1(t) - y_2(t)\|_{X^\alpha_{T_p}} + 2\|S\|_{[X^\alpha_{T_p}]^*} \sup_{0\leq\tau\leq t}\|y_1(\tau) - y_2(\tau)\|_{X^\alpha_{T_p}}\Big) \\ &\leq c \sup_{0\leq\tau\leq t}\|y_1(\tau) - y_2(\tau)\|_{X^\alpha_{T_p}}\end{aligned} \tag{3.5}$$

for all $y_1, y_2 \in \overline{B_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})}(y_0, r)}$ and $t \in \overline{J_T}$. Consider the mapping

$$\Phi_{y_0,u}(y)(t) := e^{-T_p t}y_0 + \int\limits_0^t e^{-T_p(t-s)}\left[f\left(y(s), \mathcal{W}[Sy](s)\right) + u(s)\right]\,ds.$$

That $\Phi_{y_0,u}$ is well defined on $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ can be shown as in [MS15, Appendix A (ii)]. We prove that $\Phi_{y_0,u}$ has a unique fixed point if $T > 0$ is small.
For $y_1, y_2 \in \overline{B_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})}(y_0, r)}$ we have by (2.24), (3.3) and (3.5) that

$$\begin{aligned}\|\Phi_{y_0,u}(y_1) - \Phi_{y_0,u}(y_2)\|_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})} &\leq \int_0^T \|e^{-T_p(t-s)}\|_{\mathcal{L}(X,X^\alpha_{T_p})}\,ds\|F[y_1] - F[y_2]\|_{\mathrm{C}(\overline{J_T};X)} \\ &\leq cT^{1-\alpha}\|y_1 - y_2\|_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})} < \frac{1}{2}\|y_1 - y_2\|_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})}\end{aligned}$$

for T small enough. Consequently, in this case $\Phi_{y_0,u}$ is a $\frac{1}{2}$-contraction.
Moreover, $(e^{-T_p t} - \mathrm{Id})T^\alpha_p y_0 \to 0$ with $t \to 0$ since $\{e^{-T_p t}\}_{t\geq 0}$ is a C_0-semigroup, according to Definition 2.21. Together with (3.4) we obtain for $y \in \overline{B_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})}(y_0, r)}$ that

$$\begin{aligned}&\|\Phi_{y_0,u}(y)(t) - y_0\|_{X^\alpha_{T_p}} \leq \|\Phi_{y_0,u}(y)(t) - \Phi_{y_0,u}(y_0)(t)\|_{X^\alpha_{T_p}} + \|\Phi_{y_0,u}(y_0)(t) - y_0\|_{X^\alpha_{T_p}} \\ &\leq \frac{r}{2} + \left(\int_0^T \|e^{-T_p(t-s)}\|^{q'}_{\mathcal{L}(X,X^\alpha_{T_p})}\,ds\right)^{1/q'} \|F[y_0] + u\|_{\mathrm{L}^q(J_T;X)} + \|(e^{-T_p t} - \mathrm{Id})y_0\|_{X^\alpha_{T_p}} \\ &\leq \frac{r}{2} + c\left(T^{1/q'-\alpha}\|f\left(y_0, z_0\right) + u\|_{\mathrm{L}^q(J_T;X)} + \|(e^{-T_p t} - \mathrm{Id})T^\alpha_p y_0\|_X\right) \leq r\end{aligned}$$

if T is small enough. Note that $\mathcal{W}[Sy_0] = z_0$ since y_0 is a constant function.
Because $\overline{B_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})}(y_0,r)}$ is invariant under $\Phi_{y_0,u}$ and since $\overline{B_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})}(y_0,r)}$ is a closed subset of $\mathrm{C}(\overline{J_T};X^\alpha_{T_p})$, Banach's fixed point theorem yields a unique fixed point y of $\Phi_{y_0,u}$ in $\overline{B_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})}(y_0,r)}$. This fixed point defines a (local) mild solution of problem (3.1) in $\overline{J_T}$ in the sense of Definition 2.33.
(II) Global existence:
We show that global mild solutions for problem (3.1) exist and prove the statement about linear boundedness of the mapping G along solutions. This part requires some cautiousness because the hysteresis operator is non-local in time.
Remember that the local mild solution y of (3.1) has the form

$$y(t) = e^{-T_p t} y_0 + \int_0^t e^{-T_p(t-s)} \left[f\left(y(s), \mathcal{W}[Sy](s)\right) + u(s)\right] ds.$$

With (A2) and Theorem 2.40 we estimate for all $t \in \overline{J_T}$

$$|\mathcal{W}[Sy](t)| \leq 2\|S\|_{[X^\alpha_{T_p}]^*} \sup_{0\leq\tau\leq t} \|y(\tau)\|_{X^\alpha_{T_p}} + |z_0|.$$

Moreover, by (2.24) there holds

$$\|T_p^\alpha \exp(-T_p t)\|_{\mathcal{L}(X)} \leq C_\alpha t^{-\alpha} \exp(-\delta t).$$

Equation (3.3) yields

$$\left(\int_0^t (t-s)^{-\alpha q'}\, ds\right)^{1/q'} = \left(\frac{t^{1-\alpha q'}}{1-\alpha q'}\right)^{1/q'} = \frac{t^{1/q'-\alpha}}{(1-\alpha q')^{1/q'-\alpha}}.$$

That is, the norm of $y(t)$ for $t \in \overline{J_T}$, can be bounded from above by

$$\begin{aligned}
&\|y(t)\|_{X^\alpha_{T_p}} \\
&\leq c\left[\|y_0\|_{X^\alpha_{T_p}} + \int_0^t (t-s)^{-\alpha}\Big(1 + 3\sup_{0\leq\tau\leq s}\|y(\tau)\|_{X^\alpha_{T_p}} + |z_0|\Big)\, ds + t^{1/q'-\alpha}\|u\|_{\mathrm{L}^q(J_T;X)}\right] \\
&\leq c_0 \int_0^t (t-s)^{-\alpha} \sup_{0\leq\tau\leq s}\|y(\tau)\|_{X^\alpha_{T_p}}\, ds + c_1(T)[1 + \|y_0\|_{X^\alpha_{T_p}} + \|u\|_{\mathrm{L}^q(J_T;X)}],
\end{aligned} \tag{3.6}$$

for some constants $c_0, c_1(T) > 0$. The solution of (3.1) which exists on $[0,T[$ can be continued to a larger time interval if $\|y(t)\|_{X^\alpha_{T_p}}$ remains bounded with $t \uparrow T$. This argument has been used in [Paz83, Theorem 6.3.3]. Boundedness of $\|y(t)\|_{X^\alpha_{T_p}}$ with $t \uparrow T$ follows if

$$\sup_{0\leq\tau<T} \|y(\tau)\|_{X^\alpha_{T_p}} \leq C(T) \tag{3.7}$$

for some $C(T) > 0$. We want to use a Gronwall argument in order to show (3.7). With $y \in \mathrm{C}(\overline{J_T};X^\alpha_{T_p})$ it follows that the function $t \mapsto \sup_{0\leq\tau<t}\|y(\tau)\|_{X^\alpha_{T_p}}$ is continuous on $[0,T[$. In order to show (3.7) we apply that for $t \in J_T$ the function

$$g : \tau \mapsto \int_0^\tau (\tau-s)^{-\alpha} \sup_{0\leq\tau'\leq s} \|y(\tau')\|_{X^\alpha_{T_p}}\, ds,\ \tau \in \overline{J_t}$$

is monotone increasing.
Indeed, let $t_0 \in \overline{J_t}$ be given and suppose $\varepsilon > 0$ is arbitrary but small enough so that $t_0 + \varepsilon \in \overline{J_t}$. Then shifting the integration interval $(\varepsilon, t_0 + \varepsilon)$ to $(0, t_0)$ we obtain

$$
\begin{aligned}
&g(t_0 + \varepsilon) - g(t_0) \\
&= \int_0^{t_0+\varepsilon} (t_0 + \varepsilon - s)^{-\alpha} \sup_{0 \le \tau' \le s} \|y(\tau')\|_{X^\alpha_{T_p}} \, ds - \int_0^{t_0} (t_0 - s)^{-\alpha} \sup_{0 \le \tau' \le s} \|y(\tau')\|_{X^\alpha_{T_p}} \\
&= \int_0^{t_0} (t_0 - s)^{-\alpha} \left(\sup_{0 \le \tau' \le s+\varepsilon} \|y(\tau')\|_{X^\alpha_{T_p}} - \sup_{0 \le \tau' \le s} \|y(\tau')\|_{X^\alpha_{T_p}} \right) ds \\
&+ \int_0^{\varepsilon} (t_0 + \varepsilon - s)^{-\alpha} \sup_{0 \le \tau' \le s} \|y(\tau')\|_{X^\alpha_{T_p}} \, ds \ge 0.
\end{aligned}
$$

We make use of the fact that g is monotone increasing to take the supremum over $\tau \in [0, t]$ in (3.6) on both sides. Then the right hand side remains the same so that

$$
\sup_{0 \le \tau \le t} \|y(\tau)\|_{X^\alpha_{T_p}} \le c_0 \int_0^t (t - s)^{-\alpha} \sup_{0 \le \tau \le s} \|y(\tau)\|_{X^\alpha_{T_p}} \, ds + c_1(T)[1 + \|y_0\|_{X^\alpha_{T_p}} + \|u\|_{\mathrm{L}^q(J_T;X)}].
$$

Finally, Gronwall's lemma [Paz83, Lemma 6.7] shows (3.7) since

$$
\sup_{0 \le \tau \le t} \|y(\tau)\|_{X^\alpha_{T_p}} \le C(T)(1 + \|y_0\|_{X^\alpha_{T_p}} + \|u\|_{\mathrm{L}^q(J_T;X)})
$$

holds for some constant $C(T) > 0$ and for all $t \in \overline{J_T}$.
(III) Local Lipschitz continuity:
We apply techniques similar to Step II in order to show local Lipschitz continuity of the solution mapping G. While the linear growth condition on f holds globally, the non-linearity is only locally Lipschitz continuous. In [MS15] it is Lipschitz continuous on bounded sets. Hence, we require some additional arguments. First we prove that the function $(y(\cdot), v) \mapsto f(y(\cdot), v)$ is locally Lipschitz continuous from $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p}) \times \mathbb{R}$ to $\mathrm{C}(\overline{J_T}; X)$ with respect to the $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$-norm. To this aim, let $y \in \mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ be given. Then because $\overline{J_T} \subset \mathbb{R}$ is compact and since y is continuous, the set $y(\overline{J_T})$ is compact in $X^\alpha_{T_p}$. Moreover, $\overline{J_T}$ is separable. That is, again because y is continuous we conclude that $y(\overline{J_T})$ equipped with the subspace topology in $X^\alpha_{T_p}$ is separable as well. Hence, we can choose a dense subset $\{y_i\}_{i\in\mathbb{N}} \subset X^\alpha_{T_p} \cap y(\overline{J_T})$ of $y(\overline{J_T})$. By (A3), the function $(\tilde{y}, v) \mapsto f(\tilde{y}, v)$ is locally Lipschitz continuous from $X^\alpha_{T_p} \times \mathbb{R}$ to X. Consequently, there exist constants $\varepsilon(y_i) > 0$ such that $(\tilde{y}, v) \mapsto f(\tilde{y}, v)$ is Lipschitz continuous on $B_{X^\alpha_{T_p}}(y_i, \varepsilon(y_i)) \times \mathbb{R}$ with some modulus $L(y_i) > 0$ for all $i \in \mathbb{N}$. Remember that $\{y_i\}_{i\in\mathbb{N}}$ is dense in $y(\overline{J_T})$, so that $y(\overline{J_T}) \subset \cup_{i\in\mathbb{N}} B_{X^\alpha_{T_p}}(y_i, \varepsilon(y_i))$. Since $y(\overline{J_T})$ is compact in $X^\alpha_{T_p}$, we can choose a finite subcover $\cup_{i=1}^k B_{X^\alpha_{T_p}}(y_i, \varepsilon(y_i))$ which still contains $y(\overline{J_T})$. Notice that the function $(\tilde{y}, v) \mapsto f(\tilde{y}, v)$ is Lipschitz continuous on $\cup_{i=1}^k B_{X^\alpha_{T_p}}(y_i, \varepsilon(y_i)) \times \mathbb{R}$ with modulus $\tilde{L}(y) := \max_{i\in\{1,\dots,k\}} L(y_i)$. Consider the open set

$$
V_y := \{\tilde{y} \in \mathrm{C}(\overline{J_T}; X^\alpha_{T_p}) : \ \tilde{y}(t) \in \cup_{i=1}^k B_{X^\alpha_{T_p}}(y_i, \varepsilon(y_i)) \ \forall t \in \overline{J_T}\}
$$

which defines a neighborhood of y in $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$. Then $(\tilde{y}(\cdot), v) \mapsto f(\tilde{y}(\cdot), v)$ is Lipschitz continuous from $V_y \times \mathbb{R} \subset \mathrm{C}(\overline{J_T}; X^\alpha_{T_p}) \times \mathbb{R}$ to $\mathrm{C}(\overline{J_T}; X)$. Hence, the function

$(y(\cdot), v) \mapsto f(y(\cdot), v)$ is locally Lipschitz continuous from $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p}) \times \mathbb{R}$ to $\mathrm{C}(\overline{J_T}; X)$ with respect to the $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$-norm.
Moreover, we obtain a pointwise estimate of the form

$$\|f(y_1(t), v_1) - f(y_2(t), v_2)\|_X \le \tilde{L}(y)(\|y_1(\tau) - y_2(\tau)\|_{X^\alpha_{T_p}} + |v_1 - v_2|) \tag{3.8}$$

for all $y_1, y_2 \in V_y$, $v_1, v_2 \in \mathbb{R}$ and $t \in \overline{J_T}$.
We exploit Lipschitz continuity of $\mathcal{W}$, see Theorem 2.40, and (A2) to obtain that the function $y \mapsto F[y] = f(y, \mathcal{W}[Sy])$ is locally Lipschitz continuous from $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ to $\mathrm{C}(\overline{J_T}; X)$.
Moreover, (3.8) implies that for $y \in \mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ and the neighborhood V_y of y there exists a constant $L(y) > 0$ such that the pointwise estimate

$$\|F(y_1)(t) - F(y_2)(t)\|_X \le L(y) \sup_{0 \le \tau \le t} \|y_1(\tau) - y_2(\tau)\|_{X^\alpha_{T_p}} \tag{3.9}$$

holds for all $y_1, y_2 \in V_y$ and $t \in \overline{J_T}$. Note that we slightly overload the notation by using $L(y)$ as in (A3).
We continue by proving local Lipschitz continuity of G. To this aim, we denote by $y = G(y_0, u)$ the solution of problem (3.1) corresponding to y_0 and u. For this fixed y we choose $r > 0$ small enough so that F is Lipschitz continuous in $\overline{B_{\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})}(y, r)}$ with modulus $L(y)$.
For $R > 0$ to be chosen let $\tilde{y}_0 \in \overline{B_{X^\alpha_{T_p}}(y_0, R)}$ and $\tilde{u} \in \overline{B_{\mathrm{L}^q(J_T; X)}(u, R)}$ be arbitrary. There holds $y(0) = y_0$ and $G(\tilde{y}_0, \tilde{u})(0) = \tilde{y}_0$. Both, y and $G(\tilde{y}_0, \tilde{u})$ are continuous functions. Hence, we can find some $\tau > 0$ such that

$$\begin{aligned}
&\sup_{0 \le t < \tau} \|y(t) - G(\tilde{y}_0, \tilde{u})(t)\|_{X^\alpha_{T_p}} \\
&\qquad \le \sup_{0 \le t < \tau} (\|y(t) - y_0\|_{X^\alpha_{T_p}} + \|y_0 - \tilde{y}_0\|_{X^\alpha_{T_p}} + \|\tilde{y}_0 - G(\tilde{y}_0, \tilde{u})(t)\|_{X^\alpha_{T_p}}) < r
\end{aligned}$$

if $R > 0$ is small enough. Hence, we can apply the pointwise Lipschitz estimate (3.9) to $\|F[y] - F[G(\tilde{y}_0, \tilde{u})]\|_X$. Together with (2.24) and (3.3) we can then estimate

$$\begin{aligned}
&\|y(t) - G(\tilde{y}_0, \tilde{u})(t)\|_{X^\alpha_{T_p}} \\
&\le c_1 \|y_0 - \tilde{y}_0\|_{X^\alpha_{T_p}} + c_2(T, y) \int_0^t (t - s)^{-\alpha} [\|(F[y])(s) - (F[G(\tilde{y}_0, \tilde{u})])(s)\|_X \\
&+ \|u(s) - \tilde{u}(s)\|_X] \, ds \\
&\le L(y) c_2(T, y) \int_0^t (t - s)^{-\alpha} \sup_{0 \le \tau \le s} \|y(\tau) - G(\tilde{y}_0, \tilde{u})(\tau)\|_{X^\alpha_{T_p}} \, ds \\
&+ c_3(T, y) \left[\|y_0 - \tilde{y}_0\|_{X^\alpha_{T_p}} + \|u - \tilde{u}\|_{\mathrm{L}^q(J_T; X)} \right]
\end{aligned}$$

for $t \in [0, \tau)$ and constants $c_1, c_2(T, y), c_3(T, y) > 0$.
Similar as in Step II, taking the supremum over $\tau \in [0, t]$ on both sides leaves the right hand side unchanged. Hence, Gronwall's lemma yields some constant $C(T, y) > 0$ such that

$$\sup_{0 \le t \le \tau} \|y(t) - G(\tilde{y}_0, \tilde{u})(t)\|_{X^\alpha_{T_p}} \le C(T, y) \left[\|y_0 - \tilde{y}_0\|_{X^\alpha_{T_p}} + \|u - \tilde{u}\|_{\mathrm{L}^q(J_T; X)} \right] < r$$

if R is chosen small enough, since $\tilde{y}_0 \in \overline{B_{X^\alpha_{T_p}}(y_0, R)}$ and $\tilde{u} \in \overline{B_{\mathrm{L}^q(J_T;X)}(u, R)}$. Repeating the argument shows that

$$\sup_{0 \le t \le T} \|y(t) - G(\tilde{y}_0, \tilde{u})(t)\|_{X^\alpha_{T_p}} \le C(T, y) \left[\|y_0 - \tilde{y}_0\|_{X^\alpha_{T_p}} + \|u - \tilde{u}\|_{\mathrm{L}^q(J_T;X)} \right] < r$$

for some appropriate $R > 0$ and all $\tilde{y}_0 \in \overline{B_{X^\alpha_{T_p}}(y_0, R)}, \tilde{u} \in \overline{B_{\mathrm{L}^q(J_T;X)}(u, R)}$. Consequently, G maps $\overline{B_{X^\alpha_{T_p}}(y_0, R)} \times \overline{B_{\mathrm{L}^q(J_T;X)}(u, R)}$ into $\overline{B_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})}(y, r)}$ and F is Lipschitz continuous on $\overline{B_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})}(y, r)}$. A similar computation then yields a constant $C(T, y) > 0$ such that for arbitrary $y_0^{(1)}, y_0^{(2)} \in \overline{B_{X^\alpha_{T_p}}(y_0, R)}$ and $u_1, u_2 \in \overline{B_{\mathrm{L}^q(J_T;X)}(u, R)}$ there holds

$$\sup_{0 \le t \le T} \|G(y_0^{(1)}, u_1)(t) - G(y_0^{(2)}, u_2)(t)\|_{X^\alpha_{T_p}} \le C(T, y) \left[\|y_0^{(1)} - y_0^{(2)}\|_{X^\alpha_{T_p}} + \|u_1 - u_2\|_{\mathrm{L}^q(J_T;X)} \right].$$

Consequently, G is Lipschitz continuous on $\overline{B_{X^\alpha_{T_p}}(y_0, R)} \times \overline{B_{\mathrm{L}^q(J_T;X)}(u, R)}$. Because y_0 and u_0 were arbitrary, this proves that G is locally Lipschitz continuous from $X^\alpha_{T_p} \times \mathrm{L}^q(J_T; X)$ to $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$.

(IV) Uniqueness:

Uniqueness of the mild solution is already implied in local Lipschitz continuity of G since $G(y_0^{(1)}, u_1) = G(y_0^{(2)}, u_2)$ if one inserts $y_0^{(1)} = y_0^{(2)}$ and $u_1 = u_2$ into the Lipschitz estimate in Step III.

(V) Higher regularity:

For the last statement of the theorem assume that T_p satisfies maximal parabolic Sobolev regularity on X, see Definition 2.34, and suppose $y_0 \in X^\beta_{T_p}$ for some $\beta \in [\alpha, 1]$. Consider $y = G(y_0, u)$ and let $s \in (1, q] \cap (1, \infty)$ be arbitrary first. Then we apply $(\frac{d}{dt} + T_p)^{-1}$ to $F[y] + u \in \mathrm{L}^s(J_T; X)$ to obtain a function $\tilde{y} \in Y_{T_p,s}$ which solves the evolution equation

$$\frac{d}{dt}\tilde{y}(t) + T_p\tilde{y}(t) = F[y](t) + u(t) \quad \text{for } t > 0, \quad \tilde{y}(0) = 0.$$

Furthermore, by Theorem 2.22 the mapping $\tilde{y}_0 : t \mapsto e^{-T_p t} y_0 \in X$ is differentiable for all $t > 0$ with

$$\left(\frac{d}{dt} + T_p \right) e^{-T_p t} y_0 = 0.$$

Consequently, there holds $y = \tilde{y}_0 + \tilde{y}$. $\tilde{y}$ is contained in $Y_{T_p,s}$ and $\tilde{y}_0$ is absolutely continuous with values in X, i.e. $\tilde{y}_0 \in \mathrm{C}(\overline{J_T}; X)$. For the derivative of $\tilde{y}_0$ there holds

$$\begin{aligned} \int_0^T \left\| \frac{d}{dt} e^{-T_p t} y_0 \right\|_X^s \, dt &= \int_0^T \left\| T_p e^{-T_p t} y_0 \right\|_X^s \, dt \\ &= \int_0^T \left\| T_p^{1-\beta} e^{-T_p t} T_p^\beta y_0 \right\|_X^s \, dt \\ &\le \int_0^T \|T_p^{1-\beta} e^{-T_p t}\|_{\mathcal{L}(X)}^s \|y_0\|_{X^\beta_{T_p}}^s \, dt. \end{aligned} \tag{3.10}$$

If $\beta = 1$ then the last expression can be estimated in the form $c^s T \|y_0\|^s_{X^\beta_{T_p}}$. This value is finite for any $s \in (1, \infty)$ since $y_0 \in X^\beta_{T_p}$. In this case, $\tilde{y}_0 \in Y_{T_p,s}$ and then also $y \in Y_{T_p,s}$ for any

$s \in (1,q] \cap (1,\infty)$. For $\beta \in [\alpha,1)$ let $s \in \left(1, \frac{1}{1-\beta}\right) \cap (1,q]$ be arbitrary. Note that there holds $-s(1-\beta) > -1$. Hence, with (2.24) and (3.3) we obtain

$$\int_0^T \left\| T_p^{1-\beta} e^{-T_p t} \right\|_{\mathcal{L}(X)}^s \|y_0\|_{X_{T_p}^\beta}^s \, dt \leq c \int_0^T t^{-s(1-\beta)} \, dt \|y_0\|_{X_{T_p}^\beta}^s < \infty.$$

That is, $\tilde{y}_0 \in Y_{T_p,s}$ and therefore $y \in Y_{T_p,s}$ for any $s \in \left(1, \frac{1}{1-\beta}\right) \cap (1,q]$. Observe that $s = q$ is not possible for $\beta = \alpha$ since $q > \frac{1}{1-\alpha}$ by assumption. We prove that the linear bound (3.2) holds with $\mathrm{C}(\overline{J_T}; X_{T_p}^\alpha)$ replaced by $Y_{T_p,s}$ and $\|y_0\|_{X_{T_p}^\alpha}$ replaced by $\|y_0\|_{X_{T_p}^\beta}$. To this aim, we choose s arbitrary in the interval $(1,q] \cap (1,\infty)$ if $\beta = 1$ and in $\left(1, \frac{1}{1-\beta}\right) \cap (1,q]$ if $\beta \in [\alpha, 1)$. Remember that $\mathcal{W}[Sy](t) \in [a,b]$ for all $t \in \overline{J_T}$. Hence, by (A3) and (3.2) there holds

$$\begin{aligned}
\|F[y] + u\|_{\mathrm{L}^s(J_T;X)} &\leq M \left(\int_0^T (1 + \|y(t)\|_{X^\alpha} + \|\mathcal{W}[Sy](t)\| + \|u\|_X)^s \, dt \right)^{1/s} \\
&\leq c_0(1 + \|y\|_{\mathrm{L}^s(J_T;X_{T_p}^\alpha)} + \|u\|_{\mathrm{L}^s(J_T;X)}) \leq c_0(1 + \|y\|_{\mathrm{C}(\overline{J_T};X_{T_p}^\alpha)} + \|u\|_{\mathrm{L}^s(J_T;X)}) \\
&\leq c_0(C(T)+1)(1 + \|y_0\|_{X_{T_p}^\alpha} + \|u\|_{\mathrm{L}^s(J_T;X)})
\end{aligned}$$

for constants $c_0, C(T) > 0$. Moreover, notice that $(\frac{d}{dt} + T_p)^{-1}$ is linear and bounded as an operator from $\mathrm{L}^s(J_T;X)$ to $Y_{T_p,s,0}$ with some constant $c_{T_p} > 0$, see Remark 2.35. That is, together with (3.10) and since $\|y_0\|_{X_{T_p}^\alpha} \leq \|y_0\|_{X_{T_p}^\beta}$ and $\|u\|_{\mathrm{L}^s(J_T;X)} \leq \|u\|_{\mathrm{L}^q(J_T;X)}$ we obtain

$$\begin{aligned}
\|G(y_0,u)\|_{Y_{T_p,s}} &\leq \|\tilde{y}\|_{Y_{T_p,s}} + \|\tilde{y}_0\|_{Y_{T_p,s}} \\
&= \left\| \left(\frac{d}{dt} + T_p \right)^{-1} (F[y] + u) \right\|_{Y_{T_p,s}} + \|\tilde{y}_0\|_{Y_{T_p,s}} \\
&\leq c_1(C(T)+1)(1 + \|y_0\|_{X_{T_p}^\alpha} + \|u\|_{\mathrm{L}^s(J_T;X)}) + c_2\|y_0\|_{X_{T_p}^\beta} \\
&\leq c_3(T)(1 + \|y_0\|_{X_{T_p}^\beta} + \|u\|_{\mathrm{L}^q(J_T;X)})
\end{aligned}$$

for constants $c_1 = c_0 c_{T_p}$, $c_2 > 0$ and $c_3(T) = c_1(C(T)+1) + c_2$. To prove that G is locally Lipschitz continuous into $Y_{T_p,s}$ let $V_0 \times V$ be a neighborhood of (y_0,u) in $X_{T_p}^\beta \times \mathrm{L}^q(J_T;X)$ such that G is Lipschitz continuous from the embedding $V_0 \times V \hookrightarrow X_{T_p}^\alpha \times \mathrm{L}^q(J_T;X)$ into $\mathrm{C}(\overline{J_T}; X_{T_p}^\alpha)$ with modulus $L_1(y_0,u) > 0$. We can choose $V_0 \times V$ small enough such that F is Lipschitz continuous on the set $G(V_0,V) \subset \mathrm{C}(\overline{J_T}; X_{T_p}^\alpha)$ with modulus $L_2(y_0,u) := L_2(G(y_0,u)) > 0$. Let $y_0^{(1)}, y_0^{(2)} \in V_0$ and $u_1, u_2 \in V$ be given and denote $y_i := G(y_0^{(i)}, u_i)$ for $i \in \{1,2\}$. Then

$$\begin{aligned}
\|F[y_1] - F[y_2] + u_1 - u_2\|_{\mathrm{L}^s(J_T;X)} &\leq L_2(y_0,u)\|y_1 - y_2\|_{\mathrm{C}(\overline{J_T};X_{T_p}^\alpha)} + \|u_1 - u_2\|_{\mathrm{L}^s(J_T;X)} \\
&\leq (1 + L_1(y_0,u)L_2(y_0,u))(\|y_0^{(1)} - y_0^{(2)}\|_{X_{T_p}^\alpha} + \|u_1 - u_2\|_{\mathrm{L}^q(J_T;X)}).
\end{aligned}$$

Similar as for the proof of linear boundedness of G we write $\tilde{y}_i := \left(\frac{d}{dt} + T_p\right)^{-1} (F[y_i] + u_i)$ and $\tilde{y}_0^{(i)} : t \mapsto e^{-T_p t} \tilde{y}_0^{(i)}$ for $i \in \{1,2\}$ and obtain

$$\|\tilde{y}_1 - \tilde{y}_2\|_{Y_{T_p,s}} \leq L_3(y_0,u))(\|y_0^{(1)} - y_0^{(2)}\|_{X_{T_p}^\alpha} + \|u_1 - u_2\|_{\mathrm{L}^q(J_T;X)})$$

for $L_3(y_0,u) := c_{T_p}(1 + L_1(y_0,u)L_2(y_0,u))$. This together with (3.10) and the embedding $X^\beta_{T_p} \hookrightarrow X^\alpha_{T_p}$ yields

$$\begin{aligned}\|y_1 - y_2\|_{Y_{T_p,s}} &= \|\tilde{y}_1 - \tilde{y}_2 + \tilde{y}_0^{(1)} - \tilde{y}_0^{(2)}\|_{Y_{T_p,s}} \le \|\tilde{y}_1 - \tilde{y}_2\|_{Y_{T_p,s}} + \|\tilde{y}_0^{(1)} - \tilde{y}_0^{(2)}\|_{Y_{T_p,s}} \\ &\le L_4(y_0,u))(\|y_0^{(1)} - y_0^{(2)}\|_{X^\beta_{T_p}} + \|u_1 - u_2\|_{\mathrm{L}^q(J_T;X)})\end{aligned}$$

for some modulus $L_4(y_0,u)) > 0$.

□

Remark 3.3. As an alternative to the proof of Step V in Theorem 3.2, we make use of Proposition [Ama05, Proposition] which states that maximal parabolic $\mathrm{L}^s(J_T;X)$-regularity of T_p is equivalent to the fact that $\left(\frac{d}{dt} + T_p, \gamma_0\right)$ is an isomorphism from $Y_{T_p,s}$ to $\mathrm{L}^s(J_T;X) \times (X, X^1_{T_p})_{s',s}$. Here, γ_0 is defined as $\gamma_0(y) := y(0)$ for $y \in Y_{T_p,s}$. We refer to [Ama95, Chapter III.4] for further details. By [Haa06, Corollary 6.6.3] there holds

$$(X, X^1_{T_p})_{\beta,1} \hookrightarrow X^\beta_{T_p} \hookrightarrow (X, X^1_{T_p})_{\beta,\infty}$$

for all $\beta \in (0,1)$. Moreover, by [Ama05, Section 1] we have

$$(X, X^1_{T_p})_{\beta,\infty} \hookrightarrow (X, X^1_{T_p})_{s',1} \hookrightarrow (X, X^1_{T_p})_{s',s}$$

if $s' < \beta \Leftrightarrow s < \frac{1}{1-\beta}$. Hence, for $\beta \in [\alpha, 1)$ and any $s \in \left(1, \frac{1}{1-\beta}\right) \cap (1,q]$, $X^\beta_{T_p} \hookrightarrow (X, X^1_{T_p})_{s',s}$ and $\left(\frac{d}{dt} + T_p, \gamma_0\right)^{-1}$ is bounded as an operator from $\mathrm{L}^s(J_T;X) \times X^\beta_{T_p}$ into $Y_{T_p,s}$.
Now in the notation of Step V in Theorem 3.2 there holds $\left(\frac{d}{dt} + T_p, \gamma_0\right)^{-1}(0, y_0) = \tilde{y}_0$ and $\left(\frac{d}{dt} + T_p, \gamma_0\right)^{-1}(F[y] + u, 0) = \tilde{y}$ so that

$$G(y_0,u) = \left(\frac{d}{dt} + T_p, \gamma_0\right)^{-1}(F[y] + u, y_0) = \tilde{y}_0 + \tilde{y} \in Y_{T_p,s}.$$

The statement about local Lipschitz continuity then follows from boundedness of the operator $\left(\frac{d}{dt} + T_p, \gamma_0\right)^{-1}$ and local Lipschitz continuity of F in $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ and of $G : X^\alpha_{T_p} \times \mathrm{L}^q(J_T; X_{T_p}) \to \mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$. The statement about the linear bound for G follows from (3.2) and boundedness of $\left(\frac{d}{dt} + T_p, \gamma_0\right)^{-1}$.
Note that we need β to be strictly larger than α in order to obtain $s = q$. Indeed, $q \in \left(\frac{1}{1-\alpha}, \infty\right]$ by assumption, so that $q \notin \left(1, \frac{1}{1-\alpha}\right) = \left(1, \frac{1}{1-\beta}\right) \cap (1,q]$ if $\beta = \alpha$.

The following Corollary 3.4 is a generalization of [Mün17a, Theorem 3.1] from zero initial value to initial values $y_0 \in X^\beta_{A_p+1}$.

Corollary 3.4. Let Assumption 3.1 hold. Instead of (A1) suppose that A_p is a diffusion operator in the sense of Definition 2.17 for some $p \in \mathrm{J} \cap [2,\infty)$, with J as in Corollary 2.19. Let $q \in \left(\frac{1}{1-\alpha}, \infty\right]$ and s be arbitrary in the interval $\left(1, \frac{1}{1-\beta}\right) \cap (1,q]$ if $\beta \in [\alpha, 1)$ and in the interval $(1,q] \cap (1,\infty)$ if $\beta = 1$. Then for all $y_0 \in X^\beta_{A_p+1}$ and $u \in \mathrm{L}^q(J_T;X)$ the evolution equation

$$\begin{aligned}\frac{d}{dt}y(t) + (A_p y)(t) &= (F[y])(t) + u(t) \quad \text{in } X = \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega) \text{ for } t > 0, \\ y(0) &= y_0 \in X,\end{aligned} \tag{3.11}$$

has a unique solution $y = y(y_0, u)$ in $Y_{A_p,s}$. The solution mapping

$$G : (y_0, u) \mapsto y(y_0, u),\ X^\beta_{A_p+1} \times \mathrm{L}^q(J_T; X) \to Y_{A_p,s}$$

is locally Lipschitz continuous. G is linearly bounded with values in $Y_{A_p,s}$, i.e. for some $C(T) > 0$ there holds

$$\|G(y_0, u)\|_{Y_{A_p,s}} \leq C(T)(1 + \|y_0\|_{X^\beta_{A_p+1}} + \|u\|_{\mathrm{L}^q(J_T;X)}) \tag{3.12}$$

for all $y_0 \in X^\beta_{A_p+1}$ and $u \in \mathrm{L}^q(J_T; X)$ and $C(T)$ is independent of y_0 and u.

Proof. The proof follows the lines of Theorem 3.2. By Remark 2.35, A_p satisfies maximal parabolic Sobolev regularity on X. The only difference to Theorem 3.2 is that (3.11) contains only the diffusion operator A_p instead of the sectorial operator $T_p = A_p + 1$, cf. (3.1). Hence, the estimates in (2.21) have to be exchanged by the estimates

$$\begin{aligned} \|\exp(-A_p t)\|_{\mathcal{L}(X)} &= \|\exp(-(A_p + 1)t)\exp(t)\|_{\mathcal{L}(X)} \leq C\exp((1-\delta)t), \\ \|(A_p + 1)\exp(-A_p t)\|_{\mathcal{L}(X)} &= \|(A_p + 1)\exp(-(A_p + 1)t)\exp(t)\|_{\mathcal{L}(X)} \\ &\leq \frac{C}{t}\exp((1-\delta)t). \end{aligned} \tag{3.13}$$

Similarly, (2.24)–(2.25) have to be replaced by

$$\|(A_p + 1)^\alpha \exp(-A_p t)\|_X = e^t\|(A_p + 1)^\alpha \exp(-(A_p + 1))\|_X \leq C_\alpha t^{-\alpha}\exp((1-\delta)t), \tag{3.14}$$

for $\alpha \geq 0$ and

$$\|(\exp(-A_p t) - \mathrm{Id})u\|_X \leq \exp((1-\delta)t)\left(\frac{1}{\alpha}C_{1-\alpha}t^\alpha + Ct\right)\|(A_p + 1)^\alpha u\|_X \tag{3.15}$$

for $\alpha \in (0, 1]$ and $u \in X^\alpha_{A_p+1}$. That the function $t \mapsto \exp((1-\delta)t)$ is monotone increasing for $\delta \in (0, 1)$ is not a problem, since the time interval J_T is a priori fixed and bounded. However, several more constants in the proof depend on $T > 0$ now. □

3.3 Hadamard differentiability of the solution operator for the evolution equation

Well-posedness of the evolution equation (3.1) and Lipschitz continuity of the solution operator G have been shown in Theorem 3.2. The next natural question is whether G is differentiable in the forcing term u and the initial value y_0 in some sense. Moreover, with respect to the optimal control problem to be studied in Section 4, a chain rule would be desirable. The stop operator $\mathcal{W}$ is not differentiable in the classical sense, and hence G cannot be so either. Nevertheless, it turns out that G is Hadamard directionally differentiable in the sense of [BS00; BK15].
Hadamard directional differentiability of G for the case of a diffusion operator A_p and trivial initial value, i.e. for the mapping $u \mapsto G(0, u)$, $u \in \mathrm{L}^q(J_T; X)$, has already appeared in [Mün17a, Section 4]. In this section, we extend those results to apply to general elliptic operators T_p and non-trivial initial values, i.e. for G defined on $X^\beta_{T_p} \times \mathrm{L}^q(J_T; X)$, $\beta \in [\alpha, 1)$, $q \in \left(\frac{1}{1-\alpha}, \infty\right)$.
First of all, we define the notion of directional differentiability of a mapping $g : U \subset Z \to Y$ which is used in this work. Here, $U \subset Z$ is an open set in Z and Z and Y are normed vector spaces.

Definition 3.5. [BS00, Definition 2.44] Let Z, Y be normed vector spaces. We call g directionally differentiable at $x \in U \subset Z$ in the direction $h \in Z$ if

$$g'[x;h] := \lim_{\lambda \downarrow 0} \frac{g(x+\lambda h) - g(x)}{\lambda}$$

exits in Y. If g is directionally differentiable at x in every direction h we call g directionally differentiable at x.

Based on this definition, we introduce the concept of Hadamard directional differentiability:

Definition 3.6. [BS00, Definition 2.45] If g is directionally differentiable at $x \in U$ and if in addition for all functions $r : [0, \lambda_0) \to Z$ with $\lim\limits_{\lambda \to 0} \frac{r(\lambda)}{\lambda} = 0$ there holds

$$g'[x;h] = \lim_{\lambda \downarrow 0} \frac{g(x+\lambda h + r(\lambda)) - g(x)}{\lambda}$$

for all directions $h \in Z$, we call $g'[x;h]$ the Hadamard directional derivative of g at x in the direction h.
Note that $g(x+\lambda h+r(\lambda))$ is only well defined if λ is already small enough so that $x+\lambda h+r(\lambda) \in U$.

As already mentioned, the chain rule applies for Hadamard directionally differentiable mappings. The latter will be crucial not only for the application in an optimal control problem, but already in order to prove differentiability of the composed mapping G.

Lemma 3.7. [BS00, Proposition 2.47] Suppose that $g : U \subset Z \to Y$ is Hadamard directionally differentiable at $x \in U$ and that $f : V \subset g(U) \to Z$ is Hadamard directionally differentiable at $g(x) \in V$. Then $f \circ g : U \to Z$ is Hadamard directionally differentiable at x and

$$(f \circ g)'[x;h] = f'\left[g(x); g'[x;h]\right].$$

The following lemma provides the main tool to prove Hadamard directional differentiability.

Lemma 3.8. [BS00, Proposition 2.49] Suppose that $g : U \subset Z \to Y$ is directionally differentiable at $x \in U$ and in addition Lipschitz continuous with modulus $c(x)$ in a neighborhood of x. Then g is Hadamard directionally differentiable at x and $g'[x;\cdot]$ is Lipschitz continuous on Z with modulus $c(x)$.

Remark 3.9. [cf. Mün17a, Remark 4.5] Remember Theorem 2.40, which states that the stop operator $\mathcal{W}$ is Hadamard directionally differentiable as a mapping $\mathrm{C}[0,T] \to \mathrm{L}^q(0,T)$.
Furthermore, note that all the results in this section remain valid if the stop operator is exchanged by $\mathcal{P}$ or by some other hysteresis operator with appropriate properties. We decided for $\mathcal{W}$ with regard to the application in Chapter 4. There, equation (3.1) will serve as the state equation for an optimal control problem. The derivation of an adjoint system for this problem will be based on a regularization of the variational inequalities (2.27)–(2.29) which represent $\mathcal{W}$.

In order to prove Hadamard directional differentiability of the solution operator for problem (3.1) we need a further assumption on the non-linearity f.

Assumption 3.10. [cf. Mün17a, Assumption 4.6] In addition to Assumption 3.1 we assume that f is directionally differentiable and therefore Hadamard directionally differentiable.

The following theorem on Hadamard directional differentiability of G resembles [MS15, Theorem 3.2]. However, different techniques for the proof are required due to the hysteresis operator and since f is only locally Lipschitz continuous. Moreover, other than [MS15, Theorem 3.2], Theorem 3.11 below applies for non-trivial initial value. Theorem 3.11 is a generalization of [Mün17a, Theorem 4.7] from diffusion operators A_p and zero initial value to general elliptic operators T_p and initial values $y_0 \in X^\beta_{T_p}$, $\beta \in [\alpha, 1]$.

Theorem 3.11. *Let Assumption 3.10 hold and consider the notation from Theorem 3.2. Then for any $q \in \left(\frac{1}{1-\alpha}, \infty\right)$, the solution operator $G : X^\alpha_{T_p} \times \mathrm{L}^q(J_T; X) \to \mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ of problem (3.1) is Hadamard directionally differentiable. Its derivative $y^{(y_0,u),(h_0,h)} := G'[(y_0, u); (h_0, h)]$ at $(y_0, u) \in X^\alpha_{T_p} \times \mathrm{L}^q(J_T; X)$ in direction $(h_0, h) \in X^\alpha_{T_p} \times \mathrm{L}^q(J_T; X)$ is given by the unique mild solution $\zeta \in \mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ of*

$$\begin{aligned} \dot\zeta(t) + (T_p\zeta)(t) &= F'[y; \zeta](t) + h(t) \quad \text{in } J_T, \\ \zeta(0) &= h_0, \end{aligned} \tag{3.16}$$

where $F'[y; \zeta](t) = f'[(y(t), \mathcal{W}[Sy](t)); (\zeta(t), \mathcal{W}'[Sy; S\zeta](t))]$ and $y = G(y_0, u)$, see Theorem 3.2. In particular, there holds

$$\zeta(t) = e^{-tT_p} h_0 + \int_0^t e^{-T_p(t-s)}(F'[y; \zeta](s) + h(s))\, ds \quad \text{for } t > 0.$$

Moreover, the mapping $(h_0, h) \mapsto G'[(y_0, u); (h_0, h)]$ is Lipschitz continuous from $X^\alpha_{T_p} \times \mathrm{L}^q(J_T; X)$ to $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ with a modulus of continuity $C = C(G(y_0, u), T)$.
If T_p satisfies maximal parabolic regularity on X and if $\beta \in [\alpha, 1]$, then all statements on G remain valid for $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ replaced by $Y_{T_p,s}$ and with $X^\beta_{T_p}$ instead of $X^\alpha_{T_p}$. Here, s is arbitrary in the interval $\left(1, \frac{1}{1-\beta}\right) \cap (1, q]$ if $\beta \in [\alpha, 1)$ and $s = q$ if $\beta = 1$.

We divide the proof of Theorem 3.11 into several lemmas. Those extend the five steps in the proof of [Mün17a, Theorem 4.7] to our generalized framework.

Lemma 3.12 (Nemytski operator of f)**.** Consider the assumptions and notation from Theorem 3.11. Then the function

$$\tilde F : \mathrm{C}(\overline{J_T}; X^\alpha_{T_p}) \times \mathrm{L}^q(J_T) \to \mathrm{L}^q(J_T; X), \quad (y, v) \mapsto [t \mapsto f(y(t), v(t))]$$

is Hadamard directionally differentiable.

Proof. The proof is divided into four steps.
(I) Well-posedness of $\tilde F$:
We show that $\tilde F$ is well defined. The function $x \mapsto x^q$, $x \in \mathbb{R}$, is convex since $q > 1$. Hence, there holds

$$(x_1 + x_2)^q = 2^q \left(\frac{x_1}{2} + \frac{x_2}{2}\right)^q \le 2^{q-1}(x_1^q + x_2^q) \quad \text{for } x_1, x_2 \in \mathbb{R}_+. \tag{3.17}$$

Let $(y, v) \in \mathrm{C}(\overline{J_T}; X^\alpha_{T_p}) \times \mathrm{L}^q(J_T)$ be arbitrary. Measurability of $\tilde F(y, v)$ is a consequence of measurability of y and v and continuity of f in both components. Furthermore, (3.17) implies

$$\|f(y(s), v(s))\|^q_X \le M^q(\|y(s)\|_{X^\alpha_{T_p}} + |v(s)| + 1)^q \le M^q 2^{q-1}[(\|y(s)\|_{X^\alpha_{T_p}} + 1)^q + |v(s)|^q]$$

for a.e. $s \in J_T$ with M from (A3) in Assumption 3.1. Consequently, $\tilde{F}(y,v) \in \mathrm{L}^q(J_T;X)$ and $\tilde{F}$ is well defined.
(II) Local Lipschitz continuity of $\tilde{F}$:
We show that $\tilde{F}$ is locally Lipschitz continuous with respect to the $\mathrm{C}(\overline{J_T};X^\alpha_{T_p})$-norm. Remember Step III in the proof of Theorem 3.2 were we proved local Lipschitz continuity of the mapping $(y(\cdot),v) \mapsto f(y(\cdot),v)$ from $\mathrm{C}(\overline{J_T};X^\alpha_{T_p}) \times \mathbb{R}$ to $\mathrm{C}(\overline{J_T};X)$ with respect to the $\mathrm{C}(\overline{J_T};X^\alpha_{T_p})$-norm. Hence, for any $y \in \mathrm{C}(\overline{J_T};X^\alpha_{T_p})$ we may choose constants $r, L(y) > 0$ such that the latter function is Lipschitz continuous in $\overline{B_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})}(y,r)} \times \mathbb{R}$ with modulus $L(y)$. Moreover, by (3.8) the corresponding Lipschitz estimates hold pointwise in time.
Let $y_1, y_2 \in \overline{B_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})}(y,r)}$ and $v_1, v_2 \in \mathrm{L}^q(J_T)$ be arbitrary. Then (3.8) implies that for a.e. $s \in J_T$

$$\|\tilde{F}(y_1,v_1)(s) - \tilde{F}(y_2,v_2)(s)\|_X \le L(y)\left[\|y_1(s)-y_2(s)\|_{X^\alpha_{T_p}} + |v_1(s)-v_2(s)|\right].$$

We integrate over $s \in J_T$, apply Minkowski's inequality and estimate

$$\|y_1 - y_2\|_{\mathrm{L}^q(J_T;X^\alpha_{T_p})} \le T^{1/q}\|y_1-y_2\|_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})}$$

to obtain

$$\begin{aligned}\|\tilde{F}(y_1,v_1) - \tilde{F}(y_2,v_2)\|_{\mathrm{L}^q(J_T;X)} &\le L(y)\left[\|y_1-y_2\|_{\mathrm{L}^q(J_T;X^\alpha_{T_p})} + \|v_1-v_2\|_{\mathrm{L}^q(J_T)}\right] \\ &\le L(y)\left[T^{1/q}\|y_1-y_2\|_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})} + \|v_1-v_2\|_{\mathrm{L}^q(J_T)}\right] \\ &\le L(y)(1+T^{1/q})\left[\|y_1-y_2\|_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})} + \|v_1-v_2\|_{\mathrm{L}^q(J_T)}\right].\end{aligned} \tag{3.18}$$

This proves local Lipschitz continuity of $\tilde{F}$.
(III) Directional differentiability of $\tilde{F}$:
We show that $\tilde{F}$ is directionally differentiable. Let $y \in \mathrm{C}(\overline{J_T};X^\alpha_{T_p})$, r and $L(y)$ be chosen as in Step II and consider any $v \in \mathrm{L}^q(J_T)$. Moreover, let $(h,l) \in \mathrm{C}(\overline{J_T};X^\alpha_{T_p}) \times \mathrm{L}^q(J_T)$ be arbitrary and choose $\lambda_0 > 0$ small enough such that $y + \lambda h \in \overline{B_{\mathrm{C}(\overline{J_T};X^\alpha_{T_p})}(y,r)}$ for all $\lambda \in (0,\lambda_0]$.
For $\lambda \in (0,\lambda_0]$ consider the differential quotient

$$\tilde{F}_\lambda := \frac{1}{\lambda}[\tilde{F}(y+\lambda h, v+\lambda l) - \tilde{F}(y,v)] \in \mathrm{L}^q(J_T;X).$$

By Assumption 3.10, f is directionally differentiable. Hence,

$$\lim_{\lambda\to 0} \tilde{F}_\lambda(s) = f'[(y(s),v(s));(h(s),l(s))] \in X \quad \text{for a.e. } s \in J_T.$$

Moreover, for a.e. $s \in J_T$ and since λ_0 is small enough, Step II implies

$$\|\tilde{F}_\lambda(s)\|_X \le L(y)\left[\|h(s)\|_{X^\alpha_{T_p}} + |l(s)|\right].$$

Because the right side is contained in $\mathrm{L}^q(J_T)$, we conclude from Lebesgue's dominated convergence theorem that $\tilde{F}_\lambda$ converges to the function

$$s \mapsto f'[(y(s),v(s));(h(s),l(s))] \in \mathrm{L}^q(J_T;X)$$

in $\mathrm{L}^q(J_T;X)$ with $\lambda \to 0$, i.e. that $\tilde{F}$ is directionally differentiable.

Note that this step was shown as [MS15, Lemma 3.1]. Steps I–II required additional work.
(IV) Hadamard directional differentiability of $\tilde{F}$:
We exploit Lemma 3.8. $\tilde{F}$ is locally Lipschitz continuous and directionally differentiable by Steps II–III. Hence, Lemma 3.8 implies that $\tilde{F}$ is Hadamard directionally differentiable. Moreover, it follows that the mapping $(h,l) \mapsto \tilde{F}'[(y,v);(h,l)]$ is Lipschitz continuous. $\square$

We apply the chain rule Lemma 3.7 and Lemma 3.12 to conclude that F is Hadamard directionally differentiable.

Lemma 3.13 (Hadamard differentiability of F)**.** Consider the assumptions and notation as in Theorem 3.11 and Lemma 3.12. Then the function

$$F : \mathrm{C}(\overline{J_T}; X^\alpha_{T_p}) \to \mathrm{L}^q(J_T; X), \quad (F[y])(t) = f(y(t), \mathcal{W}[Sy](t)) = \tilde{F}(y(t), \mathcal{W}[Sy](t))$$

is Hadamard directionally differentiable and locally Lipschitz continuous. Moreover, for any $y \in \mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ the mapping

$$\begin{aligned} F'[y;\cdot] : \mathrm{C}(\overline{J_T}; X^\alpha_{T_p}) &\to \mathrm{L}^q(J_T; X), \\ h &\mapsto F'[y;h], && \text{where} \\ F'[y;h](t) &= f'[(y(t), \mathcal{W}[Sy](t)); (h(t), \mathcal{W}'[Sy;Sh](t))] && \text{for a.e. } t \in J_T, \end{aligned}$$

is Lipschitz continuous with a modulus $C(y)$.

Proof. First of all, the identity mapping Id on $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ and the operator $S : \mathrm{C}(\overline{J_T}; X^\alpha_{T_p}) \to \mathrm{C}(\overline{J_T})$ are linear and continuous and hence Fréchet differentiable. The derivatives are given by Id and S. Moreover, $\mathcal{W} : \mathrm{C}(\overline{J_T}) \to \mathrm{L}^q(J_T)$ is Hadamard directionally differentiable by Theorem 2.40. Hence, the chain rule Lemma 3.7 implies Hadamard directional differentiability of the mapping

$$y \mapsto (y, \mathcal{W}[Sy])$$

from $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ into $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p}) \times \mathrm{L}^q(J_T)$ with derivative

$$h \mapsto (h, \mathcal{W}'[Sy;Sh]).$$

Since $\tilde{F}$ is Hadamard directionally differentiable by Lemma 3.12 and because $F = \tilde{F} \circ (\mathrm{Id}, \mathcal{W} \circ S)$, another application of Lemma 3.7 implies that F is Hadamard directionally differentiable with

$$F'[y;h](t) = f'[(y(t), \mathcal{W}[Sy](t)); (h(t), \mathcal{W}'[Sy;Sh](t))]$$

for $y, h \in \mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ and a.e. $t \in J_T$.
In Step III in the proof of Theorem 3.2 we have shown that F is locally Lipschitz continuous from $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ to $\mathrm{C}(\overline{J_T}, X)$ and hence from $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ to $\mathrm{L}^q(J_T; X)$ as well. Lemma 3.8 therefore yields Lipschitz continuity of the mapping $h \to F'[y;h]$ from $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ to $\mathrm{L}^q(J_T; X)$ for any $y \in \mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$. $\square$

Lemma 3.14 (Mild solutions for (3.16))**.** Consider the assumptions and notation as in Theorem 3.11. Then for arbitrary $y \in \mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ and $(h_0, h) \in X^\alpha_{T_p} \times \mathrm{L}^q(J_T; X)$ there exists a unique mild solution $\zeta = \zeta(y, h_0, h)$ of (3.16). Moreover, for y fixed the mapping

$$\zeta(y, \cdot, \cdot) : X^\alpha_{T_p} \times \mathrm{L}^q(J_T; X) \to \mathrm{C}(\overline{J_T}; X^\alpha_{T_p}), \quad (h_0, h) \mapsto \zeta(y, h_0, h)$$

is Lipschitz continuous with a modulus $C = C(y,T)$.

Proof. We have to show that for arbitrary $y \in \mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ and $(h_0, h) \in X^\alpha_{T_p} \times \mathrm{L}^q(J_T; X)$ there exists a unique fixed point $\zeta(y, h_0, h)$ in $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ of the integral equation

$$\zeta(t) = e^{-T_p t} h_0 + \int_0^t e^{-T_p(t-s)} [F'[y;\zeta](s) + h(s)]\, ds.$$

Lipschitz continuity of the corresponding fixed point mapping g (see below) yields Lipschitz continuity of the mapping $(h_0, h) \mapsto \zeta(y, h_0, h)$ for fixed y, provided that $\zeta(y, h_0, h)$ is well defined. To prove the latter, we exploit the concrete Lipschitz modulus of g to prove the existence of ζ first locally and continue the mild solution to the whole interval $\overline{J_T}$ by induction. Remember Lemma 3.13, namely that the function

$$\begin{aligned} F'[y;\cdot] : \mathrm{C}(\overline{J_T}; X^\alpha_{T_p}) &\to \mathrm{L}^q(J_T; X), \\ \zeta &\mapsto F'[y;\zeta], && \text{where} \\ F'[y;\zeta](t) &= f'[(y(t), \mathcal{W}[Sy](t)); (\zeta(t), \mathcal{W}'[Sy; S\zeta](t))] && \text{for a.e. } t \in J_T, \end{aligned}$$

is Lipschitz continuous with a modulus $C(y)$.
Moreover, recall (2.24) and (3.3) so that

$$\begin{aligned} \|T_p^\alpha \exp(-T_p t)\|_{\mathcal{L}(X)} &\le C_\alpha t^{-\alpha} \exp(-\delta t), && \text{and} \\ \left(\int_0^t (t-s)^{-\alpha q'}\, ds\right)^{1/q'} &= \left(\frac{t^{1-\alpha q'}}{1-\alpha q'}\right)^{1/q'} = \frac{t^{1/q'-\alpha}}{(1-\alpha q')^{1/q'-\alpha}}. \end{aligned}$$

Then as in Step II in Theorem 3.2 we apply Minkowski's inequality to prove that for arbitrary $\tilde{T} \in (0, T]$ the function

$$g : \zeta \mapsto \left[t \mapsto e^{-T_p t} h_0 + \int_0^t e^{-T_p(t-s)} [F'[y;\zeta](s) + h(s)]\, ds \right]$$

is well defined on $\mathrm{C}([0, \tilde{T}]; X^\alpha_{T_p})$ and Lipschitz continuous with a modulus of the form

$$L(\tilde{T}) = \tilde{C}(y) \tilde{T}^{1/q'-\alpha}.$$

Similar as Steps IV–V of Theorem 3.2 about local Lipschitz continuity and uniqueness of G, this observation and a Gronwall argument already imply uniqueness and the statement about Lipschitz continuity for fixed y, assuming that the fixed point mapping $(h_0, h) \mapsto \zeta(y, h_0, h)$ is well defined.
It remains to show that g has a fixed point in $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$. To this aim, we choose $k \in \mathbb{N}$ with

$$L\left(\frac{T}{k}\right) := C(y) \left(\frac{T}{k}\right)^{1/q'-\alpha} < \frac{1}{2}.$$

Moreover, we define $t_j := \frac{jT}{k}$ for $1 \le j \le k$ and

$$H(t) := e^{-T_p t} h_0 + \int_0^t e^{-T_p(t-s)} h(s)\, ds, \qquad t \in J_T.$$

We begin by proving that g has a fixed point in $\mathrm{C}(\overline{J_{t_1}}; X^\alpha_{T_p}) = \mathrm{C}([0,t_1]; X^\alpha_{T_p})$. Denote by $g_1 := g : \mathrm{C}(\overline{J_{t_1}}; X^\alpha_{T_p}) \to \mathrm{C}(\overline{J_{t_1}}; X^\alpha_{T_p})$ the restriction of g to $\mathrm{C}(\overline{J_{t_1}}; X^\alpha_{T_p})$.
Consider $N_0 := \|H\|_{\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})}$ and the closed set $\overline{B_{\mathrm{C}(\overline{J_{t_1}}; X^\alpha_{T_p})}(H, N_0)}$. Note that g_1 is a contraction on $\mathrm{C}(\overline{J_{t_1}}; X^\alpha_{T_p})$ so that we can apply Banach's fixed point theorem if $\overline{B_{\mathrm{C}(\overline{J_{t_1}}; X^\alpha_{T_p})}(H, N_0)}$ is invariant under g_1. By definition of g and H there holds $g(0) = H$. Hence, for any $\zeta \in \overline{B_{\mathrm{C}(\overline{J_{t_1}}; X^\alpha_{T_p})}(H, N_0)}$, because $L\left(\frac{T}{k}\right) < \frac{1}{2}$, we can estimate

$$
\begin{aligned}
\|g_1(\zeta)(t) - H(t)\|_{X^\alpha_{T_p}} &= \|g(\zeta)(t) - g(0)(t)\|_{X^\alpha_{T_p}} \le \frac{1}{2}\|\zeta(t)\|_{X^\alpha_{T_p}} \\
&\le \frac{1}{2}\|\zeta(t) - H(t)\|_{X^\alpha_{T_p}} + \frac{1}{2}\|H(t)\|_{X^\alpha_{T_p}} \le N_0.
\end{aligned}
$$

This proves that $\overline{B_{\mathrm{C}(\overline{J_{t_1}}; X^\alpha_{T_p})}(H, N_0)}$ is invariant under g_1, and we obtain a (unique) fixed point $\zeta_1 \in \overline{B_{\mathrm{C}(\overline{J_{t_1}}; X^\alpha_{T_p})}(H, N_0)}$ of g_1.
We continue by induction. Consider $N_j := 2N_{j-1} + N_0$ for $2 \le j \le k$. Assuming that the unique fixed point ζ_{j-1} of g_{j-1} exists from the previous step let the mapping $g_j : \mathrm{C}(\overline{J_{t_j}}, X^\alpha_{T_p}) \to \mathrm{C}(\overline{J_{t_j}}, X^\alpha_{T_p})$ be defined as

$$
g_j(\zeta)(t) := \begin{cases} \zeta_{j-1}(t) & \text{if } t \in [0, t_{j-1}], \\ \zeta_{j-1}(t_{j-1}) + e^{-T_p(t-t_{j-1})}h_0 + \int\limits_{t_{j-1}}^{t} e^{-T_p(t-s)}[F'[y;\zeta](s) + h(s)]\, ds & \text{if } t \in [t_{j-1}, t_j] \end{cases}.
$$

We have to show that each g_j has a fixed point.
As for g_1 it follows from the definition that $g_j(0) = \zeta_{j-1}(t_{j-1}) + H - H(t_{j-1})$ in $\mathrm{C}([t_{j-1}, t_j]; X^\alpha_{T_p})$. Moreover, g_j is a $\frac{1}{2}$-contraction on $\mathrm{C}(\overline{J_{t_j}}; X^\alpha_{T_p})$. Hence, Banach's fixed point theorem implies existence of a fixed point ζ_j of g_j if we can prove that g_j maps $\overline{B_{\mathrm{C}(\overline{J_{t_j}}; X^\alpha_{T_p})}(H, N_j)}$ into itself.
Let $\zeta \in \overline{B_{\mathrm{C}(\overline{J_{t_j}}; X^\alpha_{T_p})}(H, N_j)}$ be arbitrary. Then on $[0, t_{j-1}]$ we can estimate

$$
\|g_j(\zeta)(t) - H(t)\|_{X^\alpha_{T_p}} = \|\zeta_{j-1}(t) - H(t)\|_{X^\alpha_{T_p}} \le N_{j-1} \le N_j
$$

by induction and definition of $N_j = 2N_{j-1} + N_0$. For $t \in [t_{j-1}, t_j]$ there holds

$$
\begin{aligned}
\|g_j(\zeta)(t) - H(t)\|_{X^\alpha_{T_p}} &= \|\zeta_{j-1}(t_{j-1}) - H(t_{j-1}) + g_j(\zeta)(t) - g_j(0)(t)\|_{X^\alpha_{T_p}} \\
&\le \|\zeta_{j-1}(t_{j-1}) - H(t_{j-1})\|_{X^\alpha_{T_p}} + \frac{1}{2}\|\zeta(t)\|_{X^\alpha_{T_p}} \\
&\le N_{j-1} + \frac{1}{2}\|\zeta(t) - H(t)\|_{X^\alpha_{T_p}} + \frac{1}{2}\|H(t)\|_{X^\alpha_{T_p}} \\
&\le N_{j-1} + \frac{N_j}{2} + \frac{1}{2}\|H\|_{\mathrm{C}(\overline{J_T}, X^\alpha_{T_p})} \\
&= N_{j-1} + \frac{2N_{j-1} + \|H\|_{\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})}}{2} + \frac{1}{2}\|H\|_{\mathrm{C}(\overline{J_T}, X^\alpha_{T_p})} \\
&= 2N_{j-1} + \|H\|_{\mathrm{C}(\overline{J_T}, X^\alpha_{T_p})} = N_j.
\end{aligned}
$$

Consequently, $\overline{B_{\mathrm{C}(\overline{J_{t_j}}; X^\alpha_{T_p})}(H, N_j)}$ is invariant under g_j and we obtain a (unique) fixed point $\zeta_j \in \overline{B_{\mathrm{C}(\overline{J_{t_j}}; X^\alpha_{T_p})}(H, N_j)}$ of g_j.

Lastly, we prove that ζ_k provides a fixed point of g. By definition of g_1 there holds

$$\zeta_2(t) = \zeta_1(t) = g_2(\zeta_1)(t)$$

for $t \in \overline{J_{t_1}}$ and by definition of g_2 we obtain

$$\begin{aligned}\zeta_2(t) =& e^{-T_p t_1} h_0 + \int_0^{t_1} e^{-T_p(t-s)}[F'[y;\zeta_1](s) + h(s)]\,ds \\ &+ e^{-T_p(t-t_1)} h_0 + \int_{t_1}^{t} e^{-T_p(t-s)}[F'[y_0;\zeta_2](s) + h(s)]\,ds\end{aligned}$$

for $t \in [t_1, t_2]$. Hence,

$$\zeta_2(t) = e^{-T_p t} h_0 + \int_0^t e^{-T_p(t-s)}[F'[y;\zeta_2](s) + h(s)]\,ds = g_2(\zeta_2)(t) \quad \text{for } t \in \overline{J_{t_2}}.$$

Inductively, it follows that $\zeta = \zeta(h_0, h) := \zeta_k$ is the unique solution in $\mathrm{C}(\overline{J_T}, X^\alpha_{T_p})$ of the integral equation

$$\zeta(t) = e^{-T_p t} h_0 + \int_0^t e^{-T_p(t-s)}[F'[y;\zeta](s) + h(s)]\,ds,$$

i.e. a fixed point of g. $\square$

Lemmas 3.12–3.14 provide us the necessary background to prove Theorem 3.11.

Proof of Theorem 3.11. We prove the theorem in two steps.
(I): Hadamard directional differentiability into $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$:
Let any $(y_0, u) \in X^\alpha_{T_p} \times \mathrm{L}^q(J_T; X)$ be given and consider $y = G(y_0, u)$. For $(h_0, h) \in X^\alpha_{T_p} \times \mathrm{L}^q(J_T; X)$ and $\lambda > 0$ we denote $y_\lambda := G(y_0 + \lambda h_0, u + \lambda h)$. Let $\zeta = \zeta(y, h_0, h) \in \mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ be the unique mild solution of (3.16) which exists according to Lemma 3.14.
We make use of (2.24) and (3.3) so that

$$\|T_p^\alpha \exp(-T_p t)\|_{\mathcal{L}(X)} \le C_\alpha t^{-\alpha} \exp(-\delta t), \qquad \text{and}$$

$$\left(\int_0^t (t-s)^{-\alpha q'}\,ds\right)^{1/q'} = \left(\frac{t^{1-\alpha q'}}{1-\alpha q'}\right)^{1/q'} = \frac{t^{1/q'-\alpha}}{(1-\alpha q')^{1/q'-\alpha}}.$$

Moreover, estimate (3.9) in Step III of the proof of Theorem 3.2 about local Lipschitz continuity of F from $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ to $\mathrm{C}(\overline{J_T}, X)$ and local Lipschitz continuity of G from $\mathrm{L}^q(J_T; X)$ to $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$ (see Theorem 3.2) lead to an estimate of the form

$$\begin{aligned}&\left\|\frac{(F[y+\lambda\zeta])(t) - (F[y_\lambda])(t)}{\lambda}\right\|_X \\ &= \left\|\frac{(F[G(y_0,u)+\lambda\zeta])(t) - (F[G(y_0+\lambda h_0, u+\lambda h)](t)}{\lambda}\right\|_X \\ &\le L(y) \sup_{0\le\tau'\le t} \left\|\frac{y_\lambda(\tau') - y(\tau')}{\lambda} - \zeta(\tau')\right\|_{X^\alpha_{T_p}}\end{aligned} \tag{3.19}$$

for a.e. $t \in J_T$ and λ small enough. Similar as in [MS15, Theorem 3.2], we employ the representation formulas for y, y_λ and ζ to estimate

$$
\begin{aligned}
&\left\| \frac{y_\lambda(t) - y(t)}{\lambda} - \zeta(t) \right\|_{X^\alpha_{T_p}} \\
&\leq c_1 \max_{\tau \in \overline{J_t}} \{e^{-\delta\tau}\} \int\limits_0^t (t-s)^{-\alpha} \left(\left\| \frac{(F[y+\lambda\zeta])(s) - (F[y])(s)}{\lambda} - F'[y;\zeta](s) \right\|_X \right. \\
&+ \left. \left\| \frac{(F[y+\lambda\zeta])(s) - (F[y_\lambda])(s)}{\lambda} \right\|_X \right) ds \\
&\leq c_2 \left(t^{1/q'-\alpha} \left\| \frac{F[y+\lambda\zeta] - F[y]}{\lambda} - F'[y;\zeta] \right\|_{\mathrm{L}^q(J_T;X)} \right. \\
&\left. + L(y) \int\limits_0^t (t-s)^{-\alpha} \sup_{0 \leq \tau' \leq s} \left\| \frac{y_\lambda(\tau') - y(\tau')}{\lambda} - \zeta(\tau') \right\|_{X^\alpha_{T_p}} ds \right)
\end{aligned}
$$

for constants $c_1, c_2 > 0$ and $\lambda > 0$ small enough. The first term converges to zero with $\lambda \to 0$ by Lemma 3.13. The estimate of the second term holds by (3.19).
We take the supremum over all $\tau \in \overline{J_t}$ on both sides, which leaves the right hand side unchanged. An application of Gronwall's lemma then implies the convergence of $\frac{y_\lambda - y}{\lambda}$ to ζ in $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$. Consequently, ζ is the directional derivative of G at (y_0, u) in direction (h_0, h).
Because G is also locally Lipschitz continuous, we finally conclude from Lemma 3.8 that the solution mapping for problem (3.1) is Hadamard directionally differentiable from $X^\alpha_{T_p} \times \mathrm{L}^q(J_T; X)$ to $\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})$.
(II) Hadamard directional differentiability into $Y_{T_p,s}$:
Assume that T_p satisfies maximal parabolic Sobolev regularity on X, see Definition 2.34, and suppose $y_0 \in X^\beta_{T_p}$ for some $\beta \in [\alpha, 1]$. Moreover, let s be arbitrary in the interval $\left(1, \frac{1}{1-\beta}\right) \cap (1, q]$ if $\beta \in [\alpha, 1)$ and $s = q$ if $\beta = 1$. As in Step I we denote $y = G(y_0, u)$ and for $(h_0, h) \in X^\beta_{T_p} \times \mathrm{L}^q(J_T; X)$ and $\lambda > 0$ we write $y_\lambda := G(y_0 + \lambda h_0, u + \lambda h)$. That the function

$$
\zeta(t) = \zeta(y, h_0, h)(t) = e^{-T_p t} h_0 + \int\limits_0^t e^{-T_p(t-s)} [F'[y;\zeta](s) + h(s)]\, ds
$$

is contained in $Y_{T_p,s}$ follows as in Theorem 3.2 from $F'[y;\zeta] + h \in \mathrm{L}^q(J_T; X)$ and $h_0 \in X^\beta_{T_p}$. Indeed, $\zeta = \left(\frac{d}{dt} + T_p, \gamma_0\right)^{-1} (F'[y;\zeta] + h, h_0)$ with γ_0 from Remark 3.3. From Step I we know that

$$
\frac{y_\lambda - y}{\lambda} \to \zeta \Leftrightarrow y + \lambda\zeta - y_\lambda \to 0 \quad \text{in } \mathrm{C}(\overline{J_T}; X^\alpha_{T_p}),
$$

with $\lambda \to 0$ and hence the convergence holds in $\mathrm{L}^s(J_T; X)$ as well. Moreover, note that

$$
\begin{aligned}
&\frac{y_\lambda - y}{\lambda} - \zeta \\
&= \left(\frac{d}{dt} + T_p, \gamma_0\right)^{-1} \left(\frac{F[y_\lambda] + u + \lambda h - (F[y] + u)}{\lambda} - (F'[y;\zeta] + h), \frac{y_0 + \lambda h_0 - y_0}{\lambda} - h_0 \right) \\
&= \left(\frac{d}{dt} + T_p, \gamma_0\right)^{-1} \left(\frac{F[y_\lambda] - F[y]}{\lambda} - F'[y;\zeta], 0 \right).
\end{aligned}
$$

These observations and Lemma 3.13 on the Hadamard directional differentiability and local Lipschitz continuity of $F : \mathrm{C}(\overline{J_T}; X^\alpha_{T_p}) \to \mathrm{L}^q(J_T;X) \hookrightarrow \mathrm{L}^s(J_T;X)$ yield that

$$
\begin{aligned}
&\left\| \frac{y_\lambda - y}{\lambda} - \zeta \right\|_{Y_{T_p,s}} \\
&= \left\| \left(\frac{d}{dt} + T_p, \gamma_0 \right)^{-1} \left(\frac{F[y_\lambda] - F[y]}{\lambda} - F'[y;\zeta], 0 \right) \right\|_{Y_{T_p,s}} \\
&\leq \left\| \left(\frac{d}{dt} + T_p, \gamma_0 \right)^{-1} \right\|_{\mathcal{L}\left(\mathrm{L}^s(J_T;X)\times X^\beta_{T_p}, Y_{T_p,s}\right)} \left\| \frac{F[y_\lambda] - F[y]}{\lambda} - F'[y;\zeta] \right\|_{\mathrm{L}^s(J_T;X)} \\
&\leq c \left(\left\| \frac{F[y+\lambda\zeta] - F[y]}{\lambda} - F'[y;\zeta] \right\|_{\mathrm{L}^s(J_T;X)} + \left\| \frac{F[y+\lambda\zeta] - F[y_\lambda]}{\lambda} \right\|_{\mathrm{L}^s(J_T;X)} \right) \\
&\leq c \left(\left\| \frac{F[y+\lambda\zeta] - F[y]}{\lambda} - F'[y;\zeta] \right\|_{\mathrm{L}^s(J_T;X)} + L(y) \left\| \frac{y_\lambda - y}{\lambda} - \zeta \right\|_{\mathrm{C}(\overline{J_T}; X^\alpha_{T_p})} \right) \to 0
\end{aligned}
$$

with $\lambda \to 0$. [Cf. MS15, Theorem 3.2] for a similar argument with zero initial values. $\square$

The following Corollary 3.15 is a generalization of [Mün17a, Theorem 4.7] from zero initial value to initial values $y_0 \in X^\beta_{A_p+1}$.

Corollary 3.15. Consider the assumptions and the notation as in Corollary 3.4. For $q \in \left(\frac{1}{1-\alpha}, \infty\right)$ let s be arbitrary in the interval $\left(1, \frac{1}{1-\beta}\right) \cap (1,q]$ if $\beta \in [\alpha, 1)$ and $s = q$ if $\beta = 1$. Then the solution operator $G : X^\beta_{A_p+1} \times \mathrm{L}^q(J_T;X) \to Y_{A_p,s}$ of problem (3.11) is Hadamard directionally differentiable. Its derivative $y^{(y_0,u),(h_0,h)} := G'[(y_0,u);(h_0,h)]$ at $(y_0,u) \in X^\beta_{A_p+1} \times \mathrm{L}^q(J_T;X)$ in direction $(h_0,h) \in X^\beta_{A_p+1} \times \mathrm{L}^q(J_T;X)$ is given by the unique solution $\zeta \in Y_{A_p,s}$ of

$$
\begin{aligned}
\dot{\zeta}(t) + (A_p\zeta)(t) &= F'[y;\zeta](t) + h(t) \text{ in } J_T, \\
\zeta(0) &= h_0,
\end{aligned}
\tag{3.20}
$$

where $F'[y;\zeta](t) = f'[(y(t), \mathcal{W}[Sy](t)); (\zeta(t), \mathcal{W}'[Sy; S\zeta](t))]$ and $y = G(y_0,u)$, see Corollary 3.4. Moreover, the mapping $(h_0,h) \mapsto G'[(y_0,u);(h_0,h)]$ is Lipschitz continuous from $X^\beta_{A_p+1} \times \mathrm{L}^q(J_T;X)$ to $Y_{A_p,s}$ with a modulus of continuity $c = C(G(y_0,u),T)$.

Proof. The proof follows the lines of Theorem 3.11. As in Corollary 3.4 note that A_p satisfies maximal parabolic Sobolev regularity on X. Again, instead of the estimates in (2.21) we have to use the estimates in (3.13) and (2.24)–(2.25) have to replaced by (3.14)–(3.15). $\square$

4 Optimal control of hysteresis-reaction-diffusion systems

In this chapter, we apply the results from Corollary 3.4 and Corollary 3.15 to an optimal control problem. In particular, we discuss the control problem from [Mün17a, Section 5] which was further analyzed in [Mün17b]. Accordingly, in the whole section, we specialize on diffusion operators A_p and zero initial value, i.e. $y_0 = 0$. We adapt the results of [Mün17a, Section 5] and [Mün17b] to the structure of this work. Moreover, several of the proofs are given in more detail. As described in the introduction, we also extend [Mün17b] by Subsection 4.4.5.

More precisely, we consider two spaces of control functions. In one case, the control u is contained in

$$U_1 := \mathrm{L}^2\left(J_T; \tilde{U}_1\right) := \mathrm{L}^2\left(J_T; [\mathrm{L}^2(\Omega)]^m\right).$$

Hence, u acts inside the domain. In the other case,

$$u \in U_2 := \mathrm{L}^2\left(J_T; \tilde{U}_2\right) := \mathrm{L}^2\left(J_T; \prod_{i=1}^{m} \mathrm{L}^2(\Gamma_{N_i}, \mathcal{H}_{d-1})\right)$$

is a Neumann boundary control.
U_1 and U_2 are embedded into $\mathrm{L}^2(J_T; X)$ by continuous operators $B_1 : \tilde{U}_1 \to X$ and $B_2 : \tilde{U}_2 \to X$, see (A6) below.
For $u \in U_i$, $i \in \{1, 2\}$, the corresponding state equation will be defined by $y = G(0, B_i u)$. The latter is well-posed by Corollary 3.4. Since the initial value will be zero in the whole chapter, we abbreviate

$$G(0, \cdot) := G(\cdot).$$

With the representation of $z = \mathcal{W}[Sy]$ as in Definition 2.38, $y = G(B_i u)$ is equivalent to

$$\begin{aligned}
\dot{y}(t) + A_p y(t) &= f(y(t), z(t)) + B_i u(t) && \text{in } X \text{ for } t \in J_T, && (4.1)\\
y(0) &= 0 && \text{in } X, &&\\
(\dot{z}(t) - S\dot{y}(t))(z(t) - \xi) &\leq 0 && \text{for } \xi \in [a, b] \text{ and } t \in J_T, && (4.2)\\
z(t) &\in [a, b] && \text{for } t \in \overline{J_T}, &&\\
z(0) &= z_0. && &&
\end{aligned}$$

For $i \in \{1, 2\}$, $y_d \in U_1$ and $\kappa > 0$ we define the tracking type optimal control problem

$$\begin{aligned}
\min_{u \in U_i} J(y, u) &:= \frac{1}{2}\|y - y_d\|_{U_1}^2 + \frac{\kappa}{2}\|u\|_{U_i}^2 \\
&= \frac{1}{2}\int_0^T \|y(s) - y_d(s)\|_{[\mathrm{L}^2(\Omega)]^m}^2 \, ds + \frac{\kappa}{2}\int_0^T \|u(s)\|_{\tilde{U}_i}^2 \, ds
\end{aligned} \tag{4.3}$$

subject to (4.1)–(4.2).

Remark 4.1. The functions $y = G(B_i u) \in Y_{2,0} \hookrightarrow \mathrm{L}^2(J_T; \mathrm{dom}(A_p))$ in (4.3) are identified with $I_p^{-1} G(B_i u) \in \mathrm{L}^2(J_T; \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)) \hookrightarrow U_1$ for $u \in U_i$, according to Corollary 2.30, see also Remark 2.32. We often write $G(B_i u)$ for both functions.

The outline of this chapter is the following:
The main assumption and some further notation in addition to that of Section 3.1 are given in Section 4.1.
In Section 4.2, we apply the continuity results on $G(B_i \cdot)$ from Corollary 3.4 and prove existence of an optimal control $\overline{u}$ for (4.1)–(4.3), together with the optimal state $\overline{y} = G(B_i \overline{u})$ and $\overline{z} = \mathcal{W}[S\overline{y}]$, see Theorem 4.6, cf. [Mün17a, Section 5].
Afterwards, we continue with the results from [Mün17b].
Our first main interest here is to derive an adjoint system (p, q) and first order necessary optimality conditions for problem (4.1)–(4.3). Corollary 3.15 yields Hadamard directional differentiability of the reduced cost function, but since the mapping $h \mapsto G'[B_i \overline{u}; B_i h]$ is not linear, we can not apply standard techniques to obtain (p, q).

Hence, in Section 4.3, we regularize the non-linearity f and the stop operator $\mathcal{W}$ in dependence of a parameter $\varepsilon > 0$ and consider a regularized control problem first. Since the regularization $y \mapsto Z_\varepsilon(Sy)$ of $y \mapsto \mathcal{W}[Sy]$ as well as f_ε are Gâteaux-differentiable, Corollary 3.15 implies that the same holds for the control-to-state operator $u \mapsto G_\varepsilon(B_i u)$, $i \in \{1, 2\}$. Moreover, since f_ε and Z_ε satisfy the properties of f and $\mathcal{W}$ (they are even smoother), we exploit Theorem 4.6 and conclude the existence of optimal solutions $\overline{u}_\varepsilon$, $\overline{y}_\varepsilon = G_\varepsilon(B_i\overline{u}_\varepsilon)$ and $\overline{z}_\varepsilon = Z_\varepsilon(S\overline{y}_\varepsilon)$ of the regularized problems, see Corollary 4.11.

In Section 4.4, we drive the regularization parameter to zero, i.e. we consider the limit $\varepsilon \to 0$. In particular, we use weak compactness arguments to derive a subsequence $(\overline{u}_{\varepsilon_k}, \overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k})$ which converges to a solution $(\overline{u}, \overline{y}, \overline{z})$ of the original problem. Nevertheless, even to derive optimality systems for the regularized problems remains challenging. The main problem here appears to characterize the adjoint operator $[G'_\varepsilon[u; \cdot]]^*$, since the mapping $G'_\varepsilon[u; \cdot]$ is only defined implicitly via the solution of a reaction-diffusion system. The adjoint equations, i.e. the evolution equations of p_ε and q_ε and an optimality condition for the regularized problem are stated in Theorem 4.20. In particular, Theorem 4.20 provides a relation between $(p_\varepsilon, q_\varepsilon)$ and $\overline{u}_\varepsilon$ and $\overline{u}$. Once the adjoint systems $(p_\varepsilon, q_\varepsilon)$ are obtained, driving the regularization parameter to zero yields an adjoint system (p, q) for problem (4.1)–(4.3). The evolution equation for p is obtained without much effort. Most of the difficulties are due to the low regularity of the adjoint variable q which belongs to $\overline{z}$. Unfortunately, uniform-in-ε bounds for $\dot{q}_\varepsilon$ only hold in $\mathrm{L}^1(J_T)$. Consequently, we have to pass to weak star convergence in $\mathrm{C}(\overline{J_T})^*$. Accordingly, q is only contained in $\mathrm{BV}(J_T)$, the space of functions with bounded total variation in $\overline{J_T}$. Hence, we cannot expect a time derivative of q, but only obtain a measure $dq \in \mathrm{C}(\overline{J_T})^*$ which describes the evolution of q. In order to complete the optimality system, we study q and dq in more detail. A lot of the properties of q and the corresponding measure dq can be proven. But the measure equation for dq still depends on an abstract measure $d\mu \in \mathrm{C}(\overline{J_T})^*$. Furthermore, the optimality conditions for problem (4.1)–(4.3) include $d\mu$, which makes it even more interesting to characterize the measure completely. Although it is shown in [Mün17b] that $d\mu$ has its support only in a part of $\overline{J_T}$, a complete characterization could not be established. With an additional regularity Assumption 4.30, the support of $d\mu$ can be further reduced. In extension to [Mün17b], we provide an example in which Assumption 4.30 is satisfied, see Example 4.32.

In Subsection 4.4.3, we summarize the results from Subsections 4.4.1–4.4.2. In particular, Theorem 4.38 contains the optimality conditions for problem (4.1)–(4.3) for $i \in \{1, 2\}$ and for general f. Corollary 4.39 states the optimality conditions for continuously differentiable f.

In the remaining part of Section 4.4, we concentrate on problem (4.1)–(4.3) with distributed control functions.

In Subsection 4.4.4, the optimality conditions are improved for $i = 1$, see Corollary 4.40. Moreover, we show uniqueness of p, q and $d\mu$ in Corollary 4.41. To this aim, we exploit that B_1 has dense range for appropriate $p \geq 2$ which does not hold for B_2 and which implies that the operator B_1^* is one-to-one.

In Subsection 4.4.5, we extend the results from [Mün17b] on $d\mu$. In particular, we investigate in the analysis of $d\mu$ for the case $i = 1$. We characterize the sign of $d\mu(E)$ for subsets $E \subset \overline{J_T}$ of different categories and prove upper bounds for $|d\mu(E)|$, see Lemma 4.46 and Theorem 4.47. With help of the measure equation for dq we conclude sign conditions and bounds for dq on subsets $E \subset \overline{J_T}$ of different categories, see Corollary 4.48. Finally, we characterize the continuity properties of q in Corollary 4.49.

In Section 4.5, we return to the general control problem (4.1)–(4.3) with $i \in \{1, 2\}$. We take advantage of relation between (p, q) and $\overline{u}$ and exploit the continuity properties of the adjoint variables to prove higher regularity of $\overline{u}$, $\overline{y}$ and $\overline{z}$, see Theorem 4.51. An application of Theorem 4.51 is given in Example 4.52.

In Section 4.6, we study the perturbed control problem (1.5)-(1.7), cf. [Mün17b, Section 6]. In particular, the set of admissible control functions is restricted to a subset $C \subset U_i$, $i \in \{1,2\}$, and (4.1)–(4.3) is perturbed by a function $r \in U_i$. In Theorem 4.54, we prove lower semi-continuity of the corresponding optimal value function $v : U_i \to C$ if C is convex and closed. If C is also compact, we show that v is continuous, and that the corresponding optimal set function $V : U_i \rightrightarrows C$ is upper semi-continuous.

4.1 Main assumption and further notation

Assumption 4.2. Cf. [Mün17b, Assumption 2.10] We adapt (A1)-(A3) in Assumption 3.1, repeat Assumption 3.10 and introduce two more assumptions for the optimal control problem (4.1)–(4.3):

(A0) $\Omega \subset \mathbb{R}^d$ for some $d \geq 2$.

(A1)' Instead of (A1) suppose that A_p is a diffusion operator in the sense of Definition 2.17, where $p \in \mathrm{J} \cap [2,\infty)$ with J as in Corollary 2.19. Moreover, assume $2 \geq p\left(1-\frac{1}{d}\right)$.

(A2)' Scalar projection: The function $w \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)\backslash\{0\}$ in (A2) which defines the operator $S \in [\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)]^*$ by $Sy = \langle y, w\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)}$ $\forall y \in \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ is contained in the space $\mathrm{dom}([(1+A_p)^{1-\alpha}]^*)$.

(A3)' (A3) holds for a coefficient $\alpha \in \left(0, \frac{1}{2}\right)$. This assumption is needed in the proof of Lemma 4.5.

(A4) In addition to (A3)', f is directionally differentiable and therefore Hadamard directionally differentiable.

(A5) B_1 is defined by

$$B_1 : [\mathrm{L}^2(\Omega)]^m \to X, \quad \langle B_1 u, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} := \int_\Omega u \cdot v \, dx \quad \forall v \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega).$$

Since $2 \geq p\left(1-\frac{1}{d}\right)$, the embeddings $\mathrm{L}^2(\Gamma_{N_j}, \mathcal{H}_{d-1}) \hookrightarrow \mathrm{W}^{-1,p}_{\Gamma_{D_j}}(\Omega)$, $j \in \{1,\dots,m\}$, are continuous [Hal+15, Remark 5.11]. Therefore,

$$B_2 : \prod_{j=1}^m \mathrm{L}^2(\Gamma_{N_j}, \mathcal{H}_{d-1}) \to X, \quad \langle B_2 y, v\rangle_{\mathbb{W}^{1,p'}(\Omega)} := \sum_{j=1}^m \int_{\Gamma_{N_j}} y_j v_j \, d\mathcal{H}_{d-1} \quad \forall v \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$$

is continuous.

(A6) The desired state y_d in (4.3) is contained in U_1 and $\kappa > 0$ is given.

Remark 4.3. We prove that B_1 in Assumption 4.2 is well defined. To this aim, note that $p \in [2,\infty) \Leftrightarrow p' \in (1,2]$. Hence, either $d = 2 = p'$ or $p' < 2 < \frac{dp'}{d-p'}$. By Remark 2.7 and the Riesz representation $[\mathrm{L}^2(\Omega)]^m$ of $[\tilde{U}_1]^*$ we obtain the compact embedding

$$\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega) \hookrightarrow\hookrightarrow [\mathrm{L}^2(\Omega)]^m = [\tilde{U}_1]^* \quad \text{and hence} \quad \tilde{U}_1 \hookrightarrow \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega).$$

More precisely, each function $u \in \tilde{U}_1$ defines an element in $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ by the assignment

$$\langle B_1 u, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} = \int_\Omega u \cdot v \, dx \quad \forall v \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega).$$

B_1 is well defined, since for $u \in \tilde{U}_1$ and $\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$, the embedding $\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega) \hookrightarrow [\mathrm{L}^2(\Omega)]^m$ together with Hölders inequality imply

$$\begin{aligned}\left|\langle B_1 u, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)}\right| &= \left|\int_\Omega u \cdot v \, dx\right| \le \|u\|_{[\mathrm{L}^2(\Omega)]^m} \|v\|_{[\mathrm{L}^2(\Omega)]^m} \le \|u\|_{[\mathrm{L}^2(\Omega)]^m} \|v\|_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} \\ &= \|u\|_{\tilde{U}_1} \|v\|_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)}.\end{aligned}$$

Since p and A_p are fixed by (A1)', we extend the notation (N1)-(N5) in Section 3.1 to

(N1) For the particular p from Assumption 3.1 we set

$$X := \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$$

with $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ from Definition 2.8. We sometimes identify elements $v \in X^*$ with their representation in $\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$, i.e.

$$\langle v, y\rangle_X = \langle y, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} \ \forall y \in X.$$

(N2)' We write $X^\beta := X^\beta_{A_p+1}$ for $\beta \ge 0$.

(N3)' For $t \in [0, T]$ we denote

$$\begin{aligned} Y_q &:= Y_{A_p+1,q} = \mathrm{W}^{1,q}((0,T); X) \cap \mathrm{L}^q((0,T); \mathrm{dom}(A_p)), \\ Y_{q,t} &:= Y_{A_p+1,q,t} = \{y \in Y_{A_p+1,q} : \ y(t) = 0\}, \\ Y^*_{q,t} &:= Y^*_{A_p+1,q,t} = \{y \in \mathrm{W}^{1,q}(0,T; [\mathrm{dom}(A_p)]^*) \cap \mathrm{L}^q((0,T); \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)) : \ y(t) = 0\}. \end{aligned}$$

Note that $\mathrm{dom}(A_p)$ is equipped with the operator norm of A_p, because A_p is not necessarily one-to-one. Moreover, $\mathrm{dom}(A_p)$ can be identified with $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$, see Remark 2.32.

(N4) $\mathcal{W}$ is a scalar stop operator as defined in Definition 2.38 for some prescribed initial value $z_0 \in [a, b]$.

(N5) We abbreviate $J_T = (0, T)$.

(N6) We write $G(\cdot) := G(0, \cdot)$ for G from Corollary 3.4.

Remark 4.4. [Mün17a, Remark 5.2] Corollary 3.15 yields Hadamard directional differentiability of $G \circ B_i : U_i \to Y_{2,0}$ for $i \in \{1, 2\}$ and $(y, z) = (G(B_i u), \mathcal{W}[SG(B_i u)])$ solves (4.1)–(4.2) for $u \in U_i$. Therefore, the reduced cost function $\mathcal{J} : U_i \to \mathbb{R}$, $\mathcal{J}(u) := J(G(B_i u), u)$ is Hadamard directionally differentiable. Remember Remark 4.1, namely that $G(B_i u)$ is identified with $I_p^{-1} G(B_i u)$ for $u \in U_i$.

4.2 Existence of an optimal control

We prove that an optimal control for problem (4.1)–(4.3) exists. To this aim, we show that the mapping $u \mapsto G(B_i u)$ is weakly continuous from U_i into $Y_{2,0}$ and weak-strong continuous from U_i into $\mathrm{C}(\overline{J_T}; X^\alpha)$.

Lemma 4.5. [Mün17a, Lemma 5.3], cf. [BK13, Lemma 2.3]. Let Assumption 4.2 hold. Suppose that for $\{u_n\}_{n\in\mathbb{N}} \subset U_i$ it holds $u_n \rightharpoonup u$ in U_i with $i \in \{1, 2\}$. Then $y_n = G(B_i u_n) \to G(B_i u)$ weakly in $Y_{2,0}$ and strongly in $\mathrm{C}(\overline{J_T}; X^\alpha)$ and $z_n = \mathcal{W}[S y_n] \to \mathcal{W}[SG(B_i u)]$ weakly in $\mathrm{H}^1(J_T)$ and strongly in $\mathrm{C}(\overline{J_T})$. If the convergence of u_n is strong then $y_n \to G(B_i u)$ in $Y_{2,0}$ strongly.

Proof. The proof is a combination of the proofs of [MS15, Lemma 2.10] and [BK13, Lemma 2.3]. Let $u_n \rightharpoonup u$ in U_i with $n \to \infty$. By (A3)' in Assumption 4.2 we have $\alpha \in (0, \frac{1}{2})$ so that $\frac{1}{1-\alpha} < 2 = q$. We apply Corollary 3.4 and Corollary 3.15 with u and h replaced by $B_i u$ and $B_i h$ and with $\mathrm{L}^2(J_T; X)$ replaced by U_i. In particular, by (3.12) there exists some $C(T) > 0$ such that

$$\|y_n\|_{Y_{2,0}} \leq C(T)(1 + \|B_i u_n\|_{\mathrm{L}^2(J_T;X)}).$$

By uniform boundedness of $\{u_n\}_{n\in\mathbb{N}}$ and weak compactness, there exists a subsequence y_{n_k} which weakly converges in $Y_{2,0}$ to some y with $k \to \infty$. Lemma 2.36 and $\alpha < \frac{1}{2}$ entail that $Y_{2,0}$ is compactly embedded into the space $\mathrm{C}(\overline{J_T}; [X, \mathrm{dom}(A_p)]_\alpha)$ and Remark 2.32 yields $\mathrm{C}(\overline{J_T}; [X, \mathrm{dom}(A_p)]_\alpha) \simeq \mathrm{C}(\overline{J_T}; X^\alpha)$. Hence, $y_{n_k} \to y$ in $\mathrm{C}(\overline{J_T}; X^\alpha)$ strongly with $k \to \infty$. Moreover, Sy_{n_k} converges weakly to Sy in $\mathrm{H}^1(J_T)$ because $S \in X^*$ by (A2)'. By Theorem 2.40, $\mathcal{W}$ is weakly continuous on $\mathrm{H}^1(J_T) \hookrightarrow \mathrm{C}(\overline{J_T})$. Consequently, weak convergence of Sy_{n_k} to Sy implies weak convergence of z_{n_k} to $\mathcal{W}[Sy] = z$ in $\mathrm{H}^1(J_T)$ with $k \to \infty$ and then also strong convergence in $\mathrm{C}(\overline{J_T})$. It remains to prove $y = G(B_i u)$. Since $\frac{d}{dt}$, A_p and B_i are weakly continuous by linearity [MS15, Lemma 2.10], we obtain

$$\frac{d}{dt} y_{n_k} + A_p y_{n_k} \rightharpoonup \frac{d}{dt} y + A_p y \quad \text{and } B_i u_{n_k} \rightharpoonup B_i u \quad \text{in } \mathrm{L}^2(J_T; X) \quad \text{with } k \to \infty.$$

By strong convergence of $y_{n_k} \to y$ in $\mathrm{C}(\overline{J_T}; X^\alpha)$ with $k \to \infty$ and local Lipschitz continuity of $f : \mathrm{C}(\overline{J_T}; X^\alpha) \times \mathbb{R} \to \mathrm{C}(\overline{J_T}; X)$ with respect to the $\mathrm{C}(\overline{J_T}; X^\alpha)$-norm (see Step III in the proof of Theorem 3.2), we can find $k_0 > 0$ such that

$$\|f(y_{n_k}, z_{n_k}) - f(y, z)\|_{\mathrm{C}(\overline{J_T};X)} \leq L(y)(\|y_{n_k} - y\|_{\mathrm{C}(\overline{J_T};X^\alpha)} + \|z_{n_k} - z\|_{\mathrm{C}(\overline{J_T})})$$

for all $k \geq k_0$. Consequently, $f(y_{n_k}(\cdot), z_{n_k}(\cdot))$ converges to $f(y(\cdot), z(\cdot))$ in $\mathrm{C}(\overline{J_T}; X)$ with $k \to \infty$. We pass to the limit $k \to \infty$ in (3.11) to see that y solves (3.11) with forcing term $B_i u$. This implies $y = G(B_i u)$ and $z = \mathcal{W}[Sy]$. Since $y = G(B_i u)$ and $z = \mathcal{W}[Sy]$ are uniquely determined, we conclude (weak) convergence of the whole sequence.
To complete the proof, assume now that $u_n \to u$ strongly in U_i with $n \to \infty$. In this case,

$$B_i u_n \to B_i u \quad \text{in } \mathrm{L}^2(J_T; X) \quad \text{with } n \to \infty,$$

and hence, by maximal parabolic regularity of A_p (see Remark 2.35),

$$\|y_n - y\|_{Y_{2,0}} \leq \left\| \left(\frac{d}{dt} + A_p\right)^{-1} \right\|_{\mathcal{L}(\mathrm{L}^2(J_T;X), Y_{2,0})} \big(\|B_i(u_n - u)\|_{\mathrm{L}^2(J_T;X)} + \|F[y_n] - F[y]\|_{\mathrm{L}^2(J_T;X)}\big).$$

Since the right side converges to zero with $n \to \infty$, we conclude $y_n \to y$ in $Y_{2,0}$ with $n \to \infty$. □

With the preceding lemma, we can show existence of an optimal control.

Theorem 4.6. *[Mün17a, Theorem 5.4] Let Assumption 4.2 hold. Then for $i \in \{1, 2\}$, there exists an optimal control $\overline{u} \in U_i$ for the optimal control problem (4.1)–(4.3). This means that $\overline{u}$, together with the optimal state $\overline{y} = G(\overline{u})$, which solves (4.1), are a solution of the minimization problem (4.3). The solution of (4.2) is given by $\overline{z} = \mathcal{W}[S\overline{y}]$.*

Proof. The proof uses Lemma 4.5 and is similar to the proof of [MS15, Proposition 2.11]:
By definition, the cost function J is bounded from below by zero. Moreover, by the form of the reduced cost function

$$\mathcal{J}(u) = J(G(B_i u), u) = \frac{1}{2}\|G(B_i u) - y_d\|^2_{U_1} + \frac{\kappa}{2}\|u\|^2_{U_i},$$

every minimizing sequence $\{u_n\}$ of $\mathcal{J}$ is bounded in U_i and thus has a weakly converging subsequence, i.e. $u_{n_k} \rightharpoonup \overline{u} \in U_i$ with $k \to \infty$ for some $\overline{u} \in U_i$. By Lemma 4.5, $y_{n_k} = G(B_i u_{n_k}) \to G(B_i \overline{u}) =: \overline{y}$ weakly in $Y_{2,0}$ and strongly in $\mathrm{C}(\overline{J_T}; X^\alpha)$ with $k \to \infty$. Moreover, $Y_{2,0} \hookrightarrow \mathrm{L}^2(\overline{J_T}; \mathrm{dom}(A_p)) \hookrightarrow U_1$ according to Corollary 2.30, see also Remark 2.32. Note that $G(B_i u) \in Y_{2,0}$ is always identified with $I_p^{-1} G(B_i u) \in U_1$ here for $u \in U_i$, but we often write $G(B_i u)$ for both functions. The cost function $J : U_1 \times U_i \to \mathbb{R}$ is weakly lower semi-continuous. Consequently,

$$\mathcal{J}(\overline{u}) = J(\overline{y}, \overline{u}) \leq \liminf_{k \to \infty} J(y_{n_k}, u_{n_k}) = \min_{u \in U_i} J(G(B_i u), u) = \min_{u \in U_i} \mathcal{J}(u)$$

as required. □

4.3 Regularized control problem

As mentioned in the beginning of the chapter, the mapping $h \mapsto G'[B_i \overline{u}; B_i h]$ is not linear so that we can not apply standard theory in order to derive an adjoint system for problem (4.1)–(4.3). As a remedy, we fix an optimal control $\overline{u}$ and regularize the cost function as well as the state equation in dependence of a parameter $\varepsilon > 0$ to obtain an optimal control problem for which we can derive an adjoint system. Hence, we regularize the variational inequality which defines $\mathcal{W}$ and the non-linearity f and therewith obtain a regularization of the solution operator $G(B_i \cdot)$. In order to regularize $\mathcal{W}$ we apply techniques from singular perturbation theory as in [BK13, Section 3]. Our regularization of the semi-linear parabolic state equations is inspired by [MS15, Section 4].

With regard to the limit $\varepsilon \to 0$, we derive uniform-in-ε-estimates for the norms of the solutions of the regularized state equations in terms of the forcing term u. This is done in the end of Subsection 4.3.1.

In Subsection 4.3.2, we investigate in the dynamics of the regularized state equations in dependence of ε. For any weakly converging sequence u_ε, $\varepsilon \to 0$, the estimates from Subsection 4.3.1 together with a weak compactness argument yield us weakly converging subsequences y_{ε_k} and z_{ε_k}.

In Subsection 4.3.3, we introduce regularized control problems. We apply the results from Subsection 4.3.2 to conclude convergence of the corresponding solutions to an optimal solution of problem (4.1)–(4.3) with $\varepsilon \to 0$, see Theorem 4.16.

In Subsection 4.3.4, we prove Gâteaux differentiability of the regularized solution operators G_ε, see Corollary 4.17.

Subsection 4.3.5 is contains the adjoint equations for the solutions of the regularized control problems with $\varepsilon > 0$ fixed. The main result here is Theorem 4.20.

In Subsection 4.3.6, we prove uniform-in-ε bounds for the norms of the adjoint variables $(p_\varepsilon, q_\varepsilon)$ from Theorem 4.20.

The key step $\varepsilon \to 0$ is established in Section 4.4. The uniform bounds on $(p_\varepsilon, q_\varepsilon)$ from Subsection 4.3.6 give rise to weakly converging subsequences p_{ε_k} and q_{ε_k} and finally yield an adjoint system for (4.1)–(4.3).

We begin with several assumptions on the functions which will enter the regularized problems.

Assumption 4.7 (Regularization)**.** [Mün17b, Assumption 3.1] For $\varepsilon_* > 0$ and $\varepsilon \in (0, \varepsilon_*]$ we assume that:

$(A1)_\varepsilon$ $f_\varepsilon : X^\alpha \times \mathbb{R} \to X$ is Gâteaux differentiable.

$(A2)_\varepsilon$ $\sup_{(y,z) \in X^\alpha \times \mathbb{R}} \|f_\varepsilon(y,z) - f(y,z)\|_X \to 0$ as $\varepsilon \to 0$.

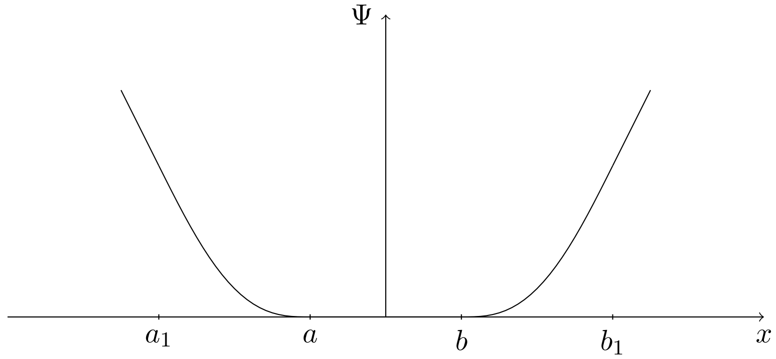

Figure 1: [Mün17b, Figure 1] Graph of Ψ

$(A3)_\varepsilon$ f_ε is locally Lipschitz continuous with respect to the X^α-norm and all the neighborhoods and Lipschitz constants are equal to the ones of f in (A2) in Assumption 4.2, independently of ε. The growth condition $\|f_\varepsilon(y,x)\|_X \leq M\,(1 + \|y\|_{X^\alpha} + |x|)$ holds for all $y \in X^\alpha$ and $x \in \mathbb{R}$, with M from (A3) in Assumption 3.1.

$(A4)_\varepsilon$ Following the ideas of [BK13], we introduce a convex function $\Psi : \mathbb{R} \to \mathbb{R}$ with $\Psi(x) \equiv 0$ for $x \in [a,b]$ and $\Psi(x) > 0$ for $x \in \mathbb{R}\backslash[a,b]$. We assume that Ψ is twice continuously differentiable and $\Psi'(x) \leq m_1|x-a|$ for some $m_1 > 0$ and all $x \in \mathbb{R}$. Moreover, $\Psi''(x) \leq m_2$ for some $m_2 > 0$ and all $x \in \mathbb{R}$ and Ψ'' is assumed to be locally Lipschitz continuous. If a concrete representation of Ψ is needed, we assume Ψ to be defined according to Remark 4.8 below.

Remark 4.8 (Construction of Ψ)**.** Cf. [Mün17b, Remark 3.2] We construct a function Ψ which satisfies $(A4)_\varepsilon$ in Assumption 4.7. The concrete structure of Ψ will be useful in several proofs. In particular, Ψ is the concatenation of four functions, namely

$$\Psi = \chi_{(-\infty,a_1]}\Psi_{-2} + \chi_{(a_1,a]}\Psi_{-1} + \chi_{(b,b_1]}\Psi_1 + \chi_{(b_1,\infty)}\Psi_2,$$

where $a_1 < a < b < b_1$. Here, χ denotes the characteristic function. The functions Ψ_{-2} and Ψ_2 are affine linear and Ψ_{-1} and Ψ_1 are polynomials of order four with roots in a and b which are at the same time saddle points and with turning points in a_1 and b_1.
More precisely, we choose $a_1 := a - 2$, $b_1 := b + 2$ and define a function similar to Figure 1 by

$$\begin{aligned}
\Psi_{-2}(x) &:= -16(x - 1 - a),\\
\Psi_{-1}(x) &:= -(x-a)^3(4 - a + x),\\
\Psi_1(x) &:= (x-b)^3(4 + b - x),\\
\Psi_2(x) &:= 16(x - 1 - b).
\end{aligned}$$

With this definition, Ψ fulfills the conditions $(A4)_\varepsilon$ in Assumption 4.7. Note that local Lipschitz continuity of Ψ'' holds also in the points where the functions $\Psi_{-2}, \Psi_{-1}, \Psi_1, \Psi_2$ are glued together. Hence, Ψ'' is Lipschitz continuous.

For $i \in \{1,2\}$ and $\varepsilon > 0$, the original state equations (4.1)–(4.2) are regularized as follows:

$$\dot{y}(t) + (A_p y)(t) = f_\varepsilon(y(t), z(t)) + (B_i u)(t) \text{ in } X \text{ for } t \in J_T,\ y(0) = 0 \text{ in } X, \tag{4.4}$$

$$\dot{z}(t) - S\dot{y}(t) = -\frac{1}{\varepsilon}\Psi'(z(t)) \qquad \text{for } t \in J_T,\ z(0) = z_0. \tag{4.5}$$

4.3.1 Regularization of (3.11) and uniform-in-ε estimates

In this subsection, we replace equation (3.11) by a regularized abstract evolution equation. The latter is essentially equivalent to the regularized state equations (4.4)–(4.5), but with forcing term $u \in \mathrm{L}^q(J_T; X)$ instead of $u \in U_i$. Hence, it can be transformed into (4.4)–(4.5) by choosing the forcing term as $B_i u$, $u \in U_i$. Once the appropriate regularity of all regularized functions has been shown, well-posedness follows as for (3.11). Afterwards, we continue to estimate the norms of the solutions in terms of the forcing term and independently of ε. The ideas for many of the steps in this subsection were inspired by [BK13, Subsection 3.1].

Definition 4.9 (Regularized stop)**.** Cf. [Mün17b, Definition 3.3] For $\varepsilon \in (0, \varepsilon_*]$ we denote by $Z_\varepsilon : v \mapsto Z_\varepsilon(v)$ the solution operator of

$$\dot{z}(t) - \dot{v}(t) = -\frac{1}{\varepsilon}\Psi'(z(t)) \quad \text{for } t \in J_T,\ z(0) = z_0,$$

or of the corresponding integral equation

$$z(t) = z_0 - v(0) + v(t) - \int_0^t \frac{1}{\varepsilon}\Psi'(z(s))\, ds \quad \text{for } t \in J_T.$$

The input v is a function defined on $\overline{J_T}$.

In extension to [Mün17b], we prove regularity results on Z_ε in Definition 4.9 which are required.

Lemma 4.10 (Regularized stop)**.** Cf. [Mün17b, Remark 3.4] Z_ε in Definition 4.9 is well defined and continuously differentiable on $\mathrm{C}(\overline{J_T})$. Its derivative at v in direction h is given by the unique solution $Z'_\varepsilon[v; h] = z$ of the integral equation

$$z(t) = -h(0) + h(t) - \int_0^t \frac{1}{\varepsilon}\Psi''(Z_\varepsilon(v)(s))z(s)ds \quad \text{for } t \in J_T.$$

Z_ε is bounded on $\mathrm{W}^{1,q}(J_T)$ for all $q \in (1, \infty)$ with derivative $Z'_\varepsilon[v; h] = z$,

$$\dot{z}(t) - \dot{h}(t) = -\frac{1}{\varepsilon}\Psi''(Z_\varepsilon(v)(t))z(t) \quad \text{for } t \in J_T,\ z(0) = 0.$$

Proof. We show the lemma in three steps.
(I) Well-posedness and Lipschitz continuity:
Because Ψ'' is globally bounded by $(A4)_\varepsilon$, the mean value theorem entails that Ψ' is Lipschitz continuous with modulus m. Now it follows with Gronwall's lemma and Schauder's fixed point theorem, that for each $v \in \mathrm{C}(\overline{J_T})$ there exists a unique local solution

$$z_\varepsilon = Z_\varepsilon(v) \in \{q \in \mathrm{C}([0, T_0]) : q(0) = v(0) \in \mathbb{R}\},$$

satisfying the integral equation

$$z_\varepsilon(t) = z_0 - v(0) + v(t) - \int_0^t \frac{1}{\varepsilon}\Psi'(z_\varepsilon(s))\, ds.$$

Because $\Psi'(z_\varepsilon(s)) \le m_1|z_\varepsilon(s) - a|$ on the interval of existence of z_ε, this solution can be extended to the whole interval $\overline{J_T}$. That is, $Z_\varepsilon : \mathrm{C}(\overline{J_T}) \to \mathrm{C}(\overline{J_T}), v \mapsto Z_\varepsilon(v)$ is well defined.
Again by $(A4)_\varepsilon$ and Gronwall's lemma, Z_ε is Lipschitz continuous with modulus $e^{mT/\varepsilon}$ and because $\Psi(z_0) = 0$ we have

$$|z_\varepsilon(t)| \le e^{mT/\varepsilon} \sup_{0\le s\le t} |v(t)| + |z_0| \quad \text{for } t \in \overline{J_T}. \tag{4.6}$$

(II) Differentiability:
The mapping $v \mapsto Z_\varepsilon(v)$ is continuously differentiable on $\mathrm{C}(\overline{J_T})$:
Let $v, h \in \mathrm{C}(\overline{J_T})$ be given. Note that the set

$$E_{v,h} := \cup_{t\in\overline{J_T}} E_{v,h}(t) := \cup_{t\in\overline{J_T}}\{x \in \mathbb{R} : \ x \text{ between } Z_\varepsilon(v)(t) \text{ and } Z_\varepsilon(v+h)(t)\}$$

is a bounded subset of $\mathbb{R}$, so that Ψ'' is Lipschitz continuous on this set. Moreover, Step I entails that for all $s \in \overline{J_T}$ there holds

$$|Z_\varepsilon(v+h)(s) - Z_\varepsilon(v)(s)| \le e^{mT/\varepsilon} \sup_{0\le s'\le s} |h(s)|. \tag{4.7}$$

Let $q(h)$ be the unique solution in $C(\overline{J_T})$ of the equation

$$q(t) = -h(0) + h(t) - \int_0^t \frac{1}{\varepsilon}\Psi''(Z_\varepsilon(v)(s))q(s)\,ds \qquad \text{for } t \in \overline{J_T}.$$

Existence of this solution follows from boundedness and local Lipschitz continuity of Ψ''.
By the mean value theorem we have

$$-\frac{1}{\varepsilon}[\Psi'(Z_\varepsilon(v+h)(t)) - \Psi'(Z_\varepsilon(v)(t))] = -\frac{1}{\varepsilon}\Psi''(\xi(t))(Z_\varepsilon(v+h)(t) - Z_\varepsilon(v)(t))$$

for all $t \in \overline{J_T}$ and for values $\xi(t) \in E_{v,h}(t)$. Hence, we can split the following expression

$$\begin{aligned}
&\frac{Z_\varepsilon(v+h)(t) - Z_\varepsilon(v)(t) - q(t)}{\|h\|_{\mathrm{C}(\overline{J_T})}} \\
&= -\int_0^t \frac{1}{\varepsilon}\Psi''(Z_\varepsilon(v)(s))\left(\frac{Z_\varepsilon(v+h)(s) - Z_\varepsilon(v)(s) - q(s)}{\|h\|_{\mathrm{C}(\overline{J_T})}}\right) ds \\
&- \int_0^t \frac{1}{\varepsilon}(\Psi''(\xi(s)) - \Psi''(Z_\varepsilon(v)(s)))\left(\frac{Z_\varepsilon(v+h)(s) - Z_\varepsilon(v)(s)}{\|h\|_{\mathrm{C}(\overline{J_T})}}\right) ds \\
&=: J_1 + J_2.
\end{aligned}$$

Lipschitz continuity of Ψ'' on bounded sets together with (4.7) and $\xi(s) \in E_{v,h}(s)$ yield

$$|J_2| \le \frac{c}{\|h\|_{\mathrm{C}(\overline{J_T})}} \int_0^t \frac{1}{\varepsilon}|Z_\varepsilon(v)(s) - Z_\varepsilon(v+h)(s)|^2 ds \le ce^{2mT/\varepsilon}T\|h\|_{\mathrm{C}(\overline{J_T})}$$

for some $c > 0$. Gronwalls's inequality hence implies that

$$\frac{|Z_\varepsilon(v)(t) - Z_\varepsilon(v+h)(t) - q(t)|}{\|h\|_{\mathrm{C}(\overline{J_T})}} \to 0$$

with $\|h\|_{\mathrm{C}(\overline{J_T})} \to 0$, uniformly in $t \in \overline{J_T}$. Because the mapping $h \mapsto q(h) = Z'_\varepsilon[v;h]$ is linear and bounded on $\mathrm{C}(\overline{J_T})$, it follows that Z_ε is continuously differentiable on $\mathrm{C}(\overline{J_T})$.
(III) Z_ε is bounded on $\mathrm{W}^{1,q}(J_T)$ for all $q \in (1,\infty)$:
First of all, note that $\mathrm{W}^{1,q}(J_T) \hookrightarrow \mathrm{C}(\overline{J_T})$. Hence, (4.6) implies

$$\|Z_\varepsilon(v)\|_{\mathrm{L}^q(J_T)} \le c\|Z_\varepsilon(v)\|_{\mathrm{C}(\overline{J_T})} \le c_1(\varepsilon,T)(\|v\|_{\mathrm{C}(\overline{J_T})} + |z_0|) \le c_2(\varepsilon,T)(1 + \|v\|_{\mathrm{W}^{1,q}(J_T)})$$

for constants $c, c_1(\varepsilon,T), c_2(\varepsilon,T) > 0$. Using this, $(A4)_\varepsilon$ and again (4.6) we estimate

$$\begin{aligned}
\|\dot{Z}_\varepsilon(v)\|_{\mathrm{L}^q(J_T)} &= \left\|\dot{v} - \frac{1}{\varepsilon}\Psi'(Z_\varepsilon(v))\right\|_{\mathrm{L}^q(J_T)} \le \|\dot{v}\|_{\mathrm{L}^q(J_T)} + \frac{1}{\varepsilon}m_1\|Z_\varepsilon(v) - a\|_{\mathrm{L}^q(J_T)} \\
&\le c_3(\varepsilon,T)(1 + \|v\|_{\mathrm{W}^{1,q}(J_T)})
\end{aligned}$$

for some constant $c_3(\varepsilon,T) > 0$. □

As for F in Section 3.2, we introduce the function $(F_\varepsilon(y))(t) := f_\varepsilon(y(t), Z_\varepsilon(Sy)(t))$ and study the abstract evolution equation

$$\begin{aligned} \dot{y}(t) + (A_p y)(t) &= (F_\varepsilon(y))(t) + u(t) \quad \text{in } X \quad \text{for } t > 0, \\ y(0) &= 0 \in X. \end{aligned} \tag{4.8}$$

Corollary 4.11 (Existence of regularized problem)**.** [Mün17b, Corollary 3.5] Let Assumption 4.2 and Assumption 4.7 hold and let $\varepsilon \in (0, \varepsilon_*]$ be arbitrary. Furthermore, assume $q \in (\frac{1}{1-\alpha}, \infty]$ and let $s \in (1, q] \cap (1, \infty)$ be arbitrary. Then for all $u \in \mathrm{L}^q(J_T; X)$ problem (4.8) has a unique solution $y_\varepsilon(u)$ in $Y_{s,0}$. The solution mapping $G_\varepsilon : u \mapsto y_\varepsilon(u) =: y_\varepsilon^u$ is locally Lipschitz continuous from $\mathrm{L}^q(J_T; X)$ to $\mathrm{C}(\overline{J_T}; X^\alpha)$ and to $Y_{s,0}$. We denote $z_\varepsilon^u := z_\varepsilon(u) := Z_\varepsilon(Sy_\varepsilon^u)$.

Proof. Existence of unique solutions of (4.8) and local Lipschitz continuity of the solution operator G_ε are a consequence of Lemma 4.10, $(A1)_\varepsilon$ and $(A3)_\varepsilon$. In particular, the functions Z_ε and f_ε satisfy the properties of $\mathcal{W}$ and f so that Theorem 4.11 can be applied. $\square$

Having shown well-posedness of (4.8), we derive uniform-in-ε-estimates for the solutions $G_\varepsilon(u)$ in terms of the forcing function $u \in \mathrm{L}^q(J_T; X)$. Those translate directly into uniform estimates for the solutions of (4.4)–(4.5) if u is replaced by $B_i u$.

Lemma 4.12 (Uniform bounds)**.** [Mün17b, Lemma 3.6] Adopt the assumptions and the notation from Corollary 4.11. There exists a constant $c > 0$ which is independent of ε and u such that the following holds true. For all $q \in (\frac{1}{1-\alpha}, \infty]$ and $\varepsilon \in (0, \varepsilon_*]$ we have

$$\|y_\varepsilon^u\|_{Y_{s,0}} \leq c(1 + \|u\|_{\mathrm{L}^q(J_T;X)}) \quad \text{and} \quad \|z_\varepsilon^u\|_{\mathrm{C}(\overline{J_T})} \leq c(1 + \|u\|_{\mathrm{L}^q(J_T;X)}) \tag{4.9}$$

with arbitrary $s \in (1, q] \cap (1, \infty)$. Moreover, there holds

$$0 \leq \int_0^T |\dot{z}_\varepsilon^u(s)|^2 \, ds + \sup_{t \in \overline{J_T}} \frac{1}{\varepsilon} \Psi(z_\varepsilon^u(t)) \leq c(1 + \|u\|_{\mathrm{L}^2(J_T;X)})^2. \tag{4.10}$$

Proof. For any $v \in \mathrm{W}^{1,s}(J_T)$ and $t \in J_T$ we rewrite $|Z_\varepsilon(v)(t) - z_0| - |Z_\varepsilon(v)(0) - z_0|$ as

$$|Z_\varepsilon(v)(t) - z_0| - |Z_\varepsilon(v)(0) - z_0| = \int_0^t \frac{d}{ds} |Z_\varepsilon(v) - z_0| ds = \int_0^t \frac{\frac{d}{ds}(Z_\varepsilon(v))(Z_\varepsilon(v) - z_0)}{|Z_\varepsilon(v) - z_0|} ds.$$

According to Definition 4.9, we can replace z_0 by $Z_\varepsilon(v)(0)$ and $\frac{d}{ds}(Z_\varepsilon(v))$ by $\dot{v} - \frac{1}{\varepsilon}\Psi'(Z_\varepsilon(v))$. Moreover, Ψ is convex and $z_0 \in [a, b]$ implies $\Psi'(z_0) = 0$ by $(A4)_\varepsilon$. Hence,

$$\Psi'(x)(x - z_0) = [\Psi'(x) - \Psi'(z_0)](x - z_0) \geq 0 \quad \forall x \in \mathbb{R}.$$

All together, we arrive at

$$\begin{aligned} 0 &\leq |Z_\varepsilon(v)(t)| + \frac{1}{\varepsilon} \int_0^t \frac{\Psi'(Z_\varepsilon(v))(Z_\varepsilon(v) - z_0)}{|Z_\varepsilon(v) - z_0|} ds \\ &\leq |Z_\varepsilon(v)(t) - z_0| + |z_0| + \frac{1}{\varepsilon} \int_0^t \frac{\Psi'(Z_\varepsilon(v))(Z_\varepsilon(v) - z_0)}{|Z_\varepsilon(v) - z_0|} ds \\ &= |z_0| + \frac{1}{\varepsilon} \int_0^t \frac{\dot{v}(Z_\varepsilon(v) - z_0)}{|Z_\varepsilon(v) - z_0|} ds \leq |z_0| + \int_0^t |\dot{v}(s)| ds \qquad \forall t \in \overline{J_T}, \end{aligned}$$

and the second term on the left side is greater or equal than zero. Accordingly, we choose $v = Sy_\varepsilon^u$ and $z_\varepsilon^u = Z_\varepsilon(Sy_\varepsilon^u)$ and conclude

$$0 \leq |z_\varepsilon^u(t)| \leq |z_0| + \int_0^t |S\dot{y}_\varepsilon^u(s)| ds \quad \forall t \in \overline{J_T}. \tag{4.11}$$

To prove (4.9) we require a bound for $SA_py^u_\varepsilon$ since (4.4) includes the term $A_py^u_\varepsilon$. This is possible due to the enforced assumption (A3)' in Assumption 4.2 on the scalar projector S. In particular, the representation w which defines S is contained in $\mathrm{dom}([(A_p+1)^{1-\alpha}]^*)$. Hence, for any $y \in \mathrm{dom}(A_p)$ we can estimate

$$\begin{aligned} |SA_py| =&|S(A_p+1)y - Sy| \le |\langle w, (A_p+1)y\rangle_X| + \|S\|_{[X^\alpha]^*}\|y\|_{X^\alpha} \\ =&|\langle w, (A_p+1)^{1-\alpha}(A_p+1)^\alpha y\rangle_X| + \|S\|_{[X^\alpha]^*}\|y\|_{X^\alpha} \\ =&|\langle [(A_p+1)^{1-\alpha}]^* w, (A_p+1)^\alpha y\rangle_X + \|S\|_{[X^\alpha]^*}\|y\|_{X^\alpha} \\ \le&(\|[(A_p+1)^{1-\alpha}]^* w\|_{X^*} + \|S\|_{[X^\alpha]^*})\|y\|_{X^\alpha} =: c_1\|y\|_{X^\alpha}. \end{aligned} \tag{4.12}$$

Since $Sy^u_\varepsilon(t) \in \mathrm{dom}(A_p)$ for a.e. $t \in J_T$, we can choose $y = Sy^u_\varepsilon(t)$ in (4.12) for a.e. $t \in J_T$ and rewrite $S\dot{y}^u_\varepsilon(t)$ according to (4.8). Finally, the triangle inequality yields

$$|S\dot{y}^u_\varepsilon(t)| \le c_1\|y^u_\varepsilon(t)\|_{X^\alpha} + |Sf_\varepsilon(y^u_\varepsilon(t), z^u_\varepsilon(t))| + |Su(t)| \quad \text{for a.e. } t \in J_T.$$

The second term is bounded by the linear growth condition on f_ε, see $(A3)_\varepsilon$ in Assumption 4.7. Hence, we continue to estimate (4.11) by

$$\begin{aligned} |z^u_\varepsilon(t)| &\le |z_0| + \int_0^t c_1\|y^u_\varepsilon(s)\|_{X^\alpha} + |Sf_\varepsilon(y^u_\varepsilon(s), z^u_\varepsilon(s))| + |Su(s)|ds \\ &\le |z_0| + \int_0^t (M\|S\|_{X^*} + c_1)\left[\|y^u_\varepsilon(s)\|_{X^\alpha} + |z^u_\varepsilon(s)| + 1\right] + \|S\|_{X^*}\|u(s)\|_X ds \quad \forall t \in \overline{J_T}. \end{aligned}$$

It remains to prove a bound for $\|y^u_\varepsilon(t)\|_{X^\alpha}$ before we can apply Gronwall's lemma. Equation (4.8) implies that the initial value $y^u_\varepsilon(0) = 0$ is fixed independently of $\varepsilon \in (0, \varepsilon_*]$. Moreover, the mild solution y^u_ε of (4.8) is determined by $y^u_\varepsilon(t) = \int_0^t e^{-A_p(t-s)}[f_\varepsilon(y^u_\varepsilon(s), z^u_\varepsilon(s)) + u(s)]ds$ according to Definition 2.33. Consequently, with help of the semigroup estimate (3.14) and the linear growth condition $(A3)_\varepsilon$ on f_ε we arrive at

$$\begin{aligned} \|y^u_\varepsilon(t)\|_{X^\alpha} &= \left\| \int_0^t e^{-A_p(t-s)}[f_\varepsilon(y^u_\varepsilon(s), z^u_\varepsilon(s)) + u(s)]ds \right\|_{X^\alpha} \\ &\le C_\alpha e^{(1-\delta)T} \int_0^t (t-s)^{-\alpha}[M(\|y^u_\varepsilon(s)\|_{X^\alpha} + |z^u_\varepsilon(s)| + 1) + \|u(s)\|_X]ds \quad \forall t \in \overline{J_T}. \end{aligned}$$

In order to use Hölder's inequality for the last expression above remember that $q < \frac{1}{1-\alpha} \Leftrightarrow -\alpha q' > -1$. In particular,

$$\left(\int_0^t (t-s)^{-\alpha q'}\, ds\right)^{1/q'} = \left(\frac{t^{1-\alpha q'}}{1-\alpha q'}\right)^{1/q'} = \frac{t^{1/q'-\alpha}}{(1-\alpha q')^{1/q'-\alpha}},$$

so that

$$\|y^u_\varepsilon(t)\|_{X^\alpha} \le c_3 \left(\int_0^t (t-s)^{-\alpha}[M(\|y^u_\varepsilon(s)\|_{X^\alpha} + |z^u_\varepsilon(s)| + 1)\, ds + \frac{t^{1/q'-\alpha}}{(1-\alpha q')^{1/q'-\alpha}}\|u\|_{\mathrm{L}^q(J_T;X)} \right)$$

holds for $c_3 = C_\alpha e^{(1-\delta)T}$ and any $t \in \overline{J_T}$. Consequently, Gronwall's lemma applied to the sum of the estimates for $|z^u_\varepsilon(t)|$ and $\|y^u_\varepsilon(t)\|_{X^\alpha}$ yields

$$\|y^u_\varepsilon\|_{\mathrm{C}(\overline{J_T};X^\alpha)} \le c_4(1 + \|u\|_{\mathrm{L}^q(J_T;X)}) \quad \text{and} \quad \|z^u_\varepsilon\|_{\mathrm{C}(\overline{J_T})} \le c_4(1 + \|u\|_{\mathrm{L}^q(J_T;X)})$$

for all $q \in (\frac{1}{1-\alpha}, \infty]$. Here, $c_4 > 0$ depends on the fixed parameters T, q' and α but not on ε and u. Since y_ε^u is the solution of (4.8), with maximal parabolic regularity of A_p according to Remark 2.35 and assumption $(A3)_\varepsilon$ on $F_\varepsilon[y_\varepsilon^u] = f_\varepsilon[y_\varepsilon^u, z_\varepsilon^u]$ we conclude

$$\begin{aligned}
\|y_\varepsilon^u\|_{Y_{s,0}} &\leq \left\| \left(\frac{d}{dt} + A\right)^{-1} \right\|_{\mathcal{L}(\mathrm{L}^q(J_T;X),Y_s)} \|F_\varepsilon[y_\varepsilon^u] + u\|_{\mathrm{L}^q(J_T;X)} \\
&\leq c_5\big(1 + \|y_\varepsilon^u\|_{\mathrm{L}^q(J_T;X^\alpha)} + \|z_\varepsilon^u\|_{\mathrm{L}^q(J_T)} + \|u\|_{\mathrm{L}^q(J_T;X)}\big) \\
&\leq c_6\big(1 + \|u\|_{\mathrm{L}^q(J_T;X)}\big)
\end{aligned}$$

for all $s \in (1, q] \cap (1, \infty)$, and again $c_5, c_6 > 0$ are independent of ε and u. This proves (4.9). To prove (4.10) remember $2 > \frac{1}{1-\alpha}$ as a consequence of (A2)' in Assumption 4.2. Moreover, $S \in X^*$ and hence (4.9) implies $\|S\dot{y}_\varepsilon^u\|_{\mathrm{L}^2(J_T)} \leq c_7(1+\|u\|_{\mathrm{L}^2(J_T;X)})$ for $c_7 = c_6\|S\|_{X^*}$. We test the equation for $\dot{z}_\varepsilon^u$ according to Definition 4.9 by $\dot{z}_\varepsilon^u$ and integrate over $\overline{J_t}$ for $t \in \overline{J_T}$. By Young's inequality this yields

$$\begin{aligned}
&\int_0^t |\dot{z}_\varepsilon^u(s)|^2 \, ds = \int_0^t S\dot{y}_\varepsilon^u(s)\dot{z}_\varepsilon^u(s)ds - \frac{1}{\varepsilon}\int_0^t \Psi'(z_\varepsilon^u(s))\dot{z}_\varepsilon^u(s)ds \\
&\leq \frac{1}{2}\int_0^t |\dot{z}_\varepsilon^u(s)|^2 \, ds + \frac{1}{2}\|S\dot{y}_\varepsilon^u\|^2_{\mathrm{L}^2(J_T)} - \frac{1}{\varepsilon}[\Psi(z_\varepsilon^u(t)) - \Psi(z_\varepsilon^u(0))] \\
&\leq \frac{1}{2}\int_0^t |\dot{z}_\varepsilon^u(s)|^2 \, ds + \frac{c_7^2}{2}(1 + \|u\|_{\mathrm{L}^2(J_T;X)})^2 - \frac{1}{\varepsilon}[\Psi(z_\varepsilon^u(t)) - \Psi(z_\varepsilon^u(0))].
\end{aligned}$$

Note that $\Psi(z_\varepsilon^u(0)) = \Psi(z_0) = 0$ and remember $\Psi \geq 0$ by $(A4)_\varepsilon$. Hence, we conclude the proof with

$$0 \leq \int_0^T |\dot{z}_\varepsilon^u(s)|^2 \, ds + 2 \sup_{t \in \overline{J_T}} \frac{1}{\varepsilon}\Psi(z_\varepsilon^u(t)) \leq c_7^2(1 + \|u\|_{\mathrm{L}^2(J_T;X)})^2.$$

□

Estimates (4.9)–(4.10) are the key argument for the proofs on the dynamics of $(y_\varepsilon^{u_\varepsilon}, z_\varepsilon^{u_\varepsilon})$ in Subsection 4.3.2.

4.3.2 Dynamics of the regularized states

We apply techniques from [MS15, Section 4] and [BK13, Section 3.1] in this subsection. With help of the uniform bounds from Lemma 4.12, we prove that G_ε is weakly continuous for fixed $\varepsilon \in (0, \varepsilon_*]$, see Lemma 4.13. In Lemma 4.14, we apply weak continuity of G_ε and a weak compactness argument. Particularly, for an arbitrary weakly converging sequence $\{u_\varepsilon\} \subset X$, we prove that a subsequence $y_{\varepsilon_k}^{u_{\varepsilon_k}}$ and $z_{\varepsilon_k}^{u_{\varepsilon_k}}$ converges weakly with $k \to \infty$. By weak continuity of B_i, $i \in \{1, 2\}$, all results apply for subsequences of $\{u_\varepsilon\} \subset U_i$, $i \in \{1, 2\}$ and the corresponding solutions of the regularized state equations (4.4)–(4.5).
A regularization of the control problem (4.1)–(4.3) is defined in Subsection 4.3.3. Lemma 4.14 is crucial to prove that optimal solutions of the regularized problems converge to an optimal solution of problem (4.1)–(4.3) in the limit $\varepsilon \to 0$.
The following lemma is proved as Lemma 4.5.

Lemma 4.13 (Weak continuity of G_ε)**.** [Mün17b, Lemma 3.7] Let Assumption 4.2 and Assumption 4.7 hold and consider the notation from Corollary 4.11. Suppose that $u_n \rightharpoonup u$ in $\mathrm{L}^2(J_T; X)$ with $n \to \infty$ for some sequence $\{u_n\} \subset \mathrm{L}^2(J_T; X)$. For $\varepsilon \in (0, \varepsilon_*]$ fixed consider the solutions $y_\varepsilon^{u_n}$ and y_ε^u of (4.8), together with $z_\varepsilon^{u_n}$ and z_ε^u. Then $y_\varepsilon^{u_n} \to y_\varepsilon^u$ with $n \to \infty$ weakly in $Y_{2,0}$ and strongly in $\mathrm{C}(\overline{J_T};X^\alpha)$ and $z_\varepsilon^{u_n} \to z_\varepsilon^u$ with $n\to\infty$ weakly in $\mathrm{H}^1(J_T)$ and strongly in $\mathrm{C}(\overline{J_T})$. If the convergence of $\{u_n\}$ is strong then the convergence of $\{y_\varepsilon^{u_n}\}$ in $Y_{2,0}$ is also strong. The same holds if $\mathrm{L}^2(J_T;X)$ is replaced by U_i for $i\in\{1,2\}$ and if u_n and u are replaced by B_iu_n and B_iu. In this case, $(y_\varepsilon^{B_iu_n}, z_\varepsilon^{B_iu_n})$ and $(y_\varepsilon^{B_iu}, z_\varepsilon^{B_iu})$ are the solutions of (4.4)–(4.5).

Proof. For fixed $\varepsilon\in(0,\varepsilon_*]$, G_ε satisfies the same properties as G in Corollary 3.4. Hence, the proof is analogous to that of Lemma 4.5, using Corollary 4.11 and Lemma 4.12. □

In the next lemma, we proof that any joint limit $\lim\limits_{\varepsilon\to0} G_\varepsilon(u_\varepsilon)$ of a weakly converging sequence $\{u_\varepsilon\}$ corresponds to a solution $G(\overline{u})$ of (3.1).

Lemma 4.14. [Mün17b, Lemma 3.8] Let Assumption 4.2 and Assumption 4.7 hold and consider the notation from Corollary 4.11. Suppose that $u_\varepsilon \rightharpoonup u$ in $\mathrm{L}^2(J_T;X)$ as $\varepsilon\to0$. Consider the solutions $y_\varepsilon^{u_\varepsilon}$, together with $z_\varepsilon^{u_\varepsilon}$. Then $y_\varepsilon^{u_\varepsilon}\to y^u$ with $\varepsilon\to0$ weakly in $Y_{2,0}$ and strongly in $\mathrm{C}(\overline{J_T};X^\alpha)$ and $z_\varepsilon^{u_\varepsilon}\to\mathcal{W}[Sy^u]$ with $\varepsilon\to0$ weakly in $\mathrm{H}^1(J_T)$ and strongly in $\mathrm{C}(\overline{J_T})$. If the convergence of $\{u_\varepsilon\}$ is strong then also the convergence of $\{y_\varepsilon^{u_\varepsilon}\}$ in $Y_{2,0}$ is strong. The same holds if $\mathrm{L}^2(J_T;X)$ is replaced by U_i for $i\in\{1,2\}$ and if u_ε and u are replaced by B_iu_ε and B_iu. In this case, $(y_\varepsilon^{B_iu_\varepsilon}, z_\varepsilon^{B_iu_\varepsilon})$ are the solutions of (4.4)–(4.5).

Proof. The proof combines the proofs of [BK13, Lemma 3.2] and Lemma 4.5.
In Lemma 4.12, the bound for $y_\varepsilon^{u_\varepsilon}$ in $Y_{2,0}$ and for $z_\varepsilon^{u_\varepsilon}$ in $\mathrm{H}^1(J_T)$ is uniform in $\varepsilon\in(0,\varepsilon_*]$. Hence, since both spaces are reflexive, we can extract a subsequence $\{\varepsilon_k\}$ of the sequence $\{\varepsilon\}$ and find functions $\tilde{y}\in Y_{2,0}$ and $\tilde{z}\in\mathrm{H}^1(J_T)$ such that

$$y_{\varepsilon_k}(u_{\varepsilon_k})\rightharpoonup\tilde{y}\quad\text{in } Y_{2,0}\quad\text{and}\quad z_{\varepsilon_k}(u_{\varepsilon_k})\rightharpoonup\tilde{z}\quad\text{in } H^1(J_T)\quad\text{with } k\to\infty.$$

Due the compact embeddings $Y_{2,0}\hookrightarrow \mathrm{C}(\overline{J_T};X^\alpha)$, see Lemma 2.36, and $\mathrm{H}^1(J_T)\hookrightarrow\mathrm{C}(\overline{J_T})$ the convergence is strong in $\mathrm{C}(\overline{J_T};X^\alpha)$ and in $\mathrm{C}(\overline{J_T})$ respectively. In the following, we abbreviate

$$y_{\varepsilon_k}:=y_{\varepsilon_k}(u_{\varepsilon_k})\quad\text{and}\quad z_{\varepsilon_k}:=z_{\varepsilon_k}(u_{\varepsilon_k}).$$

We prove $\tilde{z}=\mathcal{W}[S\tilde{y}]$. To this aim, we have to show that the conditions (2.27)–(2.29) hold for $v=S\tilde{y}$, i.e. that

$$(\dot{\tilde{z}}(t)-S\dot{\tilde{y}}(t))(\tilde{z}(t)-\xi)\le0\qquad\text{for }\xi\in[a,b]\text{ and }t\in(0,T),\tag{4.13}$$
$$\tilde{z}(t)\in[a,b]\text{ for }t\in[0,T],\tag{4.14}$$
$$\tilde{z}(0)=z_0.\tag{4.15}$$

Equation (4.15) follows from Definition 4.9 since $z_0=z_{\varepsilon_k}(0)\to\tilde{z}(0)$ with $k\to\infty$. The uniform bound (4.10) implies $\Psi(z_{\varepsilon_k}(t))\to0$ with $k\to\infty$ for $t\in\overline{J_T}$. Due to the assumptions on Ψ in $(A4)_\varepsilon$ we conclude $\tilde{z}(t)\in[a,b]$ for $t\in\overline{J_T}$ which shows (4.14). Again by $(A4)_\varepsilon$, Ψ is convex and $\Psi'(\xi)=0$ for $\xi\in[a,b]$. Consequently, for any $x\in\mathbb{R}$ and $\xi\in[a,b]$ there holds $\Psi'(x)(x-\xi)\ge0$. With those observations in mind, we choose $\xi\in[a,b]$ arbitrary and insert the evolution equation for z_{ε_k} according to Definition 4.9 with $v=Sy_{\varepsilon_k}$ and obtain

$$\int_0^T(\dot{z}_{\varepsilon_k}(t)-S\dot{y}_{\varepsilon_k}(t))(z_{\varepsilon_k}(t)-\xi)\,dt=\int_0^T-\frac{1}{\varepsilon}\Psi'(z_{\varepsilon_k}(u_{\varepsilon_k}(t))(z_{\varepsilon_k}(t)-\xi)\,dt\le0.$$

Taking the limit $k\to\infty$ yields that $\tilde{z}$ solves the variational inequality (4.13) which implies $\tilde{z}=\mathcal{W}[S\tilde{y}]$.
Next, we show $\tilde{y}=G(\overline{u})$. Weak continuity of $\frac{d}{dt}$ and A_p from $Y_{2,0}$ to $\mathrm{L}^2(J_T;X)$ implies

$$\frac{d}{dt}y_{\varepsilon_k}+A_py_{\varepsilon_k}\rightharpoonup\frac{d}{dt}\tilde{y}+A_p\tilde{y}\quad\text{in }\mathrm{L}^2(J_T;X)\qquad\text{with }k\to\infty.$$

For ε_k small enough, $(A3)_\varepsilon$ yields the estimate

$$\begin{aligned}
&\|F_{\varepsilon_k}[y_{\varepsilon_k}] - F[\tilde y]\|_{\mathrm{C}(\overline{J_T};X)} = \|f_{\varepsilon_k}(y_{\varepsilon_k}(\cdot), z_{\varepsilon_k}(\cdot)) - f(\tilde y(\cdot), \tilde z(\cdot))\|_{\mathrm{C}(\overline{J_T};X)} \\
&\le \|f_{\varepsilon_k}(y_{\varepsilon_k}(\cdot), z_{\varepsilon_k}(\cdot)) - f_{\varepsilon_k}(\tilde y(\cdot), \tilde z(\cdot))\|_{\mathrm{C}(\overline{J_T};X)} + \|f_{\varepsilon_k}(\tilde y(\cdot), \tilde z(\cdot)) - f(\tilde y(\cdot), \tilde z(\cdot))\|_{\mathrm{C}(\overline{J_T};X)} \\
&\le L(\tilde y)(\|y_{\varepsilon_k} - \tilde y\|_{\mathrm{C}(\overline{J_T};X^\alpha)} + \|z_{\varepsilon_k} - \tilde z\|_{\mathrm{C}(\overline{J_T})}) + \|f_{\varepsilon_k}(\tilde y(\cdot), \tilde z(\cdot)) - f(\tilde y(\cdot), \tilde z(\cdot))\|_{\mathrm{C}(\overline{J_T};X)}.
\end{aligned}$$

Note that the right side converges to zero. Hence, we conclude that $F_{\varepsilon_k}[y_{\varepsilon_k}]$ converges to $F[\tilde y]$ in $\mathrm{C}(\overline{J_T};X)$ with $k \to \infty$. Because $\tilde z = \mathcal{W}[S\tilde y]$, this proves $\tilde y = G(\overline{u})$. Since the limit $G(\overline{u})$ is unique, we conclude convergence of the whole sequence. The statement about strong convergence follows by maximal parabolic Sobolev regularity of A_p, just as in Lemma 4.5. □

4.3.3 The regularized optimal control problem

This subsection is inspired by [BK13, Section 3.2] and [MS15, Section 4]. We introduce regularized optimal control problems for (4.1)–(4.3). In the subsequent sections, we will exploit first order optimality conditions of those problems as a tool to derive an adjoint system for problem (4.1)–(4.3).

Although proving adjoint equations of the regularized control problems remains challenging, see Subsection 4.3.5 below, linearity of the derivatives of the solution operators of (4.4)–(4.5) allows for a direct approach.

Remember that optimal solutions for problem (4.1)–(4.3) exist by Theorem 4.6. That is, let $i \in \{1, 2\}$ be given and consider an optimal control $\overline{u} \in U_i$ of problem (4.1)–(4.3) together with the state $\overline{y} = G(B_i\overline{u})$ and $\overline{z} = \mathcal{W}[S\overline{y}]$. For $\varepsilon \in (0, \varepsilon_*]$, we define the regularized optimal control problem

$$\min_{u\in U_i} J_{\mathrm{reg}}(y, u; \overline{u}) := \min_{u\in U_i} J(y, u) + \frac{1}{2}\|u - \overline{u}\|^2_{U_i} \tag{4.16}$$

subject to (4.4)–(4.5).

Remark 4.15. As for problem (4.1)–(4.3), in view of the embedding $Y_{2,0} \hookrightarrow \mathrm{L}^2(J_T; \mathrm{dom}(A_p))$, for $u \in U_i$, the functions $y = G_\varepsilon(B_iu) \in Y_{2,0}$ in (4.16) are identified with $I_p^{-1}G_\varepsilon(B_iu) \in \mathrm{L}^2(J_T; \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)) \hookrightarrow U_1$ according to Corollary 2.30, see also Remark 2.32. We often write $G_\varepsilon(B_iu)$ for both functions.

Theorem 4.16 (Convergence of regularized solutions)**.** *[Mün17b, Theorem 3.9] Let Assumption 4.2 and Assumption 4.7 hold. For $i \in \{1, 2\}$ suppose that $\overline{u} \in U_i$ is an optimal control for problem* (4.1)*–*(4.3)*. Then for all $\varepsilon \in (0, \varepsilon_*]$ problem* (4.4)*,*(4.5)*,*(4.16) *has an optimal control $\overline{u}_\varepsilon \in U_i$. This means that $\overline{u}_\varepsilon$, together with $\overline{y}_\varepsilon = G_\varepsilon(B_i\overline{u}_\varepsilon)$ and $\overline{z}_\varepsilon = Z_\varepsilon(S\overline{y}_\varepsilon)$ (see Definition 4.9), are a solution of the minimization problem* (4.16)*. Furthermore, $\overline{u}_\varepsilon \to \overline{u}$ in U_i, $\overline{y}_\varepsilon \to \overline{y} = G(B_i\overline{u})$ in $Y_{2,0}$ and in $\mathrm{C}(\overline{J_T}; X^\alpha)$ and $\overline{z}_\varepsilon \to \overline{z} = \mathcal{W}[S\overline{y}]$ weakly in $\mathrm{H}^1(J_T)$ and strongly in $\mathrm{C}(\overline{J_T})$ with $\varepsilon \to 0$.*

Proof. First of all, note that the embedding $Y_{2,0} \hookrightarrow U_1$ is continuous, because $\mathrm{dom}(A_p) \simeq \mathbb{W}^{1,p}_{\Gamma_D}(\Omega) \hookrightarrow [\mathrm{L}^2(\Omega)]^m$, see Corollary 2.30. Existence of $\overline{u}$ has been shown in Theorem 4.6. That optimal controls $\overline{u}_\varepsilon$ for the regularized problem (4.4),(4.5),(4.16) exist is shown in the same way as Theorem 4.6. The proof requires weak continuity of G_ε which was shown in Lemma 4.13 and that J_{reg} is bounded from below and weakly lower semi-continuous and coercive in u. For all $\varepsilon \in (0, \varepsilon_*]$, optimality of $(\overline{y}_\varepsilon, \overline{z}_\varepsilon, \overline{u}_\varepsilon)$ for problem (4.4),(4.5),(4.16) and of $(\overline{y}, \overline{z}, \overline{u})$ for problem (4.1)–(4.3) implies

$$\begin{aligned}
J(G_\varepsilon(B_i\overline{u}), \overline{u}) &= J_{\mathrm{reg}}(G_\varepsilon(B_i\overline{u}), \overline{u}; \overline{u}) \ge J_{\mathrm{reg}}(\overline{y}_\varepsilon, \overline{u}_\varepsilon; \overline{u}) \\
&= J(\overline{y}_\varepsilon, \overline{u}_\varepsilon) + \frac{1}{2}\|\overline{u}_\varepsilon - \overline{u}\|^2_{U_i} \ge J(\overline{y}, \overline{u}).
\end{aligned} \tag{4.17}$$

Moreover, for $\varepsilon \in (0, \varepsilon_*]$, the uniform bound (4.9) from Lemma 4.12 for $G_\varepsilon(B_i\overline{u}) \in Y_{2,0}$ yields

$$J(G_\varepsilon(B_i\overline{u}), \overline{u}) = \frac{1}{2}\|G_\varepsilon(B_i\overline{u}) - y_d\|_{U_1}^2 + \frac{\kappa}{2}\|\overline{u}\|_{U_i}^2 \leq c_1\|G_\varepsilon(B_i\overline{u})\|_{Y_{2,0}}^2 + c_2 \leq c_3$$

for some constants $c_1, c_2, c_3 > 0$. Note that c_3 depends on $\overline{u}$ but this function is fixed. Hence, with (4.17) we obtain

$$c_3 \geq J_{\mathrm{reg}}(\overline{y}_\varepsilon, \overline{u}_\varepsilon; \overline{u}) = \frac{1}{2}\|\overline{y}_\varepsilon - y_d\|_{U_1}^2 + \frac{\kappa}{2}\|\overline{u}_\varepsilon\|_{U_i}^2 + \frac{1}{2}\|\overline{u}_\varepsilon - \overline{u}\|_{U_i}^2.$$

This yields a bound for $\overline{u}_\varepsilon$ in U_i which is independent of $\varepsilon \in (0, \varepsilon_*]$. Consequently, we can extract a subsequence $\{\overline{u}_{\varepsilon_k}\}$ which converges weakly in U_i to some $\tilde{u}$ with $k \to \infty$. Lemma 4.14 implies $\overline{y}_{\varepsilon_k} = G_{\varepsilon_k}(B_i\overline{u}_{\varepsilon_k}) \to G(B_i\tilde{u})$ with $k \to \infty$ weakly in $Y_{2,0}$. Also by Lemma 4.14, $G_\varepsilon(B_i\overline{u}) \to \overline{y}$ with $\varepsilon \to 0$ strongly in $Y_{2,0}$. Remember the embedding $Y_{2,0} \hookrightarrow U_1$ and that J_{reg} is weakly lower semi-continuous. Hence, with (4.17) we conclude

$$J(\overline{y}, \overline{u}) = \lim_{k\to\infty} J(G_{\varepsilon_k}(B_i\overline{u}), \overline{u}) \geq \liminf_{k\to\infty} J_{\mathrm{reg}}(\overline{y}_{\varepsilon_k}, \overline{u}_{\varepsilon_k}; \overline{u}) \geq J(\tilde{y}, \tilde{u}) + \frac{1}{2}\|\tilde{u} - \overline{u}\|_{U_i}^2 \geq J(\overline{y}, \overline{u}).$$

But this implies $\tilde{u} = \overline{u}$ and that the convergence of $\{u_{\varepsilon_k}\}$ in U_i is strong. Since the limit is uniquely determined by $\overline{u}$, the whole sequence $\{u_\varepsilon\}$ converges to $\overline{u}$ in U_i with $\varepsilon \to 0$. All results then follow by applying the statement about strong convergence in Lemma 4.14. □

4.3.4 Gâteaux differentiability of the solution operator of the regularized state equation

In this subsection, we apply Corollary 3.15 and show that G_ε is Gâteaux differentiable for all $\varepsilon \in (0, \varepsilon_*]$.

Corollary 4.17 (Gâteaux differentiability of G_ε)**.** [Mün17b, Lemma 3.10] Let Assumption 4.2 and Assumption 4.7 hold and take the notation from Corollary 4.11. Then for any $\varepsilon \in (0, \varepsilon_*]$ and $q \in (\frac{1}{1-\alpha}, \infty)$ the solution operator $G_\varepsilon : \mathrm{L}^q(J_T; X) \to Y_{q,0}$ of problem (4.8) is Gâteaux differentiable. The derivative $G'_\varepsilon[u; h]$ at $u \in \mathrm{L}^q(J_T; X)$ in direction $h \in \mathrm{L}^q(J_T; X)$ is given by $y_\varepsilon^{u,h}$, where $y_\varepsilon^{u,h}$ together with $z = z_\varepsilon^{u,h} = Z'_\varepsilon[Sy_\varepsilon^u; Sy_\varepsilon^{u,h}] \in \mathrm{W}^{1,q}(J_T)$ are the unique solution of

$$\begin{aligned}
\dot{y}(t) + (A_p y)(t) &= \frac{\partial}{\partial y} f_\varepsilon(y_\varepsilon^u(t), z_\varepsilon^u(t))y(t) + \frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u(t), z_\varepsilon^u(t))z(t) + h(t) \text{ for } t \in J_T, && (4.18)\\
y(0) &= 0,\\
\dot{z}(t) - S\dot{y}(t) &= -\frac{1}{\varepsilon}\Psi''(z_\varepsilon^u(t))z(t) \qquad \text{for } t \in J_T, && (4.19)\\
z(0) &= 0.
\end{aligned}$$

For $i \in \{1, 2\}$ and $u, h \in U_i$ the derivative of the solution mapping $u \mapsto G_\varepsilon(B_i u)$ at u in direction h is given by $y_\varepsilon^{B_i u, B_i h}$, i.e. by the unique solution of (4.18) with h replaced by $B_i h$ and $z = z_\varepsilon^{B_i u, B_i h} = Z'_\varepsilon[Sy_\varepsilon^{B_i u}; Sy_\varepsilon^{B_i u, B_i h}]$.

Proof. G_ε is Hadamard directionally differentiable because Lemma 4.10 implies that Z_ε satisfies the properties of $\mathcal{W}$ in Corollary 3.15. Gâteaux differentiability and the representation (4.18)–(4.19) then follows from linearity of all the derivatives. The representation for $z_\varepsilon^{u,h} = Z'_\varepsilon[Sy_\varepsilon^u; Sy_\varepsilon^{u,h}]$ according to (4.19) and the regularity $z_\varepsilon^{u,h} \in \mathrm{W}^{1,q}(J_T)$ have been shown in Lemma 4.10. □

4.3.5 Adjoint system for the regularized problem

In this subsection, we derive adjoint systems for the regularized problems (4.4),(4.5),(4.16) with $\varepsilon \in (0, \varepsilon_*]$ and $i \in \{1, 2\}$, see Theorem 4.20 below. We proceed in a similar way as in [BK13, Sections 3.3 and 3.5] and [MS15, Section 4]. In the following lemma, we prove Gâteaux differentiability of the Nemitskii operator of f_ε and show crucial (pointwise-in-time) estimates for the derivatives f'_ε which are uniform in ε.

Lemma 4.18. [Mün17b, Lemma 3.11] Let Assumption 4.2 and Assumption 4.7 hold. With a little abuse of notation we use the same symbol for the Nemitskii operator of f_ε, i.e. we write $f_\varepsilon : (y, z) \mapsto f_\varepsilon(y(\cdot), z(\cdot))$. Then f_ε is locally Lipschitz continuous and Gâteaux differentiable from $\mathrm{C}(\overline{J_T}; X^\alpha) \times \mathrm{L}^q(J_T)$ to $\mathrm{L}^q(J_T; X)$ for all $\varepsilon \in (0, \varepsilon_*]$ and $q \in (\frac{1}{1-\alpha}, \infty)$.
Moreover, the derivative $f'_\varepsilon[(y, z); (\cdot, \cdot)]$ at $(y, z) \in \mathrm{C}(\overline{J_T}; X^\alpha) \times \mathrm{L}^q(J_T)$ is Lipschitz continuous with a modulus of the form $K(y) = L(y)(1 + T^{1/q})$, where $L(y) > 0$ only depends on $y \in \mathrm{C}(\overline{J_T}; X^\alpha)$. $K(y)$ and $L(y)$ are independent of ε and remain the same in a sufficiently small neighborhood of y. For $(v, h) \in \mathrm{C}(\overline{J_T}; X^\alpha) \times \mathrm{L}^q(J_T)$ we can estimate

$$\left\| \frac{\partial}{\partial y} f_\varepsilon(y, z)v \right\|_{\mathrm{L}^q(J_T;X)} + \left\| \frac{\partial}{\partial z} f_\varepsilon(y, z)h \right\|_{\mathrm{L}^q(J_T;X)} \le K(y)(\|v\|_{\mathrm{C}(\overline{J_T};X^\alpha)} + \|h\|_{\mathrm{L}^q(J_T)}) \tag{4.20}$$

For a.e. $t \in J_T$, there also holds the pointwise estimate

$$\left\| \frac{\partial}{\partial y} f_\varepsilon(y(t), z(t))v(t) \right\|_X + \left\| \frac{\partial}{\partial z} f_\varepsilon(y(t), z(t))h(t) \right\|_X \le K(y)(\|v(t)\|_{X^\alpha} + |h(t)|). \tag{4.21}$$

Furthermore, $\frac{\partial}{\partial y} f_\varepsilon(y, z) = \frac{\partial}{\partial y} f_\varepsilon(y(\cdot), z(\cdot))$ is bounded by $K(y)$ in $\mathrm{L}^\infty(J_T; \mathcal{L}(X^\alpha, X))$. Moreover, $\frac{\partial}{\partial z} f_\varepsilon(y, z) = \frac{\partial}{\partial z} f_\varepsilon(y(\cdot), z(\cdot))$ is bounded by $K(y)$ in $\mathrm{L}^\infty(J_T; X)$.

Proof. We prove the lemma in two steps.
(I): For all $\varepsilon \in (0, \varepsilon_*]$ and for any $q \in (\frac{1}{1-\alpha}, \infty)$ the function $f_\varepsilon : \mathrm{C}(\overline{J_T}; X^\alpha) \times \mathrm{L}^q(J_T) \to \mathrm{L}^q(J_T; X)$ is locally Lipschitz continuous and Gâteaux differentiable:
The proof follows the lines of Lemma 3.12. We recap the important steps:
Well-posedness of $f_\varepsilon : \mathrm{C}(\overline{J_T}; X^\alpha) \times \mathrm{L}^q(J_T) \to \mathrm{L}^q(J_T; X)$ follows with help of the linear growth estimate in $(A3)_\varepsilon$ in Assumption 4.7.
Again $(A3)_\varepsilon$ yields that the mapping $(y(\cdot), v) \mapsto f_\varepsilon(y(\cdot), v)$ is locally Lipschitz continuous from $\mathrm{C}(\overline{J_T}; X^\alpha) \times \mathbb{R}$ to $\mathrm{C}(\overline{J_T}; X)$ with respect to the $\mathrm{C}(\overline{J_T}; X^\alpha)$-norm. Indeed, for $y \in \mathrm{C}(\overline{J_T}; X^\alpha)$ and some appropriate neighborhood $\overline{B_{\mathrm{C}(\overline{J_T};X^\alpha)}(y, r)}$ of y, there even holds a pointwise-in-time estimate of the form

$$\|f_\varepsilon(y_1(t), z_1) - f(y_2(t), z_2)\|_X \le L(y)(\|y_1(t) - y_2(t)\|_{X^\alpha} + |z_1 - z_2|)$$

for all $y_1, y_2 \in \overline{B_{\mathrm{C}(\overline{J_T};X^\alpha)}(y, r)}$, $z_1, z_2 \in \mathbb{R}$ and $t \in \overline{J_T}$, where $L(y) > 0$ is a Lipschitz modulus which depends on $y \in \mathrm{C}(\overline{J_T}; X^\alpha)$. This local estimate implies the pointwise-in-time estimate

$$\|f_\varepsilon(y_1, z_1)(s) - f_\varepsilon(y_2, z_2)(s)\|_X \le L(y) \left[\|y_1(s) - y_2(s)\|_{X^\alpha} + |z_1(s) - z_2(s)|\right]$$

for a.e. $s \in J_T$, for any $y_1, y_2 \in \overline{B_{\mathrm{C}(\overline{J_T};X^\alpha)}(y, r)}$ and $z_1, z_2 \in \mathrm{L}^q(J_T)$. With help of Minkowski's inequality this yields local Lipschitz continuity of $f_\varepsilon : \mathrm{C}(\overline{J_T}; X^\alpha) \times \mathrm{L}^q(J_T) \to \mathrm{L}^q(J_T; X)$ with respect to the $\mathrm{C}(\overline{J_T}; X^\alpha)$-norm. The Lipschitz constants are of the form $K(y) = L(y)(1+T^{1/q})$.
In a second step one shows that f_ε is directionally differentiable, just as for f in Step III in the proof of Lemma 3.12. Pointwise-in-time convergence of the difference quotients,

$$\lim_{\lambda \to 0} \frac{f_\varepsilon(y(s) + \lambda v(s), z(s) + \lambda h(s))}{\lambda} = f'_\varepsilon[(y(s), z(s)); (v(s), h(s))] \in X$$

for a.e. $s \in J_T$ and $(y,z),(v,h) \in \mathrm{C}(\overline{J_T}; X^\alpha) \times \mathrm{L}^q(J_T)$, holds by Gâteaux differentiability of f_ε, see $(A3)_\varepsilon$ in Assumption 4.7. Lebesgue's dominated convergence theorem finally yields directional differentiability of $f_\varepsilon : \mathrm{C}(\overline{J_T}; X^\alpha) \times \mathrm{L}^q(J_T) \to \mathrm{L}^q(J_T; X)$ and the bounds (4.20) and (4.21) for $f'_\varepsilon[(y,z);(\cdot,\cdot)]$. Since $f'_\varepsilon[(y,z);(\cdot,\cdot)]$ is a bounded and linear operator, Gâteaux differentiability of f_ε follows.
(II) L^∞-bounds:
Let $\tilde{y} \in X^\alpha$ with $\|\tilde{y}\|_{X^\alpha} \leq 1$ be arbitrary. We test (4.21) with the constant function $v \in \mathrm{C}(\overline{J}_T; X^\alpha)$, $v(t) = \tilde{y}$ for $t \in \overline{J_T}$ and $h = 0 \in \mathrm{L}^q(J_T)$. This implies

$$\left\| \frac{\partial}{\partial y} f_\varepsilon(y,z) \right\|_{\mathrm{L}^\infty(J_T; \mathcal{L}(X^\alpha, X))} = \operatorname*{ess\,sup}_{t \in J_T} \sup_{\tilde{y} \in \overline{B_{X^\alpha}(0,1)}} \left\| \frac{\partial}{\partial y} f_\varepsilon(y(t), z(t)) \tilde{y} \right\|_X \leq K(y)$$

as required. Then we test (4.21) with $v = 0 \in \mathrm{C}(\overline{J}_T; X^\alpha)$ and the constant function $h \in \mathrm{L}^q(J_T)$, $h(t) = c > 0$ for $t \in \overline{J_T}$ and divide by c on both sides. Hence, we conclude

$$\left\| \frac{\partial}{\partial z} f_\varepsilon(y,z) \right\|_{\mathrm{L}^\infty(J_T; X)} = \operatorname*{ess\,sup}_{t \in J_T} \left\| \frac{\partial}{\partial z} f_\varepsilon(y(t), z(t)) \right\|_X \leq K(y).$$

□

In the following lemma, we prove a key result on the way to an adjoint system for the regularized problem (4.4),(4.5),(4.16). In particular, we derive the evolution equations of the adjoint system corresponding to problem (4.18)–(4.19), see Corollary 4.17.
The most challenging part in the following proof is to find an explicit expression of the adjoint operator $[G'_\varepsilon[u;\cdot]]^{-*} : \mathrm{L}^{q'}(J_T; X^*) \to Y^*_{q,0}$ of $G'_\varepsilon[u;\cdot]^{-1}$ which does not involve any abstract solution operators. The reason is that $G'_\varepsilon[u;\cdot]$ is defined as the mapping which assigns to each $h \in \mathrm{L}^q(J_T; X)$ the solution $y^{u,h}_\varepsilon = G'_\varepsilon[u;h] \in Y_{q,0}$ of (4.18). The latter contains the solution $z^{u,h}_\varepsilon = Z'_\varepsilon[Sy^u_\varepsilon; Sy^{u,h}_\varepsilon]$ of (4.19) only implicitly. Moreover, it turns out that $z^{u,h}_\varepsilon \in \mathrm{W}^{1,q}(J_T)$ has to be interpreted as a function in $\mathrm{L}^q(J_T)$. One carefully has to keep track of the correct spaces which the involved operators are defined on.

Lemma 4.19. (Adjoint operators of the regularized solution operators) [Mün17b, Lemma 3.12] Let Assumption 4.2 and Assumption 4.7 hold and adopt the notation from Corollary 4.17. For $\varepsilon \in (0, \varepsilon_*]$ and any $q \in (\frac{1}{1-\alpha}, \infty)$, $h \in \mathrm{L}^q(J_T; X)$ and $\nu \in \mathrm{L}^{q'}(J_T; [\mathrm{dom}(A_p)]^*)$ there holds

$$\langle \nu, y^{u,h}_\varepsilon \rangle_{\mathrm{L}^q(J_T; \mathrm{dom}(A_p))} = \langle p^\nu_\varepsilon + S q^\nu_\varepsilon, h \rangle_{\mathrm{L}^q(J_T; X)}, \tag{4.22}$$

where $p^\nu_\varepsilon \in Y^*_{q',T}$ and $q^\nu_\varepsilon \in \mathrm{L}^{q'}(J_T)$ are the unique solution of

$$\begin{aligned}
-\dot{p} + A^*_p p &= \left[\frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon) \right]^* p + S \left[-A_p + \frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon) \right] q + \nu \text{ for } t \in J_T, \\
p(T) &= 0, \\
-\dot{q} &= \langle p, \frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon) \rangle_X + S \frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon) q - \frac{1}{\varepsilon} \Psi''(z^u_\varepsilon) q \quad \text{for } t \in J_T, \\
q(T) &= 0,
\end{aligned}$$

and where $y^{u,h}_\varepsilon \in Y_{q,0}$ and $z^{u,h}_\varepsilon \in \mathrm{W}^{1,q}(J_T)$ are the unique solution of (4.18)–(4.19). Moreover,

$$\|y^{u,h}_\varepsilon\|_{Y_{q,0}} \leq C(y^u_\varepsilon) \|h\|_{\mathrm{L}^q(J_T; X)} \text{ and } \|z^{u,h}_\varepsilon\|_{\mathrm{C}(\overline{J_T})} \leq C(y^u_\varepsilon) \|h\|_{\mathrm{L}^q(J_T; X)} \tag{4.23}$$

for some constant $C(y^u_\varepsilon) > 0$. $C(y^u_\varepsilon)$ remains the same in a sufficiently small neighborhood of y^u_ε.

Proof. We prove the lemma in four steps. Let $q \in (\frac{1}{1-\alpha}, \infty)$ be arbitrary.
(I) Auxiliary problem:
We introduce an auxiliary problem similar to (4.18)–(4.19). First of all, by Assumption 4.7, the solution operator which maps any $v \in \mathrm{L}^q(J_T)$ to the solution $z \in \mathrm{W}^{1,q}(J_T)$ of the Cauchy problem

$$\dot{z}(t) = v(t) + \left(S\frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon(t), z^u_\varepsilon(t)) - \frac{1}{\varepsilon}\Psi''(z^u_\varepsilon(t)) \right) z(t) \quad \text{for } t \in J_T,\ z(0) = 0,$$

is well defined. We denote by $T^u_{z,\varepsilon} : \mathrm{L}^q(J_T) \to \mathrm{L}^q(J_T),\ v \mapsto T^u_{z,\varepsilon} v$ the corresponding solution operator on $\mathrm{L}^q(J_T)$, i.e. we write $T^u_{z,\varepsilon}$ for the solution operator of the integral equation

$$z(t) = \int_0^t v(s) + \left(S\frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon(s), z^u_\varepsilon(s)) - \frac{1}{\varepsilon}\Psi''(z^u_\varepsilon(s)) \right) z(s)\, ds \quad \text{for } t \in J_T.$$

Note that both mappings actually coincide, but we have to interpret $T^u_{z,\varepsilon}$ as a mapping into $\mathrm{L}^q(J_T)$ for the construction of the following operator. With Assumption 4.7 and the definition of $T^u_{z,\varepsilon}$ in mind, we introduce the operator

$$\begin{aligned}
&T^u_{y,\varepsilon} : Y_{q,0} \to \mathrm{L}^q(J_T; X),\\
&T^u_{y,\varepsilon} := A_p - \frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon) - \frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon) T^u_{z,\varepsilon} S \left(-A_p + \frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon) \right).
\end{aligned}$$

Similar to (4.18)–(4.19), we then consider the problem

$$\dot{y}(t) + (T^u_{y,\varepsilon} y)(t) = h(t) \text{ for } t \in J_T,\ y(0) = 0, \tag{4.24}$$

$$z = T^u_{z,\varepsilon} S \left(-A_p + \frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon) \right) y. \tag{4.25}$$

Note that equations (4.24)–(4.25) are almost equivalent to (4.18)–(4.19), but the h-term is missing in the second equation. However, similar as for (4.18)–(4.19), it is shown that for each $h \in \mathrm{L}^q(J_T; X)$ there exists a unique couple of solutions $(\tilde{y}^{u,h}_\varepsilon, \tilde{z}^{u,h}_\varepsilon)$ in $Y_{q,0} \times \mathrm{L}^q(J_T)$ of (4.24)–(4.25). Hence, the operator $\left(\frac{d}{dt} + T^u_{y,\varepsilon}\right)^{-1}$ is bijective from $\mathrm{L}^q(J_T; X)$ to $Y_{q,0}$.
(II) Uniform estimates:
Even though $T^u_{z,\varepsilon}$ is defined as a mapping with values in $\mathrm{L}^q(J_T)$, $\tilde{z}^{u,h}_\varepsilon \in \mathrm{L}^q(J_T)$ can be identified with the corresponding function in $\mathrm{W}^{1,q}(J_T)$. We estimate the norms of $(\tilde{y}^{u,h}_\varepsilon, \tilde{z}^{u,h}_\varepsilon)$. By definition of $\tilde{z}^{u,h}_\varepsilon$ there holds

$$|\tilde{z}^{u,h}_\varepsilon(t)| = \int_0^t \frac{\dot{\tilde{z}}^{u,h}_\varepsilon(s) \tilde{z}^{u,h}_\varepsilon(s)}{|\tilde{z}^{u,h}_\varepsilon(s)|} ds = \int_0^t \frac{-S[(T^u_{y,\varepsilon}\tilde{y}^{u,h}_\varepsilon)(s)]\tilde{z}^{u,h}_\varepsilon(s)}{|\tilde{z}^{u,h}_\varepsilon(s)|} ds - \frac{1}{\varepsilon}\int_0^t \Psi''(z^u_\varepsilon(s))|\tilde{z}^{u,h}_\varepsilon(s)|\, ds$$

for any $t \in \overline{J_T}$. Remember the pointwise-in-time estimate (4.21) in Lemma 4.18. Moreover, the definition of S by $w \in \mathrm{dom}([(1 + A_p)^{1-\alpha}]^*)$ implies that $\|SA_p y\|_X$ can be estimated by $\|y\|_{X^\alpha}$ for all $y \in \mathrm{dom}(A_p)$, see also (4.12). It follows

$$\begin{aligned}
0 &\le |\tilde{z}^{u,h}_\varepsilon(t)| + \frac{1}{\varepsilon}\int_0^t \Psi''(z^u_\varepsilon(s))|\tilde{z}^{u,h}_\varepsilon(s)| ds\\
&\le \int_0^t |SA_p\tilde{y}^{u,h}_\varepsilon(s)| + \left| S\frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon(s), z^u_\varepsilon(s))\tilde{y}^{u,h}_\varepsilon(s)\right| + \left| S\frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon(s), z^u_\varepsilon(s))\tilde{z}^{u,h}_\varepsilon(s)\right| ds\\
&\le (c + \|S\|_{X^*} K(y^u_\varepsilon)) \int_0^t \|\tilde{y}^{u,h}_\varepsilon(s)\|_{X^\alpha} + |\tilde{z}^{u,h}_\varepsilon(s)| ds
\end{aligned}$$

for a constant $c > 0$ which is independent of ε. Note that $\Psi''(z_\varepsilon^u(s)) \geq 0$ because Ψ is convex by $(A4)_\varepsilon$. To estimate $\tilde{y}_\varepsilon^{u,h}$, we exploit the representation as a mild solution and apply (3.14) and (4.21). This implies

$$
\begin{aligned}
&\|\tilde{y}_\varepsilon^{u,h}(t)\|_{X^\alpha} \\
&= \left\| \int_0^t e^{-A_p(t-s)} \left[\frac{\partial}{\partial y} f_\varepsilon(y_\varepsilon^u(s), z_\varepsilon^u(s)) \tilde{y}_\varepsilon^{u,h}(s) + \frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u(s), z_\varepsilon^u(s)) \tilde{z}_\varepsilon^{u,h}(s) + h(s) \right] ds \right\|_{X^\alpha} \\
&\leq C_\alpha (1 + K(y_\varepsilon^u)) e^{(1-\delta)T} \int_0^t (t-s)^{-\alpha} [\|\tilde{y}_\varepsilon^{u,h}(s)\|_{X^\alpha} + |\tilde{z}_\varepsilon^{u,h}(s)| + \|h(s)\|_X] ds.
\end{aligned}
$$

Finally, Gronwall's lemma yields a constant $C_1(y_\varepsilon^u) > 0$ which depends only on $y_\varepsilon^u \in \mathrm{C}(\overline{J_T}; X^\alpha)$ such that

$$
\|\tilde{y}_\varepsilon^{u,h}\|_{\mathrm{C}(\overline{J_T};X^\alpha)} \leq C_1(y_\varepsilon^u) \|h\|_{\mathrm{L}^q(J_T;X)} \quad \text{and} \quad \|\tilde{z}_\varepsilon^{u,h}\|_{\mathrm{C}(\overline{J_T})} \leq C_1(y_\varepsilon^u) \|h\|_{\mathrm{L}^q(J_T;X)}
$$

for $q \in (\frac{1}{1-\alpha}, \infty)$. Moreover, the only constant which depends on y_ε^u is $K(y_\varepsilon^u)$. Hence, by Lemma 4.18 there holds $C_1(y_\varepsilon^u) = C_1(y)$ for ε small enough if $\{y_\varepsilon^u\}$ converges to y with $\varepsilon \to 0$. Hence, $C_1(\overline{y}_\varepsilon) = C_1(\overline{y})$ for small enough ε by Theorem 4.16. As several times before we use maximal parabolic regularity of A_p and (4.20) in

$$
\tilde{y}_\varepsilon^{u,h} = \left(\frac{d}{dt} + A_p \right)^{-1} \left[\frac{\partial}{\partial y} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u) \tilde{y}_\varepsilon^{u,h} + \frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u) \tilde{z}_\varepsilon^{u,h} + h \right]
$$

to conclude

$$
\|\tilde{y}_\varepsilon^{u,h}\|_{Y_{q,0}} \leq C_2(y_\varepsilon^u) \|h\|_{\mathrm{L}^q(J_T;X)},
$$

where $C_2(y_\varepsilon^u) > 0$ has the same dependence on y_ε^u as $C_1(y_\varepsilon^u)$. The inequalities in (4.23) are shown analogous to the estimates which we derived for $(\tilde{y}_\varepsilon^{u,h}, \tilde{z}_\varepsilon^{u,h})$. We also conclude from $\tilde{y}_\varepsilon^{u,h} = \left(\frac{d}{dt} + T_{y,\varepsilon}^u\right)^{-1} h$ that there exists a constant $C(y_\varepsilon^u) > 0$ with

$$
\left\| \left(\frac{d}{dt} + T_{y,\varepsilon}^u \right)^{-1} \right\|_{\mathcal{L}(\mathrm{L}^q(J_T;X), Y_{q,0})} \leq C(y_\varepsilon^u).
$$

This proves maximal parabolic $\mathrm{L}^q(J_T;X)$-regularity of $T_{y,\varepsilon}^u$ for $q \in (\frac{1}{1-\alpha}, \infty)$. Again, the values $C(y_\varepsilon^u)$ can be chosen independently of ε for $\varepsilon \in (0, \varepsilon_*]$ small enough and if $\{y_\varepsilon^u\}$ converges to some function y with $\varepsilon \to 0$, which is the case for the sequence $\{\overline{y}_\varepsilon\}$ in Theorem 4.16.

(III) Adjoint operators and representation of $[T_{y,\varepsilon}^u]^*$:

Maximal parabolic $\mathrm{L}^q(J_T;X)$-regularity of $T_{y,\varepsilon}^u$ for $q \in (\frac{1}{1-\alpha}, \infty)$ implies maximal parabolic $\mathrm{L}^{q'}(J_T; [\mathrm{dom}(A_p)]^*)$-regularity of $[T_{y,\varepsilon}^u]^*$ [MS15, Lemma 4.10], see also [HMS15, Lemma 36]. We have to find a representation of $[T_{y,\varepsilon}^u]^*$. To this aim, we derive the adjoint mappings of the single components which define $T_{y,\varepsilon}^u$. Lemma 4.18 yields that multiplication with $\frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u)$ is well defined as a mapping from $\mathrm{L}^q(J_T)$ into $\mathrm{L}^q(J_T;X)$. Moreover, there holds $\left[\frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u)\right]^* = \langle \cdot, \frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u) \rangle_X$. Again by Lemma 4.18, $\frac{\partial}{\partial y} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u)$ is a linear continuous mapping from $\mathrm{L}^q(J_T;X^\alpha)$ into $\mathrm{L}^q(J_T;X)$. The mapping $\left[S \frac{\partial}{\partial y} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u)\right]^* : \mathrm{L}^{q'}(J_T) \to \mathrm{L}^{q'}(J_T; [X^\alpha]^*)$ is given by multiplication with $S \frac{\partial}{\partial y} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u)$. The adjoint operator of $T_{z,\varepsilon}^u$ maps any $v \in \mathrm{L}^{q'}(J_T)$ to the function $q \in \mathrm{L}^{q'}(J_T)$ which may be identified with the unique solution of

$$
-\dot{q}(t) = v(t) + S \frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u(t), z_\varepsilon^u(t)) q(t) - \frac{1}{\varepsilon} \Psi''(z_\varepsilon^u(t)) q(t) \quad \text{for } t \in J_T, \ q(T) = 0.
$$

The operators S^* and $[SA_p]^*$ are defined by multiplication with S and SA_p respectively. All bounds are independent of ε for the optimal states $\overline{y}_\varepsilon$ and $\overline{z}_\varepsilon$ according to Theorem 4.16 and if ε is small enough. All together, we obtain the representation

$$\begin{aligned}
[T^u_{y,\varepsilon}]^* &= A_p^* - \left[\frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\right]^* - \left[\frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon) T^u_{z,\varepsilon} S\left(-A_p + \frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\right)\right]^* \\
&= A_p^* - \left[\frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\right]^* + \left[[SA_p]^* - \left[S\frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\right]^*\right] [T^u_{z,\varepsilon}]^* \left[\frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\right]^* \qquad (4.26) \\
&= A_p^* - \left[\frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\right]^* + S\left[A_p - \frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\right] [T^u_{z,\varepsilon}]^* \langle \cdot, \frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\rangle_X.
\end{aligned}$$

(IV) Adjoint system and adjoint equation:
Due to maximal parabolic $\mathrm{L}^{q'}(J_T; [\mathrm{dom}(A_p)]^*)$-regularity of $[T^u_{y,\varepsilon}]^*$ there exists for each $\nu \in \mathrm{L}^{q'}(J_T; [\mathrm{dom}(A_p)]^*)$ a unique function $p^\nu_\varepsilon \in Y^*_{q',T}$ with $\left(-\frac{d}{dt} + [T^u_{\varepsilon,y}]^*\right) p = \nu$. For fixed $\nu \in \mathrm{L}^{q'}(J_T; [\mathrm{dom}(A_p)]^*)$ we define $q^\nu_\varepsilon := [T^u_{z,\varepsilon}]^* \langle p^\nu_\varepsilon, \frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\rangle_X$. As seen in Step III, q^ν_ε is the representative in $\mathrm{L}^{q'}(J_T)$ of the solution of

$$\begin{aligned}
&-\dot{q}(t) = \langle p^\nu_\varepsilon(t), \frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon(t), z^u_\varepsilon(t))\rangle_X + S\frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon(t), z^u_\varepsilon(t)) q(t) - \frac{1}{\varepsilon}\Psi''(z^u_\varepsilon(t)) q(t) \text{ for } t \in J_T, \\
&q(T) = 0.
\end{aligned}$$

Consider the solutions $(y^{u,h}_\varepsilon, z^{u,h}_\varepsilon)$ of (4.18)–(4.19) for some given $h \in \mathrm{L}^q(J_T; X)$. Then (4.18) and partial integration in time imply

$$\begin{aligned}
&\int_0^T \langle p^\nu_\varepsilon + S q^\nu_\varepsilon, h\rangle_X dt \\
&= \int_0^T \langle p^\nu_\varepsilon, \dot{y}^{u,h}_\varepsilon + A_p y^{u,h}_\varepsilon - \frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon) y^{u,h}_\varepsilon - \frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon) z^{u,h}_\varepsilon\rangle_X + \langle S q^\nu_\varepsilon, h\rangle_X dt \\
&= \int_0^T \langle -\dot{p}^\nu_\varepsilon + A_p^* p^\nu_\varepsilon - \left[\frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\right]^* p^\nu_\varepsilon, y^{u,h}_\varepsilon\rangle_{\mathrm{dom}(A_p)} + \langle S q^\nu_\varepsilon, h\rangle_X \\
&- \langle p^\nu_\varepsilon, \frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\rangle_X z^{u,h}_\varepsilon dt.
\end{aligned}$$

By definition of q^ν_ε we can replace the last term on the right side by

$$\int_0^T \left(\dot{q}^\nu_\varepsilon + S\frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon) q^\nu_\varepsilon - \frac{1}{\varepsilon}\Psi''(z^u_\varepsilon) q^\nu_\varepsilon\right) z^{u,h}_\varepsilon dt.$$

Hence, after another partial integration in time together with the evolution equation (4.19) of $z^{u,h}_\varepsilon$ we arrive at

$$\begin{aligned}
\int_0^T \langle p^\nu_\varepsilon + S q^\nu_\varepsilon, h\rangle_X dt &= \int_0^T \langle -\dot{p}^\nu_\varepsilon + A_p^* p^\nu_\varepsilon - \left[\frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\right]^* p^\nu_\varepsilon, y^{u,h}_\varepsilon\rangle_{\mathrm{dom}(A_p)} + \langle S q^\nu_\varepsilon, h\rangle_X \\
&- q^\nu_\varepsilon S\left[\left(-A_p + \frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\right) y^{u,h}_\varepsilon + h\right] dt \\
&= \int_0^T \langle -\dot{p}^\nu_\varepsilon + A_p^* p^\nu_\varepsilon - \left[\frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\right]^* p^\nu_\varepsilon, y^{u,h}_\varepsilon\rangle_{\mathrm{dom}(A_p)} \\
&+ S\left[A_p - \frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\right] q^\nu_\varepsilon y^{u,h}_\varepsilon dt.
\end{aligned}$$

Since $q_\varepsilon^\nu = [T_{z,\varepsilon}^u]^* \langle p_\varepsilon^\nu, \frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u)\rangle_X$, this together with the representation (4.26) for $[T_{y,\varepsilon}^u]^*$ from Step III and $\left(-\frac{d}{dt} + [T_{\varepsilon,y}^u]^*\right) p = \nu$ finally yields

$$\int_0^T \langle p_\varepsilon^\nu + S q_\varepsilon^\nu, h\rangle_X dt = \int_0^T \langle -\dot{p}_\varepsilon^\nu + [T_{y,\varepsilon}^u]^* p_\varepsilon^\nu, y_\varepsilon^{u,h}\rangle_{\mathrm{dom}(A_p)} dt = \int_0^T \langle \nu, y_\varepsilon^{u,h}\rangle_{\mathrm{dom}(A_p)} dt.$$

□

Lemma 4.19 provides us the main tool towards an optimality system for the optimal control problem (4.4),(4.5),(4.16).

Theorem 4.20 (Optimality system for the regularized problem)**.** *[Mün17b, Theorem 3.13] Adopt the assumptions of Theorem 4.16 and the notation from Lemma 4.19. For $i \in \{1,2\}$ and $\varepsilon \in (0, \varepsilon_*]$ let $\overline{u}_\varepsilon \in U_i$ be an optimal control for problem* (4.4),(4.5),(4.16)*. Then the adjoint variables for $\overline{y}_\varepsilon \in Y_{2,0}$ and $\overline{z}_\varepsilon \in \mathrm{H}^1(J_T)$ are given by $p_\varepsilon := p_\varepsilon^{\overline{y}_\varepsilon - y_d} \in Y_{2,T}^*$ and $q_\varepsilon := q_\varepsilon^{\overline{y}_\varepsilon - y_d} \in \mathrm{H}^1(J_T)$. There holds the optimality condition $B_i^*(p_\varepsilon + S q_\varepsilon) = -(\kappa+1)\overline{u}_\varepsilon + \overline{u}$ in U_i and the following system of evolution equations is satisfied by p_ε and q_ε:*

$$\begin{aligned} -\dot{p}_\varepsilon + A_p^* p_\varepsilon &= \left[\frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{y}_\varepsilon)\right]^* p_\varepsilon + S\left[-A_p + \frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon)\right] q_\varepsilon + \overline{y}_\varepsilon - y_d \quad \textit{for } t \in J_T, \qquad (4.27)\\ p_\varepsilon(T) &= 0,\\ -\dot{q}_\varepsilon &= \langle p_\varepsilon, \frac{\partial}{\partial z} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon)\rangle_X + S\frac{\partial}{\partial z} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon) q_\varepsilon - \frac{1}{\varepsilon}\Psi''(\overline{z}_\varepsilon) q_\varepsilon \qquad \textit{for } t \in J_T, \qquad (4.28)\\ q_\varepsilon(T) &= 0. \end{aligned}$$

Proof. By (A3)' in Assumption 4.2 there holds $2 > \frac{1}{1-\alpha} \Leftrightarrow \alpha < \frac{1}{2}$. Hence, Lemma 4.19 holds for $q = q' = 2$. We have to characterize the Gâteaux derivative of the reduced cost function $\mathcal{J}_{\mathrm{reg}}(u)$. To this aim, we prove that

$$\langle \overline{y}_\varepsilon - y_d, y_\varepsilon^{B_i \overline{u}_\varepsilon, B_i h}\rangle_{\mathrm{L}^2(J_T; \mathrm{dom}(A_p))} = \int_0^T \int_\Omega (I_p^{-1}\overline{y}_\varepsilon - y_d) \cdot I_p^{-1} y_\varepsilon^{B_i \overline{u}_\varepsilon, B_i h} dx dt$$

is well defined. By Corollary 2.30, the embedding $Y_{2,0} \hookrightarrow U_1$ is continuous because $\mathrm{dom}(A_p) \simeq \mathbb{W}_{\Gamma_D}^{1,p}(\Omega) \hookrightarrow [\mathrm{L}^2(\Omega)]^m$, see also Remark 2.32. In particular, there holds $I_p^{-1}\overline{y}_\varepsilon - y_d \in U_1$. Note that B_1 is the extension of I_p from $\mathbb{W}_{\Gamma_D}^{1,p}(\Omega)$ to $[\mathrm{L}^2(\Omega)]^m$. Hence, $I_p^{-1}\overline{y}_\varepsilon - y_d$ can be identified with $\overline{y}_\varepsilon - B_1 y_d \in \mathrm{L}^2(J_T; X)$.
By Corollary 4.17 there holds $y_\varepsilon^{B_i \overline{u}_\varepsilon, B_i h} \in Y_{2,0} \hookrightarrow \mathrm{L}^2(J_T; \mathrm{dom}(A_p))$. Again according to Corollary 2.30, this allows us to identify $y_\varepsilon^{B_i \overline{u}_\varepsilon, B_i h}$ with $I_p^{-1} y_\varepsilon^{B_i \overline{u}_\varepsilon, B_i h} \in \mathrm{L}^2(J_T; \mathbb{W}_{\Gamma_D}^{1,p}(\Omega))$. Since $p' \le 2 \le p$, $I_p^{-1} y_\varepsilon^{B_i \overline{u}_\varepsilon, B_i h}$ is contained in $\mathrm{L}^2(J_T; \mathbb{W}_{\Gamma_D}^{1,p'}(\Omega)) \simeq \mathrm{L}^2(J_T; X^*)$.
Furthermore, Corollary 2.19 entails $(\mathcal{A}_p + I_p)^{-1} \in \mathcal{L}\big(X, \mathbb{W}_{\Gamma_D}^{1,p}(\Omega)\big)$. Consequently, for some $c > 0$ and for a.e. $t \in J_T$, there holds

$$\begin{aligned} &\left\| I_p^{-1} y_\varepsilon^{B_i \overline{u}_\varepsilon, B_i h}(t)\right\|_{\mathbb{W}_{\Gamma_D}^{1,p'}(\Omega)} \le c \left\| I_p^{-1} y_\varepsilon^{B_i \overline{u}_\varepsilon, B_i h}(t)\right\|_{\mathbb{W}_{\Gamma_D}^{1,p}(\Omega)}\\ &= c \left\| (\mathcal{A}_p + I_p)^{-1} (\mathcal{A}_p + I_p) I_p^{-1} y_\varepsilon^{B_i \overline{u}_\varepsilon, B_i h}(t)\right\|_{\mathbb{W}_{\Gamma_D}^{1,p}(\Omega)}\\ &\le c \left\| (\mathcal{A}_p + I_p)^{-1}\right\|_{\mathcal{L}\left(X, \mathbb{W}_{\Gamma_D}^{1,p}(\Omega)\right)} \left\| y_\varepsilon^{B_i \overline{u}_\varepsilon, B_i h}(t)\right\|_{\mathrm{dom}(A_p)}. \end{aligned}$$

All together, we obtain

$$\begin{aligned}&\left|\int_0^T\int_\Omega (I_p^{-1}\overline{y}_\varepsilon - y_d)\cdot I_p^{-1}y_\varepsilon^{B_i\overline{u}_\varepsilon,B_ih}dxdt\right| = \left|\int_0^T \langle \overline{y}_\varepsilon - B_1y_d, I_p^{-1}y_\varepsilon^{B_i\overline{u}_\varepsilon,B_ih}\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)}dt\right|\\ &\leq \|\overline{y}_\varepsilon - B_1y_d\|_{\mathrm{L}^2(J_T;X)}\left\|(\mathcal{A}_p+I_p)^{-1}\right\|_{\mathcal{L}\left(X,\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)\right)}\|y_\varepsilon^{B_i\overline{u}_\varepsilon,B_ih}\|_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))}.\end{aligned}$$

The Gâteaux derivative of the reduced cost function

$$\mathcal{J}_{\mathrm{reg}}(u) := J_{\mathrm{reg}}(I_p^{-1}G_\varepsilon(B_iu),u;\overline{u}) = J(I_p^{-1}G_\varepsilon(B_iu),u) + \frac{1}{2}\|u-\overline{u}\|^2_{U_i}$$

with respect to u has to be zero at $\overline{u}_\varepsilon$ by optimality.
Let J_1, J_2 be defined by

$$\begin{aligned}&J_1: Y_{2,0}\to\mathbb{R},\quad y\mapsto \frac{1}{2}\|I_p^{-1}y-y_d\|^2_{U_1} = \frac{1}{2}\langle I_p^{-1}y-y_d, I_p^{-1}y-y_d\rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))},\\ &J_2: U_i\to\mathbb{R},\qquad u\mapsto \frac{\kappa}{2}\|u\|^2_{U_i}+\frac{1}{2}\|u-\overline{u}\|^2_{U_i}.\end{aligned}$$

Then $J_{\mathrm{reg}}(y,u) = J_1(y)+J_2(u)$. Consider $y\in Y_{2,0}$ and $\tilde{y}\in Y_{2,0}\backslash\{0\}$. The difference quotient $\frac{J_1(y+\tilde{y})-J_1(y)}{\|\tilde{y}\|_{Y_{2,0}}}$ satisfies

$$\frac{J_1(y+\tilde{y})-J_1(y)-\langle y-y_d,\tilde{y}\rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))}}{\|\tilde{y}\|_{Y_{2,0}}} = \frac{\langle\tilde{y},\tilde{y}\rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))}}{2\|\tilde{y}\|_{Y_{2,0}}}. \tag{4.29}$$

Moreover, similar as for $\langle\overline{y}_\varepsilon - y_d, y_\varepsilon^{B_i\overline{u}_\varepsilon,B_ih}\rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))}$ we estimate

$$\langle\tilde{y},\tilde{y}\rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))}\leq c\|\tilde{y}\|_{\mathrm{L}^2(J_T;X)}\left\|(\mathcal{A}_p+I_p)^{-1}\right\|_{\mathcal{L}\left(X,\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)\right)}\|\tilde{y}\|_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))}\leq c\|\tilde{y}\|^2_{Y_{2,0}}$$

for some $c>0$. For fixed y, similar arguments imply that the function

$$\tilde{y}\mapsto\langle y-y_d,\tilde{y}\rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))}$$

defines a linear and continuous functional on $Y_{2,0}$. That is, letting $\|\tilde{y}\|\to 0$ in (4.29), it follows that $J_1'[y;\tilde{y}] = \langle y-y_d,\tilde{y}\rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))}$ is the derivative of J_1 at y in direction $\tilde{y}$. By standard techniques one shows $J_2'[u;h] = \kappa\langle u,h\rangle_{U_i} + \langle u-\overline{u},h\rangle_{U_i}$ for $u,h\in U_i$. Hence, for $h\in U_i$, the chain rule and optimality of $\overline{u}_\varepsilon$ implies

$$\begin{aligned}0 = \mathcal{J}_{\mathrm{reg}}'[\overline{u}_\varepsilon;h] &= J_1'[\overline{y}_\varepsilon;G_\varepsilon'[\overline{u}_\varepsilon;h]] + J_2'[\overline{u}_\varepsilon;h]\\ &= \langle\overline{y}_\varepsilon - y_d, y^{B_i\overline{u}_\varepsilon,B_ih}\rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))} + \kappa\langle\overline{u},h\rangle_{U_i} + \langle\overline{u}_\varepsilon-\overline{u},h\rangle_{U_i}.\end{aligned} \tag{4.30}$$

We replace the first term in (4.30) by the adjoint equation (4.22) with $\nu = \overline{y}_\varepsilon - y_d$ and compute

$$\begin{aligned}0 &= \mathcal{J}_{\mathrm{reg}}'[\overline{u}_\varepsilon;h] = \langle\overline{y}_\varepsilon - y_d, y_\varepsilon^{B_i\overline{u}_\varepsilon,B_ih}\rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))} + \kappa\langle\overline{u}_\varepsilon,h\rangle_{U_i} + \langle\overline{u}_\varepsilon-\overline{u},h\rangle_{U_i}\\ &= \langle p_\varepsilon + Sq_\varepsilon, B_ih\rangle_{\mathrm{L}^2(J_T;X)} + \langle(\kappa+1)u_\varepsilon - \overline{u},h\rangle_{U_i} = \langle B_i^*(p_\varepsilon+Sq_\varepsilon)+(\kappa+1)u_\varepsilon-\overline{u},h\rangle_{U_i}.\end{aligned}$$

Since $h\in U_i$ was arbitrary, this concludes the proof. □

Corollary 4.21. Adopt the assumptions of Theorem 4.16 and the notation from Theorem 4.20. There holds

$$\langle\overline{y}_\varepsilon - y_d, y_\varepsilon^{B_i\overline{u}_\varepsilon,h}\rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))} = \langle p_\varepsilon + Sq_\varepsilon, h\rangle_{\mathrm{L}^2(J_T;X)}\quad\forall h\in\mathrm{L}^2(J_T;X), \tag{4.31}$$

where

$$\langle\overline{y}_\varepsilon - y_d, y_\varepsilon^{B_i\overline{u}_\varepsilon,h}\rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))} = \int_0^T\int_\Omega (I_p^{-1}\overline{y}_\varepsilon - y_d)\cdot I_p^{-1}y_\varepsilon^{B_i\overline{u}_\varepsilon,B_ih}dxdt.$$

In particular, (4.31) holds for all B_ih, $h\in U_i$.

Proof. (4.31) follows from (4.22) in Lemma 4.19. □

4.3.6 Estimates for the adjoints of the regularized problem

In this subsection, we proceed similar as in [BK13, Section 3.5] and [MS15, Lemma 4.14]. In particular, we estimate the norms of the adjoint states p_ε and q_ε from Theorem 4.20 independently of ε and of the norms of the optimal controls $\overline{u}_\varepsilon$. In Section 4.4, we drive the regularization parameter ε to zero. As already for optimal solutions $(\overline{u}_\varepsilon, \overline{y}_\varepsilon, \overline{z}_\varepsilon)$ of problem (4.4),(4.5),(4.16), a weak compactness argument yields weakly (star) converging subsequences of p_ε and q_ε. As a consequence, we obtain an adjoint system for problem (4.1)–(4.3), see Theorem 4.38 below.

Lemma 4.22 (Uniform bounds)**.** [Mün17b, Lemma 3.14] Adopt the assumptions and the notation of Theorem 4.20. There exists a constant $c > 0$ which is independent of ε and some $\varepsilon_0 \in (0, \varepsilon_*]$ such that the following holds true. If $\varepsilon \in (0, \varepsilon_0)$, then

$$0 \leq \|q_\varepsilon\|_{\mathrm{C}(\overline{J_T})} + \frac{1}{\varepsilon}\int_0^T \Psi''(\overline{z}_\varepsilon(s))|q_\varepsilon(s)|ds \leq c, \tag{4.32}$$

$$\int_0^T |\dot{q}_\varepsilon(s)|ds \leq c, \tag{4.33}$$

$$\|p_\varepsilon\|_{Y^*_{2,T}} \leq c, \tag{4.34}$$

$$\left\| \left[\frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon)\right]^* p_\varepsilon \right\|_{\mathrm{L}^2(J_T;[X^\alpha]^*)} \leq c, \tag{4.35}$$

$$\left\| S\frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon) q_\varepsilon \right\|_{\mathrm{L}^2(J_T;[X^\alpha]^*)} \leq c, \text{ as well as} \tag{4.36}$$

$$\|SA_p q_\varepsilon\|_{\mathrm{C}(\overline{J_T};[X^\alpha]^*)} \leq c. \tag{4.37}$$

Proof. Remember Theorem 4.16, where we proved that $\overline{u}_\varepsilon \to \overline{u}$ in U_i, $\overline{y}_\varepsilon \to \overline{y}$ in $Y_{2,0}$ and in $\mathrm{C}(\overline{J_T}; X^\alpha)$ and $\overline{z}_\varepsilon \to \overline{z}$ weakly in $\mathrm{H}^1(J_T)$ and strongly in $\mathrm{C}(\overline{J_T})$. As seen in the proof of Theorem 4.20, $\overline{y}_\varepsilon - y_d$ is well defined as an element of $\mathrm{L}^2(J_T; [\mathrm{dom}(A_p)]^*)$ with the assignment

$$\langle \overline{y}_\varepsilon - y_d, v\rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))} = \int_0^T \int_\Omega (I_p^{-1}\overline{y}_\varepsilon - y_d) \cdot I_p^{-1} v \, dxdt \quad \forall v \in \mathrm{L}^2(J_T; \mathrm{dom}(A_p)).$$

Moreover, there holds

$$\begin{aligned} \|\overline{y}_\varepsilon - y_d\|_{\mathrm{L}^2(J_T;[\mathrm{dom}(A_p)]^*)} &= \sup_{v \in B_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))}(0,1)} \langle \overline{y}_\varepsilon - y_d, v\rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p)} \\ &\leq c\|\overline{y}_\varepsilon - B_1 y_d\|_{\mathrm{L}^2(J_T;X)} \left\|(\mathcal{A}_p + I_p)^{-1}\right\|_{\mathcal{L}\left(X;\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)\right)} =: c_0. \end{aligned}$$

Since $\overline{y}_\varepsilon$ converges to $\overline{y}$ in $Y_{2,0}$ and because $Y_{2,0} \hookrightarrow \mathrm{L}^2(J_T; X)$, we can choose the constant $c_0 > 0$ independently of ε. Remember the definition $(p_\varepsilon, q_\varepsilon) = (p_\varepsilon^{\overline{y}_\varepsilon - y_d}, q_\varepsilon^{\overline{y}_\varepsilon - y_d})$. Hence, for any $\xi \in \mathrm{L}^2(J_T; X)$, we make use of equation (4.31) in Corollary 4.21 and the uniform estimate (4.23) in Lemma 4.19 applied to $y_\varepsilon^{B_i\overline{u}_\varepsilon,\xi}$. We obtain

$$\langle p_\varepsilon + Sq_\varepsilon, \xi\rangle_{\mathrm{L}^2(J_T;X)} = \langle \overline{y}_\varepsilon - y_d, y_\varepsilon^{B_i\overline{u}_\varepsilon,\xi}\rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))} \leq c_0 C(\overline{y}_\varepsilon)\|\xi\|_{\mathrm{L}^2(J_T;X)}.$$

Moreover, because $\overline{y}_\varepsilon \to \overline{y}$ in $Y_{2,0}$ we can find some $\varepsilon_0 > 0$ such that $C(\overline{y}_\varepsilon) = C(\overline{y})$ for all $\varepsilon \in (0, \varepsilon_0)$. Reflexivity of $\mathrm{L}^2(J_T; X)$ finally yields

$$\|p_\varepsilon + Sq_\varepsilon\|_{\mathrm{L}^2(J_T;X^*)} \leq c_0 C(\overline{y}) =: c_1 \quad \forall \varepsilon \in (0, \varepsilon_0). \tag{4.38}$$

In the next step we prove (4.32). To this aim, we test the evolution equation (4.28) with $q_\varepsilon/|q_\varepsilon|$. Afterwards, for arbitrary $t \in J_T$ we integrate over (t,T). With the bound (4.20) from Lemma 4.18 for the derivative $f_\varepsilon'[\overline{y}_\varepsilon, \overline{z}_\varepsilon]$ and (4.38) we arrive at

$$\begin{aligned} |q_\varepsilon(t)| + \frac{1}{\varepsilon}\int_t^T \Psi''(\overline{z}_\varepsilon(s))|q_\varepsilon(s)|ds &= \int_t^T \langle p_\varepsilon(s) + Sq_\varepsilon(s), \frac{\partial}{\partial z} f_\varepsilon(\overline{y}_\varepsilon(s), \overline{z}_\varepsilon(s))\rangle_X \frac{q_\varepsilon(s)}{|q_\varepsilon(s)|}\, ds \\ &\le c_1 \left\| \frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon, \overline{z}_\varepsilon)\right\|_{\mathrm{L}^2(J_T;X)} \le c_1 K(\overline{y}_\varepsilon). \end{aligned} \tag{4.39}$$

Again by Lemma 4.18 and the convergence of $\overline{y}_\varepsilon$ there holds $c_1 K(\overline{y}_\varepsilon) = c_1 K(\overline{y}) =: c_2$ for all $\varepsilon \in (0, \varepsilon_0)$. W.l.o.g. we can choose ε_0 as above. Convexity of Ψ implies $\Psi''(\overline{z}_\varepsilon) \ge 0$, see $(A4)_\varepsilon$. This proves (4.32) because

$$0 \le \|q_\varepsilon\|_{\mathrm{C}(\overline{J_T})} + \frac{1}{\varepsilon}\int_0^T \Psi''(\overline{z}_\varepsilon(s))|q_\varepsilon(s)|ds \le c_2 \quad \forall \varepsilon \in (0, \varepsilon_0). \tag{4.40}$$

With $S \in X^*$, we conclude $Sq_\varepsilon \in \mathrm{L}^2(J_T; X^*)$. Hence, by (4.38) also $p_\varepsilon \in \mathrm{L}^2(J_T; X^*)$ holds and both of the norms $\|p_\varepsilon\|_{\mathrm{L}^2(J_T;X^*)}$ and $\|Sq_\varepsilon\|_{\mathrm{L}^2(J_T;X^*)}$ are bounded independently of $\varepsilon \in (0, \varepsilon_0)$.
We continue with the proof of (4.33). The representation (4.28) for $\dot{q}_\varepsilon$ yields

$$\int_0^T |\dot{q}_\varepsilon(s)|ds \le \int_0^T |\langle p_\varepsilon(s) + Sq_\varepsilon(s), \frac{\partial}{\partial z} f_\varepsilon(\overline{y}_\varepsilon(s), \overline{z}_\varepsilon(s))\rangle_X|ds + \frac{1}{\varepsilon}\int_0^T \Psi''(\overline{z}_\varepsilon(s))|q_\varepsilon(s)|ds.$$

From (4.39) we deduce that the right side is bounded by $2c_2$ and hence (4.33) follows from $\int_0^T |\dot{q}_\varepsilon(s)|ds \le 2c_2 =: c_3$ for $\varepsilon \in (0, \varepsilon_0)$.
It remains to prove the estimates concerning p_ε. To show (4.34), note that maximal parabolic regularity of A_p on X implies maximal parabolic $\mathrm{L}^2(J_T; [\mathrm{dom}(A_p)]^*)$-regularity of A_p^* [HMS15, Lemma 36]. With the evolution equation (4.27) for p_ε we estimate

$$\begin{aligned} \|p_\varepsilon\|_{Y_{2,T}^*} \le & \left\| \left(-\frac{d}{dt} + A_p^*\right)^{-1} \right\|_{\mathcal{L}(\mathrm{L}^2(J_T;[\mathrm{dom}(A_p)]^*), Y_{2,T}^*)} \\ & \left\| \left[\frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon)\right]^* p_\varepsilon + S\left[-A_p + \frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon)\right] q_\varepsilon + \overline{y}_\varepsilon - y_d \right\|_{\mathrm{L}^2(J_T;[\mathrm{dom}(A_p)]^*)} \\ \le & \left\| \left(-\frac{d}{dt} + A_p^*\right)^{-1} \right\|_{\mathcal{L}(\mathrm{L}^2(J_T;[\mathrm{dom}(A_p)]^*), Y_{2,T}^*)} \\ & \left(\left\| \frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon) \right\|_{\mathcal{L}(\mathrm{L}^2(J_T;X^\alpha), \mathrm{L}^2(J_T;X))} \|p_\varepsilon + Sq_\varepsilon\|_{\mathrm{L}^2(J_T;X^*)} \right. \\ & \left. + \|SA_p\|_{[X^\alpha]^*} \|q_\varepsilon\|_{\mathrm{C}(\overline{J_T})} + \|\overline{y}_\varepsilon - y_d\|_{\mathrm{L}^2(J_T;[\mathrm{dom}(A_p)]^*)} \right). \end{aligned}$$

Note that we used $[X^\alpha]^* \hookrightarrow [\mathrm{dom}(A_p)]^*$ which follows from $\mathrm{dom}(A_p) \hookrightarrow X^\alpha$, see Corollary 2.30. We apply the bound (4.20) from Lemma 4.18 for $f_\varepsilon'[\overline{y}_\varepsilon, \overline{z}_\varepsilon]$ together with the estimate (4.38) and remember (4.40) and $\|\overline{y}_\varepsilon - y_d\|_{\mathrm{L}^2(J_T;[\mathrm{dom}(A_p)]^*)} \le c_0$. Finally, we conclude (4.34) from

$$\|p_\varepsilon\|_{Y_{2,T}^*} \le \left\| \left(-\frac{d}{dt} + A_p^*\right)^{-1} \right\|_{\mathcal{L}(\mathrm{L}^2(J_T;[\mathrm{dom}(A_p)]^*), Y_{2,T}^*)} \big(c_1 K(\overline{y}) + \|SA_p\|_{[X^\alpha]^*} c_2 + c_0\big) =: c_4$$

for $\varepsilon \in (0, \varepsilon_0)$. The proofs of (4.35)–(4.37) are similar. They require the estimates

$$\left\| \left[\frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon) \right]^* p_\varepsilon \right\|_{\mathrm{L}^2(J_T;[X^\alpha]^*)} \leq \left\| \frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon) \right\|_{\mathcal{L}(\mathrm{L}^2(J_T;X^\alpha),\mathrm{L}^2(J_T;X))} \|p_\varepsilon\|_{\mathrm{L}^2(J_T;X^*)},$$

$$\left\| S \frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon) q_\varepsilon \right\|_{\mathrm{L}^2(J_T;[X^\alpha]^*)} \leq \left\| \frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon) \right\|_{\mathcal{L}(\mathrm{L}^2(J_T;X^\alpha),\mathrm{L}^2(J_T;X))} \|S q_\varepsilon\|_{\mathrm{L}^2(J_T;X^*)}, \quad \text{and}$$

$$\|S A_p q_\varepsilon\|_{\mathrm{C}(\overline{J_T};[X^\alpha]^*)} \leq \|S A_p\|_{[X^\alpha]^*} \|q_\varepsilon\|_{\mathrm{C}(\overline{J_T})}.$$

□

4.4 Adjoint system and optimality conditions for the optimal control problem

In Theorem 4.16 we proved convergence of solutions $(\overline{u}_\varepsilon, \overline{y}_\varepsilon, \overline{z}_\varepsilon)$ of problem (4.4)–(4.16) to solutions $(\overline{u}, \overline{y}, \overline{z})$ of problem (4.1)–(4.3) with $\varepsilon \to 0$.
In this section, we consider the limit $\varepsilon \to 0$ in Theorem 4.20 in order to derive first order optimality conditions for the original optimization problem (4.1)–(4.3) from the adjoint systems $(p_\varepsilon, q_\varepsilon)$ of the regularized control problems. Moreover, we clarify the type of convergence and analyze the limit system (p, q). The proceeding is oriented at [BK13, Section 4] and [MS15, Theorem 4.15].
In Subsections 4.4.1–4.4.2 we study the general problem with spatially distributed or boundary controls, i.e. with $i \in \{1, 2\}$. In the first part of Subsection 4.4.1 we derive an adjoint system (p, q) for problem (4.1)–(4.3) for the optimal control $\overline{u}$ from Theorem 4.16. The main result here is Lemma 4.23. While the evolution of p is determined by the limit $\varepsilon \to 0$ of the evolution equations for p_ε, the characterization of the limit function q is more involved. Hence, in the second part of Subsection 4.4.1, we analyze the evolution behaviour and the continuity properties of q.
In Subsection 4.4.2, we complete the discussion of the general problem $i \in \{1, 2\}$ by proving optimality conditions for problem (4.1)–(4.3) for the optimal control $\overline{u}$ in terms of p and q, see Lemma 4.37.
The results from Subsections 4.4.1–4.4.2 are summarized in Theorem 4.38 in Subsection 4.4.3. Moreover, an improvement of the optimality condition (4.48) from Theorem 4.38 for the particular case when f is continuously differentiable is derived in Corollary 4.39.
Both optimality conditions (4.48) and (4.56) are weak in the sense that they are restricted to test functions $y^{B_i\overline{u}, B_i h}$ with $h \in U_i$, $i \in \{1, 2\}$.
Accordingly, the next question to ask is whether they hold in a strong way, i.e. if the functions $y^{B_i\overline{u}, B_i h}$ can be replaced by arbitrary elements $v \in \operatorname{dom}(A_p)$ and if the corresponding inequalities hold a.e. in time? This is not possible without B_i having dense range. Furthermore, the evolution of q in Subsections 4.4.1–4.4.2 is not completely understood. In particular, on the subset $I_\partial \subset \overline{J_T}$ of times where the hysteresis $\overline{z}$ touches the boundary points $\{a, b\}$ of the interval $[a, b]$, the measure $dq \in \mathrm{C}(\overline{J_T})$ which determines q still depends on a measure $d\mu$ which we cannot characterize completely for $i \in \{1, 2\}$.
That is, in Subsections 4.4.4–4.4.5, we focus on the control problem with $i = 1$ for which the control functions are distributed in Ω.
Indeed, in Subsection 4.4.4, we exploit the fact that B_1 has dense range for appropriate $p \geq 2$ in order to improve the optimality conditions from Theorem 4.38 and Corollary 4.39. More precisely, while the non-locality in time of $\mathcal{W}$ hinders us to extend the variational inequalities (4.48) and (4.56) to strong maximum conditions, we are able to prove optimality conditions for test functions of the form $v\varphi$ with $v \in \operatorname{dom}(A_p)$, $Sv > 0$ and $\varphi \in \mathrm{C}_0^\infty(J_T)$, see Corollary 4.40. Dividing by Sv on both sides yields, at least in the case (4.56), an optimality condition with

arbitrary test functions $\varphi \in C_0^\infty(J_T)$ and independent of v. The result is a variational inequality in time only. For $i = 1$, we also apply injectivity of B_1^* to prove uniqueness of p and q if f is continuously differentiable, see Corollary 4.41.

In Subsection 4.4.5, we return to the question of characterizing $d\mu$ and therewith of dq. This part extends the results in [Mün17b]. We introduce different categories of times $E \subset I_\partial$, see Definition 4.24. Depending on the category of the set E, we characterize the sign of $d\mu(E)$ and prove upper bounds for $|d\mu(E)|$, see Lemma 4.46 and Theorem 4.47. Afterwards, we exploit the relation between $d\mu$ and dq and conclude sign conditions and bounds for $dq(E)$, see Corollary 4.48. Finally, we characterize the continuity properties of q in Corollary 4.49.

4.4.1 Adjoint system for distributed or boundary controls

In this subsection, we derive an adjoint system (p, q) for problem (4.1)–(4.3) for $i \in \{1, 2\}$. The evolution equation and the regularity properties of p are obtained by a standard compactness argument as the limit $\varepsilon \to 0$ of p_ε, see Lemma 4.23. The limiting procedure for q_ε is more involved, since Lemma 4.22 provides a uniform-in-ε-bound of the norm of $\dot{q}_\varepsilon$ only in $L^1(J_T)$. Hence, we only obtain weak star convergence of q_ε, see Lemma 4.23. Low regularity of the limit $q \in BV(J_T)$ complicates the characterization of its time evolution. To get more insight into the behaviour of q, it turns out useful to split the interval J_T into the set I_0 of times t where the limit $\overline{z}(t)$ is contained in the open interval (a, b) and the rest I_∂ where $\overline{z}(t) \in \{a, b\}$. Indeed, in open subintervals of I_0, q is an H^1-function and can be described by an evolution equation, see Lemma 4.25 below. On I_∂, there remains a measure $d\mu \in C(\overline{J_T})^*$ which corresponds to the limit $\varepsilon \to 0$ of $\frac{1}{\varepsilon}\Psi''(\overline{z}_\varepsilon)q_\varepsilon$ in the evolution equation of q_ε. The result is an equality for dq in the sense of measures on I_∂ which depends on $d\mu$, see Lemma 4.28. Although the abstract measure $d\mu$ has its support only in I_∂, this is not quite satisfying since I_∂ is a-priori unknown. Moreover, $d\mu$ appears in the optimality conditions for problem (4.1)–(4.3). Hence, to characterize $d\mu$ would not only complete the description of q, but also help to interpret the optimality condition. With this regard, we introduce a regularity Assumption 4.30 which essentially supposes that $S\overline{y}(t)$ is strictly monotone for $t \in I_\partial$. In extension to [Mün17b], we provide an example for a case in which Assumption 4.30 is satisfied, see Example 4.32. With Assumption 4.30 we can shrink the support of $d\mu$ to a subset of I_∂. In particular, we are able to describe the evolution of q in open subintervals of I_∂ and we can prove that q is continuous at so-called $(0, \partial)$-switching times, see Lemma 4.35. Since a deeper analysis of $d\mu$ is more involved and technical, we dedicate the whole Subsection 4.4.5 to this question.

Lemma 4.23 (Adjoint system in the limit)**.** [Mün17b, Lemma 4.1] Adopt the assumptions and the notation of Theorem 4.20. For $i \in \{1, 2\}$ let $\overline{u} \in U_i$, $\overline{y} = G(\overline{u})$ and $\overline{z} = \mathcal{W}[S\overline{y}]$ be defined as in Theorem 4.16. Then every sequence $\{\varepsilon\}$ with $\varepsilon \to 0$ has a subsequence $\{\varepsilon_k\}$ such that the following holds true. There exist functions functions $p \in Y_{2,T}^*$ and $\lambda_1, \lambda_2 \in L^2(J_T; [X^\alpha]^*)$ such that as $k \to \infty$, $p_{\varepsilon_k} \rightharpoonup p$ in $Y_{2,T}^*$ and

$$\left[\frac{\partial}{\partial y} f_{\varepsilon_k}(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k})\right]^* p_{\varepsilon_k} \rightharpoonup \lambda_1 \quad \text{in } L^2(J_T; [X^\alpha]^*),$$
$$S\frac{\partial}{\partial y} f_{\varepsilon_k}(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k}) q_{\varepsilon_k} \rightharpoonup \lambda_2 \quad \text{in } L^2(J_T; [X^\alpha]^*).$$

Moreover, there exists a function q which has bounded variation, i.e. $q \in BV(J_T)$, such that q_{ε_k} converges pointwise to q with $k \to \infty$. There holds $\mathrm{Var}(q) \le \liminf_{\varepsilon_k \to 0} \mathrm{Var}(q_{\varepsilon_k})$. Alternatively, $\dot{q}_{\varepsilon_k} \to dq$ weak star in $C(\overline{J_T})^*$ with $k \to \infty$ for some signed regular Borel measure $dq \in C(\overline{J_T})^*$. The relation between q and dq is given by $q(t-) - q(s+) = dq((s, t))$ and $q(t+) - q(s-) = dq([s, t])$ for $[s, t] \subset \overline{J_T}$.

The function p solves the evolution equation

$$-\dot{p} + A_p^* = \lambda_1 + \lambda_2 - SA_p q + \overline{y} - y_d \quad \text{for } t \in J_T,\ p(T) = 0. \tag{4.41}$$

If f is continuously differentiable from $X^\alpha \times \mathbb{R}$ into X then $\lambda_1 = \left[\frac{\partial}{\partial y} f(\overline{y}, \overline{z})\right]^* p$ and $\lambda_2 = S\frac{\partial}{\partial y} f(\overline{y}, \overline{z})q$. Furthermore,

$$B_i^*(p + Sq) = -\kappa\overline{u} \quad \text{in } U_i. \tag{4.42}$$

Proof. We exploit the uniform estimates in Lemma 4.22 and apply a weak compactness argument. First of all, remember the convergence results of Theorem 4.16, i.e. that $u_\varepsilon \to \overline{u}$ in U_i, $\overline{y}_\varepsilon \to \overline{y}$ in $Y_{2,0}$ and in $\mathrm{C}(\overline{J_T}; X^\alpha)$ and $\overline{z}_\varepsilon \to \overline{z}$ uniformly and weakly in $\mathrm{H}^1(J_T)$ with $\varepsilon \to 0$. The spaces $Y_{2,0}^*$ and $\mathrm{L}^2(J_T; [X^\alpha]^*)$ are reflexive. Hence, because the estimates (4.34)–(4.36) in Lemma 4.22 are uniform in ε, there exist functions $p \in Y_{2,0}^*$ and $\lambda_1, \lambda_2 \in \mathrm{L}^2(J_T; [X^\alpha]^*)$ together with a subsequence $\{\varepsilon_k\}$ such that $p_{\varepsilon_k} \rightharpoonup p$ in $Y_{2,T}^*$, $\left[\frac{\partial}{\partial y} f_{\varepsilon_k}(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k})\right]^* p_{\varepsilon_k} \rightharpoonup \lambda_1$ in $\mathrm{L}^2(J_T; [X^\alpha]^*)$ and $S\frac{\partial}{\partial y} f_{\varepsilon_k}(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k})q_{\varepsilon_k} \rightharpoonup \lambda_2$ in $\mathrm{L}^2(J_T; [X^\alpha]^*)$ with $k \to \infty$. Note that $p(T) = 0$ by definition of $Y_{2,T}^*$. Estimate (4.33) provides us a uniform-in-ε bound of $\dot{q}_\varepsilon$ in $\mathrm{L}^1(J_T)$. This implies that q_ε has uniformly bounded variation, i.e. $q_\varepsilon \in \mathrm{BV}(J_T)$, with a norm which is bounded independently of ε. A weak form of Helly's theorem in Banach spaces [BP12, Theorem 1.126] implies that (w.l.o.g. the same) subsequence q_{ε_k} converges pointwise to some $q \in \mathrm{BV}(J_T)$ with $k \to \infty$ and $\mathrm{Var}(q) \le \liminf_{\varepsilon_k \to 0} \mathrm{Var}(q_{\varepsilon_k})$. Alternatively, by Alaoglu's compactness theorem, $\dot{q}_{\varepsilon_k} \to dq$ weak star in $\mathrm{C}(\overline{J_T})^*$ with $k \to \infty$ for some signed regular Borel measure $dq \in \mathrm{C}(\overline{J_T})^*$. The relation between q and dq is given by $q(t-) - q(s+) = dq((s,t))$ and $q(t+) - q(s-) = dq([s,t])$ for $[s,t] \subset \overline{J_T}$, see [BP12, Chapter 1.3.3] or [BK13, Section 4]. The operator $-\frac{d}{dt} + A_p^*$ is linear and continuous from $Y_{2,T}^*$ to $\mathrm{L}^2(J_T; [\mathrm{dom}(A_p)]^*)$ and hence weakly continuous. Therefore, we obtain

$$\begin{aligned}
0 &= -\dot{p}_{\varepsilon_k} + A_p^* p_{\varepsilon_k} - \left[\frac{\partial}{\partial y} f_{\varepsilon_k}(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k})\right]^* p_{\varepsilon_k} + SA_p q_{\varepsilon_k} - S\frac{\partial}{\partial y} f_\varepsilon(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k})q_{\varepsilon_k} - (\overline{y}_{\varepsilon_k} - y_d) \\
&\rightharpoonup -\dot{p} + A_p^* - \lambda_1 - \lambda_2 + SA_p q - (\overline{y} - y_d)
\end{aligned}$$

in $\mathrm{L}^2(J_T; [\mathrm{dom}(A_p)]^*)$ with $k \to \infty$. Consequently, $p \in Y_{2,T}^*$ solves equation (4.41).
Assume now that f is continuously differentiable from $X^\alpha \times \mathbb{R}$ into X, so that we can define $f_\varepsilon \equiv f$ in Assumption 4.7. Remember the strong convergence of $\overline{y}_{\varepsilon_k}$ to $\overline{y}$ in $\mathrm{C}(\overline{J_T}; X^\alpha)$ and of $\overline{z}_{\varepsilon_k}$ to $\overline{z}$ in $\mathrm{C}(\overline{J_T})$. Moreover, $p_{\varepsilon_k} \rightharpoonup p$ in $Y_{2,T}^*$ and $q_{\varepsilon_k} \to q$ pointwise and hence $Sq_{\varepsilon_k} \rightharpoonup Sq$ weakly in $\mathrm{L}^2(J_T; X^*)$ with $k \to \infty$. Since f is continuously differentiable, there holds $\left[\frac{\partial}{\partial y} f(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k})\right]^* \to \left[\frac{\partial}{\partial y} f(\overline{y}, \overline{z})\right]^*$ in $\mathcal{L}(\mathrm{L}^2(J_T; X^*), \mathrm{L}^2(J_T; [X^\alpha]^*))$ with $k \to \infty$. This implies

$$\left[\frac{\partial}{\partial y} f(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k})\right]^* p_{\varepsilon_k} \rightharpoonup \left[\frac{\partial}{\partial y} f(\overline{y}, \overline{z})\right]^* p \qquad \text{and}$$

$$\left[\frac{\partial}{\partial y} f(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k})\right]^* Sq_{\varepsilon_k} \rightharpoonup \left[\frac{\partial}{\partial y} f(\overline{y}, \overline{z})\right]^* Sq = S\frac{\partial}{\partial y} f(\overline{y}, \overline{z})q$$

in $\mathrm{L}^2(J_T; [X^\alpha]^*)$ with $k \to \infty$. Consequently, $\lambda_1 = \left[\frac{\partial}{\partial y} f(\overline{y}, \overline{z})\right]^* p$ and $\lambda_2 = S\frac{\partial}{\partial y} f(\overline{y}, \overline{z})q$ if f is continuously differentiable.
In the general case, weak continuity of B_i^* implies

$$0 = B_i^*(p_{\varepsilon_k} + Sq_{\varepsilon_k}) + (\kappa + 1)u_{\varepsilon_k} - \overline{u} \rightharpoonup B_i^*(p + Sq) + \kappa\overline{u} \quad \text{in } U_i \quad \text{with } k \to \infty.$$

This proves (4.42). $\square$

In order to study q it turns out that a splitting of the interval $\overline{J_T}$ as in [BK13, Section 4] is helpful.

Definition 4.24 (Partition of J_T)**.** [Mün17b, Definition 4.2] Let $\overline{z}$ be as in Theorem 4.16. We split $\overline{J_T}$ into $I_0 := \{t \in \overline{J_T} : \overline{z}(t) \in (a,b)\}$ and $I_\partial := \overline{J_T} \backslash I_0 = \{t \in \overline{J_T} : \overline{z}(t) \in \{a,b\}\}$. We further introduce $I_\partial^a := \{t \in \overline{J_T} : \overline{z}(t) = a\}$ and $I_\partial^b := \{t \in \overline{J_T} : \overline{z}(t) = b\}$.

Observe that the set I_0 is open as the pre-image of the continuous function $\overline{z}$. Hence, every time $t \in \overline{I_0} \cap I_\partial$ on the boundary between $\overline{I_0}$ and I_∂ is contained in I_∂ but not in I_0.

Lemma 4.25 (q in I_0)**.** [Mün17b, Lemma 4.3] Adopt the assumptions and the notation of Lemma 4.23 and consider the subdivision of $\overline{J_T}$ from Definition 4.24. For any interval $(c,d) \subset I_0$ the limit q in Lemma 4.23 belongs to $\mathrm{H}^1(c,d)$ and there exist $\nu_1, \nu_2 \in \mathrm{L}^2(J_T)$ such that $-\dot{q} = \nu_1 + \nu_2$ in $\mathrm{L}^2(c,d)$. If f is continuously differentiable from $X^\alpha \times \mathbb{R}$ into X then $\nu_1 = \langle p, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X$ and $\nu_2 = \langle Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X$.

Proof. Remember that $\overline{z}_\varepsilon \to \overline{z}$ uniformly in $\overline{J_T}$, see Theorem 4.16. Let $(c,d) \subset I_0$ be an arbitrary but fixed interval and consider any closed subinterval $[s,t] \subset (c,d)$. By definition of I_0, uniform convergence of $\overline{z}_\varepsilon$ to $\overline{z}$ implies $\overline{z}_\varepsilon([s,t]) \subset (a,b)$ for all ε small enough. We assume w.l.o.g that this is the case for all $\varepsilon \in (0, \varepsilon_0)$ with ε_0 from Lemma 4.22. But then $(A4)_\varepsilon$ in Assumption 4.7 implies $\Psi''(\overline{z}_\varepsilon) \equiv 0$ on $[s,t]$ for all $\varepsilon \in (0,\varepsilon_0)$. Hence, for $\varepsilon \in (0,\varepsilon_0)$, this term drops out in the evolution equation (4.28) of q_ε in Theorem 4.20. We integrate over $[s,t]$ and obtain

$$q_\varepsilon(t) - q_\varepsilon(s) = \int_s^t -\langle p_\varepsilon(s) + Sq_\varepsilon(s), \frac{\partial}{\partial z} f_\varepsilon(\overline{y}_\varepsilon(s), \overline{z}_\varepsilon(s))\rangle_X ds \quad \forall \varepsilon \in (0, \varepsilon_0).$$

Consider the subsequence $\{\varepsilon_k\}$ from Lemma 4.23 and let $k_0 > 0$ be chosen such that $\varepsilon_k < \varepsilon_0$ for all $k > k_0$. Note that ε_0 in Lemma 4.22 was chosen such that $\frac{\partial}{\partial z} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon)$ is bounded uniformly in $\mathrm{L}^\infty(J_T; X)$ for all $\varepsilon \in (0,\varepsilon_0)$, see also Lemma 4.18. This together with the estimates from Lemma 4.22 and Hölder's inequality then implies uniform boundedness of $\langle p_\varepsilon, \frac{\partial}{\partial z} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon)\rangle_X$ and $\langle Sq_\varepsilon, \frac{\partial}{\partial z} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon)\rangle_X$ in $\mathrm{L}^2(J_T)$ if $\varepsilon \in (0,\varepsilon_0)$. Hence, by reflexivity of $\mathrm{L}^2(J_T)$, there exist functions $\nu_1, \nu_2 \in \mathrm{L}^2(J_T)$ along with a subsequence of $\{\varepsilon_k\}$ (still denoted by $\{\varepsilon_k\}$) such that

$$\langle p_{\varepsilon_k}, \frac{\partial}{\partial z} f_{\varepsilon_k}(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k})\rangle_X \rightharpoonup \nu_1 \quad \text{and} \quad \langle Sq_{\varepsilon_k}, \frac{\partial}{\partial z} f_{\varepsilon_k}(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k})\rangle_X \rightharpoonup \nu_2 \quad \text{in } \mathrm{L}^2(J_T) \quad \text{with } k \to \infty.$$

If f is continuously differentiable from $X^\alpha \times \mathbb{R}$ into X then a similar argument as for λ_1, λ_2 in the proof of Lemma 4.23 implies $\nu_1 = \langle p, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X$ and $\nu_2 = \langle Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X$.
In the general case we obtain

$$q_{\varepsilon_k}(t) - q_{\varepsilon_k}(s) = \int_0^T -\langle p_{\varepsilon_k} + Sq_{\varepsilon_k}, \frac{\partial}{\partial z} f_{\varepsilon_k}(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k})\rangle_X \chi_{[s,t]} ds \to \int_s^t -\nu_1 - \nu_2 ds \quad \text{with } k \to \infty.$$

This implies that $-(\nu_1 + \nu_2)$ is the weak derivative of q in $\mathrm{L}^2(c,d)$. $\square$

Lemma 4.25 describes the evolution of q in I_0. In a next step, we investigate in the analysis of q in I_∂. It turns out that q and $\frac{d}{dt}\mathcal{P}[S\overline{y}]$ are pointwise orthogonal.

Lemma 4.26 (q in I_∂: Relation to $\mathcal{P}(S\overline{y})$)**.** [Mün17b, Lemma 4.4] Adopt the assumptions and the notation of Lemma 4.23 and consider the subdivision of $\overline{J_T}$ from Definition 4.24. With $\mathcal{P} = \mathrm{Id} - \mathcal{W}$ according to Theorem 2.40, there holds $\left[\frac{d}{dt}\mathcal{P}[S\overline{y}](t)\right] q(t) = 0$ for a.e. $t \in I_\partial$.

Proof. By $(A4)_\varepsilon$, Ψ is defined according to Remark 4.8. We denote by $c > 0$ and $\varepsilon_0 > 0$ the constants from Lemma 4.22. Theorem 4.16 implies $\overline{z}_\varepsilon \to z$ uniformly in $\overline{J_T}$. Hence,

$$\overline{z}_\varepsilon(t) \to b \quad \text{for} \quad t \in I_\partial^b \quad \text{and} \quad \overline{z}_\varepsilon(t) \to a \quad \text{for} \quad t \in I_\partial^a \quad \text{with} \quad \varepsilon \to 0.$$

We show the statement for I_∂^b first. Since $\overline{z}_\varepsilon(t) \to b$ for $t \in I_\partial^b$, there exists some $\varepsilon_1 \in (0, \varepsilon_0]$ such that

$$a < \overline{z}_\varepsilon(t) < b + 1 \quad \text{for } t \in I_\partial^b, \quad \forall \varepsilon \in (0, \varepsilon_1). \tag{4.43}$$

By Remark 4.8, Ψ satisfies $\Psi(x) = \Psi_1(x) = (x-b)^3(4+b-x)$ for $x \in (b, b+2]$ and $\Psi \equiv 0$ on $[a, b]$. For $\varepsilon \in (0, \varepsilon_1)$ and $t \in I_\partial^b$, this yields

$$\begin{aligned}
\Psi'(\overline{z}_\varepsilon(t)) &= \Psi_1'(\overline{z}_\varepsilon(t))\chi_{\{b<\overline{z}_\varepsilon \le b+2\}}(t) \\
&= 4(3 - (\overline{z}_\varepsilon(t) - b))(\overline{z}_\varepsilon(t) - b)^2 \chi_{\{b<\overline{z}_\varepsilon \le b+2\}}(t), && (4.44) \\
\Psi''(\overline{z}_\varepsilon(t)) &= 12(\overline{z}_\varepsilon(t) - b)[2 - (\overline{z}_\varepsilon(t) - b)]\chi_{\{b<\overline{z}_\varepsilon \le b+2\}}(t). && (4.45)
\end{aligned}$$

We insert (4.45) into the uniform estimate (4.32) from Lemma 4.22 and apply (4.43) to obtain

$$\begin{aligned}
c &\geq \frac{1}{\varepsilon}\int_0^T \Psi''(\overline{z}_\varepsilon(s))|q_\varepsilon(s)|ds \geq \frac{1}{\varepsilon}\int_{I_\partial^b} \Psi''(\overline{z}_\varepsilon(s))|q_\varepsilon(s)|ds \\
&= \frac{1}{\varepsilon}\int_{I_\partial^b} 12(\overline{z}_\varepsilon(t) - b)[2 - (\overline{z}_\varepsilon(t) - b)]\chi_{\{b<\overline{z}_\varepsilon \le b+2\}}|q_\varepsilon(s)|ds && (4.46) \\
&\geq \frac{1}{\varepsilon}\int_{I_\partial^b} 12(\overline{z}_\varepsilon(t) - b)\chi_{\{b<\overline{z}_\varepsilon \le b+2\}}|q_\varepsilon(s)|ds \qquad \forall \varepsilon \in (0, \varepsilon_1).
\end{aligned}$$

Remember the representation $\mathcal{W} + \mathcal{P} = \mathrm{Id}$ from Theorem 2.40 and the evolution equation (4.5) of $\overline{z}_\varepsilon = Z_\varepsilon(S\overline{y})$. Weak convergence of $S\overline{y}_\varepsilon \rightharpoonup S\overline{y}$ and of $\overline{z}_\varepsilon \rightharpoonup \overline{z} = \mathcal{W}[S\overline{y}]$ in $\mathrm{H}^1(J_T)$ with $\varepsilon \to 0$ according to Theorem 4.16 hence yields

$$\frac{1}{\varepsilon}\Psi'(\overline{z}_\varepsilon) = S\dot{\overline{y}}_\varepsilon - \dot{\overline{z}}_\varepsilon \rightharpoonup S\dot{\overline{y}} - \dot{\overline{z}} = \frac{d}{dt}(S\overline{y} - \mathcal{W}[S\overline{y}]) = \frac{d}{dt}\mathcal{P}[S\overline{y}] \quad \text{in} \quad \mathrm{L}^2(J_T) \quad \text{with} \quad \varepsilon \to 0.$$

Furthermore, Lemma 4.23 implies the strong convergence $|q_{\varepsilon_k}| \to |q|$ in $\mathrm{L}^2(J_T)$ with $k \to \infty$. Moreover, by the variational inequality (4.2) which determines $\overline{z} = \mathcal{W}[S\overline{y}]$ and because of the definition of I_∂^b there holds

$$0 \leq S\dot{\overline{y}} = S\dot{\overline{y}} - \dot{\overline{z}} = \frac{d}{dt}\mathcal{P}[S\overline{y}] \quad \text{a.e. in } I_\partial^b,$$

and hence $\frac{d}{dt}\mathcal{P}[S\overline{y}] = \left|\frac{d}{dt}\mathcal{P}[S\overline{y}]\right|$ a.e. in I_∂^b. This together with (4.44) and (4.46) yields

$$\begin{aligned}
0 &\leq \int_{I_\partial^b} \left|\frac{d}{dt}\mathcal{P}[S\overline{y}]\right| |q(s)|ds = \lim_{k\to\infty} \frac{1}{\varepsilon_k}\int_{I_\partial^b} \Psi'(\overline{z}_{\varepsilon_k}(s))|q_{\varepsilon_k}(s)|ds \\
&= \lim_{k\to\infty} \frac{1}{\varepsilon_k}\int_{I_\partial^b} 4(3 - (\overline{z}_\varepsilon(t) - b))(\overline{z}_\varepsilon(t) - b)^2\chi_{\{b<\overline{z}_{\varepsilon_k} \le b+2\}}|q_{\varepsilon_k}(s)|ds \\
&\leq \lim_{k\to\infty} \frac{12}{\varepsilon_k}\int_{I_\partial^b} (\overline{z}_{\varepsilon_k}(s) - b)^2\chi_{\{b<\overline{z}_{\varepsilon_k} \le b+2\}}|q_{\varepsilon_k}(s)|ds \leq c \lim_{k\to\infty} \sup_{s\in I_\partial^b}(\overline{z}_{\varepsilon_k}(s) - b) = 0.
\end{aligned}$$

It follows $\left[\frac{d}{dt}\mathcal{P}[S\overline{y}](t)\right] q(t) = 0$ for a.e. $t \in I_\partial^b$.
Similar estimates for I_∂^a conclude the proof because $I_\partial = I_\partial^a \cup I_\partial^b$. □

Remark 4.27. Note that Lemma 4.26 can be trivially extended to I_0. Indeed, it follows from the variational inequality (4.2) which determines $\overline{z} = \mathcal{W}[S\overline{y}]$ and from the definition $I_0 = \{t \in \overline{J_T} : \overline{z}(t) \in (a,b)\}$ that

$$\frac{d}{dt}\mathcal{P}[S\overline{y}](t) = S\dot{\overline{y}}(t) - \dot{\overline{z}}(t) = 0 \quad \text{for a.e.} \quad t \in I_0.$$

Hence, $\left[\frac{d}{dt}\mathcal{P}[S\overline{y}](t)\right] q(t) = 0$ for a.e. $t \in I_0$.

Remember the low regularity of $q \in \mathrm{BV}(J_T)$. Hence, in general q has no weak derivative on the whole interval J_T. The next lemma provides an equation for dq in the sense of measures on I_∂.

Lemma 4.28 (q in I_∂: Relation to $d\mu$)**.** [Mün17b, Lemma 4.5] Adopt the assumptions and the notation of Lemma 4.23 and let ν_1 and ν_2 be as in Lemma 4.25. Consider the subdivision of $\overline{J_T}$ from Definition 4.24. We denote $d\mu_\varepsilon := \frac{1}{\varepsilon}\Psi''(\overline{z}_\varepsilon)q_\varepsilon$. There exists a measure $d\mu \in \mathrm{C}(\overline{J_T})^*$, such that a subsequence $\{d\mu_{\varepsilon_k}\}$ (w.l.o.g we may consider $\{\varepsilon_k\}$ from Lemma 4.23) converges weak star to $d\mu$ in $\mathrm{C}(\overline{J_T})^*$ with $k \to \infty$. The support of $d\mu$ is contained in I_∂. For any $\varphi \in \mathrm{C}(\overline{J_T})$ there holds

$$\int_0^T -\varphi(t)dq(t) + \int_{I_\partial} \varphi(t)d\mu(t) = \int_0^T \varphi(t)(\nu_1(t) + \nu_2(t))dt. \tag{4.47}$$

This implies $d\mu = dq + (\nu_1 + \nu_2)dt$ as measures on I_∂.

Proof. Estimate (4.32) in Lemma 4.22 provides a uniform-in-ε bound in $\mathrm{L}^1(J_T)$ for the functions $d\mu_\varepsilon$ for all $\varepsilon \in (0, \varepsilon_0)$. Consequently, by Anaoglu's compactness theorem [W05, Korollar VIII.3.12], we can extract a subsequence of $\{d\mu_\varepsilon\}$ which converges weakly star in $\mathrm{C}(\overline{J_T})^*$ to some measure $d\mu$. Let $\varphi \in \mathrm{C}(\overline{J_T})$ have compact support in I_0. Then the uniform convergence of $\overline{z}_\varepsilon$ to $\overline{z}$ implies the existence of some $\varepsilon_\varphi \in (0, \varepsilon_0)$ such that $\overline{z}_\varepsilon(t) \in (a,b)$ for all $t \in \mathrm{supp}(\varphi) \Subset I_0$ and $\varepsilon \in (0, \varepsilon_\varphi)$. By $(A4)_\varepsilon$ in Assumption 4.7, this implies $\varphi\,\frac{1}{\varepsilon}\Psi''(\overline{z}_\varepsilon)q_\varepsilon \equiv 0$ on J_T for $\varepsilon \in (0, \varepsilon_\varphi)$. Since $\varphi \in \mathrm{C}(\overline{J_T})$ was arbitrary, the support of the limit measure $d\mu$ is contained in I_∂ [BK13, p.343]. The other statements are shown similar as [BK13, Lemma 4.6] and [BK13, Lemma 4.7]. Indeed, testing (4.28) with an arbitrary test function $\varphi \in \mathrm{C}(\overline{J_T})$ and taking the limit $\varepsilon \to 0$ implies (4.47). To see that $d\mu = dq + (\nu_1 + \nu_2)dt$ as measures on I_∂, choose $\varphi \in \mathrm{C}(I_\partial)$ arbitrary and consider any extension $\tilde{\varphi} \in \mathrm{C}(\overline{J_T})$ of φ. Then we define $\varphi_k(t) := \max\{0, 1 - k\mathrm{dist}(t, I_\partial)\}\tilde{\varphi}(t)$ for $k \in \mathbb{N}$. All φ_k are uniformly bounded in $\mathrm{C}(\overline{J_T})$. Moreover, $\varphi_k(t) \to 0$ for $t \in I_0$ and $\varphi_k(t) \to \varphi(t)$ for $t \in I_\partial$ with $k \to \infty$. Hence, testing (4.47) with φ_k and taking the limit $k \to \infty$ implies

$$\int_{I_\partial} -\varphi(t)dq(t) + \int_{I_\partial} \varphi(t)d\mu(t) = \int_{I_\partial} \varphi(t)(\nu_1(t) + \nu_2(t))dt.$$

□

In the next lemma, we study the continuity properties of q. Moreover, we prove that the absolute value of q can only jump downwards in reverse time.

Lemma 4.29 (Discontinuity properties of q)**.** [Mün17b, Lemma 4.6] Adopt the assumptions and notation of Lemma 4.23. The absolute value of q can only jump downwards in reverse time. Consequently, for any $t \in \overline{J_T}$ there holds $|q(t-)| \le |q(t+)|$ and $q(T-) = q(T) = 0$. Moreover, q is right-continuous in $[0,T)$ and left-continuous at T.

Proof. Lemma 4.23 implies the existence of some subsequence $\{\varepsilon_k\}$ such that q_{ε_k} converges to q in $\mathrm{L}^1(J_T)$ and $dq_{\varepsilon_k} = \dot{q}_{\varepsilon_k}dt$ to dq weak star in $\mathrm{C}(\overline{J_T})^*$ respectively with $k \to \infty$. Equivalently, by [Vis13, Chapter XII.7], there exist representatives q_{ε_k} and q in the space $\mathrm{BV}_r(J_T)$ of functions with bounded total variation which are right-continuous in $[0,T)$ and left-continuous at T such that q_{ε_k} converges weak star to q in $\mathrm{BV}_r(J_T)$. Moreover, $\mathrm{BV}_r(J_T)$ is isometric to $\mathrm{BV}(J_T)$. Hence, it follows that q is bounded in $\mathrm{BV}(J_T)$, right-continuous in $[0,T)$ and left-continuous at T. The rest of the statements are shown just as [BK13, Lemma 4.4]: Let $t \in J_T$ be arbitrary. Similar as in (4.39), we test the evolution equation (4.28) with $q_\varepsilon/|q_\varepsilon|$. Afterwards, we integrate over (s_1,s_2) with $t \in (s_1,s_2)$ and obtain

$$\begin{aligned}|q_\varepsilon(s_1)| - |q_\varepsilon(s_2)| &+ \frac{1}{\varepsilon}\int_{s_1}^{s_2} \Psi''(\overline{z}_\varepsilon(s))|q_\varepsilon(s)|ds \\ &= \int_{s_1}^{s_2} \langle p_\varepsilon(s) + Sq_\varepsilon(s), \frac{\partial}{\partial z} f_\varepsilon(\overline{y}_\varepsilon(s), \overline{z}_\varepsilon(s))\rangle_X \frac{q_\varepsilon(s)}{|q_\varepsilon(s)|}\, ds.\end{aligned}$$

The second term on the left side is non-negative and $\langle p_\varepsilon(s) + Sq_\varepsilon, \frac{\partial}{\partial z} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon)\rangle_X$ is bounded in $\mathrm{L}^1(J_T)$ by (4.20) and (4.38). Hence, in the limit $s_1 \uparrow t$ and $s_2 \downarrow t$ we obtain $|q(t-)| \le |q(t+)|$. $q(T-) = q(T) = 0$ follows similarly and $q(t+) = q(0)$ holds because q is right-continuous at 0. □

Even though the abstract measure $d\mu$ in Lemma 4.28 has its support only in I_∂, it still remains present in the equation $d\mu = dq + (\nu_1 + \nu_2)dt$ of measures on I_∂. Hence, even if f is continuously differentiable so that $\nu_1 = \langle p, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X$ and $\nu_2 = \langle Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X$ according to Lemma 4.25, the characterization of dq is still not fully understood. That is, with Lemma 4.26 in mind, we make the following regularity assumption in order to analyze q also in I_∂, cf. [BK13, p.344]:

Assumption 4.30 (Regularity assumption)**.** [Mün17b, Assumption 4.7] Let $\overline{y}$ be as in Theorem 4.16 and consider the subdivision of $\overline{J_T}$ from Definition 4.24. We suppose that the function $\mathcal{P}[S\overline{y}]$ satisfies $\frac{d}{dt}\mathcal{P}[S\overline{y}] \neq 0$ a.e. in I_∂. Equivalently, $S\dot{\overline{y}} > 0$ a.e. in I_∂^b and $S\dot{\overline{y}} < 0$ a.e. in I_∂^a.

Remark 4.31. [Cf. Mün17b, Remark 4.8] Assumption 4.30 is reasonable if $S\overline{y}$ is the size of interest. Assume for instance that S computes approximately the mean value of $I_p^{-1}y$, $y \in \mathrm{dom}(A_p)$. More precisely, let w in (A2)' in Assumption 4.2 have the form $w = \frac{1}{m|\Omega|}\varphi$, where the components φ_j, $j \in \{1,\dots,m\}$, of $\varphi \in \prod_{j=1}^m \mathrm{C}^\infty_{\Gamma_{D_j}}(\Omega)$ are equal to 1 within most of Ω, except for a neighborhood of Γ_{D_j} of measure $0 < \varepsilon << 1$. For

$$y \in \mathrm{dom}(A_p) = \mathrm{ran}\,(I_p) = I_p(\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)) \subset \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega),$$

this implies

$$Sy = \langle y, w\rangle_X = \frac{1}{m|\Omega|}\sum_{j=1}^m \int_\Omega (I_p^{-1}y)_j \varphi_j dx.$$

Hence, Sy is approximately the mean value of the function $I_p^{-1}y$ in Ω. If the optimal control problem (4.1)–(4.3) enforces this value to vary (in I_∂), then it becomes very unlikely that $S\dot{\overline{y}} = 0$ in a subset of I_∂ with positive measure, so that Assumption 4.30 is justified.

In extension to [Mün17b, Remark 4.8], we provide an example for Remark 4.31, when S computes approximately the mean value of $I_p^{-1}y$ for $y \in \mathrm{dom}(A_p)$. Assumption 4.30 can be enforced by a special choice of the tracking term y_d.

Example 4.32. We assume $i = m = |\Omega| = T = 1$ in this example. Note that the solution mapping $G : \mathrm{L}^2(J_T;X) \to Y_{2,0}$ is surjective. Indeed, if $y \in Y_{2,0}$ is arbitrary, then $h := (\frac{d}{dt} + A_p)y - F[y]$ is contained in $\mathrm{L}^2(J_T;X)$ and $G(h) = y$. As will be shown in Subsection 4.4.4, B_1 has dense range in X for $\frac{1}{2} > 1 - \frac{1}{p} - \frac{1}{d}$ which we assume in this example. Hence, the mapping

$u \in U_1 \mapsto G(B_1 u)$ has dense range in $Y_{2,0}$. For $0 < \varepsilon << 1$, let $\varphi \in \mathrm{C}^\infty_{\Gamma_D}(\Omega)$ with $|\varphi| \leq 1$ be chosen such that

$$|\Omega_\varepsilon| = |\{x \in \Omega : \varphi(x) = 1\}| = 1 + O(\varepsilon^3) \quad \text{and } |\Omega \backslash \Omega_\varepsilon| = |\{x \in \Omega : \varphi(x) \neq 1\}| < \varepsilon^3.$$

Moreover, let w in (A2)' in Assumption 4.2 be defined by $w = \varphi$ and choose the tracking term $y_d(x,t) := t\varphi(x)$ for $(x,t) \in \Omega \times \overline{J_T}$. Since $\mathrm{C}^\infty_{\Gamma_D}(\Omega) \subset \mathrm{W}^{1,p}_{\Gamma_D}(\Omega) \hookrightarrow \mathrm{dom}(A_p)$, there holds $I_p y_d \in Y_{2,0}$. Hence, there exists some function $u \in U_1$ such that $\|G(B_1 u) - y_d\|^2_{U_1} \leq \varepsilon^3$. We further choose κ of order $O(\varepsilon^3)$. Consequently, the minimal value in (4.1)–(4.3) is of order $O(\varepsilon^3)$. Let $\overline{y}$ be an optimal state for problem (4.1)–(4.3) and assume that there exist $k > 0$ disjoint intervals $(t_j, t_{j+1}) \subset I_\partial$, $j \in \{1, \dots, k\}$, such that $S\dot{\overline{y}} = 0$ a.e. in $\cup_{j=1}^k (t_j, t_{j+1})$ and $S\dot{\overline{y}} \neq 0$ a.e. in $I_\partial \backslash \cup_{j=1}^k (t_j, t_{j+1})$. For $j \in \{1, \dots, k\}$, we denote by $m_j := Sy(t_j)$ the value of Sy in $[t_j, t_{j+1}]$. We prove that $|t_{j+1} - t_j|$ is of order $O(\varepsilon)$ for each $j \in \{1, \dots, k\}$.
Let $j \in \{1, \dots, k\}$ be given. Since $w = \varphi$, there holds

$$Sy(t) = \langle y(t), \varphi \rangle_X = \int_\Omega I_p^{-1} y(x,t)\varphi(x)\,dx = \int_{\Omega_\varepsilon} I_p^{-1} y(x,t)\,dx + O(\varepsilon^3) \quad \text{for a.e. } t \in J_T.$$

Hence, for a.e. $t \in (t_j, t_{j+1})$, we can estimate

$$\int_{\Omega_\varepsilon} |\overline{y}(x,t)|^2\,dx \geq \left| \int_{\Omega_\varepsilon} I_p^{-1}\overline{y}(x,t)\,dx \right|^2 = |S\overline{y}(t)|^2 + O(\varepsilon) = m_j^2 + O(\varepsilon^3).$$

This implies

$$\begin{aligned}
\|I_p^{-1}\overline{y} - y_d\|^2_{U_1} &\geq \int_{(t_j,t_{j+1})} \int_{\Omega_\varepsilon} |[I_p^{-1}y](x,t) - y_d(x,t)|^2\,dxdt + O(\varepsilon^3) \\
&= \int_{(t_j,t_{j+1})} \int_{\Omega_\varepsilon} |[I_p^{-1}\overline{y}](x,t)|^2 - 2t[I_p^{-1}\overline{y}](x,t) + t^2\,dxdt + O(\varepsilon^3) \\
&\geq \int_{(t_j,t_{j+1})} m_j^2 - 2tm_j + t^2|\Omega_\varepsilon|\,dt + O(\varepsilon^3) \\
&= \int_{(t_j,t_{j+1})} (m_j - t)^2\,dt + O(\varepsilon^3).
\end{aligned}$$

Note that

$$\begin{aligned}
&\int_{(t_j,t_{j+1})} (m_j - t)^2\,dt \\
&= \int_{(m_j-\varepsilon, m_j+\varepsilon)\cap(t_j,t_{j+1})} (m_j - t)^2\,dt + \int_{(t_j,t_{j+1})\backslash(m_j-\varepsilon,m_j+\varepsilon)} (m_j - t)^2\,dt \\
&\geq \int_{(m_j-\varepsilon, m_j+\varepsilon)\cap(t_j,t_{j+1})} (m_j - t)^2\,dt + \varepsilon^2|(t_j, t_{j+1})\backslash(m_j - \varepsilon, m_j + \varepsilon)|.
\end{aligned}$$

The first term is of order $O(\varepsilon^3)$. Hence,

$$\|I_p^{-1}\overline{y} - y_d\|^2_{U_1} \geq \varepsilon^2|(t_j, t_{j+1})\backslash(m_j - \varepsilon, m_j + \varepsilon)| + O(\varepsilon^3),$$

where the left side is of order $O(\varepsilon^3)$. Consequently, $|(t_j, t_{j+1})\backslash(m_j - \varepsilon, m_j + \varepsilon)|$ is of order $O(\varepsilon)$ and because $|(t_j, t_{j+1}) \cap (m_j - \varepsilon, m_j + \varepsilon)| \leq 2\varepsilon$ we conclude that $|t_{j+1} - t_j|$ is of order $O(\varepsilon)$.
Under the assumption that I_∂ is decomposed into intervals in which $S\dot{\overline{y}} = 0$ almost surely and intervals in which $S\dot{\overline{y}} \neq 0$ almost surely, a small choice of ε implies that $S\dot{\overline{y}} \neq 0$ can only be violated in a small subset of every connected component of I_∂.

Remark 4.33. Note that under Assumption 4.30, Lemma 4.26 implies $q(t) = 0$ for a.e. $t \in I_\partial$. Clearly, this does not mean that q vanishes everywhere in I_∂. But since q is right-continuous in $[0, T)$ and left-continuous at T, we conclude $q(t) = 0$ for all $t \in [c, d)$ for arbitrary subintervals $[c, d] \subset I_\partial$. Hence, for a full characterization of q, the only remaining question is how dq behaves on the set of switching times between I_0 and I_∂, i.e. on $\overline{I_0} \cap I_\partial$?

With regard to Remark 4.33, we introduce the following categories of times as in [BK13]:

Definition 4.34 (Switching times)**.** [Mün17b, Definition 4.9] Consider the subdivision of $\overline{J_T}$ from Definition 4.24. We call a time $t \in \overline{J_T}$ a $(0, \partial)$-switching time if $t \in \overline{I_0} \cap I_\partial$ and if there is some $\varepsilon > 0$ such that $(t - \varepsilon, t) \subset I_0$ and $[t, t + \varepsilon) \subset I_\partial$. We say that t is a $(\partial, 0)$-switching time if $t \in \overline{I_0} \cap I_\partial$ and if for some $\varepsilon > 0$ we have $(t - \varepsilon, t] \subset I_\partial$ and $(t, t + \varepsilon) \subset I_0$.

With the same argument as in Remark 4.33, we can characterize dq at $(0, \partial)$-switching times:

Lemma 4.35 (q at $(0, \partial)$-switching times)**.** [Mün17b, Lemma 4.10] Adopt the assumptions and the notation of Lemma 4.23. If t is a $(0, \partial)$-switching time in the sense of Definition 4.34 and if Assumption 4.30 holds then there exits some $\varepsilon > 0$ such that $q \equiv 0$ on $[t, t + \varepsilon)$. Moreover, q is continuous at t with $t = 0$. Furthermore, for every open interval $(c, d) \subset I_\partial$ there holds that $q \equiv 0$ in $[c, d)$.

Proof. Let Assumption 4.30 hold. As seen in Remark 4.33, this implies $q \equiv 0$ in $[c, d)$ for any subinterval $[c, d] \subset I_\partial$. Hence, for any subinterval $[\beta, \gamma] \subset (c, d) \subset I_\partial$ we obtain $0 = q(\gamma-) - q(\beta+) = dq((\beta, \gamma))$, see Lemma 4.23. This implies $dq = 0$ as a measure on (c, d). We are left to prove that q is continuous at $(0, \partial)$-switching times. To this aim, remember that the absolute value of q can only jump downwards in reverse time by Lemma 4.29. Furthermore, for any interval $(e, c) \subset I_0$ there holds $q \in \mathrm{H}^1(e, c) \hookrightarrow \mathrm{C}([e, c])$ according to Lemma 4.25. Consequently, whenever there exist intervals $(e, c) \subset I_0$ and $[c, d] \subset I_\partial$, then $|q(c-)| \leq |q(c+)| = |q(c)| = 0$ and q is right-continuous at e so that q is absolutely continuous on $[e, d)$ with $q(c) = 0$.
If t is a $(0, \partial)$-switching time then $(t - \varepsilon, t) \subset I_0$ and $[t, t + \varepsilon) \subset I_\partial$ for some $\varepsilon > 0$. Hence, continuity of q at t follows from the general case with $e = t - \varepsilon$, $c = t$ and $d = t + \varepsilon$.

□

Remark 4.36. Even if Assumption 4.30 implies continuity of q at $(0, \partial)$-switching times, the characterization of dq on $\overline{I_0} \cap I_\partial$ is still not complete. Indeed, there might exist $(\partial, 0)$-switching times for example, see Definition 4.34, and other categories of times in $\overline{I_0} \cap I_\partial$ are possible. Those include isolated times $t \in \overline{I_0} \cap I_\partial$ for example.
Moreover, if Assumption 4.30 does not hold then we can not apply Lemma 4.26 to show that q vanishes on half-open intervals $[c, d)$ where $[c, d] \subset I_\partial$. That is, it would be interesting to understand the behaviour of dq on so-called waiting slots, i.e. on subintervals of I_∂ in which $S\dot{y} = 0$ a.e.
We will return to these open questions in Subsection 4.4.5, since we can only answer them for the case of distributed control functions, i.e. for $i = 1$. Also the definition of isolated times and waiting slots can be found in that subsection, see Definition 4.42.

4.4.2 Optimality conditions for distributed or boundary controls

In this subsection, we prove an optimality condition for problem (4.1)–(4.3) for the optimal control $\overline{u} \in U_i$, $i \in \{1, 2\}$, with help of the adjoint system (p, q) from Lemma 4.23. As already explained in the beginning of Section 4.4, we begin with the general case $i \in \{1, 2\}$ and derive optimality conditions of weak type. Those will be improved for $i = 1$ in Subsection 4.4.4 below.

Lemma 4.37 (Optimality condition)**.** [Mün17b, Lemma 4.12] Adopt the assumptions and the notation of Lemma 4.23 and let ν_1 and ν_2 be as in Lemma 4.25. For any $h \in U_i$, $y^{B_i\overline{u},B_ih} = G'[B_i\overline{u}; B_ih]$ and $F'[\overline{y}; y^{B_i\overline{u},B_ih}](t) = f'[(\overline{y}(t), \mathcal{W}[S\overline{y}](t)); (y^{B_i\overline{u},B_ih}(t), \mathcal{W}'[S\overline{y}; Sy^{B_i\overline{u},B_ih}](t))]$ (see Corollary 3.15), there holds the optimality condition

$$\begin{aligned}
&\int_0^T \langle \lambda_1 + \lambda_2 + S(\nu_1 + \nu_2), y^{B_i\overline{u},B_ih} \rangle_{\mathrm{dom}(A_p)} dt \\
&\leq \int_{I_\partial} Sy^{B_i\overline{u},B_ih} d\mu + \int_0^T \langle p + Sq, F'[\overline{y}; y^{B_i\overline{u},B_ih}] \rangle_X dt.
\end{aligned} \tag{4.48}$$

Proof. We denote $y^{B_i\overline{u},B_ih} = G'[B_i\overline{u}; B_ih] \in Y_{2,0}$ as in Corollary 3.15.
Since $\overline{u}$ solves the minimization problem $\min\limits_{u \in U_i} J(I_p^{-1}G(u), u)$, the directional derivative of the reduced cost function $\mathcal{J}(u) := J(I_p^{-1}G(u), u)$ has to be greater or equal than zero in each direction. The derivative of $\mathcal{J}$ can be computed with the same techniques as in the proof of Theorem 4.20. For arbitrary $h \in U_i$ there holds

$$0 \leq \mathcal{J}'[\overline{u}; h] = \langle \overline{y} - y_d, y^{B_i\overline{u},B_ih} \rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))} + \kappa \langle \overline{u}, h \rangle_{U_i}. \tag{4.49}$$

Moreover, $y^{B_i\overline{u},B_ih}$ solves the evolution equation (3.20) in Corollary (3.15) with y replaced by $\overline{y}$ and with h replaced by B_ih. If we test that equation with $p + Sq$ and integrate over J_T, then equation (4.42) implies

$$\begin{aligned}
&\int_0^T \langle p + Sq, \dot{y}^{B_i\overline{u},B_ih} + A_p y^{B_i\overline{u},B_ih} \rangle_X dt - \int_0^T \langle p + Sq, F'[\overline{y}; y^{B_i\overline{u},B_ih}] \rangle_X dt \\
&= \int_0^T \langle p + Sq, B_ih \rangle_X dt = -\kappa \langle \overline{u}, h \rangle_{U_i}.
\end{aligned} \tag{4.50}$$

Note that the right side in (4.50) is just the negative partial derivative of J with respect to u, evaluated at $\overline{u}$ and in direction h. We integrate the first term on the left side of (4.50) by parts and insert the evolution equation (4.41) for p, see Lemma 4.23. Moreover, we replace $-dq$ by the measure $-d\mu + (\nu_1 + \nu_2)dt$ according to Lemma 4.28. Finally, we end up with

$$\begin{aligned}
&\int_0^T \langle p + Sq, \dot{y}^{B_i\overline{u},B_ih} + A_p y^{B_i\overline{u},B_ih} \rangle_X dt \\
&= \int_0^T \langle \lambda_1 + \lambda_2 - SA_pq + \overline{y} - y_d, y^{B_i\overline{u},B_ih} \rangle_{\mathrm{dom}(A_p)} dt \\
&- \int_0^T Sy^{B_i\overline{u},B_ih} dq + \int_0^T \langle SA_pq, y^{B_i\overline{u},B_ih} \rangle_{\mathrm{dom}(A_p)} dt \\
&= \int_0^T \langle \lambda_1 + \lambda_2 + \overline{y} - y_d, y^{B_i\overline{u},B_ih} \rangle_{\mathrm{dom}(A_p)} dt - \int_0^T Sy^{B_i\overline{u},B_ih} dq \\
&= \int_0^T \langle \lambda_1 + \lambda_2 + \overline{y} - y_d, y^{B_i\overline{u},B_ih} \rangle_{\mathrm{dom}(A_p)} dt - \int_{I_\partial} Sy^{B_i\overline{u},B_ih} d\mu \\
&+ \int_0^T (\nu_1 + \nu_2) Sy^{B_i\overline{u},B_ih} dt.
\end{aligned} \tag{4.51}$$

Now we replace $\kappa\langle \overline{u}, h \rangle_{U_i}$ in (4.49) according to (4.50) and insert the right hand side of (4.51) for $\int_0^T \langle p + Sq, \dot{y}^{B_i\overline{u},B_ih} + A_p y^{B_i\overline{u},B_ih}\rangle_X dt$ to obtain

$$
\begin{aligned}
0 &\leq \int_0^T \langle \overline{y} - y_d, y^{B_i\overline{u},B_ih}\rangle_{\mathrm{dom}(A_p)} dt + \kappa\langle \overline{u}, h\rangle_{U_i} \\
&= -\int_0^T \langle \lambda_1 + \lambda_2, y^{B_i\overline{u},B_ih}\rangle_{\mathrm{dom}(A_p)} dt + \int_{I_\partial} S y^{B_i\overline{u},B_ih} d\mu - \int_0^T (\nu_1 + \nu_2) S y^{B_i\overline{u},B_ih} dt \\
&+ \int_0^T \langle p + Sq, F'[\overline{y}; y^{B_i\overline{u},B_ih}]\rangle_X dt.
\end{aligned}
$$

□

4.4.3 Summary: Adjoint system and optimality conditions for distributed- or boundary controls

In this subsection, we summarize the results about the optimality system for problem (4.1)–(4.3) for $i \in \{1, 2\}$. The optimality condition can be improved if f is continuously differentiable, see Corollary 4.39 below.

Theorem 4.38 (Adjoint system and optimality condition)**.** *[Mün17b, Theorem 4.13] Let Assumption 4.2 and Assumption 4.7 hold. For $i \in \{1, 2\}$ suppose that $\overline{u} \in U_i$ is an optimal control for problem (4.1)–(4.3) together with the optimal state $\overline{y} \in \mathrm{Y}_{2,0}$ and $\overline{z} = \mathcal{W}[S\overline{y}] \in \mathrm{H}^1(J_T)$. Consider the subdivision of $\overline{J_T}$ from Definition 4.24. Then there exist adjoint states $p \in \mathrm{Y}_{2,T}^*$ and $q \in \mathrm{BV}(J_T)$ of the following kind: There holds*

$$B_i^*(p + Sq) = -\kappa\overline{u} \quad \text{in } U_i. \tag{4.52}$$

For some functions $\lambda_1, \lambda_2 \in \mathrm{L}^2(J_T; [X^\alpha]^*)$ we have

$$-\dot{p} + A_p^* p = \lambda_1 + \lambda_2 - SA_p q + \overline{y} - y_d \quad \text{for } t \in J_T,\ p(T) = 0.$$

q is right-continuous in J_T, left-continuous at T and absolutely continuous in I_0. There exist $\nu_1, \nu_2 \in \mathrm{L}^2(J_T)$ such that q solves $-\dot{q} = \nu_1 + \nu_2$ in every open subinterval of I_0. $\frac{d}{dt}\mathcal{P}[S\overline{y}](t)q(t) = 0$ for a.e. $t \in I_\partial$ and there is a measure $d\mu \in \mathrm{C}(\overline{J_T})^*$ with support in I_∂ such that $d\mu = dq + (\nu_1 + \nu_2)dt$ as measures on I_∂. For all $h \in U_i$ and with $y^{B_i\overline{u},B_ih} = G'[B_i\overline{u}; B_ih]$ (see Theorem 3.11) there holds the optimality condition

$$
\begin{aligned}
&\int_0^T \langle \lambda_1 + \lambda_2 + S(\nu_1 + \nu_2), y^{B_i\overline{u},B_ih}\rangle_{\mathrm{dom}(A_p)} dt \\
&\leq \int_{I_\partial} S y^{B_i\overline{u},B_ih} d\mu + \int_0^T \langle p + Sq, F'[\overline{y}; y^{B_i\overline{u},B_ih}]\rangle_X dt,
\end{aligned}
\tag{4.53}
$$

where $F'[\overline{y}; y^{B_i\overline{u},B_ih}](t) = f'[(\overline{y}(t), \mathcal{W}[S\overline{y}](t)); (y^{B_i\overline{u},B_ih}(t), \mathcal{W}'[S\overline{y}; Sy^{B_i\overline{u},B_ih}](t))]$. The absolute value of q can only jump downwards in reverse time so that $q(T-) = q(T) = 0$ and $|q(t-)| \leq |q(t+)|$ for all $t \in \overline{J_T}$. If the regularity Assumption 4.30 is valid then q is continuous at every $(0, \partial)$-switching time t (see Definition 4.34) with $q(t) = 0$. In this case, for every open interval $(c, d) \subset I_\partial$ it follows $q \equiv 0$ on $[c, d)$.

If f is continuously differentiable, then the functions λ_1, λ_2 and ν_1, ν_2 have been computed in Lemma 4.23 and Lemma 4.25. With the concrete form of $F'[\overline{y}; y^{B_i\overline{u},B_ih}]$, λ_1, λ_2 in (4.53) cancel out and we obtain an improved optimality condition:

Corollary 4.39 (Adjoint system and optimality condition for regular f)**.** [Mün17b, Corollary 4.14] Let Assumption 4.2 and Assumption 4.7 hold. Moreover, suppose that f is continuously differentiable from $X^\alpha \times \mathbb{R}$ into X. For $i \in \{1,2\}$ assume that $\overline{u} \in U_i$ is an optimal control for problem (4.1)–(4.3) together with the optimal state $\overline{y} \in Y_{2,0}$ and $\overline{z} = \mathcal{W}[S\overline{y}] \in \mathrm{H}^1(J_T)$. Consider the subdivision of $\overline{J_T}$ from Definition 4.24. Then there exist adjoint states $p \in Y_{2,T}^*$ and $q \in \mathrm{BV}(J_T)$ of the following kind: There holds

$$B_i^*(p + Sq) = -\kappa\overline{u} \quad \text{in } U_i. \tag{4.54}$$

We have

$$-\dot{p} + A_p^* p = \left[\frac{\partial}{\partial y} f(\overline{y}, \overline{z})\right]^* (p + Sq) - SA_p q + \overline{y} - y_d \quad \text{for } t \in J_T,\ p(T) = 0. \tag{4.55}$$

q is right-continuous in J_T, left-continuous at T and absolutely continuous in I_0. q solves the evolution equation $-\dot{q} = \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X$ in every open subinterval of I_0. $\frac{d}{dt}\mathcal{P}[S\overline{y}](t)q(t) = 0$ for a.e. $t \in I_\partial$ and there is a measure $d\mu \in \mathrm{C}(\overline{J_T})^*$ with support in I_∂ such that $d\mu = dq + \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X dt$ as measures on I_∂. For all $h \in U_i$ and with $y^{B_i\overline{u},B_ih} = G'[B_i\overline{u}; B_ih]$ (see Corollary 3.15) and $\mathcal{P} = \mathrm{Id} - \mathcal{W}$ (see Theorem 2.40) there holds the optimality condition

$$\int_0^T \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X \mathcal{P}'[S\overline{y}; Sy^{B_i\overline{u},B_ih}]dt \leq \int_{I_\partial} Sy^{B_i\overline{u},B_ih} d\mu. \tag{4.56}$$

The absolute value of q can only jump downwards in reverse time so that $q(T-) = q(T) = 0$ and $|q(t-)| \leq |q(t+)|$ for all $t \in \overline{J_T}$. If the regularity Assumption 4.30 is valid then q is continuous at every $(0, \partial)$-switching time t (see Definition 4.34) with $q(t) = 0$. In this case, for every open interval $(c, d) \subset I_\partial$ it follows $q \equiv 0$ on $[c, d)$.

Proof. Since f is continuously differentiable, the functions λ_1, λ_2 and ν_1, ν_2 in Theorem 4.38 can be computed. In particular, by Lemma 4.23 there holds $\lambda_1 = \left[\frac{\partial}{\partial y} f(\overline{y}, \overline{z})\right]^* p$ and $\lambda_2 = S\frac{\partial}{\partial y} f(\overline{y}, \overline{z})q$. Moreover, Lemma 4.25 entails $\nu_1 = \langle p, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X$ and $\nu_2 = \langle Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X$. It remains to prove (4.56). With $\lambda_1, \lambda_2, \nu_1, \nu_2$ as above, (4.53) takes the form

$$\begin{aligned}
&\int_0^T \langle \left[\frac{\partial}{\partial y} f(\overline{y}, \overline{z})\right]^* (p + Sq), y^{B_i\overline{u},B_ih}\rangle_{\mathrm{dom}(A_p)} + \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X Sy^{B_i\overline{u},B_ih} dt \\
&\leq \int_{I_\partial} Sy^{B_i\overline{u},B_ih} d\mu + \int_0^T \langle p + Sq, \frac{\partial}{\partial y} f(\overline{y}, \overline{z}) y^{B_i\overline{u},B_ih} + \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \mathcal{W}'[S\overline{y}; Sy^{B_i\overline{u},B_ih}]\rangle_X dt.
\end{aligned}$$

Note that the first term on the left side and the second term on the right side cancel out. Rearranging yields

$$\langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \left(Sy^{B_i\overline{u},B_ih} - \mathcal{W}'[S\overline{y}; Sy^{B_i\overline{u},B_ih}]\right)\rangle_X dt \leq \int_{I_\partial} Sy^{B_i\overline{u},B_ih} d\mu.$$

Now remember $\mathcal{P} = \mathrm{Id} - \mathcal{W}$ (see Theorem 2.40). Consequently,

$$Sy^{B_i\overline{u},B_ih} - \mathcal{W}'[S\overline{y}; Sy^{B_i\overline{u},B_ih}] = \mathcal{P}'[S\overline{y}; Sy^{B_i\overline{u},B_ih}],$$

and the optimality condition (4.56) follows. $\square$

4.4.4 Improved optimality conditions and uniqueness for distributed controls

In this subsection, we study the question whether the weak optimality conditions (4.53) and (4.56) in Theorem 4.38 and Corollary 4.39 hold in a strong sense? In particular, we would like to replace $y^{B_i\overline{u},B_ih}$ in (4.53) and (4.56) by arbitrary functions $v \in \mathrm{dom}(A_p)$ and prove that both conditions hold a.e. in $t \in J_T$. As described in the beginning of Section 4.4, this is not possible since the hysteresis operator $\mathcal{W}$ acts non-local in time. Nevertheless, we follow the strategy of [MS15, Section 5] as long as possible. In a first step, we replace $y^{B_i\overline{u},B_ih}$ in (4.53) and (4.56) by an arbitrary function $y \in Y_{2,0}$. Unfortunately, this requires that B_i has dense range, which is not the case for B_2.

Hence, throughout this subsection, we consider problem (4.1)–(4.3) with $i = 1$, i.e. the control problem with distributed controls $u \in U_1$.

For appropriate choice of p in (A1)' in Assumption 4.2, the operator B_1 which maps $U_1 = \mathrm{L}^2(J_T; [\mathrm{L}^2(\Omega)]^m)$ into $\mathrm{L}^2(J_T; X)$ has dense range. Indeed, in (A1)' in Assumption 4.2 one can choose p with $\frac{1}{2} > 1 - \frac{1}{p} - \frac{1}{d}$. Equivalently, $2 < \frac{dp'}{d-p'}$. Hence, in this case Remark 2.7 yields $\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega) \hookrightarrow [\mathrm{L}^2(\Omega)]^m$. Moreover, the embedding is one-to-one which implies that the embedding $[\mathrm{L}^2(\Omega)]^m \hookrightarrow \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega) = X$ is dense so that B_1 has dense range.

In Subsection 4.4.4.1 we improve the optimality conditions (4.53) and (4.56) from Theorem 4.38 and Corollary 4.39 for $i = 1$, see Corollary 4.40. If f is continuously differentiable, this results in a variational inequality in time only.

The latter can be used to prove uniqueness of p, q and $d\mu$. This is done in Subsection 4.4.4.2, see Corollary 4.41 below.

4.4.4.1 Improved optimality conditions

The following corollary states an improvement of the optimality conditions (4.53) and (4.56) in Theorem 4.38 and Corollary 4.39 for $i = 1$.

Corollary 4.40 (Optimality condition for distributed controls)**.** [Mün17b, Corollary 4.15] Let Assumption 4.2 and Assumption 4.7 hold and let $\frac{1}{2} > 1 - \frac{1}{p} - \frac{1}{d}$. Assume that $\overline{u} \in U_1$ is a solution of problem (4.1)–(4.3) with $i = 1$, together with the state $\overline{y} \in Y_{2,0}$ and $\overline{z} = \mathcal{W}[S\overline{y}] \in \mathrm{H}^1(J_T)$. Let $v \in \mathrm{dom}(A_p)$ with $Sv > 0$ and $\varphi \in \mathrm{C}_0^\infty(J_T)$ be arbitrary. Then in addition to (4.53) in Theorem 4.38 there holds

$$\begin{aligned}
&\int_0^T \langle \lambda_1 + \lambda_2, \frac{v}{Sv}\varphi\rangle_{\mathrm{dom}(A_p)} + (\nu_1 + \nu_2)\varphi dt \\
&\leq \int_{I_\partial} \varphi d\mu + \int_0^T \langle p + Sq, f'[(\overline{y}, \overline{z}); ((v/Sv)\varphi, \mathcal{W}'[S\overline{y}; \varphi])]\rangle_X dt.
\end{aligned}$$

If f is continuously differentiable then in addition to (4.56) in Corollary 4.39 there holds

$$\int_0^T \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X \mathcal{P}'[S\overline{y}; \varphi] dt \leq \int_{I_\partial} \varphi \, d\mu \quad \text{for all } \varphi \in \mathrm{C}_0^\infty(J_T).$$

Proof. As seen in Subsection 4.4.4, $\frac{1}{2} > 1 - \frac{1}{p} - \frac{1}{d} \Leftrightarrow 2 < \frac{dp'}{d-p'}$ implies that B_1 has dense range according to Remark 2.7. As in [MS15, Lemma 5.2], this fact can be used to prove that the set $\{y^{B_1\overline{u},B_1h} : h \in U_1\}$ is dense in $Y_{2,0}$. We adapt the proof to our framework: For arbitrary $\eta \in Y_{2,0}$ one defines ξ such that $\eta = y^{B_1\overline{u},\eta} = G'[B_1\overline{u}; \xi]$, see Corollary 3.15, i.e.

$$\begin{aligned}
\dot{\eta}(t) + (A_p\eta)(t) &= F'[y; \eta](t) + \xi(t) \text{ in } J_T, \\
\eta(0) &= 0.
\end{aligned} \tag{4.57}$$

Since B_1 has dense range, one can approximate ξ by a sequence of functions $\{B_1 h_n\}_n$, $h_n \in U_1$ for $n \in \mathbb{N}$, which means $B_1 h_n \to \xi$ with $n \to \infty$. Because $G'[B_1\overline{u}; \cdot]$ is Lipschitz continuous by Corollary 3.15, it follows

$$y^{B_1\overline{u}, B_1 h_n} = G'[B_1\overline{u}; B_1 h_n] \to G'[B_1\overline{u}; \xi] = y^{B_1\overline{u},\xi} = \eta \quad \text{with} \quad n \to \infty.$$

In [MS15, Theorem 5.3], density of $\{y^{B_1\overline{u}, B_1 h} : h \in U_1\}$ in $Y_{2,0}$ is applied to proof a pointwise optimality condition which holds for all $v \in \operatorname{dom}(A_p)$ and a.e. $t \in J_T$. This is not possible in our case: For $\zeta \in \mathrm{C}(\overline{J_T}; X^\alpha)$, the values of the function $\mathcal{W}'[S\overline{y}; S\zeta](t)$ at time $t \in J_T$ depend on the previous values $\mathcal{W}'[S\overline{y}; S\zeta](s)$, $s \in \overline{J_t} = [0, t]$. Hence, the function is non-local in time and this translates to the composition

$$F'[y; \zeta](t) = f'[(y(t), \mathcal{W}[Sy](t)); (\zeta(t), \mathcal{W}'[Sy; S\zeta](t))].$$

Nevertheless, both functions $\mathcal{W}'[S\overline{y}; \cdot]$ and $f'[(y(t), \mathcal{W}[Sy](t)); \cdot]$, and hence also $F'[y; \cdot]$, are positive homogeneous. Let $\eta \in Y_{2,0}$ be arbitrary. Then there exists a sequence $\{\eta_n\} := \{y^{B_1\overline{u}, B_1 h_n}\} \subset \{y^{B_1\overline{u}, B_1 h} : h \in U_1\}$ which converges to η with $n \to \infty$. Since $F'[y; \cdot]$ is Lipschitz continuous from $\mathrm{C}(\overline{J_T}; X^\alpha)$ to $\mathrm{L}^2(J_T; X)$ according to Lemma 3.13, we can pass to the limit $n \to \infty$ in the sequence of inequalities (4.53), where $y^{B_1\overline{u}, B_1 h}$ is replaced by $y^{B_1\overline{u}, B_1 h_n}$. Hence, there holds

$$\int_0^T \langle \lambda_1 + \lambda_2 + S(\nu_1 + \nu_2), \eta \rangle_{\operatorname{dom}(A_p)} dt \leq \int_{I_\partial} S\eta d\mu + \int_0^T \langle p + Sq, F'[\overline{y}; \eta] \rangle_X dt,$$

where $F'[\overline{y}; \eta](t) = f'[(\overline{y}(t), \mathcal{W}[S\overline{y}](t)); (\eta(t), \mathcal{W}'[S\overline{y}; S\eta](t))]$. For arbitrary $v \in \operatorname{dom}(A_p)$ with $Sv > 0$ and $\varphi \in \mathrm{C}_0^\infty(J_T)$, the product $v\varphi$ is contained in $Y_{2,0}$. Moreover, since $\mathcal{W}'[S\overline{y}; \cdot]$ and $f'[(y(t), \mathcal{W}[Sy](t)); \cdot]$ are positive homogeneous, there holds $\mathcal{W}'[S\overline{y}; S(v\varphi)] = Sv\mathcal{W}'[S\overline{y}; \varphi]$ and then $f'[(\overline{y}, \overline{z}); (v\varphi, Sv\mathcal{W}'[S\overline{y}; \varphi])] = Svf'[(\overline{y}, \overline{z}); (v\varphi/Sv, \mathcal{W}'[S\overline{y}; \varphi])]$. Hence, choosing $\eta = \varphi v$ and rearranging some terms we obtain

$$\begin{aligned}
&\int_0^T \langle \lambda_1 + \lambda_2, \varphi v \rangle_{\operatorname{dom}(A_p)} + \varphi(\nu_1 + \nu_2) Sv dt \\
&\leq \int_{I_\partial} \varphi Sv d\mu + \int_0^T \langle p + Sq, Svf'[(\overline{y}, \overline{z}); (v\varphi/Sv, \mathcal{W}'[S\overline{y}; \varphi])] \rangle_X dt.
\end{aligned}$$

Dividing by Sv on both sides yields the first inequality in the corollary. The proof of the second inequality if f is continuously differentiable follows the same lines. $\square$

4.4.4.2 Uniqueness of the adjoint variables

In the following corollary, we exploit density of B_1 for appropriate $p \geq 2$ to prove that the adjoint system (p, q) and the measure $d\mu$ are unique in the case $i = 1$ and when f is continuously differentiable.

Corollary 4.41 (Unique adjoint system for distributed controls)**.** [Mün17b, Corollary 4.16] Let Assumption 4.2 and Assumption 4.7 hold and let $\frac{1}{2} > 1 - \frac{1}{p} - \frac{1}{d}$. Moreover, suppose that f is continuously differentiable from $X^\alpha \times \mathbb{R}$ into X. Assume that $\overline{u} \in U_1$ is a solution of problem (4.1)–(4.3) with $i = 1$, together with the state $\overline{y} \in Y_{2,0}$ and $\overline{z} = \mathcal{W}[S\overline{y}] \in \mathrm{H}^1(J_T)$. Then in the setting of Corollary 4.39 the adjoint couple $p \in Y_{2,T}^*$ and $q \in \mathrm{BV}(J_T)$ together with the measure $d\mu$ in $\mathrm{C}(\overline{J_T})^*$ is unique.

Proof. Remember that B_1 has dense range because $\frac{1}{2} > 1 - \frac{1}{p} - \frac{1}{d}$. Therefore, $\ker(B_1^*) = \overline{\operatorname{ran}(B_1)}^{\perp} = \{0\}$, i.e. B_1^* is one-to-one. That is, we can take the inverse $(B_1^*)^{-1}$ on both sides in equation (4.54), which was

$$B_1^*(p + Sq) = -\kappa\overline{u} \quad \text{in } U_1,$$

see Corollary 4.39. Hence, there holds

$$p + Sq = -\kappa(B_1^*)^{-1}\overline{u} \quad \text{in } X^* \text{ a.e. in } J_T, \tag{4.58}$$

cf. [MS15, Theorem 4.15]. Now suppose that there are two adjoint couples (p_1, q_1) and (p_2, q_2) which both satisfy the conditions of Corollary 4.39.
First we prove $p_1 = p_2$. To this aim let $\zeta \in \mathrm{L}^2(J_T; \operatorname{dom}(A_p))$ be an arbitrary test function. We subtract the evolution equation (4.55) for p_1 from that of p_2 and test the result with ζ. Note that the term

$$(p_1 + Sq_1) - (p_2 + Sq_2)$$

cancels out by (4.58). We end up with

$$\begin{aligned}
&\langle \dot p_2 - \dot p_1, \zeta\rangle_{\mathrm{L}^2(J_T;\operatorname{dom}(A_p))} \\
&= \langle \left[\frac{\partial}{\partial y} f(\overline{y}, \overline{z})\right]^* (p_1 + Sq_1 - (p_2 + Sq_2)) - A_p^*(p_2 - p_1) - SA_p(q_2 - q_1), \zeta\rangle_{\mathrm{L}^2(J_T;\operatorname{dom}(A_p))} \\
&= \langle p_1 + Sq_1 - (p_2 + Sq_2), \frac{\partial}{\partial y} f(\overline{y}, \overline{z})\zeta\rangle_{\mathrm{L}^2(J_T;X)} - \langle p_2 + Sq_2 - (p_1 + Sq_1), A_p\zeta\rangle_{\mathrm{L}^2(J_T;X)} = 0.
\end{aligned}$$

Since $\zeta \in \mathrm{L}^2(J_T; \operatorname{dom}(A_p))$ was arbitrary, this implies $\dot p_2 = \dot p_1$ in $\mathrm{L}^2(J_T; [\operatorname{dom}(A_p)]^*)$. Moreover, by definition of $Y_{2,T}^*$ there holds $p_1(T) = p_2(T) = 0 \in [\operatorname{dom}(A_p)]^*$. Because the functions $p_j \in \mathrm{L}^2(J_T; [\operatorname{dom}(A_p)]^*)$, $j \in \{1, 2\}$, satisfy

$$p_j(t) = p_j(T) - \int_t^T \dot p_j(s)\, ds \quad \text{for} \quad t \in J_T,$$

this shows $p_1 = p_2$ in $\mathrm{L}^2(J_T; [\operatorname{dom}(A_p)]^*)$. Density of the embedding $\operatorname{dom}(A_p) \hookrightarrow X$ implies that the embedding of X^* into $[\operatorname{dom}(A_p)]^*$ is one-to-one, which translates to the embedding $\mathrm{L}^2(J_T; X^*) \hookrightarrow \mathrm{L}^2(J_T; [\operatorname{dom}(A_p)]^*)$. But this implies $p_1 = p_2$ also in $\mathrm{L}^2(J_T; X^*)$. By the definition of $Y_{2,T}^*$, we conclude $p_1 = p_2 \in Y_{2,T}^*$.
We are left to prove $q_1 = q_2$. Since $p_1 = p_2$ in X^* a.e. in J_T, (4.58) yields $S(q_1 - q_2) = 0$ in X^* a.e. in J_T. Now let $v \in \operatorname{dom}(A_p)$ with $Sv > 0$ be arbitrary but fixed. Then we obtain

$$q_1 - q_2 = \frac{(q_1 - q_2)Sv}{Sv} = \frac{\langle S(q_1 - q_2), v\rangle_X}{Sv} = 0 \quad \text{in } \mathbb{R} \text{ a.e. in } J_T.$$

Hence, $q_1 = q_2$ in $\mathrm{L}^1(J_T)$. That is, there holds

$$\int_{[0,T]} |dq_1 - dq_2| = \sup\left\{\int_0^T (q_1 - q_2)\dot\varphi dt : \varphi \in \mathrm{C}_0^1(J_T),\ |\varphi| \le 1\right\} = 0.$$

Consequently, $dq_1 - dq_2 = 0$ as measures on $\overline{J_T}$, cf. [Vis13, p. XII.7]. We conclude $q_1 = q_2 \in \mathrm{BV}(0, T)$. Subtracting the equality of measures of dq_1 in Corollary 4.39 from that of dq_2 finally yields $d\mu_1 = d\mu_2$. □

4.4.5 Properties of the measures $d\mu$ and dq for distributed controls

In this subsection, we return to the question of c haracterizing the measures $d\mu$ and dq. As mentioned in the introduction of Section 4.4, this part extends the results in [Mün17b]. All the proofs rely on the improved optimality condition in Corollary 4.40 for continuously differentiable f and $i = 1$. That is, throughout this subsection we consider the control problem (4.1)–(4.3) with distributed control functions.
In Definition 4.42 below, we define the category of isolated times $t \in I_\partial$ in addition to the switching times from Definition 4.34. Moreover, we introduce the category of waiting slots, which includes all isolated times and switching times. In Lemma 4.46, we prove sign properties and bounds for $d\mu\{t\}$ for all isolated times $t \in I_\partial$. Afterwards, we generalize Lemma 4.46 by proving similar results for $d\mu$ on the category of waiting slots, see Theorem 4.47. With help of the measure equation for dq according to Corollary 4.39 we conclude the corresponding behaviour of dq on the category of waiting slots, see Corollary 4.48. In case of the regularity Assumption 4.30, all waiting slots are either isolated times or switching times. Hence, in this case Corollary 4.48 determines the direction of a jump of q at any time $t \in I_\partial \cap \overline{I_0}$. Moreover, we can derive an upper bound for the absolute value of a jump of q at time t. The results on the continuity properties of q are summarized in Corollary 4.49.

Definition 4.42. Consider the setting of Corollary 4.39. Let $(0, \partial)$-switching times and $(\partial, 0)$-switching times be defined as in Definition 4.34. We introduce two more categories of times in I_∂:

- We call $t \in I_\partial$ an isolated time if there exists some open interval $(c, d) \subset J_T$ with $t \in (c, d)$ such that $(c, d) \cap I_\partial = t$. The time $t = 0$ is called isolated if $[0, d) \cap I_\partial = 0$ for some $d \in J_T$ and $t = T$ is isolated if $(c, T] \cap I_\partial = T$ for some $c \in J_T$.

- A time interval $[d, e] \subset I_\partial$ is called a waiting slot if $S\overline{y} = 0$ a.e. in (d, e). A waiting slot $[d, e]$ is called isolated from below if there exists some constant $\varepsilon > 0$ such that $(d - \varepsilon, d) \subset I_0$. $[d, e]$ is called isolated from above if there exists some constant $\varepsilon > 0$ such that $(e, e+\varepsilon) \subset I_0$. If $[d, e]$ is isolated from below and above, then the waiting slot is called isolated. All waiting slots are defined maximal in the sense that there exists no waiting slot $[d', e']$ with $[d, e] \subset (d', e')$. The case $[d, e] = [d, d] := \{d\}$ is included in this definition, but we only call $\{d\}$ a waiting slot if $\{d\}$ is isolated from below and/or from above.

We suppose in the whole subsection that I_∂ is good natured in the following sense.

Assumption 4.43. The set $I_\partial = \{t \in \overline{J_T} : \ \overline{z}(t) \in \{a, b\}\}$ from Definition 4.24 consists only of waiting slots and intervals in which $S\dot{\overline{y}} \neq 0$ almost surely. There are at most countably many waiting slots in I_∂. Note that $(0, \partial)$-switching times, $(\partial, 0)$-switching times and isolated times are included in the definition of a waiting slot.

For the characterization of $d\mu$ we need one more definition:

Definition 4.44. Consider the subdivision of I_∂ from Definition 4.42 and let $t \in \overline{J_T}$ be given. Then we define the unique time $t^+ = t^+(t) \in \overline{J_T}$ according to the following hierarchical distinction of cases:

- $t^+ \in I_\partial^a$ and $(t, t^+) \cap I_\partial$ is empty or contains only waiting slots in I_∂^b.
- t^+ is the smallest time in $I_\partial^b \cap [t, T]$ such that $S\dot{\overline{y}} > 0$ a.e. in $(t^+, t^+ + \varepsilon)$ for some $\varepsilon > 0$.
- $t^+ = T$.

Similarly, we define $t^- = t^-(t)$ by the following hierarchical distinction of cases:

- $t^- \in I_\partial^b$ and $(t, t^-) \cap I_\partial$ is empty or contains only waiting slots in I_∂^a.
- t^- is the smallest time in $I_\partial^a \cap [t, T]$ such that $S\dot{\overline{y}} < 0$ a.e. in $(t^-, t^- + \varepsilon)$ for some $\varepsilon > 0$.
- $t^- = T$.

We will frequently apply two properties which are characteristic for the play operator $\mathcal{P}$:

Lemma 4.45. Let $\mathcal{P}$ denote the play operator from Definition 2.38. Then $\mathcal{P}$ satisfies the concatenation property

$$\mathcal{P}[v](t_2) = \mathcal{P}[v, v(0) - z_0](t_2) = \mathcal{P}[v(t_1 + .), \mathcal{P}[v](t_1)](t_2 - t_1) \tag{4.59}$$

for all $v \in \mathrm{C}(\overline{J_T})$ and all $0 \leq t_1 \leq t_2 \leq T$. Here, the second input variable denotes the initial value, see Definition 2.38. Moreover, the play operator satisfies the monotonicity property: For $v_1, v_2 \in \mathrm{C}(\overline{J_T})$ and for all $s \in \overline{J_T}$ for which $v_1(t) \leq v_2(t)$ for all $t \in \overline{J_s} = [0, s]$ there holds

$$\mathcal{P}[v_1, v_1(0) - z_0](t) \leq \mathcal{P}[v_2, v_2(0) - z_0](t) \quad \forall t \in \overline{J_s}. \tag{4.60}$$

Proof. (4.59) follows for example from [Vis13, III. (1.4)] and (4.60) is for example shown in the comment before [Vis13, III. Proposition 2.5]. □

4.4.5.1 Isolated times and waiting slots

In the following Lemma 4.46, we study $d\mu$ at isolated times. Subsequently, we generalize Lemma 4.46 to the more general category of waiting slots, see Theorem 4.47. In Corollary 4.48, we then conclude the behaviour of dq on the category of waiting slots.

Lemma 4.46 ($d\mu$ at isolated times)**.** With the assumptions as in Corollary 4.40 suppose that f is continuously differentiable and that Assumption 4.43 holds. Let $t \in I_\partial$ be an isolated time in the sense of Definition 4.42 and let $t^+ = t^+(t)$ and $t^- = t^-(t)$ be defined according to Definition 4.44.
If $d\mu(\{t\}) = 0$ then q is absolutely continuous in some interval $[t - \varepsilon, t + \varepsilon)$.
$t \in I_\partial^b$ is only possible if

$$\int_t^{t^+} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X ds \leq d\mu(\{t\}) \leq 0.$$

If in addition $(t, t^+) \subset I_0$, then $q(t+) < q(t^+-)$ if $d\mu(\{t\}) < 0$ and $q(t) \leq q(t^+-)$ if $d\mu(\{t\}) = 0$.
$t \in I_\partial^a$ is only possible if

$$\int_t^{t^-} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X ds \geq d\mu(\{t\}) \geq 0.$$

If in addition $(t, t^-) \subset I_0$, then $q(t+) > q(t^--)$ if $d\mu(\{t\}) > 0$ and $q(t) \geq q(t^+-)$ if $d\mu(\{t\}) = 0$.

Proof. Let $t \in I_\partial$ be an isolated time. We prove the lemma for $t \in I_\partial^b \cap J_T$. The prove for $t \in I_\partial^a$ and for $t \in I_\partial \cap \{0, T\}$ is analogous. By the definition of isolated times there exists a constant $\varepsilon_0 > 0$ such that $(t - \varepsilon_0, t) \cup (t, t + \varepsilon_0) \subset I_0$. Hence, continuity of $S\overline{y}$ and $\mathcal{P}[S\overline{y}]$ implies the existence of a constant $c > 0$ such that

$$\begin{aligned} S\overline{y}(\tau) - \mathcal{P}[S\overline{y}](\tau) = \mathcal{W}[S\overline{y}](\tau) \in (a + c, b) \quad \text{for} \quad \tau \in (t - \varepsilon_0, t) \cup (t, t + \varepsilon_0) \quad \text{and} \\ S\overline{y}(t) - \mathcal{P}[S\overline{y}](t) = \mathcal{W}[S\overline{y}](t) = b. \end{aligned} \tag{4.61}$$

Consequently,

$$S\overline{y} - \mathcal{P}[S\overline{y}] \in (a+c, b) \quad \text{for a.e.} \quad \tau \in (t-\varepsilon_0, t+\varepsilon_0), \tag{4.62}$$

and (2.30) implies that $\mathcal{P}[S\overline{y}]$ is constant in $[t-\varepsilon_0, t+\varepsilon_0]$.
Let $\varphi \in \mathrm{C}_0^\infty(J_T)$ be chosen with $\operatorname{supp}(\varphi) \subset (t-\varepsilon_0, t+\varepsilon_0)$. Moreover, consider

$$\mathcal{P}_\lambda[S\overline{y};\varphi] := \frac{1}{\lambda}(\mathcal{P}[S\overline{y}+\lambda\varphi] - \mathcal{P}[S\overline{y}]) \quad \text{for} \quad \lambda > 0.$$

$\mathcal{P} : \mathrm{C}(\overline{J_T}) \to \mathrm{L}^2(J_T)$ is Hadamard directionally differentiable according to Theorem 2.40. Hence, $\mathcal{P}_\lambda[S\overline{y};\varphi] \to \mathcal{P}'[S\overline{y};\varphi]$ in $\mathrm{L}^2(J_T)$ with $\lambda \to 0$.
Let $s \geq t+\varepsilon_0$ be arbitrary. For $v := S\overline{y} + \lambda\varphi$, $t_1 := t+\varepsilon_0$ and $t_2 := s$, the concatenation property (4.59) and $\operatorname{supp}(\varphi) \subset (t-\varepsilon_0, t+\varepsilon_0)$ imply

$$\mathcal{P}[S\overline{y}+\lambda\varphi](\tau) = \mathcal{P}[S\overline{y}] \qquad \text{for } \tau \in [0, t-\varepsilon_0] \qquad \text{and} \tag{4.63}$$

$$\begin{aligned}\mathcal{P}[S\overline{y}+\lambda\varphi](s) &= \mathcal{P}[(S\overline{y}+\lambda\varphi)(t+\varepsilon_0+\cdot), \mathcal{P}[S\overline{y}+\lambda\varphi](t+\varepsilon_0)](s-(t+\varepsilon_0)) \\ &= \mathcal{P}[S\overline{y}(t+\varepsilon_0+\cdot), \mathcal{P}[S\overline{y}+\lambda\varphi](t+\varepsilon_0)](s-(t+\varepsilon_0)).\end{aligned} \tag{4.64}$$

We have to distinguish three cases.
(I) $d\mu(\{t\}) = 0$:
If $d\mu(\{t\}) = 0$ then $dq(\{t\}) = 0$ by the equation of measures for dq, see Corollary 4.39. Moreover, $q|_{(t-\varepsilon_0,t)} \in \mathrm{H}^1(t-\varepsilon_0, t)$, $q|_{(t,t+\varepsilon_0)} \in \mathrm{H}^1(t, t+\varepsilon_0)$ and $q(t+) = q(t)$ since q is right-continuous. Hence, q is absolutely continuous in $[t-\varepsilon_0, t)$ and $[t, t+\varepsilon_0)$. Since $q(t-) = q(t+) - dq(\{t\}) = q(t)$, we conclude that q is continuous at t and hence absolutely continuous in $[t-\varepsilon_0, t+\varepsilon_0)$.
(II) $d\mu(\{t\}) > 0$:
We assume $d\mu(\{t\}) > 0$ and prove that this contradicts the maximum condition in Corollary 4.40. Let $\varphi \in \mathrm{C}_0^\infty(J_T)$ be chosen such that

$$\operatorname{supp}(\varphi) \subset (t-\varepsilon_0, t+\varepsilon_0), \quad \varphi \leq 0 \quad \text{and} \quad \varphi(t) < 0.$$

We prove $\mathcal{P}'[S\overline{y};\varphi] = 0 \in \mathrm{L}^2(J_T)$. To this aim, we show

$$\mathcal{P}[S\overline{y}+\lambda\varphi](\tau) = \mathcal{P}[S\overline{y}](\tau) = \mathcal{P}[S\overline{y}](t-\varepsilon_0)$$

for all $\tau \in [t-\varepsilon_0, t+\varepsilon_0]$ if λ is small enough.
$\mathcal{P}$ satisfies the monotonicity property (4.60) and $\mathcal{P}$ is Lipschitz continuous with modulus 1 according to (2.36) in Theorem 2.40. Hence, $\varphi \leq 0$ implies

$$\mathcal{P}[S\overline{y}](\tau) - \mathcal{P}[S\overline{y}+\lambda\varphi](\tau) \in [0, \lambda \max_{s\in[t-\varepsilon_0,\tau]} |\varphi(s)|] \quad \forall \tau \in \overline{J_T}.$$

We choose $\lambda_0 > 0$ such that $c + \lambda \min_{\tau\in(t-\varepsilon_0,t+\varepsilon_0)} \varphi(\tau) > 0$ for all $\lambda \in (0, \lambda_0)$. By (4.62) we can estimate

$$\begin{aligned} a &< a + c + \lambda\varphi(\tau) + \mathcal{P}[S\overline{y}](\tau) - \mathcal{P}[S\overline{y}+\lambda\varphi](\tau) \\ &< S\overline{y}(\tau) - \mathcal{P}[S\overline{y}](\tau) + \lambda\varphi(\tau) + \mathcal{P}[S\overline{y}](\tau) - \mathcal{P}[S\overline{y}+\lambda\varphi](\tau) \\ &= S\overline{y}(\tau) + \lambda\varphi(\tau) - \mathcal{P}[S\overline{y}+\lambda\varphi](\tau)\end{aligned}$$

for all $\tau \in [t-\varepsilon_0, t+\varepsilon_0]$ and for all $\lambda \in (0, \lambda_0)$. By (2.30), this implies $\frac{d}{dt}(\mathcal{P}[S\overline{y}+\lambda\varphi]) \geq 0$ and hence $\mathcal{P}[S\overline{y}+\lambda\varphi](\tau) \geq \mathcal{P}[S\overline{y}+\lambda\varphi](t-\varepsilon_0) = \mathcal{P}[S\overline{y}](t-\varepsilon_0)$ for all $\tau \in [t-\varepsilon_0, t+\varepsilon_0]$ and all $\lambda \in (0, \lambda_0)$. But since $\mathcal{P}[S\overline{y}+\lambda\varphi](\tau) \leq \mathcal{P}[S\overline{y}](\tau) = \mathcal{P}[S\overline{y}](t-\varepsilon_0)$ for all $\tau \in [t-\varepsilon_0, t+\varepsilon_0]$, we conclude

$$\mathcal{P}[S\overline{y}+\lambda\varphi](\tau) = \mathcal{P}[S\overline{y}+\lambda\varphi](t-\varepsilon_0) = \mathcal{P}[S\overline{y}](t-\varepsilon_0) = \mathcal{P}[S\overline{y}](\tau)$$

for all $\tau \in [t-\varepsilon_0, t+\varepsilon_0]$ and for all $\lambda \in (0,\lambda_0)$. But then the equations (4.63)–(4.64) yield $\mathcal{P}[S\overline{y}+\lambda\varphi] = \mathcal{P}[S\overline{y}]$ in $\overline{J_T}$ and hence $\mathcal{P}_\lambda[S\overline{y};\varphi] = 0$ in $\overline{J_T}$ for all $\lambda \in (0,\lambda_0)$.
Because $\mathcal{P}_\lambda[S\overline{y};\varphi] \to \mathcal{P}'[S\overline{y};\varphi]$ in $\mathrm{L}^2(J_T)$ with $\lambda \to 0$, we conclude $\mathcal{P}'[S\overline{y};\varphi] = 0 \in \mathrm{L}^2(J_T)$. Consequently, the maximum condition in Corollary 4.40 implies

$$0 \leq \int_{I_\partial} \varphi d\mu = \varphi(t) d\mu(\{t\}) < 0,$$

which is a contradiction.
(III) $d\mu(\{t\}) \leq 0$:
This time we chose a function $\varphi \in \mathrm{C}_0^\infty(J_T)$ with $\operatorname{supp}(\varphi) \subset [t_\varphi, \tilde{t_\varphi}] \subset (t-\varepsilon_0, t+\varepsilon_0)$ such that

$$\varphi \geq 0, \quad \dot{\varphi} > 0 \quad \text{in} \quad (t_\varphi, t), \quad \dot{\varphi} < 0 \quad \text{in} \quad (t, \tilde{t_\varphi}) \quad \text{and} \quad \varphi(t) = 1.$$

As in (4.63), note that $\mathcal{P}[S\overline{y}+\lambda\varphi](\tau) = \mathcal{P}[S\overline{y}](\tau)$ for $\tau \in [0, t_\varphi]$. Recall the monotonicity property (4.60) and that $\mathcal{P}$ is Lipschitz continuous with modulus 1. Since $\varphi \geq 0$, there holds

$$\mathcal{P}[S\overline{y}+\lambda\varphi](\tau) - \mathcal{P}[S\overline{y}](\tau) \in [0, \lambda \max_{s\in[t-\varepsilon_0,\tau]} \varphi(s)] \subset [0,\lambda] \quad \text{for} \quad \tau \in [t-\varepsilon_0, T]. \tag{4.65}$$

Hence,

$$\begin{aligned} S\overline{y}(t) + \lambda\varphi(t) - \mathcal{P}[S\overline{y}+\lambda\varphi](t) &= S\overline{y}(t) - \mathcal{P}[S\overline{y}](t) + \lambda - \mathcal{P}[S\overline{y}+\lambda\varphi](t) + \mathcal{P}[S\overline{y}](t) \\ &= b + \lambda - \mathcal{P}[S\overline{y}+\lambda\varphi](t) + \mathcal{P}[S\overline{y}](t) \geq b. \end{aligned}$$

Since $S\overline{y}(\tau) + \lambda\varphi(\tau) - \mathcal{P}[S\overline{y}+\lambda\varphi](\tau) \in [a,b]$ for all $\tau \in \overline{J_T}$, this implies

$$S\overline{y}(t) + \lambda\varphi(t) - \mathcal{P}[S\overline{y}+\lambda\varphi](t) = b. \tag{4.66}$$

Recall equation (4.62) and that $\mathcal{P}[S\overline{y}]$ is constant in $[t-\varepsilon_0, t+\varepsilon_0]$. We choose $\lambda_0 \in (0,c)$ so that $c - \lambda > 0$ for all $\lambda \in (0,\lambda_0)$. Then by (4.62) and (4.65) and because $\varphi \geq 0$ we obtain

$$\begin{aligned} a &< a + c + \lambda\varphi(\tau) - \lambda \\ &\leq a + c + \lambda\varphi(\tau) + \mathcal{P}[S\overline{y}](\tau) - \mathcal{P}[S\overline{y}+\lambda\varphi](\tau) \\ &< S\overline{y}(\tau) - \mathcal{P}[S\overline{y}](\tau) + \lambda\varphi(\tau) + \mathcal{P}[S\overline{y}](\tau) - \mathcal{P}[S\overline{y}+\lambda\varphi](\tau) \\ &= S\overline{y}(\tau) + \lambda\varphi(\tau) - \mathcal{P}[S\overline{y}+\lambda\varphi](\tau) \end{aligned} \tag{4.67}$$

for all $\tau \in [t-\varepsilon_0, t+\varepsilon_0]$ and for all $\lambda \in (0,\lambda_0)$. For $\lambda \in (0,\lambda_0)$, (2.30) thus yields that $\mathcal{P}[S\overline{y}+\lambda\varphi]$ is monotone increasing in $[t-\varepsilon_0, t+\varepsilon_0]$, which implies that $\mathcal{P}_\lambda[S\overline{y};\varphi]$ is monotone increasing in $[t-\varepsilon_0, t+\varepsilon_0]$ as well because $\mathcal{P}[S\overline{y}]$ is constant in this interval.
Moreover, $\lambda\varphi \geq 0$ and $\lambda\dot{\varphi} < 0$ in $(t, \tilde{t_\varphi})$. Hence,

$$S\overline{y}(\tau) + \lambda\varphi(\tau) - \mathcal{P}[S\overline{y}+\lambda\varphi](t) < S\overline{y}(t) + \lambda\varphi(t) - \mathcal{P}[S\overline{y}+\lambda\varphi](t) = b$$

for $\tau \in (t, \tilde{t_\varphi})$ and $\lambda \in (0,\lambda_0)$. According to (2.30), this together with (4.67) implies $\mathcal{P}[S\overline{y} + \lambda\varphi](\tau) = \mathcal{P}[S\overline{y}+\lambda\varphi](t)$ for $\tau \in [t, \tilde{t_\varphi}]$ and $\lambda \in (0,\lambda_0)$.
Moreover, (4.66) yields

$$\mathcal{P}[S\overline{y}+\lambda\varphi](t) - \mathcal{P}[S\overline{y}](t) = S\overline{y}(t) + \lambda\varphi(t) - b - (S\overline{y}(t) - b) = \lambda.$$

Hence, for all $\lambda \in (0,\lambda_0)$ there holds

$$\mathcal{P}_\lambda[S\overline{y};\varphi](\tau) = 1 \quad \text{for} \quad \tau \in [t, \tilde{t_\varphi}]. \tag{4.68}$$

For $\lambda \in (0, \lambda_0)$, let $t_\lambda \in \overline{J_T}$ be defined by the maximal time in $(t, T]$ which satisfies

$$\begin{aligned} &S\overline{y}(\tau) - \mathcal{P}[S\overline{y}](\tau) > a + \lambda \quad \text{for all} \quad \tau \in [t, t_\lambda] \quad \text{and} \\ &[t, t_\lambda) \cap I_\partial \quad \text{contains at most waiting slots.} \end{aligned} \tag{4.69}$$

Then (4.65) implies for $\lambda \in (0, \lambda_0)$ and $\tau \in [t, t_\lambda]$ that

$$\begin{aligned} a &\le a + \lambda + \mathcal{P}[S\overline{y}](\tau) - \mathcal{P}[S\overline{y} + \lambda\varphi](\tau) \\ &< S\overline{y}(\tau) - \mathcal{P}[S\overline{y}](\tau) + \mathcal{P}[S\overline{y}](\tau) - \mathcal{P}[S\overline{y} + \lambda\varphi](\tau) = S\overline{y}(\tau) - \mathcal{P}[S\overline{y} + \lambda\varphi](\tau). \end{aligned}$$

Moreover, for $\lambda \in (0, \lambda_0)$ and $\tau \in [t, t_\lambda]$ there holds

$$\begin{aligned} S\overline{y}(\tau) - \mathcal{P}[S\overline{y} + \lambda\varphi](\tau) &= S\overline{y}(\tau) - \mathcal{P}[S\overline{y}](\tau) + \mathcal{P}[S\overline{y}](\tau) - \mathcal{P}[S\overline{y} + \lambda\varphi](\tau) \\ &\le S\overline{y}(\tau) - \mathcal{P}[S\overline{y}](\tau) \le b. \end{aligned}$$

According to (4.69), $S\overline{y}(\tau) - \mathcal{P}[S\overline{y}](\tau) = b$ holds only at isolated times $\tau \in [t, t_\lambda]$ or in time intervals in which $S\overline{y}$ remains constant. This implies $S\overline{y}(\tau) - \mathcal{P}[S\overline{y} + \lambda\varphi](\tau) \in (a, b)$ for $\tau \in [t, t_\lambda]$, and we can apply (2.30) to conclude that $\mathcal{P}[S\overline{y}]$, $\mathcal{P}[S\overline{y} + \lambda\varphi]$ and then also $\mathcal{P}_\lambda[S\overline{y}; \varphi]$ are constant in $[t, t_\lambda]$. Together with (4.68) we conclude

$$\mathcal{P}_\lambda[S\overline{y}; \varphi](\tau) = 1 \quad \text{for} \quad \tau \in [t, t_\lambda] \text{ and } \lambda \in (0, \lambda_0). \tag{4.70}$$

We prove that $t_\lambda = t^+$ holds for small $\lambda \in (0, \lambda_0)$. To this aim, we have to distinguish two cases. In a last step, we finally prove the statement of the lemma.
(III.i) $t^+ \in I_\partial^b$:
If $t^+ \in I_\partial^b$, then continuity of $S\overline{y} - \mathcal{P}[S\overline{y}] = \mathcal{W}[S\overline{y}]$ implies the existence of a constant $\varepsilon_1 \in (0, \lambda_0)$ such that $S\overline{y} - \mathcal{P}[S\overline{y}] > a + \varepsilon_1$ in $[t, t^+]$. Hence, for $\lambda \in (0, \varepsilon_1)$, t^+ satisfies the conditions (4.69) which implies $t_\lambda \ge t^+$. Consequently, $\mathcal{P}_\lambda[S\overline{y}; \varphi] = 1$ in $[t, t^+]$ follows from (4.70).
According to the definition of t^+, there either exists some constant $\varepsilon > 0$ such that $\dot{S\overline{y}} > 0$ a.e. in $(t^+, t^+ + \varepsilon)$ or $t^+ = T$. If $t^+ = T$, then $T \ge t_\lambda \ge t^+ = T$ for $\lambda \in (0, \varepsilon_1)$. Otherwise, $t_\lambda = t^+$ for $\lambda \in (0, \varepsilon_1)$ holds because (4.69) equals the definition of $t^+ \in I_\partial^b$.
Assume that $t^+ < T$. Then $S\overline{y}$ is continuous and strictly increasing in some interval $(t^+, t^+ + \varepsilon)$ and $S\overline{y}(t^+ + \varepsilon) > S\overline{y}(t^+) + \varepsilon_1$ holds w.l.o.g. Hence, for $\lambda \in (0, \varepsilon_1)$ there exist times $s_\lambda \in (t^+, t^+ + \varepsilon)$ such that $s_\lambda \downarrow t^+$ with $\lambda \to 0$ and

$$S\overline{y}(\tau) < S\overline{y}(t^+) + \lambda \qquad \text{for } \tau \in [t^+, s_\lambda) \qquad \text{and} \qquad S\overline{y}(s_\lambda) = S\overline{y}(t^+) + \lambda.$$

Note that $\mathcal{P}[S\overline{y}](\tau) = S\overline{y}(\tau) - b$ holds for all $\tau \in [t^+, s_\lambda] \subset I_\partial^b$. Moreover, (2.30) implies $\mathcal{P}[S\overline{y} + \lambda\varphi](\tau) = \mathcal{P}[S\overline{y} + \lambda\varphi](t^+)$ for $\tau \in (t^+, s_\lambda]$ as long as $S\overline{y}(s) - \mathcal{P}[S\overline{y} + \lambda\varphi](t^+) < b$ for all $s \in [t^+, \tau)$. But for $\lambda \in (0, \varepsilon_1)$, (4.69) implies $\mathcal{P}[S\overline{y} + \lambda\varphi](t^+) = \mathcal{P}[S\overline{y}](t^+) + \lambda$ and hence

$$S\overline{y}(s) - \mathcal{P}[S\overline{y} + \lambda\varphi](t^+) = S\overline{y}(s) - \mathcal{P}[S\overline{y}](t^+) - \lambda \;\; = S\overline{y}(s) - S\overline{y}(t^+) + b - \lambda < b$$

for all $s \in [t^+, \tau) \subset [t^+, s_\lambda]$ according to the definition of s_λ. Moreover, we obtain

$$\mathcal{P}[S\overline{y} + \lambda\varphi](s_\lambda) = \mathcal{P}[S\overline{y}](t^+) + \lambda = S\overline{y}(t^+) - b + \lambda = S\overline{y}(s_\lambda) - b = \mathcal{P}[S\overline{y}](s_\lambda).$$

Hence,

$$\mathcal{P}_\lambda[S\overline{y}; \varphi](\tau) = 0 \quad \text{for} \quad \tau \in [s_\lambda, T].$$

(III.ii) $t^+ \in I_\partial^a$:
$t_\lambda = t^+$ for $\lambda \in (0, \varepsilon_1)$ follows similar as in Step III.i. Moreover, for $t^+ < T$, analogous to Step III.i one can show the existence of times s_λ of the following kind:

There holds $s_\lambda \downarrow t^+$ with $\lambda \to 0$. For $\lambda \in (0, \varepsilon_1)$, $\mathcal{P}[S\overline{y} + \lambda\varphi]$ is strictly decreasing and $\mathcal{P}[S\overline{y}]$ is constant in (t^+, s_λ). Moreover, $\mathcal{P}_\lambda[S\overline{y}; \varphi](\tau) = 0$ for $\tau \in [s_\lambda, T]$.
(III.iii) Conclusion:
By (4.63),(4.70) and Steps III.i–III.ii we have $\mathcal{P}_\lambda[S\overline{y}; \varphi](\tau) = 1$ for $\tau \in [t, t^+]$, $\mathcal{P}_\lambda[S\overline{y}; \varphi](\tau) \leq 1$ for $\tau \in [t^+, s_\lambda]$ and $\mathcal{P}_\lambda[S\overline{y}; \varphi](\tau) = 0$ for $\tau \in [0, t - t_\varphi] \cup [s_\lambda, T]$ for all $\lambda \in (0, \varepsilon_1)$. Consequently, because $s_\lambda \downarrow t^+$ with $\lambda \to 0$, $\mathcal{P}_\lambda[S\overline{y}; \varphi]$ converges pointwise a.e. to 0 in $[0, t - t_\varphi] \cup [t^+, T]$ and to 1 in $[t, t^+]$. Since $\mathcal{P}'[S\overline{y}; \varphi]$ is the $\mathrm{L}^2(J_T)$-limit of $\mathcal{P}_\lambda[S\overline{y}; \varphi]$ we conclude

$$\mathcal{P}'[S\overline{y}; \varphi] = 0 \quad \text{a.e. in } [0, t - t_\varphi] \cup [t^+, T] \quad \text{and} \quad \mathcal{P}'[S\overline{y}; \varphi] = 1 \quad \text{a.e. in } [t, t^+]. \tag{4.71}$$

Remember that $\mathcal{P}'[S\overline{y}; \varphi] \leq 1$ holds a.e. in $\overline{J_T}$. Moreover, $\operatorname{supp}(\varphi) \subset [t_\varphi, \tilde{t_\varphi}] \subset (t - \varepsilon_0, t + \varepsilon_0)$. Hence, for arbitrary $\varepsilon_3 > 0$ we can choose $t_\varphi = t_\varphi(\varepsilon_3)$ close enough to t such that

$$\int_0^t \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X \mathcal{P}'[S\overline{y}; \varphi] ds = \int_{t_\varphi}^t \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X \mathcal{P}'[S\overline{y}; \varphi] ds < \varepsilon_3.$$

Since $\varepsilon_3 > 0$ is arbitrary, (4.71) and the maximum condition in Corollary 4.40 finally yield

$$\int_t^{t^+} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X ds \leq \varphi(t) d\mu(\{t\}) = d\mu(\{t\}).$$

If $(t, t^+) \subset I_0$, then $\langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X = -\dot{q}$ a.e. in (t, t^+). For $d\mu(\{t\}) < 0$ this implies

$$q(t+) = q(t^+-) - \int_t^{t^+} \dot{q}(s) ds = q(t^+-) + \int_t^{t^+} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X ds < q(t^+-).$$

Similarly, we obtain $q(t) = q(t+) \leq q(t^+-)$ for $d\mu(\{t\}) = 0$. $\square$

Note that isolated times are isolated waiting slots $[d, e]$ where $d = e$. Moreover, the definition of a waiting slot includes $(0, \partial)$-switching times and $(\partial, 0)$-switching times. Accordingly, we generalize Lemma 4.46 by studying $d\mu$ on the category of waiting slots in Theorem 4.47.

Theorem 4.47 ($d\mu$ on waiting slots)**.** *Adopt the assumptions of Lemma 4.46. Let $d\mu^+$ and $d\mu^-$ be the positive and negative variation of $d\mu$ together with the positive and negative sets P and N [cf. Els11, Chapter VII]. Let $[d, e] \subset I_\partial$ be a waiting slot and consider the times $t^+(e)$ and $t^-(e)$ according to Definition 4.44. If $[d, e] \subset I_\partial^b$ then $d\mu^+([d, e]) = 0$ and $-d\mu$ is a positive measure on $[d, e]$. Moreover, $[d, e] \subset I_\partial^b$ is only possible if*

$$\int_{h_1}^{t^+(e)} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X \, ds \leq d\mu([h_1, h_2]) = d\mu^-([h_1, h_2]) \leq 0 \qquad \forall [h_1, h_2] \subseteq [d, e].$$

If $[d, e] \subset I_\partial^b$ is not isolated from above then $t^+(e) = e$ and $d\mu(\{e\}) = 0$.
If $[d, e] \subset I_\partial^a$ then $d\mu^-([d, e]) = 0$ and $d\mu$ is a positive measure on $[d, e]$. Moreover, $[d, e] \subset I_\partial^a$ is only possible if

$$\int_{h_1}^{t^-(e)} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X \, ds \geq d\mu([h_1, h_2]) = d\mu^+([h_1, h_2]) \geq 0 \qquad \forall [h_1, h_2] \subseteq [d, e].$$

If $[d, e] \subset I_\partial^a$ is not isolated from above then $t^-(e) = e$ and $d\mu(\{e\}) = 0$.

Proof. We prove the theorem for $[d,e] \subset I_\partial^b$. The proof for $[d,e] \subset I_\partial^a$ is analogous.
(I) $d\mu^+((d,e)) = 0$:
We show $d\mu^+((d,e)) = 0$. Assume $A := (d,e) \cap P \neq \emptyset$ and that $d\mu^+(A) = c > 0$. By regularity of $d\mu$ and $d\mu^+$ we can find a compact set $K \subseteq A$ and an open set U with $A \subseteq U \subseteq (d,e)$ such that

$$d\mu^+(U) - \frac{c}{2} < d\mu^+(A) < d\mu^+(K) + \frac{c}{2} \quad \text{and} \quad |d\mu|(U \backslash K) < \frac{c}{2}.$$

Furthermore, by the C^∞-Urysohn-Lemma [KP99, Theorem 1.1.3] we can find a function $\tilde{\varphi} \in \mathrm{C}^\infty(\overline{J_T})$ with

$$\chi_K \leq \tilde{\varphi} \leq \chi_U.$$

We set $\varphi := -\tilde{\varphi}$. Since $\varphi \leq 0$, the same techniques as in Step II of Lemma 4.46 yield that $\mathcal{P}'[S\overline{y};\varphi] = 0 \in \mathrm{L}^2(J_T)$. Moreover, we can estimate

$$\begin{aligned}
\int_{I_\partial} \varphi d\mu &= \int_{I_\partial} \varphi d\mu^+ + \int_{I_\partial} \varphi d\mu^- = \int_U \varphi d\mu^+ + \int_{U\backslash A} \varphi d\mu^- \\
&\leq -d\mu^+(K) - d\mu^-(U\backslash A) \leq -d\mu^+(K) + \frac{c}{2} \\
&< -d\mu^+(A) + 2\frac{c}{2} = -c + c = 0.
\end{aligned}$$

All together, Corollary 4.40 yields the contradiction

$$0 \leq \int_{I_\partial} \varphi d\mu < 0.$$

Therefore, $d\mu^+((d,e)) = 0$ and $-d\mu$ is a positive measure on (d,e).
(II) $d\mu^+(\{d,e\}) = 0$:
(II.i) $[d,e]$ is isolated from below and/or above:
If $[d,e]$ is isolated from below then there exists a constant $\varepsilon > 0$ such that $(d-\varepsilon, d) \subset I_0$. In this case, $d\mu((d-\varepsilon,d)) = 0$. Accordingly, we replace the sets K, A and U by $K_d \subseteq A_d := P \cap [d,e) \subseteq U_d \subseteq (d-\varepsilon, e)$ such that

$$d\mu^+(U_d) - \frac{d\mu^+(A_d)}{2} < d\mu^+(A_d) < d\mu^+(K_d) + \frac{d\mu^+(A_d)}{2} \quad \text{and} \quad |d\mu|(U_d \backslash K_d) < \frac{d\mu^+(A_d)}{2}.$$

The rest of the proof remains as in Step I and we conclude $d\mu^+([d,e)) = 0$. Similarly, if $[d,e]$ is isolated from above then there exists $\varepsilon > 0$ such that $(e, e+\varepsilon) \subset I_0$ and $d\mu((e,e+\varepsilon)) = 0$. In this case, we replace the sets K, A and U by $K_e \subseteq A_e := P \cap (d,e] \subseteq U_e \subseteq (d, e+\varepsilon)$ such that

$$d\mu^+(U_e) - \frac{d\mu^+(A_e)}{2} < d\mu^+(A_e) < d\mu^+(K_e) + \frac{d\mu^+(A_e)}{2} \quad \text{and} \quad |d\mu|(U_e \backslash K_e) < \frac{d\mu^+(A_e)}{2}.$$

We obtain $d\mu^+((d,e]) = 0$. Finally, if $[d,e]$ is isolated, then we replace K, A and P by $K_d \cup K_e$, $A_d \cup A_e$ and $U_d \cup U_e$. Again the rest of the proof remains as in Step I and we conclude $d\mu^+([d,e]) = 0$.
(II.ii) $[d,e]$ is not isolated from below:
If $[d,e]$ is not isolated from below then there exists $\varepsilon > 0$ such that $(d-\varepsilon, d) \subset I_\partial$ and $S\dot{\overline{y}} > 0$ a.e. in $(d-\varepsilon,d)$. We choose $\varphi \in \mathrm{C}_0^\infty(\overline{J_T})$ with $\mathrm{supp}(\varphi) \subset [t_\varphi, \tilde{t}_\varphi]$ such that

$$d \in (t_\varphi, \tilde{t}_\varphi), \qquad t_\varphi \in (d-\varepsilon, d), \qquad \tilde{t}_\varphi \in (d,e), \qquad -1 \leq \varphi \leq 0 \qquad \text{and} \qquad \varphi(d) = -1.$$

	$[d,e]$ isolated from above	$[d,e]$ not isolated from above
$[h_1,h_2]\subset[d,e)$	A.1	A.2
$h_1=h_2=e$	B.1	B.2
$h_1\in[d,e)$, $h_2=e$	C.1	C.2

Table 1: Case division for $[h_1,h_2]$

With the techniques as in Step II of Lemma 4.46 one can show that $\mathcal{P}'[S\overline{y};\varphi]=0$ a.e. in $(0,t_\varphi)\cup(\tilde{t}_\varphi,T)$. Moreover, $\mathcal{P}'[S\overline{y};\varphi]\geq -1$ almost surely. Hence, the maximum condition in Corollary 4.40 yields

$$-\int_{t_\varphi}^{\tilde{t}_\varphi}|\langle p+Sq,\frac{\partial}{\partial z}f(\overline{y},\overline{z})\rangle_X|\,ds\leq\int_{t_\varphi}^{\tilde{t}_\varphi}\langle p+Sq,\frac{\partial}{\partial z}f(\overline{y},\overline{z})\rangle_X\mathcal{P}'[S\overline{y};\varphi]\,ds$$
$$\leq\int_{I_\partial}\varphi d\mu\leq -d\mu^-((t_\varphi,\tilde{t}_\varphi))-d\mu^+(\{d\}).$$

For arbitrary $\varepsilon_1>0$, regularity of $d\mu^-$ implies that $(t_\varphi,\tilde{t}_\varphi)$ can be chosen small enough such that $\int_{t_\varphi}^{\tilde{t}_\varphi}|\langle p+Sq,\frac{\partial}{\partial z}f(\overline{y},\overline{z})\rangle_X|\,ds<\frac{\varepsilon_1}{2}$ and $|d\mu^-((t_\varphi,\tilde{t}_\varphi)\backslash\{d\})|<\frac{\varepsilon_1}{2}$. Hence, if $d\mu^+(\{d\})>0$ then $d\mu^-(\{d\})=0$ and $|d\mu^-((t_\varphi,\tilde{t}_\varphi)|=|d\mu^-((t_\varphi,\tilde{t}_\varphi)\backslash\{d\})|<\frac{\varepsilon_1}{2}$.
Consequently, $d\mu^+(\{d\})=0$ or

$$-\varepsilon_1\leq-\int_{t_\varphi}^{\tilde{t}_\varphi}|\langle p+Sq,\frac{\partial}{\partial z}f(\overline{y},\overline{z})\rangle_X|\,ds+d\mu^-((t_\varphi,\tilde{t}_\varphi))\leq -d\mu^+(\{d\})<0.$$

This proves $d\mu^+(\{d\})\leq\varepsilon_1$, and since ε_1 is arbitrary we conclude $d\mu^+(\{d\})=0$.
(II.iii) $[d,e]$ is not isolated from above:
If $[d,e]$ is not isolated from above then there exists $\varepsilon>0$ such that $(e,e+\varepsilon)\subset I_\partial$ and $S\dot{\overline{y}}>0$ a.e. in $(e,e+\varepsilon)$. Hence, $q|_{[e,e+\varepsilon)}=0$ by Lemma 4.26 and because q is right-continuous. By Corollary 4.39, the absolute value of q can only jump downwards in reverse time. Consequently, $dq(\{e\})=0$. But then the measure equation for dq yields

$$0=dq(\{e\})=d\mu(\{e\})+\int_{\{e\}}\langle p+Sq,\frac{\partial}{\partial z}f(\overline{y},\overline{z})\rangle_X\,ds=d\mu(\{e\}).$$

(III) We prove the lower bound for $d\mu([h_1,h_2])$:
Let $[h_1,h_2]\subset[d,e]$ be given. By Steps I–II there holds $d\mu([d,e])=d\mu^-([d,e])\leq 0$. Moreover, there exists $\varepsilon>0$ of the following kind: Either $(d-\varepsilon,d)\subset I_0$ or $S\dot{\overline{y}}>0$ a.e. in $(d-\varepsilon,d)$. Furthermore, $(e,e+\varepsilon)\subset I_0$ or $S\dot{\overline{y}}>0$ a.e. in $(e,e+\varepsilon)$.
Consider the division of cases in Table 1.
(III.i) Case B.2:
If $[d,e]$ is not isolated from above then $d\mu\{e\}=0$ by Step II.iii so that $t^+(e)=e$. Consequently,

$$0=\int_e^e\langle p+Sq,\frac{\partial}{\partial z}f(\overline{y},\overline{z})\rangle_X\,ds=d\mu(\{e\})$$

as required.
(III.ii) Other cases:
If Case B.2 does not apply, then we proceed as follows:
For arbitrary $h_1\in[d,h_2]$ we choose $t_\varphi\in(d-\varepsilon,h_1)$ and define $u_1:=t_\varphi$, $k_1:=h_1$.
In the Cases A.1–C.1 we choose $\tilde{t}_\varphi\in(h_2,e+\varepsilon)$ and define $u_2:=\tilde{t}_\varphi$, $k_2:=h_2$.

	1	2
A	$U = (t_\varphi, \tilde{t}_\varphi)$, $K = [h_1, h_2]$	$U = (t_\varphi, h_2)$, $K = [h_1, h_2]$
B	$U = (t_\varphi, \tilde{t}_\varphi)$, $K = [h_1, h_2]$	Does not apply
C	$U = (t_\varphi, \tilde{t}_\varphi)$, $K = [h_1, h_2]$	$U = (t_\varphi, h_2)$, $K = [h_1, \tilde{t}_\varphi]$

Table 2: Definition of U and K

In Case A.2 we choose $\tilde{t}_\varphi \in (h_2, e)$. Again we define $u_2 := \tilde{t}_\varphi$, $k_2 := h_2$.
Finally, in Case C.2 we choose $\tilde{t}_\varphi \in (h_1, e)$ and define $u_2 := e = h_2$, $k_2 := \tilde{t}_\varphi$.
The definition of $U := (u_1, u_2)$ and $K := [k_1, k_2]$ is summarized in Table 2.
Let $\varepsilon_1 > 0$ be arbitrary and remember that $d\mu$ is regular. Hence, t_φ and $\tilde{t}_\varphi$ can be chosen such that $U = (u_1, u_2)$ and $K = [k_1, k_2]$ satisfy

$$|d\mu|(U \backslash K) < \varepsilon_1.$$

Let $\varphi \in C_0^\infty(\overline{J_T})$ with $\operatorname{supp}(\varphi) \subset U$ and with $0 \le \varphi \le 1$, $\dot{\varphi} > 0$ in (u_1, k_1), $\dot{\varphi} < 0$ in (k_2, u_2) and $\varphi = 1$ in K. Since $\varphi \le 1$, this implies

$$\int_{I_\partial} \varphi d\mu = \int_U \varphi d\mu = \int_{U \backslash K} \varphi d\mu + d\mu(K) \le d\mu(U \backslash K) + d\mu(K) < d\mu(K) + \varepsilon_1$$

As in Step III in the proof of Lemma 4.46 one can show that $\mathcal{P}'[S\overline{y}; \varphi] = 0$ holds a.e. in $(0, u_1) \cup (t^+(k_2), T)$ and that $\mathcal{P}'[S\overline{y}; \varphi] = 1$ holds a.e. in $(k_1, t^+(k_2))$. Hence, the maximum condition in Corollary 4.40 yields

$$\begin{aligned}
&\int_{u_1}^{k_1} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X \mathcal{P}'[S\overline{y}; \varphi]\, ds + \int_{k_1}^{t^+(k_2)} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X\, ds \\
&= \int_0^T \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X \mathcal{P}'[S\overline{y}; \varphi]\, ds \le \int_{I_\partial} \varphi d\mu < d\mu(K) + \varepsilon_1.
\end{aligned}$$

Note that $\mathcal{P}'[S\overline{y}; \varphi] \le 1$ almost surely. Consequently, for arbitrary ε_2 we can choose $u_1 = t_\varphi = t_\varphi(\varepsilon_2)$ close to k_1 to obtain

$$\left| \int_{u_1}^{k_1} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X \mathcal{P}'[S\overline{y}; \varphi]\, ds \right| < \varepsilon_2.$$

For this choice and since $h_1 = k_1$ we can therefore estimate

$$\int_{h_1}^{t^+(k_2)} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X\, ds < d\mu(K) + \varepsilon_1 + \varepsilon_2.$$

Since ε_2 was arbitrary we conclude

$$\int_{h_1}^{t^+(k_2)} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X\, ds \le d\mu(K) + \varepsilon_1.$$

Remember that $d\mu = d\mu^-$ on $[d, e]$ by Steps I–II.
In Case C.2 we have $U = (t_\varphi, h_2) = (t_\varphi, e)$ and $K = [h_1, \tilde{\varphi}]$. Consequently,

$$d\mu([h_1, \tilde{\varphi}]) = d\mu(K) < d\mu([h_1, e)) + \varepsilon_1$$

since $(t_\varphi, h_1) = (u_1, k_1) \subset U \backslash K$. Moreover, Step II.iii yields $d\mu(\{u_2\}) = d\mu(\{e\}) = 0$ if $[d,e]$ is not isolated from above. Hence, in Case C.2 there holds

$$d\mu(K) < d\mu([h_1, e]) + \varepsilon_1 = d\mu^-([h_1, e]) + \varepsilon_1$$

and $t^+(h_1) = t^+(t_{\tilde{\varphi}}) = t^+(e) = e$.
In all other cases we have $U = (t_\varphi, t_{\tilde{\varphi}})$ and $K = [h_1, h_2]$ and we obtain $t^+(h_1) = t^+(h_2) = t^+(e)$ and

$$d\mu(K) = d\mu([h_1, h_2]) = d\mu^-([h_1, h_2]).$$

Since ε_1 was arbitrary, this concludes the proof. □

Corollary 4.48 (dq on waiting slots)**.** Consider the assumptions from Theorem 4.47. Let $[d,e]$ be a waiting slot. If $[d,e] \subset I_\partial^b$ then

$$\int_{h_2}^{t^+(e)} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X \, ds \leq \min\{0, dq([h_1, h_2])\} \qquad \forall [h_1, h_2] \subseteq [d,e].$$

If $[d,e] \subset I_\partial^b$ is not isolated from above then $t^+(e) = e$, $q(e) = 0$, $dq(\{e\}) = 0$,

$$0 \leq dq([h,e]) \qquad \forall h \in [d,e] \qquad \text{and} \qquad q|_{[d,e]} \leq 0.$$

If $[d,e] \subset I_\partial^a$ then

$$\int_{h_2}^{t^-(e)} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X \, ds \geq \max\{0, dq([h_1, h_2])\} \qquad \forall [h_1, h_2] \subseteq [d,e].$$

If $[d,e] \subset I_\partial^a$ is not isolated from above then $t^-(e) = e$, $q(e) = 0$, $dq(\{e\}) = 0$,

$$0 \geq dq([h,e]) \qquad \forall h \in [d,e] \qquad \text{and} \qquad q|_{[d,e]} \geq 0.$$

Proof. We prove the Corollary for $[d,e] \subset I_\partial^b$. The proof for $[d,e] \subset I_\partial^a$ is analogous. Let $[h_1, h_2] \subseteq [d,e]$ be arbitrary. By Theorem 4.47 and the measure equation for dq from Corollary 4.39 we can estimate

$$\begin{aligned}\int_{h_2}^{t^+(e)} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X \, ds &= \int_{h_1}^{t^+(e)} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X \, ds - \int_{h_1}^{h_2} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X \, ds \\ &\leq d\mu([h_1, h_2]) - \int_{h_1}^{h_2} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X \, ds = dq([h_1, h_2]).\end{aligned}$$

Moreover,

$$\int_{h_2}^{t^+(e)} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X \, ds \leq d\mu([h_2, e]) \leq 0.$$

If $[d,e]$ is not isolated from above then $t^+(e) = e$ by definition of t^+. In this case, $q(e) = 0$ is a consequence of Lemma 4.26 because q is right-continuous. $dq(\{e\}) = 0$ then follows from the fact that the absolute value of q can only jump down in reverse time, see Corollary 4.39.

□

4.4.5.2 Continuity properties of q in I_∂ with the regularity assumption

In this subsection, we assume that the regularity Assumption 4.30 applies and we consider the assumptions of Lemma 4.46. We apply Corollary 4.48 and characterize the continuity properties of q in I_∂. Under Assumption 4.30, all waiting slots consist of a single point. Moreover, by Assumption 4.43, I_∂ decomposes into intervals in which $S\dot{\overline{y}} \neq 0$ almost surely, isolated times, $(0,\partial)$-switching times and $(\partial,0)$-switching times.

Corollary 4.49 (Continuity properties of q in I_∂)**.** With the assumptions as in Lemma 4.46 let Assumption 4.30 hold true. Then all waiting slots consist of a single point. For any time $t \in I_\partial$ consider $t^- = t^-(t)$ and $t^+ = t^+(t)$ from Definition 4.44. We set $\sum_1^0 := 0$. The index set $1 \leq i \leq k$ has to be replaced by $1 \leq i < \infty$ and k by ∞ if the number of isolated times in the following is infinite.
The properties of q and dq at $t \in I_\partial$ can be characterized as follows:
1. t is a $(0,\partial)$-switching time or not contained in a waiting slot:
In this case, q is continuous at t with $q(t) = 0$.
2. $(\partial,0)$-switching times:
If t is a $(\partial,0)$-switching time then three cases can occur.
2.1. q is continuous at t with $q(t) = 0$.
2.2. q jumps up at t:
In this case, $t \in I_\partial^a$. Moreover, all isolated times t_i, $1 \leq i \leq k$, in the interval (t, t^-) are contained in I_∂^a so that $\delta_i := dq(\{t_i\}) \geq 0$. Furthermore,

$$0 < dq(\{t\}) \leq \int_t^{t^-} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X dt \quad \text{and} \quad q(t+) > 0 = q(t-) \geq -\sum_{i=1}^k \delta_i + q(t^- -).$$

2.3 q jumps down at t:
In this case, $t \in I_\partial^b$. Moreover, all isolated times t_i, $1 \leq i \leq k$, in the interval (t, t^+) are contained in I_∂^b so that $\delta_i := -dq(\{t_i\}) \geq 0$. Furthermore,

$$0 > dq(\{t\}) \geq \int_t^{t^+} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X dt \quad \text{and} \quad q(t+) < 0 = q(t-) \leq \sum_{i=1}^k \delta_i + q(t^+ -).$$

3. Isolated times in I_∂^a:
In this case, all isolated times t_i, $1 \leq i \leq k$, in the interval (t, t^-) are contained in I_∂^a so that $\delta_i := dq(\{t_i\}) \geq 0$. Moreover, q may only jump up at t. If q is discontinuous at t then $q(t+) > 0$,

$$\begin{aligned} &0 < dq(\{t\}) \leq \int_t^{t^-} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X dt \qquad \text{and} \\ &q(t+) > q(t-) \geq \max\left\{ -q(t+), -\sum_{i=1}^k \delta_i + q(t^- -) \right\}. \end{aligned}$$

4. Isolated times in I_∂^b:
In this case, all isolated times t_i, $1 \leq i \leq k$, in the interval (t, t^+) are contained in I_∂^b so that $\delta_i := -dq(\{t_i\}) \geq 0$. Moreover, q may only jump down at t. If q is discontinuous at t then

$q(t+) < 0$ and there holds

$$0 > dq(\{t\}) \geq \int_t^{t^+} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X dt \qquad \text{and}$$

$$q(t+) < q(t-) \leq \min \left\{ -q(t+), \sum_{i=1}^{k} \delta_i + q(t^+-) \right\}.$$

Proof. We prove the corollary in three steps.
(I) t is a $(0, \partial)$-switching time or not contained in a waiting slot:
If t is not contained in a waiting slot then t is contained in an interval in which $S\dot{\overline{y}} \neq 0$ almost surely. In this case, the claim follows from Corollary 4.39. If t is a $(0, \partial)$-switching time then Corollary 4.48 proves the statement since the regularity Assumption 4.30 implies that each $(0, \partial)$-switching time t corresponds to a waiting slot $\{t\}$ which is not isolated from above.
(II) $(\partial, 0)$-switching times and isolated times:
(II.i) Statement 2.1:
The regularity Assumption 4.30 implies that each $(\partial, 0)$-switching time t corresponds to a waiting slot $\{t\}$ which is not isolated from below.
Hence, for each such time there exists a constant $\varepsilon > 0$ such that $(t - \varepsilon, t] \subset I_\partial$, $S\dot{\overline{y}} \neq 0$ almost surely in $(t - \varepsilon, t)$ and $(t, t + \varepsilon) \subset I_0$. Corollary 4.39 entails $dq((t - \varepsilon, t)) = 0$ and $q(\tau) = 0$ for all $\tau \in [t - \varepsilon, t)$. Consequently, $q(t) = 0$ if q is continuous and we conclude Statement 2.1.
(II.ii) q jumps up at t:
Let t be a $(\partial, 0)$-switching time or an isolated time. If q jumps up at t, then $dq(\{t\}) > 0$. Moreover, $t \in I_\partial^a$ and

$$\max\{0, dq(\{t\})\} \leq \int_t^{t^-} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X dt$$

follow from Corollary 4.48 with $d = e = h_1 = h_2 = t$.
This shows Statement 2.2 except of the last inequality. The latter will be proven in Step III.ii below.
(II.iii) q jumps down at t:
Let t be a $(\partial, 0)$-switching time or an isolated time. If q jumps down at t, then $dq(\{t\}) < 0$. Moreover, $t \in I_\partial^b$ and

$$\min\{0, dq(\{t\})\} \geq \int_t^{t^+} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z}) \rangle_X dt$$

follow from Corollary 4.48 with $d = e = h_1 = h_2 = t$. This proves Statement 2.3 except of the last inequality. The latter will be proven in Step III.i below.
(III) Discontinuity points of q:
The behaviour of q at discontinuity points can be described by one of the following two cases.
(III.i) t is a $(\partial, 0)$-switching time or an isolated time and q jumps down at t:
Note that t^+ is the first $(0, \partial)$-switching time after t. We denote by $t_i \in I_\partial^b$, $1 \leq i \leq k$, the i-th isolated time in $I_\partial^b \cap (t, t^+)$ after t. If there are infinitely many such times then the index set $1 \leq i < \infty$ has to be considered and k has to be replaced by ∞ in the following proof. We define by

$$\delta_i := |d\mu(\{t_i\})| = -dq(\{t_i\}) \geq 0$$

the height of the jump of q at time t_i for $1 \leq i \leq k$. Note that $dq(\{t_i\}) \leq 0$ follows from Step II.iii. Suppose that q jumps down at t so that $\delta := |dq(\{t\})| = -dq(\{t\}) > 0$. In this case

Step II.iii proves $t \in I^b_\partial$ and

$$\int_t^{t^+} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X ds \leq -\delta.$$

For $t_0 = t$ and $t_{k+1} = t^+$ there holds $(t_{i-1}, t_i) \subset I_0$ for $1 \leq i \leq k+1$. By Corollary 4.39, this implies $q \in \mathrm{H}^1(t_{i-1}, t_i)$. Hence, we compute

$$\begin{aligned} q(t-) &= q(t-) - q(t+) + q(t+) = \delta + q(t_1-) + \int_t^{t_1} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X dt \\ &= \cdots = \delta + \sum_{i=1}^k \delta_i + q(t^+-) + \int_t^{t^+} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X dt \leq \sum_{i=1}^k \delta_i + q(t^+-). \end{aligned}$$

Moreover, Corollary 4.39 implies that the absolute value of q can only jump downwards in reverse time. Consequently, q can only jump down at t if $q(t+) < 0$. All together, q can only jump down at t if $q(t+) < 0$ and in this case we have

$$q(t+) < q(t-) \leq \min\left\{ -q(t+), \sum_{i=1}^k \delta_i + q(t^+-) \right\}.$$

This shows Statement 4. If t is a $(\partial, 0)$-switching time, then $q(t-) = 0$. Hence, we also conclude Statement 2.3.
(III.ii) t is a $(\partial, 0)$-switching time or an isolated time and q jumps up at t:
The result for this case follows analogous to Step III.i and proves Statement 2.1 and Statement 3.
□

4.5 Higher regularity of the solutions of the optimal control problem

In this section, we return to the general control problem (4.1)–(4.3) with $i \in \{1, 2\}$. The following results have been published in [Mün17b] in a similar form.
The aim of this section is to apply equation (4.52) from Theorem 4.38 and the regularity of the adjoint system (p, q) to increase the regularity of the optimal control $\overline{u} \in U_i$, $i \in \{1, 2\}$. The forcing term $B_i\overline{u} \in \mathrm{L}^2(J_T; X)$ in the state-equation (4.1) is responsible for the low a-priori regularity of $\overline{y} = G(B_i\overline{u})$, see Corollary 3.4. Hence, we can use the higher regularity of B_iu in order to improve the regularity of $\overline{y}$, and therewith also of $\overline{z} = \mathcal{W}[S\overline{y}]$.
Remember the notation $U_1 = \mathrm{L}^2(J_T; \tilde{U}_1)$ and $U_2 = \mathrm{L}^2(J_T; \tilde{U}_2)$ with

$$\tilde{U}_1 := [\mathrm{L}^2(\Omega)]^m \quad \text{and} \quad \tilde{U}_2 := \prod_{j=1}^m \mathrm{L}^2(\Gamma_{N_j}, \mathcal{H}_{d-1}).$$

Equation (4.52) in Theorem 4.38 implies the pointwise-in-time condition

$$B_i^*(p + Sq) = -\kappa\overline{u} \text{ in } [\tilde{U}_i]^* \text{ a.e. in } J_T.$$

Unfortunately, we can not exploit the high time-regularity of $p+Sq$ directly. The problem is that p is continuous only as a mapping into $[\mathrm{dom}(A_p)]^*$ but not into X^*, while B_i^* is only continuous as a mapping from X^* into $[\tilde{U}_i]^*$, but not necessarily as an operator defined on $[\mathrm{dom}(A_p)]^*$, see (A5) in Assumption 4.2.
Since, the regularity of p is limited, we enforce the assumptions on B_i in order to prove higher regularity of the optimal solutions in Theorem 4.51 below. Afterwards, in Example 4.52 we provide an example in which the following Assumption 4.50 applies.

Assumption 4.50. [Mün17b, Assumption 5.1] For $i \in \{1,2\}$, the operator $B_i : \tilde{U}_i \to X$ in (A5) is also continuous as a mapping into X^γ for some $\gamma \in (0,1]$. We denote by $I_{(\gamma)}$ the canonical embedding from X^γ into X. Then the assumption is equivalent to the fact that $B_i = I_{(\gamma)}\tilde{B}_i$ for a linear and continuous function $\tilde{B}_i : \tilde{U}_i \to X^\gamma$.

Theorem 4.51 (Higher regularity)**.** *[Mün17b, Theorem 5.2] In the setting of Theorem 4.38 let Assumption 4.50 hold for some $\gamma \in (0,1]$.*
If $\gamma > \frac{1}{2}$, then $\overline{u} \in \mathrm{L}^\infty(J_T;\tilde{U}_i)$, $\overline{y} \in Y_{s,0}$ and $\overline{z} \in \mathrm{W}^{1,s}(J_T)$ for arbitrary $s \in (1,\infty)$. If $\frac{1}{2}(1+\frac{d}{p}) < 1$, which is the case when $d = 2$ and $p > 2$ in (A1)' in Assumption 4.2, this implies $\overline{y} \in \mathrm{C}(\overline{J_T};[\mathrm{L}^\infty(\Omega)]^m)$. If in addition Ω is a Lipschitz domain then $\overline{y}$ is Hölder continuous in time and space.
If $\gamma \leq \frac{1}{2}$, then $\overline{u} \in \mathrm{L}^{\frac{2}{1-2s}}(J_T;\tilde{U}_i)$, $\overline{y} \in Y_{2/(1-2s),0}$ and $\overline{z} \in \mathrm{W}^{1,\frac{2}{1-2s}}(J_T)$ for arbitrary $s \in (0,\gamma)$. This implies $\overline{y} \in \mathrm{C}(\overline{J_T};X^\theta)$ for any $\theta \in (0,\frac{1}{2}+\gamma)$. If $\gamma > \frac{d}{2p}$ applies for d and p in (A1)' in Assumption 4.2, this implies $\overline{y} \in \mathrm{C}(\overline{J_T};[\mathrm{L}^\infty(\Omega)]^m)$. If in addition Ω is a Lipschitz domain then $\overline{y}$ is Hölder continuous in time and space.

Proof. In Corollary 2.30 we proved the compact embeddings

$$X^\gamma \hookrightarrow X^\beta \hookrightarrow X \quad \text{for} \quad 0 < \beta < \gamma \leq 1.$$

Consequently, there holds

$$X^* \hookrightarrow [X^\gamma]^* \hookrightarrow [X^\beta]^* \quad \text{for} \quad 0 < \beta < \gamma \leq 1.$$

With the representation $X^\gamma \simeq [X, \mathrm{dom}(A_p)]_\gamma$ according to Remark 2.31 and Remark 2.32 and by general calculus for complex interpolation spaces [cf. Ama95, Chp. 2.5 and Chp. 2.6] we obtain

$$[[\mathrm{dom}(A_p)]^*, X^*]_{1-\gamma} \simeq [X^*, [\mathrm{dom}(A_p)]^*]_\gamma \simeq [X, \mathrm{dom}(A_p)]_\gamma^* \simeq [X^\gamma]^*. \tag{4.72}$$

(I) Suppose $\gamma > \frac{1}{2}$ in Assumption 4.50:
(I.i) Higher regularity of $\overline{u}$:
Because the embedding $\mathrm{dom}(A_p) \hookrightarrow X$ is one-to-one and dense, the the same holds for the embedding $X^* \hookrightarrow [\mathrm{dom}(A_p)]^*$. Moreover, $1-\gamma < \frac{1}{2}$. Hence, as in Lemma 2.36, [Ama95, Theorem 3] together with (4.72) yield the injective embedding

$$Y_{2,T}^* \subset \mathrm{H}^1(J_T;[\mathrm{dom}(A)]^*) \cap \mathrm{L}^2(J_T;X^*) \hookrightarrow \mathrm{C}(\overline{J_T};[[\mathrm{dom}(A)]^*,X^*]_{1-\gamma}) \simeq \mathrm{C}(\overline{J_T};[X^\gamma]^*).$$

By Theorem 4.38, the adjoint function p is contained in $Y_{2,T}^*$. Consequently, the function $\tilde{p} := I_{(\gamma)}^* p \in \mathrm{C}(\overline{J_T};[X^\gamma]^*)$ can be uniquely identified with $p \in \mathrm{L}^2(J_T;X^*)$. Hence,

$$B_i^* p = \tilde{B}_i^* I_{(\gamma)}^* p = \tilde{B}_i^* \tilde{p} \in \mathrm{C}(\overline{J_T};[\tilde{U}_i]^*),$$

so that $\tilde{B}_i^*\tilde{p} \in \mathrm{C}(\overline{J_T};[\tilde{U}_i]^*)$ is a representative of $B_i^* p \in \mathrm{L}^2(J_T;[\tilde{U}_i]^*)$. Moreover, since the adjoint function q in Theorem 4.38 has bounded total variation, q is essentially bounded in J_T. By (A2)' in Assumption 4.2, S is contained in X^*. This implies that the product Sq can be interpreted as an element of $\mathrm{L}^\infty(J_T;X^*)$, which implies $B_i^* Sq \in \mathrm{L}^\infty(J_T;[\tilde{U}_i]^*)$. All together, equation (4.52) in Theorem 4.38 yields

$$\tilde{B}_i^*\tilde{p} + B_i^* Sq = B_i^*(p + Sq) = -\kappa\overline{u} \quad \text{in } [\tilde{U}_i]^*, \text{ a.e. in } J_T.$$

We identify $[\tilde{U}_i]^*$ in this equation with $\tilde{U}_i$ according to its Riesz representation. Since the function on the left side is then contained in $\mathrm{L}^\infty(J_T;\tilde{U}_i)$, it follows that the optimal control $\overline{u} \in \mathrm{L}^2(J_T;\tilde{U}_i)$ has a representative in $\mathrm{L}^\infty(J_T;\tilde{U}_i)$.

(I.ii) Higher regularity of $\overline{y}$:
We exploit the higher regularity of $\overline{u}$ from Step I.i in order to increase the regularity of the optimal state $\overline{y}$. Since $B_i\overline{u} \in \mathrm{L}^\infty(J_T; X)$ is the forcing term of the evolution equation (4.1) of $\overline{y} = G(B_i\overline{u})$, Corollary 3.4 implies $\overline{y} \in Y_{s,0}$ for arbitrary $s \in (1, \infty)$. Hence, the embedding Lemma 2.36 yields $\overline{y} \in \mathrm{C}(\overline{J_T}; X^\theta)$ for arbitrary $\theta \in [0, 1)$. Moreover, Remark 2.32 entails that X^θ is a subset of $[\mathrm{L}^\infty(\Omega)]^m$ if $\theta > \frac{1}{2}(1+\frac{d}{p})$. Corollary 2.19 ensures that $p \in \mathrm{J} \cap [2, \infty)$ can be chosen strictly larger than 2. Hence, if $\Omega \subset \mathbb{R}^d$ for $d = 2$ and $p > 2$ then $\frac{1}{2}(1 + \frac{d}{p}) = \frac{1}{2} + \frac{1}{p} \in (0, 1)$ so that there exists some $\theta \in (0, 1)$ with $\theta > \frac{1}{2}(1+\frac{d}{p})$. In this case, we obtain $\overline{y} \in \mathrm{C}(\overline{J_T}; [\mathrm{L}^\infty(\Omega)]^m)$. Finally, again for $p \in \mathrm{J} \cap (2, \infty)$ and $\Omega \subset \mathbb{R}^2$ suppose that Ω satisfies the assumptions of [DER15, Theorem 4.5], which enforce Assumption 2.6, see Remark 2.32. Those assumptions include Lipschitz domains [DER15, Remark 2.1], see also [ER14, Theorem 1.1] and the subsequent comments. Then [DER15, Theorem 4.5] yields that $\overline{y}$ is Hölder continuous in time and space.
(II) Suppose $\gamma \leq \frac{1}{2}$ in Assumption 4.50:
(II.i) Higher regularity of $\overline{u}$:
As seen in Step I.i, the embedding $X^* \hookrightarrow [\mathrm{dom}(A_p)]^*$ is one-to-one and dense. Hence, [Ama05, Theorem 3 and (22)] together with (4.72) entail

$$Y^*_{2,T} \subset \mathrm{H}^1(J_T; [\mathrm{dom}(A)]^*) \cap \mathrm{L}^2(J_T; X^*) \hookrightarrow \mathrm{L}^{\frac{2}{1-2s}}(J_T; [[\mathrm{dom}(A)]^*, X^*]_{1-\gamma}) \simeq \mathrm{L}^{\frac{2}{1-2s}}(J_T; [X^\gamma]^*)$$

for any $s \in (0, \gamma)$. By similar arguments as in Step I.i we exploit the regularity of the adjoint function $p \in Y^*_{2,T}$ in Theorem 4.38 to uniquely identify $p \in \mathrm{L}^2(J_T; X^*)$ with $\tilde{p} = \tilde{I_{(\gamma)}}^* p \in \mathrm{L}^{\frac{2}{1-2s}}(J_T; [X^\gamma]^*)$. Hence, $\tilde{B}_i^*\tilde{p} \in \mathrm{L}^{\frac{2}{1-2s}}(J_T; [\tilde{U}_i]^*)$ is a representative of $B_i^* p \in \mathrm{L}^2(J_T; [\tilde{U}_i]^*)$. With the same proof as in Step I.i we obtain $\overline{u} \in \mathrm{L}^{\frac{2}{1-2s}}(J_T; \tilde{U}_i)$ for arbitrary $s \in (0, \gamma)$.
(II.i) Higher regularity of $\overline{y}$:
Similar to Step I.ii, $B_i\overline{u} \in \mathrm{L}^{\frac{2}{1-2s}}(J_T; X)$ and Corollary 3.4 imply $\overline{y} \in Y_{2/(1-2s),0}$ for arbitrary $s \in (0, \gamma)$. Consequently, the embedding Lemma 2.36 yields $\overline{y} \in \mathrm{C}(\overline{J_T}; X^\theta)$ for arbitrary $\theta \in \left[0, 1 - \left(\frac{2}{1-2s}\right)^{-1}\right) = [0, \frac{1}{2} + s)$. Because $s \in (0, \gamma)$ can be chosen arbitrary, $\overline{y} \in \mathrm{C}(\overline{J_T}; X^\theta)$ for all $\theta \in [0, \frac{1}{2} + \gamma)$. The remaining statements are shown analogous to those of Step I.ii. □

We close this subsection with an example in which Theorem 4.51 applies.

Example 4.52. [cf. Mün17b, Remark 5.3] We provide an example in which Assumption 4.2 is valid for B_1 and for any $\gamma \in (0, \frac{1}{2})$. Theorem 4.51 then entails that $\overline{u}$ and $\overline{y}$ are more regular.
Suppose that the domain Ω is contained in $\mathbb{R}^2$, i.e. $d = 2$. Moreover, we choose $p \in \mathrm{J} \cap (2, \infty)$ in (A1)' in Assumption 4.2, i.e. $p > 2$. Let the assumptions and the notation be the same as in Theorem 4.38.
First of all, remember that B_1 defines an embedding $\tilde{U}_1 \hookrightarrow \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$, see Remark 4.3.
In particular, there holds

$$\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega) \hookrightarrow [\mathrm{L}^2(\Omega)]^m = [\tilde{U}_1]^* \quad \text{and} \quad B_1 : \tilde{U}_1 \hookrightarrow \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega). \tag{4.73}$$

Before we can apply Theorem 4.51 we have to enforce Assumption 2.6 on the domain Ω in order to achieve that Assumption 4.50 is valid.
[Gri+02, cf. Assumption 2.2] In the setting of Assumption 2.2 we suppose for all $j \in \{1, \dots, m\}$ and any $x \in \partial\Omega$ that there is an open neighborhood U_x of x and a bi-Lipschitz mapping ϕ_x from U_x onto some open set $V \in \mathbb{R}^d$ such that $\phi_x((\Omega \cup \Gamma_{D_j}) \cap U_x)$ equals either the whole unit ball or the union of the lower unit half ball and its top surface. Moreover, the functional determinant of each bi-Lipschitz transformation ϕ_x is a.e. constant.

Note that it does not matter if we consider the unit ball or the unit cube in the above assumption on the domain. Furthermore, the two restrictions of [Gri+02, Assumption 2.2] in comparison to Assumption 2.6 are that [Gri+02, Assumption 2.2] is supposed to hold for all $x \in \partial\Omega$ and that each bi-Lipschitz transformation ϕ_x has a.e. constant functional determinant. By [Gri+02, Remark 2.3], this stronger assumption still includes all Lipschitz domains. The rest of Assumption 4.2 remains the same.
With this enforced assumption, [Gri+02, Theorem 3.1] entails the topological equivalences

$$\left[\mathbb{W}^{-1,p_1}_{\Gamma_D}(\Omega), [\mathrm{L}^2(\Omega)]^m\right]_\theta \simeq \mathbb{W}^{-\theta,p}_{\Gamma_D}(\Omega) \quad \forall \theta \in (0,1),\ \theta \neq \frac{1}{p'},\ \frac{1}{p} = \frac{1-\theta}{2} + \frac{\theta}{p_1}.$$

Together with (4.73) and $p_1' \leq 2$, this yields an embedding

$$\tilde{U}_1 \hookrightarrow \mathbb{W}^{-\theta,p}_{\Gamma_D}(\Omega) \quad \forall \theta \in (0,1),\ \theta \neq \frac{1}{p'}. \tag{4.74}$$

Moreover, [Gri+02, Theorem 3.5] implies the equivalences

$$\mathbb{W}^{-\theta,p}_{\Gamma_D}(\Omega) \simeq [\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)]_\gamma \quad \text{for} \quad \gamma = \frac{1-\theta}{2} \quad \forall \theta \in (0,1).$$

With Remark 2.31, Remark 2.32 and (4.74) we conclude

$$\mathbb{W}^{-\theta,p}_{\Gamma_D}(\Omega) \simeq [\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)]_\gamma \simeq X^\gamma \quad \text{and} \quad \tilde{U}_1 \hookrightarrow X^\gamma \quad \text{for} \quad \gamma = \frac{1-\theta}{2} \quad \forall \theta \in (0,1)\setminus\left\{\frac{1}{p'}\right\}.$$

Note that $\theta = 1-2\gamma \in (0,1)$ for arbitrary $\gamma \in (0,\frac{1}{2})$ and $\theta = 1-2\gamma \Leftrightarrow \gamma = \frac{1-\theta}{2}$. Hence, we obtain an embedding $\tilde{B}_1 : \tilde{U}_1 \hookrightarrow X^\gamma$ for any $\gamma \in (0,\frac{1}{2})\setminus\left\{\frac{1}{2p}\right\}$. Therefore, Assumption 4.50 holds for $B_1 = I_{(\gamma)}\tilde{B}_1$ for any $\gamma \in (0,\frac{1}{2})\setminus\left\{\frac{1}{2p}\right\}$. Consequently, we conclude from Theorem 4.51 the increased regularity $\overline{u} \in \mathrm{L}^{\frac{2}{1-2s}}(J_T;\tilde{U}_1)$, $\overline{y} \in Y_{1/(1-2s),0}$ and then $\overline{z} = \mathcal{W}[S\overline{y}] \in \mathrm{W}^{1,\frac{2}{1-2s}}(J_T)$ for arbitrary $s \in (0,\gamma)$. Note that because $d = 2$ and $p > 2$, $\gamma \in (0,\frac{1}{2})$ can be chosen such that $\gamma > \frac{d}{2p}$. Hence, Theorem 4.51 yields $\overline{y} \in \mathrm{C}(\overline{J_T};[\mathrm{L}^\infty(\Omega)]^m)$. As mentioned above, all Lipschitz domains still satisfy the assumption of this example. If we consider Ω to be a Lipschitz domain, then Theorem 4.51 entails that $\overline{y}$ is Hölder continuous in time and space.

4.6 The value function of a perturbed control problem

In this section, we consider a family of perturbed control problems similar to (4.1)–(4.3). All results have already been published in [Mün17b], but we provide the proofs in more details. We consider the same notation as in Section 4, and the main assumption during this whole section is Assumption 4.2. For $i \in \{1,2\}$ and $r \in U_i$, the control problem of interest is the following:

$$\min_{u \in C} J(G(B_i(u+r)), u+r) = \|G(B_i(u+r)) - y_d\|^2_{U_1} + \frac{\kappa}{2}\|u+r\|^2_{U_i}. \tag{4.75}$$

Remark 4.53. As in Remark 4.1, note that the functions $G(B_i(u+r)) \in Y_{2,0} \hookrightarrow \mathrm{L}^2(J_T;\mathrm{dom}(A_p))$ in (4.75) are identified with $I_p^{-1}G(B_i(u+r)) \in \mathrm{L}^2(J_T;\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)) \hookrightarrow U_1$ for $u+r \in U_i$, see also Corollary 2.30 and Remark 2.32. Accordingly, (4.75) has to be understood as

$$\min_{u \in C} J(I_p^{-1}G(B_i(u+r)), u+r) = \|I_p^{-1}G(B_i(u+r)) - y_d\|^2_{U_1} + \frac{\kappa}{2}\|u+r\|^2_{U_i}.$$

The set of admissible control functions C is assumed to be a convex and closed subset of U_i. Note that if $C = U_i$ and if $\overline{u}$ is an optimal control for problem (4.1)–(4.3), then $u = \overline{u} - r$ is admissible for problem (4.75) for any perturbation $r \in U_i$. Hence, in this case the minimal value in problem (4.75) for any $r \in U_i$ equals the minimal value in (4.1)–(4.3). Therefore, we assume $C \neq U_i$ throughout this section.
We are interested in the optimal value function

$$v : U_i \to \mathbb{R},\ r \mapsto v(r) := \min_{u \in C} J(G(B_i(u+r)), u+r). \tag{4.76}$$

Closely related to v is the multifunction V, which assigns to each $r \in U_i$ the set of controls $u \in U_i$ for which the minimal value $v(r)$ is obtained:

$$V : r \in U_i \mapsto V(r) := \{u \in C : \ J(G(B_i(u+r)), u+r) = v(r)\}. \tag{4.77}$$

We refer to [BS00, Chp. 4.1] for a broader introduction into sensitivity and stability analysis. Our interest is to understand the stability properties of v and V.

Theorem 4.54 (Optimal value function and optimal set function)**.** *[Mün17b, Theorem 6.1] Let Assumption 4.2 hold. For $i \in \{1, 2\}$, let $C \subset U_i$ be convex and closed. Consider the optimal control problem* (4.75) *for $r \in U_i$ together with the corresponding minimal value function v, defined by* (4.76)*, and the multifunction V from* (4.77)*. Then v is weakly lower semi-contiuous. If C is compact in U_i then v is upper semi-contiuous and therefore continuous. In this case, also the multifunction V is upper semi-continuous, i.e. for each $r_0 \in U_i$ and for any neighborhood $U_{V(r_0)}$ of $V(r_0)$ there exists a neighborhood U_{r_0} of r_0 such that $V(r) \subset U_{V(r_0)}$ for all $r \in U_{r_0}$, cf. [BS00, Chapter 4.1].*

Proof. We prove the theorem in four steps.
(I) Well-posedness of problem (4.75) for all $r \in U_i$:
That problem (4.75) is well-posed is shown in the same way as Theorem 4.6, where existence of an optimal control for the unperturbed problem (4.1)–(4.3) was proven. The latter problem is equal to the perturbed problem with $r = 0$ and $C = U_i$. In the proof, we used results about weak continuity of the solution operator G of the generalized state equation (3.11), see Lemma 4.5. The fact that C is closed and convex - and hence weakly closed - is necessary to obtain that the weak limit $\overline{u}$ of each minimizing sequence $\{u_n\} \subset C$ of (4.75) is admissible, i.e. that $\overline{u} \in C$.
(II) $v : U_i \to \mathbb{R}$ is weakly lower semi-contiuous:
Note first that weak lower semi-continuity implies strong lower semi-continuity, since every strongly convergent sequence converges in the weak sense as well. We prove that the optimal value function v is weakly lower semi-contiuous. To this aim, we show that

$$v(r_0) \leq \liminf_{n \to \infty} v(r_n) \tag{4.78}$$

for any $r_0 \in U_i$ and for each sequence $\{r_n\} \subset U_i$ for which $r_n \rightharpoonup r_0$ in U_i with $n \to \infty$. Let $r_0 \in U_i$ be arbitrary and suppose that $\{r_n\} \subset U_i$ converges weakly to r_0 with $n \to \infty$. In order to prove (4.78) we show that for any $\varepsilon > 0$ there exists some $n_0 \in \mathbb{N}$ such that

$$v(r_0) - \varepsilon \leq v(r_n) \quad \text{for all} \quad n \geq n_0.$$

First of all, note that the sequence $\{r_n\}$ is bounded by some constant $c_0 > 0$ because it converges weakly to r_0. By definition of the cost function J and the multifunction V, it follows that the union of optimal controls $\cup_{n \in \mathbb{N}} V(r_n) \subset C$ is contained in some ball $\mathrm{B}_{U_i}(0, R)$ with $R > 0$, i.e. $\cup_{n \in \mathbb{N}} V(r_n) \subset \mathrm{B}_{U_i}(0, R)$.

Suppose in contradiction to the assumption that for some $\varepsilon > 0$ there exists a subsequence $\{r_{n_k}\}$ of $\{r_n\}$, together with optimal solutions $u_{n_k} \in V(r_{n_k})$ of the perturbed problems, such that

$$v(r_0) - \varepsilon > v(r_{n_k}) = J(I_p^{-1}G(B_i(u_{n_k} + r_{n_k})), u_{n_k} + r_{n_k}) \quad \text{for all} \quad n_k \quad \text{with } k \geq 0.$$

Since C is a convex and closed subset of the reflexive space U_i, it is weakly closed and even weakly compact by Alaoglu's compactness theorem [W05, Satz VIII.3.18]. Moreover, the sequence $\{u_{n_k}\} \subset C$ is bounded by $R > 0$. Hence, there exists yet another subsequence, for which we maintain the index n_k, and some $\overline{u} \in C$ such that $u_{n_k} \rightharpoonup \overline{u}$ with $k \to \infty$. By Lemma 4.5, $G(u_{n_k} + r_{n_k})$ converges to $G(\overline{u} + r_0)$ even strongly in $Y_{2,0}$ and hence $I_p^{-1}G(u_{n_k} + r_{n_k})$ converges to $I_p^{-1}G(\overline{u} + r_0)$ in U_1 by the embedding $Y_{2,0} \hookrightarrow \mathrm{L}^2(J_T; \mathrm{dom}(A_p)) \hookrightarrow U_1$. Moreover, J is continuous and weakly lower semi-continuous on $U_1 \times U_i$. This implies that $J(I_p^{-1}G(B_i(\cdot + r_0)), \cdot)$ is continuous and weakly lower semi-continuous on U_i. Consequently, there holds

$$v(r_0) \leq J(I_p^{-1}G(B_i(\overline{u} + r_0)), \overline{u} + r_0) \leq \liminf_{k \to \infty} J(I_p^{-1}G(B_i(u_{n_k} + r_{n_k})), u_{n_k} + r_{n_k}) \leq v(r_0) - \varepsilon,$$

which is a contradiction. Therefore, $v(r_0) \leq \liminf_{n\to\infty} v(r_n)$, so that v is weakly lower semi-continuous.

(III) $v : U_i \to \mathbb{R}$ is upper semi-continuous if C is compact:
Let C be convex and compact. We apply techniques from [BS00, Proposition 4.4]. We have to show that

$$v(r_0) \geq \lim_{n \to \infty} v(r_n)$$

holds for any $r_0 \in U_i$ and for each sequence $\{r_n\} \subset U_i$ for which $r_n \to r_0$ in U_i with $n \to \infty$. Equivalently, v is upper semi-continuous if for any $\varepsilon > 0$ there exists a neighborhood U_{r_0} of r_0 such that

$$v(r) \leq v(r_0) + \varepsilon \quad \text{for all} \quad r \in U_{r_0}. \tag{4.79}$$

To prove (4.79), we show that there exist neighborhoods $U_{V(r_0)}$ of $V(r_0)$ and U_{r_0} of r_0 such that

$$J(I_p^{-1}G(B_i(u + r)), u + r) < v(r_0) + \varepsilon \quad \text{for all} \quad (u, r) \in U_{V(r_0)} \times U_{r_0}. \tag{4.80}$$

Note that by definition of v, $v(r) \leq J(I_p^{-1}G(B_i(u + r)), u + r)$ holds in (4.80) for all $u \in U_{V(r_0)}$ and $r \in U_{r_0}$, so that (4.80) indeed implies (4.79). As seen in Step II, the mapping $(r, u) \mapsto J(I_p^{-1}G(B_i(r + u)), r + u)$ is continuous on $U_i \times C$. Hence, the set

$$S_\varepsilon := \{(r, u) \in U_i \times C : \ J(I_p^{-1}G(B_i(u + r)), u + r) < v(r_0) + \varepsilon\}$$

is open. Moreover, $\{r_0\} \times V(r_0) \subset S_\varepsilon$ by definition of v and V. For each $u \in V(r_0)$, this implies the existence of neighborhoods $U_{r_0,u} \subset U_i$ of r_0 and $W_u \subset C$ of u such that $U_{r_0,u} \times W_u \subset S_\varepsilon$. The union $\cup_{u \in V(r_0)} W_u$ provides an open cover of $V(r_0)$. Since $J(I_p^{-1}G(B_i(\cdot + r_0)), \cdot)$ is continuous on U_i, $V(r_0) = \{u \in C : \ J(I_p^{-1}G(B_i(u + r_0)), u + r_0) = v(r_0)\}$ is closed and therefore compact as a closed subset of the compact set C. Hence, there exist $u_1, \ldots, u_k \in V(r_0)$ such that $U_{V(r_0)} := \cup_{j=1}^k W_{u_j}$ defines a subcover of $V(r_0)$ which is the union of finitely many open sets. But then the set $U_{r_0} := \bigcap_{j=1}^{k} U_{r_0,u_j}$ is open as the intersection of finitely many open sets. Consequently, U_{r_0} defines a neighborhood of r_0. Moreover, (4.80) holds for $U_{r_0} \times U_{V(r_0)}$ since $U_{r_0} \times U_{V(r_0)} \subset S_\varepsilon$. This proves (4.79) and that v is upper semi-continuous.

(IV) $V : U_i \rightrightarrows C \subset U_i$ is upper semi-continuous if C is compact:
The proof is oriented at [BS00, Proposition 4.4] but more detailed. Let $U_{V(r_0)}$ be a neighborhood of $V(r_0)$. We have to show that there exists a neighborhood U_{r_0} of r_0 such that $V(r) \subset U_{V(r_0)}$

for all $r \in U_{r_0}$. By (4.80) in Step III there exist $\varepsilon > 0$ and a neighborhood $U^1_{r_0}$ of r_0 such that the open set

$$U^{\varepsilon}_{V(r_0)} := \{u \in C : \ J(I_p^{-1}G(B_i(u+r), u+r)) < v(r_0) + \varepsilon \ \forall r \in U^1_{r_0}\}$$

is contained in $U_{V(r_0)}$. Note that $U^{\varepsilon}_{V(r_0)}$ defines a neighborhood of $V(r_0)$. Because of the inclusion $C \backslash U_{V(r_0)} \subset C \backslash U^{\varepsilon}_{V(r_0)}$, this implies

$$J(I_p^{-1}G(B_i(u+r)), u+r) \geq v(r_0) + \varepsilon \quad \text{for all} \quad (r,u) \in U^1_{r_0} \times C \backslash U_{V(r_0)}.$$

By Steps II–III, v is continuous because C is compact. Hence, we can find a neighborhood $U^2_{r_0}$ of r_0 such that $|v(r) - v(r_0)| \leq \frac{\varepsilon}{2}$ for all $r \in U^2_{r_0}$. We define $U_{r_0} := U^1_{r_0} \cap U^2_{r_0}$. By this choice we obtain

$$J(I_p^{-1}G(B_i(u+r)), u+r) \geq v(r) + \frac{\varepsilon}{2} \quad \text{for all} \quad (r,u) \in U_{r_0} \times C \backslash U_{V(r_0)}.$$

If $u \in V(r)$, then $J(I_p^{-1}G(B_i(u+r)), u+r) = v(r)$. Hence, there holds $V(r) \cap C \backslash U_{V(r_0)} = \emptyset$ for all $r \in U_{r_0}$. We conclude $V(r) \subset U_{V(r_0)}$ for all $r \in U_{r_0}$.

$\square$

Acknowledgments

I gratefully acknowledge the funding received towards my Ph.D. from the German Research Foundation (DFG) through the International Research Training Group IGDK 1754 “Optimization and Numerical Analysis for Partial Differential Equations with Nonsmooth Structures.”

A lot of people contributed to the success of this work.

First of all, I would like to thank my supervisor Prof. Brokate from the Technical University of Munich. He helped me finding my way into science and in particular in the field of optimization theory and hysteresis. He always had an open ear for problems and questions of any kind during my Ph.D.

I would also like to express my thanks to Prof. Fellner from the Karl-Franzens-University of Graz who acted as my second supervisor. He hosted me for my three stays in Graz and introduced me in the fields of reaction-diffusion-systems as well as dynamical systems. I appreciate our many helpful discussions and that he thoroughly proofread all of my manuscripts.

Furthermore, I would like to thank all my friends and colleagues for the great time at the chair of mathematical numerics / control theory and beyond. You always had an open door for any kinds of questions and we had a lot of fruitful discussions. Thanks also to Prof. Kühn from the Technical University of Munich for the good cooperation and to Dr. Joachim Rehberg from the Weierstrass Institute in Berlin for the helpful discussions.

I had the great opportunity to be a member of the IGDK. The generous support of the IGDK gave me the possibility to travel to workshops and conferences. This has been fundamental for my personal development not only in science. I enjoyed my time at the annual summer workshops and the graduate seminars together with the other members of the IGDK.

Finally, I would like to thank my family. Without you this would not have been possible and it would not be worth it. Thank you.

Bibliography

[AF03] R. Adams and J. Fournier. *Sobolev spaces*. 2nd ed. Vol. 140. Academic press, 2003.

[Ama95] H. Amann. *Linear and Quasilinear Parabolic Problems: Volume I: Abstract Linear Theory*. Vol. 1. Springer Science & Business Media, 1995.

[Ama05] H. Amann. "Nonautonomous parabolic equations involving measures". In: *Journal of Mathematical Sciences* 130.4 (2005), pp. 4780–4802.

[Aus+14] P. Auscher et al. "The square root problem for second-order, divergence form operators with mixed boundary conditions on L^p". In: *Journal of Evolution Equations* **15** (2014), pp. 165–208.

[BP12] V. Barbu and T. Precupanu. *Convexity and Optimization in Banach Spaces*. Vol. 161. Springer Monographs in Mathematics. Springer Netherlands, 2012. ISBN: 978-94-007-2246-0.

[BJT10] W. Barthel, C. John, and F. Tröltzsch. "Optimal boundary control of a system of reaction diffusion equations". In: *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik* 90.12 (2010), pp. 966–982.

[BC85] J.F. Bonnans and E. Casas. "On the choice of the function spaces for some state-constrained control problems". In: *Numerical Functional Analysis and Optimization* 7.4 (1985), pp. 333–348. DOI: `10.1080/01630568508816197`. eprint: `http://dx.doi.org/10.1080/01630568508816197`. URL: `http://dx.doi.org/10.1080/01630568508816197`.

[BS00] J.F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. New York: Springer, 2000.

[Bro87] M. Brokate. *Optimale Steuerung von gewöhnlichen Differentialgleichungen mit Nichtlinearitäten vom Hysteresis-Typ*. Methoden und Verfahren der mathematischen Physik. P. Lang, 1987. ISBN: 9783820498950.

[Bro88] M. Brokate. "Optimal control of ODE systems with hysteresis nonlinearities". In: *Trends in Mathematical Optimization*. Springer, 1988, pp. 25–41.

[Bro91] M. Brokate. "Optimal control of systems described by ordinary differential equations with nonlinear characteristics of the hysteresis type." In: *Autom. Remote Control* 52 (1991), pp. 1639–1681.

[BK13] M. Brokate and P. Krejčì. "Optimal control of ODE systems involving a rate independent variational inequality". In: *Discrete Continuous Dynam. Systems* 18 (2013), pp. 331–348.

[BK15] M. Brokate and P. Krejčì. "Weak Differentiability of Scalar Hysteresis Operators". In: *Discrete Continuous Dynam. Systems* 35.6 (2015), pp. 2405–2421.

[BR05] M. Brokate and D. Rachinskii. "On global stability of the scalar Chaboche models". In: *Nonlinear Analysis: Real World Applications* 6.1 (2005), pp. 67–82.

[BS96] M. Brokate and J. Sprekels. *Hysteresis and phase transitions*. Vol. 121. Applied Mathematical Sciences. Springer Science & Business Media, 1996.

[CA01] C.M. Carracedo and M. Sanz Alix. *The Theory of Fractional Powers of Operators*. North-Holland Mathematics Studies. Elsevier, 2001.

[Cas97] E. Casas. “Pontryagin’s Principle for State-Constrained Boundary Control Problems of Semilinear Parabolic Equations”. In: *SIAM Journal on Control and Optimization* 35.4 (1997), pp. 1297–1327. DOI: 10.1137/S0363012995283637. eprint: http://dx.doi.org/10.1137/S0363012995283637. URL: http://dx.doi.org/10.1137/S0363012995283637.

[CMF14] C. Castaing, M. Monteiro Marques, and P. R. de Fitte. “Some problems in Optimal Control governed by the sweeping process”. In: *Journal of Nonlinear and Convex Analysis* 15.5 (2014), pp. 1043–1070.

[CC02] C. Cavaterra and F. Colombo. “Automatic control problems for reaction-diffusion systems”. In: *Journal of Evolution Equations* 2.2 (2002), pp. 241–273.

[Col+12] G. Colombo et al. “Optimal control of the sweeping process”. In: *Dynamics of Continuous, Discrete and Impulsive Systems Series B: Applications and Algorithms* 19.1-2 (2012), pp. 117–159. ISSN: 1492-8760.

[Col+16] G. Colombo et al. “Optimal control of the sweeping process over polyhedral controlled sets”. In: *Journal of Differential Equations* 260.4 (2016), pp. 3397–3447.

[CGT16] M. Curran, P. Gurevich, and S. Tikhomirov. “Recent advances in reaction-diffusion equations with non-ideal relays”. In: *Control of Self-Organizing Nonlinear Systems*. Springer, 2016, pp. 211–234.

[DER15] K. Disser, A.F.M. ter Elst, and J. Rehberg. “Hölder estimates for parabolic operators on domains with rough boundary”. In: *eprint arXiv:1503.07035* (2015).

[DN11] G. Dudziuk and M. Niezgódka. “Closed-loop control of a reaction-diffusion system”. In: *eprint arXiv:1109.4530* (2011).

[EL14] M. Eleuteri and L. Lussardi. “Thermal control of a rate-independent model for permanent inelastic effects in shape memory materials”. In: *Evolution Equations and Control Theory* 3.3 (2014), pp. 411–427. ISSN: 2163-2480. DOI: 10.3934/eect.2014.3.411.

[ELS13] M. Eleuteri, L. Lussardi, and U. Stefanelli. “Thermal control of the Souza-Auricchio model for shape memory alloys”. In: *Discrete Cont. Dynam. Syst., Ser. S* 6 (2013), pp. 369–386.

[ER14] A.F.M. ter Elst and J. Rehberg. “Hölder estimates for second-order operators with mixed boundary conditions”. In: (2014).

[Els11] J. Elstrodt. *Maß-und Integrationstheorie*. Springer-Verlag, 2011.

[Eva10] L.C. Evans. *Partial Differential Equations*. 2nd ed. American Mathematical Society, 2010.

[Gri+02] J.A. Griepentrog et al. “Interpolation for function spaces related to mixed boundary value problems”. In: *Mathematische Nachrichten* 241.1 (2002), pp. 110–120.

[Gri03] R. Griesse. “Parametric Sensitivity Analysis for Control-Constrained Optimal Control Problems Governed by Systems of Parabolic Partial Differential Equations”. University of Bayreuth, 2003.

[GV06] R. Griesse and S. Volkwein. “Parametric sensitivity analysis for optimal boundary control of a 3D reaction-diffusion system”. In: *Large-scale nonlinear optimization*. Springer, 2006, pp. 127–149.

[GR89] K. Gröger and J. Rehberg. “Resolvent estimates in $\mathrm{W}^{-1,p}$ for second order elliptic differential operators in case of mixed boundary conditions”. In: *Mathematische Annalen* 285.1 (1989), pp. 105–113.

[GST13] P. Gurevich, R. Shamin, and S. Tikhomirov. “Reaction-diffusion equations with spatially distributed hysteresis”. In: *SIAM Journal on Mathematical Analysis* 45.3 (2013), pp. 1328–1355.

[Haa06] M. Haase. *The Functional Calculus for Sectorial Operators*. Vol. 169. Springer Science & Business Media, 2006.

[Hal+15] R. Haller-Dintelmann et al. “Elliptic and parabolic regularity for second-order divergence operators with mixed boundary conditions”. In: *Mathematical Methods in the Applied Sciences* (2015). DOI: 10.1002/mma.3484.

[Hen81] D. Henry. *Geometric theory of semilinear parabolic equations*. Lecture Notes in Mathematics. Berlin: Springer, 1981.

[HMS15] R. Herzog, C. Meyer, and A. Schäfer. “Existence of solutions of a (nonsmooth) thermoviscoplastic model and associated optimal control problems”. In: *Ergebnisberichte des Instituts für Angewandte Mathematik 521*. TU Dortmund, Faculty for Mathematics, 2015.

[HMW12] R. Herzog, C. Meyer, and G. Wachsmuth. “C-stationarity for optimal control of static plasticity with linear kinematic hardening”. In: *SIAM Journal on Control and Optimization* 50.5 (2012), pp. 3052–3082.

[HMW13] R. Herzog, C. Meyer, and G. Wachsmuth. “B-and strong stationarity for optimal control of static plasticity with hardening”. In: *SIAM Journal on Optimization* 23.1 (2013), pp. 321–352.

[HMW14] R. Herzog, C. Meyer, and G. Wachsmuth. “Optimal Control of Elastoplastic Processes: Analysis, Algorithms, Numerical Analysis and Applications”. In: *Trends in PDE Constrained Optimization*. Ed. by G. Leugering et al. Springer International Publishing, 2014, pp. 27–41. ISBN: 978-3-319-05083-6. DOI: 10.1007/978-3-319-05083-6_4. URL: http://dx.doi.org/10.1007/978-3-319-05083-6_4.

[HKR13] D. Hömberg, K. Krumbiegel, and J. Rehberg. “Optimal control of a parabolic equation with dynamic boundary condition”. In: *Applied Mathematics & Optimization* 67.1 (2013), pp. 3–31.

[HJ80] F.C. Hoppensteadt and W. Jäger. “Pattern formation by bacteria”. In: *Biological growth and spread*. Springer, 1980, pp. 68–81.

[Kat80] T. Kato. *Perturbation Theory for Linear Operators*. 2nd ed. Berlin: Springer, 1980.

[Kop06] J. Kopfová. “Hysteresis in biological models”. In: *Journal of Physics: Conference Series*. Vol. 55. 1. IOP Publishing. 2006, p. 130.

[KP99] S.G. Krantz and H.R. Parks. *The Geometry of Domains in Space*. Basler Lehrbücher. Birkhäuser Boston, 1999. ISBN: 9780817640972.

[KP12] M. A. Krasnosel’skii and A. V. Pokrovskii. *Systems with hysteresis*. Springer Science & Business Media, 2012.

[KM17] C. Kuehn and C. Münch. “Generalized Play Hysteresis Operators in Limits of Fast-Slow Systems”. In: *SIAM Journal on Applied Dynamical Systems* 16.3 (2017), pp. 1650–1685. DOI: 10.1137/17M1110584. URL: https://doi.org/10.1137/17M1110584.

[Kuh03] K. Kuhnen. “Modeling, identification and compensation of complex hysteretic nonlinearities: A modified Prandtl-Ishlinskii approach”. In: *European journal of control* 9.4 (2003), pp. 407–418.

[Lun95] A. Lunardi. *Analytic Semigroups and Optimal Regularity in Parabolic Problems*. Vol. 16. Progress in Nonlinear Differential Equations and Their Applications. Springer Science and Business Media, 1995.

[May03] I. D. Mayergoyz. *Mathematical Models of Hysteresis and their Applications*. Elsevier, 2003.

[MS15] C. Meyer and L. Susu. *Optimal Control of Nonsmooth, Semilinear Parabolic Equations*. Tech. rep. Ergebnisberichte des Instituts für Angewandte Mathematik, Nummer 524. Fakultät für Mathematik, TU Dortmund, Sept. 2015.

[Mie05] A. Mielke. "Chapter 6 - Evolution of Rate-Independent Systems". In: *Handbook of Differential Equations Evolutionary Equations*. Ed. by C.M. Dafermos and E. Feireisl. Vol. 2. Handbook of Differential Equations: Evolutionary Equations. North-Holland, 2005, pp. 461–559. DOI: `https://doi.org/10.1016/S1874-5717(06)80009-5`. URL: `http://www.sciencedirect.com/science/article/pii/S1874571706800095`.

[MR15] A. Mielke and T. Roubíček. *Rate-independent systems*. Springer, 2015.

[Mün17a] C. Münch. "Global existence and Hadamard differentiability of hysteresis reaction–diffusion systems". In: *Journal of Evolution Equations* (2017). DOI: `10.1007/s00028-017-0419-x`. URL: `https://doi.org/10.1007/s00028-017-0419-x`.

[Mün17b] C. Münch. "Optimal control of reaction-diffusion systems with hysteresis". In: *eprint arXiv:1705.11031* (2017).

[Nec12] J. Necas. *Direct Methods in the Theory of Elliptic Equations*. Springer Monographs in Mathematics. Berlin: Springer, 2012.

[Paz83] A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Applied Mathematical Sciences. New York: Springer, 1983.

[Pim+12] A. Pimenov et al. "Memory effects in population dynamics: spread of infectious disease as a case study". In: *Mathematical Modelling of Natural Phenomena* 7.3 (2012), pp. 204–226.

[RZ98] J. P. Raymond and H. Zidani. "Pontryagin's Principle for State-Constrained Control Problems Governed by Parabolic Equations with Unbounded Controls". In: *SIAM Journal on Control and Optimization* 36.6 (1998), pp. 1853–1879. DOI: `10.1137/S0363012996302470`. eprint: `http://dx.doi.org/10.1137/S0363012996302470`. URL: `http://dx.doi.org/10.1137/S0363012996302470`.

[Rin08] F. Rindler. "Optimal control for nonconvex rate-independent evolution processes". In: *SIAM Journal on Control and Optimization* 47.6 (2008), pp. 2773–2794.

[Rin09] F. Rindler. "Approximation of rate-independent optimal control problems". In: *SIAM Journal on Numerical Analysis* 47.5 (2009), pp. 3884–3909.

[SWW16] U. Stefanelli, G. Wachsmuth, and D. Wachsmuth. "Optimal control of a rate-independent evolution equation via viscous regularization". In: *arXiv:1607.00809* (2016).

[TR12] A.F.M. Ter Elst and J. Rehberg. "L^∞-estimates for divergence operators on bad domains". In: *Analysis and Applications* 10.02 (2012), pp. 207–214. DOI: `10.1142/S0219530512500091`.

[Trö10] F. Tröltzsch. *Optimal Control of Partial Differential Equations: Theory, Methods, and Applications*. Graduate studies in mathematics. American Mathematical Society, 2010. ISBN: 9780821849040.

[Vis13] A. Visintin. *Differential models of hysteresis*. Vol. 111. Springer Science & Business Media, 2013.

[W05] Dirk W. *Funktionalanalysis*. 5th ed. Berlin: Springer-Verlag, 2005.

[Wac12] G. Wachsmuth. "Optimal Control of Quasi-static Plasticity with Linear Kinematic Hardening, Part I: Existence and Discretization in Time". In: *SIAM Journal on Control and Optimization* 50.5 (2012), pp. 2836–2861.

[Wac15] G. Wachsmuth. "Optimal Control of Quasistatic Plasticity with Linear Kinematic Hardening II: Regularization and Differentiability". In: *Zeitschrift für Analysis und ihre Anwendungen* 34.4 (2015), pp. 391–418.

[Wac16] G. Wachsmuth. "Optimal control of quasistatic plasticity with linear kinematic hardening III: Optimality conditions". In: *Zeitschrift für Analysis und ihre Anwendungen* 35.1 (2016), pp. 81–118.

[Yag09] A. Yagi. *Abstract parabolic evolution equations and their applications*. Springer Science & Business Media, 2009.