ESAIM: COCV 24 (2018) 1453–1488
https://doi.org/10.1051/cocv/2018025

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

OPTIMAL CONTROL OF REACTION-DIFFUSION SYSTEMS WITH HYSTERESIS[☆]

Christian Münch[*]

Abstract. This paper is concerned with the optimal control of hysteresis-reaction-diffusion systems. We study a control problem with two sorts of controls, namely 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. We choose a variational inequality to represent the hysteresis. In this paper, we prove first order necessary optimality conditions. In particular, 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 and show uniqueness of the adjoint variables. We employ the optimality system to prove higher regularity of the optimal solutions of our problem. The specific feature of rate-independent hysteresis in the state equation leads to difficulties concerning the analysis of the solution operator. Non-locality in time of the Hadamard derivative of the control-to-state operator complicates the derivation of an adjoint system. This work is motivated by its academic challenge, as well as by its possible potential for applications such as in economic modeling.

Mathematics Subject Classification. 49J20, 47J40, 35K51

Received June 8, 2017. Accepted April 6, 2018.

1. Introduction

In this paper, we derive an adjoint system for the optimal control problem

$$\min_{u\in U_i} J(y,u) := \frac{1}{2}\|y-y_d\|^2_{U_1} + \frac{\kappa}{2}\|u\|^2_{U_i} \tag{1.1}$$

[☆] The author is supported by the DFG through the International Research Training Group IGDK 1754 "Optimization and Numerical Analysis for Partial Differential Equations with Nonsmooth Structures". The author would like to thank Prof. Brokate from the Technical University of Munich and Prof. Fellner from the Karl-Franzens University of Graz for thoroughly proofreading the manuscript.

Keywords and phrases: Optimal control, reaction-diffusion, semilinear parabolic evolution problem, hysteresis operator, stop operator, global existence, solution operator, Hadamard differentiability, optimality conditions, adjoint system.

Department of Mathematics – M6, Technical University of Munich, Boltzmannstr. 3, 85747 Garching, Germany.

[*] Corresponding author: christian.muench@ma.tum.de

subject to

$$\dot{y}(t) + (A_p y)(t) = f(y(t), z(t)) + (B_i u)(t) \text{ in } \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega) \text{ for } t \in (0,T), \tag{1.2}$$

$$y(0) = 0 \qquad \text{in } \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega),$$

$$z = \mathcal{W}[Sy], \tag{1.3}$$

where $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ is a product of dual spaces, see *e.g.* (18)–(20) from [32] for the existence theory of problem (1.1)–(1.3) and related references therein. We consider either spatially distributed controls in the space $U_1 := \mathrm{L}^2\left((0,T);[\mathrm{L}^2(\Omega)]^m\right)$, or controls which act on given Neumann boundary parts Γ_{N_j}, $j \in \{1,\dots,m\}$, of the state space, *i.e.* controls in $U_2 := \mathrm{L}^2\left((0,T);\prod_{j=1}^m \mathrm{L}^2(\Gamma_{N_j},\mathcal{H}_{d-1})\right)$. The operator $B_1 : [\mathrm{L}^2(\Omega)]^m \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ and the operator $B_2 : \prod_{j=1}^m \mathrm{L}^2(\Gamma_{N_j},\mathcal{H}_{d-1}) \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ are continuous and A_p is an unbounded diffusion operator on the space $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. With $i \in \{1,2\}$, we identify B_i with the corresponding continuous operators from U_i into $\mathrm{L}^2\left((0,T);\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)$ which act pointwise in time, *i.e.* we write $(B_i u)(t) = B_i(u(t))$ for all $t \in (0,T)$. In the same way we identify $(A_p y)(t)$ with $A_p(y(t))$ for functions $y : (0,T) \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. Moreover, S projects y to a scalar valued function. In particular, $\mathcal{W}$ is a scalar stop operator and it is well-known (see *e.g.* [9, 40]) that $\mathcal{W}$ is represented by the solution operator $z = \mathcal{W}[v]$ of the variational inequality

$$(\dot{z}(t) - \dot{v}(t))(z(t) - \xi) \leq 0 \text{ for } \xi \in [a,b] \text{ and a.e. } t \in (0,T), \tag{1.4}$$

$$z(t) \in [a,b] \text{ for } t \in [0,T],\ z(0) = z_0. \tag{1.5}$$

For $i \in \{1,2\}$, we denote by G the operator, which maps $B_i u$ to the unique solution y of (1.2) and (1.3), see Theorem 3.1 of [32]. Note that $y = G(B_i u)$ is a function of time with values in a product of dual spaces.

The first motivation for this work is of academic nature. Non-smoothness of the state equation due to the coupling between the PDE (1.2) and the hysteresis operator (1.3) leads to several technical difficulties in the analysis of first order optimality conditions for problem (1.1)–(1.3). In particular, the directional derivatives of the control-to-state operator are non-linear. Secondly, problem (1.1)–(1.3) has potential for applications in fields such as economic modeling, see also [2] for a time discrete model including the stop operator. Consider for example the following scenario: Let $y(x,t)$ denote the density of money in a domain Ω, say a country. Moreover, let $z(t) = \mathcal{W}[Sy](t) \in [0,b]$ represent some measure for the overall savings of the whole nation at time t. The bound $b > 0$ determines some maximal limit for the savings. Not more money than b can be saved, or the nation is not willing to save more than b since it feels safe and satisfied when $z(t) = b$. The reaction term $f(y,z)$ in (1.2) can incorporate a non-linear coupling between the density of money y, or the overall money Sy, and the savings z. In particular, $f(y,z)$ can model the propensity to consume in dependence of the current amount of savings. At the same time, diffusion models the money distribution over the domain due to transactions. The control u can represent some control by the state over the monetary in- and outflow (within the country Ω or on a part of its boundary), in order to steer the density of money to a desired density, modeled by the tracking term y_d. In reality, such a control can for example be established by the rise or decrease of taxes. One possible interpretation of the state equation is the following: If the overall money Sy is increasing due to transactions with other countries while there is still potential for savings, *i.e.* $z \in [0,b)$, then a part of the income is reserved until eventually the overall savings of the country reach the critical value b. Any overflow of income is then directly spent in an excessively consuming manner. If in turn Sy is decreasing and $z \in (0,b]$, then a part of the savings z are invested in order to maintain the current living standard until $z = 0$. In this context, the effect of hysteresis is the result of an almost infinitely large time scale difference between the evolution of the money density y and the overall savings of money z. Compared to the time scale of transactions, the saving of money in a bank account happens almost instantaneously.

Note that the general problem (1.1)–(1.3) does not originate from the model which was sketched above. The draft of an economic scenario should rather point out one possible field in which problem (1.1)–(1.3) can find its application. The very general formulation of problem (1.1)–(1.3), as well as the generality in the list of assumptions to be introduced in the course of this work, are interesting from a mathematical point of view. But this generality also induces a broader variety of possible applications.

Optimal control of (systems of) partial differential equations has extensively been analyzed in the literature before.

In particular, optimal control problems with state equations of *semilinear parabolic* type are part of the well-known monograph [38] and the early work [5]. Further studies in this direction are the subject of [12, 34]. We also refer to [30] for a control problem with parabolic state equation and rough boundary conditions like in our setting.

Early studies in the field of optimal control of *reaction-diffusion systems* and in particular in the direction of parameter sensitivity analysis have been performed in [22] and were further established in [23] and several more papers. Optimality conditions for a similar problem were also derived in [4].

The *non-linearities* in all the works mentioned so far are mostly *smooth* enough to obtain a (twice) continuously differentiable control-to-state operator, so that first and many times also second order optimality conditions could be derived.

In the literature, there are only few results available concerning optimal control of *infinite-dimensional rate-independent processes*. For a class of energetically driven processes, existence of optimal controls for problems of this type has first been studied in [35, 36]. Subsequently, the results were applied to (thermal) control problems in the field of shape memory materials in [18, 19]. No optimality conditions are given in these works. Optimal control of a problem of static plasticity in the infinite-dimensional setting is the subject of [26, 27]. The results were used in [28] to numerically solve a quasi-static control problem by time-discretization. Optimality conditions for time-continuous, infinite-dimensional, rate-independent control problems of quasi-static plasticity type could be derived in [41–43] by means of time-discretization. Another time-continuous, infinite-dimensional optimal control problem of a rate-independent system, which is represented by its energetic formulation, is addressed in [39]. With help of viscous regularization, a necessary optimality condition is derived.

To our knowledge, the first results for optimal control of *hysteresis* have been achieved in [6–8]. Necessary optimality conditions for the optimal control of an ODE-system with hysteresis were established. An adjoint system was derived by a time discretization approach. Optimal control of sweeping processes has been studied in [13, 15, 20]. In [9], first order optimality conditions for a control problem of an ODE-system with hysteresis of (vectorial) stop type were derived. The stop operator is represented in form of a variational inequality. The main challenge with the stop operator (as with all hysteresis operators) is the fact that hysteresis acts non-local in time so that the state $y(t)$ at each time $t \in (0,T]$ depends on the whole background $[0,t)$. Moreover, the stop operator is not differentiable in the classical sense and so the control-to-state can not be expected to be so either. Regularization techniques were used in order to derive an optimality system. Several of the ideas of this approach are useful also for us. To handle a reaction-diffusion system requires additional work though. Firstly, the state vector $y : [0,T] \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ in (1.2) is a function with values in an infinite-dimensional space and secondly, the non-linearity f in our case is not necessarily continuously differentiable but only locally Lipschitz continuous and directionally differentiable. Therefore, techniques as in [31] are required. Particularly, since the domain Ω has a rough boundary, we have to consider a product of dual spaces for the domain of the diffusion operator A_p.

The existing literature provides only few rigorous results in the field of control of *hysteresis-reaction-diffusion systems*, especially when it comes to optimal control of such systems. In [14], automatic control problems governed by reaction-diffusion systems with feedback control of relay switch and Preisach type have been studied. Global existence and uniqueness of solutions were proven. Closed-loop control of a reaction-diffusion system coupled with ordinary differential inclusions has been considered in [17]. A feedback law for the case with a finite number of control devices was derived.

Necessary conditions for the optimal control of (general) *non-smooth semilinear parabolic equations* have been established in [31]. In particular, the non-linearity is merely locally Lipschitz continuous and directionally differentiable so that the control-to-state operator is not differentiable in the classical sense. Regularization techniques have been used to derive an adjoint system. No hysteresis is considered in this paper. Nevertheless, a modification of the approach in [31] is applicable for the problem at hand. In particular, we include ideas from [9] and adapt the proof to apply to non-localities in time such as hysteresis. We refer to the references in [31] for a good overview over further contributions dealing with optimal control of non-smooth parabolic equations.

In this paper, we are interested in the optimal control of *non-smooth reaction-diffusion systems with hysteresis*. In particular, a scalar stop operator enters the non-linearity f. The function f is assumed to be locally Lipschitz continuous and directionally differentiable. Additionally, the domain Ω satisfies minimal smoothness assumptions.

The outline of the paper is as follows:

In Section 2, we introduce the framework for the rest of the work and collect results from the literature. Section 2.3 contains the main assumption and notation.

Our first main interest is to derive an adjoint system and first order necessary optimality conditions for problem (1.1)–(1.3).

In Section 3, we introduce a family of regularized control problems with ε-dependent state equations and derive adjoint systems as well as optimality conditions for those. In particular, we regularize f and the stop operator $\mathcal{W}$ in dependence of the parameter $\varepsilon > 0$ and replace the original control problem by a regularized one. The corresponding control-to-state operator $u \mapsto G_\varepsilon(B_i u)$, $i \in \{1, 2\}$, and the regularization $y \mapsto Z_\varepsilon(Sy)$ of $\mathcal{W}[S\cdot]$ are Gâteaux-differentiable and we obtain optimal solutions $\overline{u}_\varepsilon$, $\overline{y}_\varepsilon = G_\varepsilon(B_i\overline{u}_\varepsilon)$ and $\overline{z}_\varepsilon = Z_\varepsilon(S\overline{y}_\varepsilon)$ of the regularized problems. We investigate in the limit $\varepsilon \to 0$ and use standard arguments to derive a solution $(\overline{u}, \overline{y}, \overline{z})$ of the original problem. It still remains difficult to derive adjoint systems $(p_\varepsilon, q_\varepsilon)$ already for the regularized problems. The main result of Section 3 is Theorem 3.13 which contains the evolution equations of p_ε and q_ε and the adjoint equation which provides a relation between $(p_\varepsilon, q_\varepsilon)$ and $\overline{u}_\varepsilon$ and $\overline{u}$.

In Section 4, we perform the key step towards an optimality system of (1.1)–(1.3) by driving the regularization parameter to zero. We exploit the adjoint systems $(p_\varepsilon, q_\varepsilon)$ to derive necessary optimality conditions for problem (1.1)–(1.3). While the evolution equation for p follows rather straight forward, the adjoint variable q which belongs to $\overline{z}$ has lower regularity, similar as in optimal control problems with implicit state constraints of the form of variational inequalities. The function q is contained in the space $\mathrm{BV}(0, T)$ of functions with bounded total variation in $[0, T]$, and instead of a time derivative we obtain a measure $\mathrm{d}q \in \mathrm{C}([0, T])^*$. In order to complete our knowledge about the optimality system, we investigate in studying q and $\mathrm{d}q$. Indeed, we reveal a lot of the properties of q and the corresponding measure $\mathrm{d}q$. There remains an abstract measure $\mathrm{d}\mu \in \mathrm{C}([0, T])^*$ on which $\mathrm{d}q$ depends and which we cannot fully characterize. Moreover, $\mathrm{d}\mu$ appears in the optimality conditions for problem (1.1)–(1.3). Still, we are able to prove that $\mathrm{d}\mu$ has its support only in a part of $[0, T]$. With an additional regularity assumption on $S\overline{y}$, we can characterize the measure $\mathrm{d}\mu$ also in most of the parts where it does not vanish. The first main results of Section 4 are Theorem 4.12 and Corollary 4.13, which contain the existence of an adjoint system and optimality conditions for problem (1.1)–(1.3) for $i \in \{1, 2\}$. After having established the optimality system for the general problem (1.1)–(1.3), $i \in \{1, 2\}$, we continue to improve the optimality conditions for the particular case of distributed control functions, *i.e.* for $i = 1$, see Corollary 4.14. Moreover, in Corollary 4.15, we show uniqueness of p, q and $\mathrm{d}\mu$ for $i = 1$. In the proof we make explicit use of fact that B_1 has dense range which implies that the operator B_1^* in the adjoint equation is one-to-one. These together are the second main result of Section 4.

In Section 5, we prove higher regularity of the optimal control $\overline{u}$ and the optimal state $\overline{y}$ by means of the adjoint equation and the continuity properties of the adjoint variables, see Theorem 5.2. An example for a case in which Theorem 5.2 can be applied is given in Remark 5.3.

All our results are applicable for more general spaces of control functions $U = \mathrm{L}^2\big((0, T); \tilde{U}\big)$, as long as there exists a continuous operator $B : \tilde{U} \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. Also $J(y, u)$ can be exchanged by a general differentiable functional $J(y, u, z)$ if the corresponding reduced cost function remains coercive in $u \in U$. Moreover, A_p can be replaced by a semi-linear parabolic operator which satisfies maximal parabolic regularity on the space $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. We focus on the two particular control problems for U_1 and U_2 and on the operator A_p in order to give an illustration.

Notation:
We write $\mathcal{L}(X,Y)$ for the space of bounded linear operators between spaces X and Y and $\mathcal{L}(X)$ for the space of bounded linear operators on X. We also abbreviate the duality in X by $\langle x,y\rangle_{X^*,X} = \langle x,y\rangle_X$. $c>0$ denotes a generic constant which is adapted in the course of the paper. In Banach space valued evolution equations like (1.2) we sometimes omit the range space if the latter is clear from the context, *i.e.* we only write "for $t\in(0,T)$".

2. Preliminaries and assumptions

We introduce the setting for the rest of the work, collect results from the literature and state the main assumption.

2.1. Sobolev spaces including homogeneous Dirichlet boundary conditions

Definition 2.1 ([32], Def. 2.1, I-sets)**.** 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 $r\in(0,1)$.

Assumption 2.2 (Domain). ([24], Assumptions 2.3 and 4.11 or [32], Assumptions 2.2 and 2.6)**.** For some given $d\ge 2$, the domain $\Omega\subset\mathbb{R}^d$ is bounded and $\overline{\Omega}$ is a d-set. For $j\in\{1,\ldots,m\}$, the Neumann boundary part $\Gamma_{N_j}\subset\partial\Omega$ is relatively open and $\Gamma_{D_j}=\partial\Omega\backslash\Gamma_{N_j}$ is a $(d-1)$-set. For any $x\in\overline{\Gamma_{N_j}}$ 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.

We only consider real valued functions. For each component $j\in\{1,\ldots,m\}$ of the space of vector valued functions, see Definition 2.4, we decompose the boundary $\partial\Omega$ into the corresponding Dirichlet part Γ_{D_j} 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.

Remark 2.3. Assumption 2.2 allows for very general domains. The framework includes Lipschitz domains Ω with $(d-1)$-dimensional manifolds Γ_{D_j} for $j\in\{1,\ldots,m\}$, *cf.* Remark 2.3 from [32]. But also more irregular 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'.*" ([16], Sect. 1. Introduction).

We define Sobolev spaces which include the Dirichlet boundary conditions of the state equation.

Definition 2.4 (Sobolev spaces). ([24], Def. 2.4 or [32], Def. 2.4)**.** For Ω from Assumption 2.2 and $p\in[1,\infty)$ we denote by $\mathrm{W}^{1,p}(\Omega)$ the usual Sobolev space on Ω. If M is a closed subset of $\overline{\Omega}$ we define

$$\mathrm{W}^{1,p}_{\mathrm{M}}(\Omega) := \overline{\{\psi|_\Omega : \psi\in \mathrm{C}_0^\infty(\mathbb{R}^d),\, \mathrm{supp}(\psi)\cap \mathrm{M}=\emptyset\}},$$

where the closure is taken in the space $\mathrm{W}^{1,p}(\Omega)$. In the case $p\in(1,\infty)$ we denote by p' the Hölder conjugate of p. Moreover, we write

$$\mathrm{W}^{-1,p}_{\mathrm{M}}(\Omega) := \left[\mathrm{W}^{1,p'}_{\mathrm{M}}(\Omega)\right]^*$$

for the dual space $\mathrm{W}^{1,p'}_{\mathrm{M}}(\Omega)$. In the vectorial setting we introduce the product space

$$\mathbb{W}^{1,p}_{\Gamma_D}(\Omega) := \prod_{j=1}^{m} \mathrm{W}^{1,p}_{\Gamma_{D_j}}(\Omega)$$

and for $p \in (1,\infty)$ we denote by $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ the (componentwise) dual space of $\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$.

2.2. Operators and their properties

In this section, we define the operators A_p in equation (1.2), see Definition 2.5. We apply results from the literature to assure that A_p satisfies the properties which we need for the analysis of (1.2) and (1.3) for particular values of p to be chosen, see Section 6 in [24] or Section 2.2 in [32].

Definition 2.5 (Diffusion operator)**.** For $p \in (1,\infty)$ we define the continuous operators

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

and

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

With given diffusion coefficients $d_1, \dots, d_m > 0$ we define the corresponding diffusion matrix in $\mathbb{R}^{md\times md}$ by $D = \mathrm{diag}(d_1, \dots, d_1, \dots, d_m, \dots, d_m)$.

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

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

and define the unbounded operator

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

The set $\mathrm{ran}(I_p)$ denotes the range of I_p. The domain $\mathrm{dom}(A_p)$ is equipped with the graph norm.

We introduce the notion of maximal parabolic regularity as in Definition 2.12 of [32].

Definition 2.6 (Maximal parabolic regularity)**.** For $p, q \in (1,\infty)$ and $(t_0, T) \subset \mathbb{R}$ we say that A_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 $y \in \mathrm{W}^{1,q}((t_0,T); \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)) \cap \mathrm{L}^q((t_0,T); \mathrm{dom}(A_p))$ of the equation

$$\dot{y} + A_p y = g, \;\; y(t_0) = 0.$$

The time derivative is taken in the sense of distributions ([3], Def. 11.2).

For $t \in [0,T]$ we abbreviate
$Y_q := \mathrm{W}^{1,q}((0,T); \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)) \cap \mathrm{L}^q((0,T); \mathrm{dom}(A_p))$, $Y_{q,t} := \{y \in Y_q : \; y(t) = 0\}$ and
$Y^*_{q,t} := \{y \in \mathrm{W}^{1,q}(0,T; [\mathrm{dom}(A_p)]^*) \cap \mathrm{L}^q((0,T); \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)) : \; y(t) = 0\}$.

As in Remark 2.13 of [32] note the following

Remark 2.7 (Properties of A_p)**.**

(1) If Definition 2.6 applies for A_p with some $p \in (1,\infty)$ then the property of maximal parabolic regularity is independent of $q \in (1,\infty)$ and of the interval (t_0, T), so we just say that A_p satisfies maximal parabolic regularity on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ in this case.
(2) If A_p satisfies maximal parabolic regularity on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ for some $p \in (1,\infty)$ then the operator $(\frac{\mathrm{d}}{\mathrm{d}t} + A_p)^{-1}$ is bounded from $\mathrm{L}^q((0,T);\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega))$ to $Y_{q,0}$ for any $q \in (1,\infty)$.
(3) In the setting of Assumption 2.2 there exists an open interval J containing 2 such that for $p \in \mathrm{J}$ the operator $\mathcal{A}_p + I_p$ is a topological isomorphism and such that $-A_p$ generates an analytic semigroup of operators on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ ([32], Thm. 2.10 or [24], Thms. 5.6 and 5.12).
(4) If $p \in \mathrm{J}$ and if $\theta \geq 0$ is given then for $A_p + 1 := A_p + \mathrm{Id}$ the fractional power spaces $X^\theta := \mathrm{dom}([A_p+1]^\theta) \subset \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ and the unbounded operators $[A_p+1]^\theta$ in the sense of Chapter 1 in [25] are well-defined with $X^0 = \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. X^θ is equipped with the norm $\|x\|_{X^\theta} = \|(A_p+1)^\theta x\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)}$ (*cf.* [32], Rem. 2.11). Note that we can identify X^1 with the space $\mathrm{dom}(A_p)$ endowed with the graph norm.
(5) If $p \in \mathrm{J} \cap [2,\infty)$, then A_p satisfies maximal parabolic Sobolev regularity on $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ and we have the topological equivalences $[\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)]_\theta \simeq [\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), \mathrm{dom}(A_p)]_\theta \simeq X^\theta$ for all $\theta \in (0,1)$ ([11], Thm. 11.6.1). By $[\cdot,\cdot]_\theta$ we mean complex interpolation.

We will make use of the following embeddings:

Remark 2.8 (Embeddings). ([32], Rem. 2.14)**.** With $q \in (1,\infty)$ one has

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

for every $0 < \theta < \eta < 1 - 1/q$ and $0 \leq \beta < 1 - 1/q - \eta$. $(\cdot,\cdot)_{\eta,1}$ or $(\cdot,\cdot)_{\eta,q}$ respectively means real interpolation. The first embeddings are compact because $\mathrm{dom}(A_p)$ is compactly embedded into $\mathbb{W}^{-1,p}_{\Gamma_\mathrm{D}}(\Omega)$.

With $p \in \mathrm{J}$, the following estimate for the fractional powers of $A_p + 1$ and the analytic semigroup $\exp(-A_p t)$ is crucial:

Remark 2.9. ([32], Rem. 2.15). Let $p \in \mathrm{J}$ with J from Remark 2.7. For $t > 0$ and arbitrary $\gamma \in (0,1)$ and $\theta \geq 0$ there exists some $C_\theta \in (0,\infty)$ such that

$$\|(A_p+1)^\theta \exp(-A_p t)\|_{\mathcal{L}\left(\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)\right)} \leq C_\theta t^{-\theta} \exp((1-\gamma)t). \tag{2.1}$$

The stop operator has the following regularity properties.

Lemma 2.10 (Stop operator)**.** *With $T > 0$ the stop operator $\mathcal{W}$, which is represented by (1.4) and (1.5), is Lipschitz continuous as a mapping on $\mathrm{C}[0,T]$ with*

$$|\mathcal{W}[v_1](t) - \mathcal{W}[v_2](t)| \leq 2 \sup_{0\leq\tau\leq t} |v_1(\tau) - v_2(\tau)| \quad \textit{and} \quad |\mathcal{W}[v](t)| \leq 2 \sup_{0\leq\tau\leq t} |v(\tau)| + |z_0| \tag{2.2}$$

for all $v, v_1, v_2 \in \mathrm{C}[0,T]$ and $t \in [0,T]$. Note that we have to add $|z_0|$ in (2.2) because, by (1.5), $\mathcal{W}[v](0) = z_0$ for any $v \in \mathrm{C}[0,T]$. For $q \in [1,\infty)$, $\mathcal{W}$ is bounded and weakly continuous on $\mathrm{W}^{1,q}(0,T)$. $\mathcal{W} : \mathrm{C}[0,T] \to \mathrm{L}^q(0,T)$ is Hadamard directionally differentiable, see Definition 2.13. The same regularity properties hold for the operator $\mathcal{P} = \mathrm{Id} - \mathcal{W}$. $\mathcal{P}$ is a scalar play operator. More precisely, for $r = \frac{b-a}{2}$ let $\mathcal{P}_r : \mathrm{C}[0,T] \times \mathbb{R} \to \mathrm{C}[0,T]$ denote a *symmetrical scalar play operator (as in [10]). Consider the affine linear transformation* $\mathcal{T} : [-r, r] \to [a, b]$, $\mathcal{T} : x \mapsto x - \frac{b+a}{2}$. *Then for* $v \in \mathrm{C}[0, T]$ *there holds*

$$\mathcal{P}[v] = \mathcal{P}_r[\mathcal{T}(v), v(0) - z_0] \in \mathrm{C}[0, T].$$

Proof. Follows from Sections 2.4 and 4.2 in [32], see also Part 1, Chapter III from [40] and [10]. □

2.3. Assumptions and notation

Our main assumption is the following:

Assumption 2.11 (Main assumption). ([32], Assumptions 2.16, 4.6 and 5.1)**.** We always suppose that Assumption 2.2 holds. Moreover we assume:

(A1) Dimension and Sobolev exponent: $d \geq 2$ and with J from Remark 2.7 there holds $p \in \mathrm{J} \cap [2, \infty)$ and $2 \geq p\left(1 - \frac{1}{d}\right)$.

(A2) Nonlinearity locally Lipschitz + Hadamard: We will need a fractional power space $X^\alpha = \mathrm{dom}([A_p + 1]^\alpha)$ with exponent strictly smaller than one half. This fact is highlighted by a new parameter α which we use instead of $\theta \in [0, \infty)$. For some $\alpha \in (0, \frac{1}{2})$ suppose that the function $f : X^\alpha \times \mathbb{R} \to \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ is locally Lipschitz continuous with respect to the X^α-norm. This means that given any $y_0 \in X^\alpha$ there exists a constant $L(y_0) > 0$ and a neighbourhood $V(y_0) = \{y \in X^\alpha : \|y - y_0\|_{X^\alpha} \leq \delta \in (0, \infty)\}$ of y_0 such that

$$\|f(y_1, x_1) - f(y_2, x_2)\|_{\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)} \leq L(y_0)\left(\|y_1 - y_2\|_{X^\alpha} + |x_1 - x_2|\right)$$

for every $y_1, y_2 \in V(y_0)$ and all $x_1, x_2 \in \mathbb{R}$. f is assumed to be directionally differentiable and therefore Hadamard directionally differentiable, see Definition 2.13. Furthermore, the linear growth condition

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

holds for some $M > 0$.

(A3) Scalar projection: For some $w \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)\backslash\{0\}$ the operator $S \in [\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)]^*$ in equation (1.3) is given by $Sy = \langle y, w\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)}$ $\forall y \in \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$. We assume that w is even contained in the space $\mathrm{dom}([(1 + A_p)^{1-\alpha}]^*)$. Note that S belongs to $[X^\theta]^*$ for all $\theta \geq 0$ because of the embedding $X^\theta \hookrightarrow \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$.

(A4) Desired state: The desired state y_d in (1.1) is in $\mathrm{L}^2\left((0, T); [\mathrm{L}^2(\Omega)]^m\right)$.

Remark 2.12. (A1) includes the range of values for p for which the analysis in the rest of the work applies. (A2) extends the typical regularity assumption for non-linearities in semigroup theory [25, 33]. It implies unique existence of mild solutions of semi-linear parabolic equations with non-linearity f. We will show that Lipschitz continuity of the mapping $y \mapsto \mathcal{W}[Sy]$ from $\mathrm{C}([0, T]; X^\alpha)$ to $\mathrm{C}([0, T])$ together with (A2) yields local Lipschitz continuity and linear growth of the mapping $y \mapsto f(y, \mathcal{W}[Sy])$ from $\mathrm{C}([0, T]; X^\alpha)$ to $\mathrm{C}([0, T]; \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega))$. Directional differentiability of f is necessary to prove Hadamard directional differentiability of the control-to-state mapping for (1.1)–(1.3). In particular, the mapping $y \mapsto f(y, \mathcal{W}[Sy])$ is Hadamard directionally differentiable from $\mathrm{C}([0, T]; X^\alpha)$ to $\mathrm{L}^q([0, T]; \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega))$ for appropriate $q \in \left(\frac{1}{1-\alpha}, \infty\right)$. The projection in (A3) can be of very general kind. For $y \in \mathrm{ran}(I_p)$, Sy can be the approximate mean value of $I_p^{-1}y \in \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$. Note that $I_p^{-1}y$ can be identified with $y \in \mathrm{dom}(A_p)$.

We introduce some notation for the rest of the work:

(N1) For the particular p from (A1) in Assumption 2.11 we set $X := \mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ with $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ from Definition 2.4. We sometimes identify elements $v \in X^*$ with their unique representative in $\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$, *i.e.* we write $\langle v, y\rangle_X = \langle y, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)}$, $\forall y \in X$.

(N2) The operators A_p and the spaces $X^\theta = \operatorname{dom}([A_p+1]^\theta)$ are defined as in Definition 2.5 and Remark 2.7.

(N3) The spaces Y_q, $Y_{q,t}$ and $Y^*_{q,t}$ are defined as in Definition 2.6.

(N4) $\mathcal{W}$ is the scalar stop operator from Lemma 2.10.

(N5) B_1 is defined by $B_1 : [\mathrm{L}^2(\Omega)]^m \to X,\ \langle B_1 u, v\rangle_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} := \int_\Omega u \cdot v \,\mathrm{d}x,\ \forall v \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$.
Since $2 \geq p\left(1-\frac{1}{d}\right)$, the embeddings $\mathrm{L}^2(\Gamma_{N_j}, \mathcal{H}_{d-1}) \hookrightarrow \mathrm{W}^{-1,p}_{\Gamma_{D_j}}(\Omega)$ are continuous for $j \in \{1,\dots,m\}$ ([24], Rem. 5.11). Therefore also the operator
$B_2 : \prod_{j=1}^m \mathrm{L}^2(\Gamma_{N_j}, \mathcal{H}_{d-1}) \to X,\ \langle B_2 y, v\rangle_{\mathbb{W}^{1,p'}(\Omega)} := \sum_{j=1}^m \int_{\Gamma_{N_j}} y_j v_j \,\mathrm{d}\mathcal{H}_{d-1},\ \forall v \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$ is continuous.

(N6) We write $J_T = (0,T)$, $U_1 = \mathrm{L}^2\left(J_T; [\mathrm{L}^2(\Omega)]^m\right)$ and $U_2 = \mathrm{L}^2\left(J_T; \prod_{j=1}^m \mathrm{L}^2(\Gamma_{N_j}, \mathcal{H}_{d-1})\right)$.

2.4. Solution operator and optimal control

As in equation (1) of [32] we denote $F[y](t) := f(y(t), \mathcal{W}[Sy](t))$ and introduce the more general abstract evolution equation

$$\begin{aligned} \dot{y}(t) + (A_p y)(t) &= (F[y])(t) + u(t) \quad \text{in } X \quad \text{for } t > 0,\\ y(0) &= 0 \in X. \end{aligned} \tag{2.3}$$

Note that $F[y]$ is non-local in time. In order to obtain some kind of differentiability of the reduced cost function, the solution operator of the state equation has to be differentiable in a sense which allows for the chain rule. We can not expect a Fréchet derivative because of the non-smooth hysteresis operator, see [10]. But the chain rule can also be applied within the weaker concept of Hadamard directional differentiability.

Definition 2.13 (Hadamard directional differentiability)**.** Let X, Y be normed vector spaces and let $U \subset X$ be open. If $g : U \to Y$ is directionally differentiable at $x \in U$ and if in addition for all functions $r : [0,\lambda_0) \to X$ with $\lim\limits_{\lambda\to 0} \frac{r(\lambda)}{\lambda} = 0$ it holds $g'[x;h] = \lim\limits_{\lambda\to 0} \frac{g(x+\lambda h + r(\lambda)) - g(x)}{\lambda}$ for all directions $h \in X$, 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 $x + \lambda h + r(\lambda) \in U$. The chain rule applies for Hadamard directionally differentiable functions ([32], Lem. 4.3).

Hadamard directional differentiability of the solution operator G is shown in [32]. By Theorems 3.1 and 4.7 of [32] we have:

Theorem 2.14 (Solution operator for the state equation)**.** *Let Assumption 2.11 hold. Then for the fixed value $\alpha \in \left(0, \frac{1}{2}\right)$ and for all $u \in \mathrm{L}^q(J_T; X)$ with $q \in (\frac{1}{1-\alpha}, \infty]$ problem (2.3) has a unique mild solution $y = y(u) =: y^u$ in $\mathrm{C}(\overline{J_T}; X^\alpha)$. In particular, this means that $(F[y]) + u$ is contained in $\mathrm{L}^1(J_T; X)$ and that y solves the integral equation*

$$y(t) = \int_0^t \exp(-A_p(t-s))[(F[y])(s) + u(s)]\,\mathrm{d}s,\ t \in J_T.$$

The solution mapping $G : u \mapsto y(u)$, $\mathrm{L}^q(J_T; X) \to \mathrm{C}(\overline{J_T}; X^\alpha)$ is locally Lipschitz continuous. G is linearly bounded with values in $\mathrm{C}(\overline{J_T}; X^\alpha)$. All statements remain valid if $\mathrm{C}(\overline{J_T}; X^\alpha)$ is replaced by $Y_{s,0}$ where $s = q$ if $q < \infty$ and with $s \in (1,\infty)$ arbitrary if $q = \infty$. G is Hadamard directionally differentiable as a mapping into $\mathrm{C}(\overline{J_T}; X^\alpha)$ as well as into $Y_{q,0}$ for any $q \in (\frac{1}{1-\alpha}, \infty)$. Its derivative $y^{u,h} := G'[u;h]$ at $u \in \mathrm{L}^q(J_T; X)$ in direction

$h \in \mathrm{L}^q(J_T;X)$ is given by the unique solution ζ of

$$\dot{\zeta}(t) + (A_p\zeta)(t) = F'[y;\zeta](t) + h(t) \text{ for } t \in J_T,\ \zeta(0) = 0, \tag{2.4}$$

where $F'[y;\zeta](t) = f'[(y(t), \mathcal{W}[Sy](t)); (\zeta(t), \mathcal{W}'[Sy;S\zeta](t))]$ and $y = G(u)$. The mapping $h \mapsto G'[u;h]$ is Lipschitz continuous from $\mathrm{L}^q(J_T;X)$ to $\mathrm{C}(\overline{J_T};X^\alpha)$ and to $Y_{q,0}$ with a modulus $C = C(G(u),T) > 0$.

Proof. See Theorems 3.1 and 4.7 from [32]. □

Existence of an optimal control for problem (1.1)–(1.3) is shown in Theorem 5.4 from [32]:

Theorem 2.15 (Existence of optimal control)**.** *Let Assumption 2.11 hold. Then for $i \in \{1,2\}$, there exists an optimal control $\overline{u} \in U_i$ for the optimal control problem (1.1)–(1.3). This means that $\overline{u}$, together with the optimal state $\overline{y} = G(\overline{u})$, which solves (1.2), are a solution of the minimization problem (1.1). The solution of (1.3) is given by $\overline{z} = \mathcal{W}[S\overline{y}]$.*

Proof. See Theorem 5.4 from [32]. □

3. Regularized control problem

In order to derive an adjoint system for problem (1.1)–(1.3) we introduce a sequence of control problems with regularized ε-dependent state equations, for which we can derive adjoint systems. To this aim we regularize the variational inequality which defines $\mathcal{W}$ and the non-linearity f. This yields a regularization of the solution operator of (2.3). The regularization of $\mathcal{W}$ follows the techniques in Section 3 from [9] and the approach for the regularization of semilinear parabolic equations relies on Section 4 from [31].

In the end of Section 3.1, we estimate the norms of the solutions of the regularized state equations against the forcing term u, independently of ε.

The dynamics of the regularized state equations in dependence of ε is analyzed in Section 3.2. In particular, the estimates from Section 3.1 together with a weak compactness argument imply weak compactness of the regularized solution operators for fixed $\varepsilon > 0$. This result is used to derive weakly converging subsequences y_{ε_k} and z_{ε_k} for any weakly converging sequence u_ε, $\varepsilon \to 0$.

In Section 3.3, we apply the convergence result from Section 3.2 to deduce convergence of the solutions of regularized control problems, which are introduced in Section 3.3, to an optimal solution of problem (1.1)–(1.3) as $\varepsilon \to 0$, see Theorem 3.9.

In Section 3.4, we conclude Gâteaux differentiability of the regularized control-to-state operators from the results for problem (2.3).

The adjoint equations for the solutions of the regularized control problems with $\varepsilon > 0$ fixed are derived in Section 3.5, see Theorem 3.13.

In Section 3.6, we derive uniform-in-ε bounds for the norms of the adjoint variables $p_\varepsilon, q_\varepsilon$ from Theorem 3.13.

The norm bounds on $p_\varepsilon, q_\varepsilon$ from Section 3.6 give rise to weakly converging subsequences p_{ε_k} and q_{ε_k}. Taking the limit $k \to \infty$ then yields an adjoint system for (1.1)–(1.3). This step is carried out in Section 4.

Following the ideas of [9], we define a function Ψ in order to regularize the stop operator $\mathcal{W}$.

Definition 3.1. For $x \in \mathbb{R}$ we introduce the functions

$$\Psi_{-2}(x) := -16(x-1-a), \qquad \Psi_{-1}(x) := (x-a)^3(4-a+x)$$
$$\Psi_1(x) := (x-b)^3(4+b-x), \qquad \text{and} \qquad \Psi_2(x) := 16(x-1-b).$$

In particular, Ψ_{-2} and Ψ_2 are affine linear and Ψ_{-1} and Ψ_1 are polynomials of order four with roots in a respectively b which are at the same time saddle points, and with turning points in $a-2$ and $b+2$.

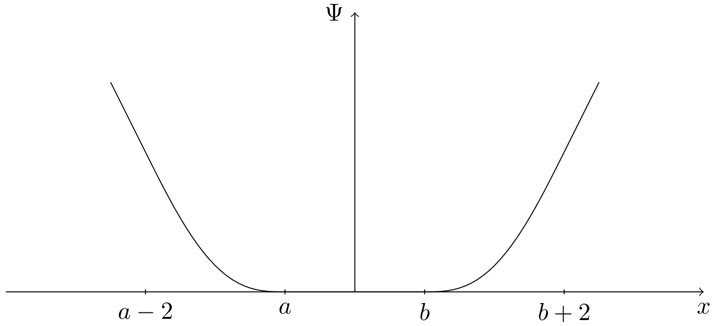

FIGURE 1. Graph of Ψ.

We define the function

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

where χ denotes the characteristic function, *cf.* Figure 1. Note that Ψ is convex with $\Psi(x) \equiv 0$ for $x \in [a,b]$ and $\Psi(x) > 0$ for $x \in \mathbb{R}\backslash[a,b]$. Moreover, Ψ is twice continuously differentiable and $\Psi'(x) \leq m_1|x-a|$ for some constant $m_1 > 0$ and for all $x \in \mathbb{R}$. There holds $\Psi''(x) \leq m_2$ for some constant $m_2 > 0$ and for all $x \in \mathbb{R}$ and Ψ'' is Lipschitz continuous.

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

Assumption 3.2 (Regularization)**.** 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 $\varepsilon \to 0$.

$(A3)_\varepsilon$ f_ε is locally Lipschitz continuous with respect to the X^α-norm and all the neighbourhoods and Lipschitz constants are equal to the ones of f in (A2) in Assumption 2.11, 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 (A2) in Assumption 2.11.

$(A4)_\varepsilon$ The function $\Psi : \mathbb{R} \to \mathbb{R}$ is defined according to Definition 3.1. In particular, Ψ is convex with $\Psi(x) \equiv 0$ for $x \in [a,b]$ and $\Psi(x) > 0$ for $x \in \mathbb{R}\backslash[a,b]$. Moreover, Ψ is twice continuously differentiable and $\Psi'(x) \leq m_1|x-a|$ for some constant $m_1 > 0$ and for all $x \in \mathbb{R}$. There holds $\Psi''(x) \leq m_2$ for some constant $m_2 > 0$ and for all $x \in \mathbb{R}$ and Ψ'' is Lipschitz continuous.

We introduce the following regularized state equations for $i \in \{1,2\}$ and $\varepsilon > 0$:

$$\dot{y}(t) + (A_p y)(t) = f_\varepsilon(y(t), z(t)) + (B_i u)(t) \text{ in } X \text{ for } t \in J_T,\ y(0) = 0 \text{ in } X, \tag{3.1}$$

$$\dot{z}(t) - S\dot{y}(t) = -\frac{1}{\varepsilon}\Psi'(z(t)) \qquad \text{for } t \in J_T,\ z(0) = z_0. \tag{3.2}$$

3.1. Regularization of (2.3) and uniform-in-ε estimates

In this section, we introduce a regularization of (2.3), similar to the regularized state equations (3.1) and (3.2) but for source terms $u \in \mathrm{L}^q(J_T; X)$. We show well-posedness and estimate the norms of the solutions independently of ε. The ideas for many of the steps in this section go back to Section 3.1 of [9].

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

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

or of the corresponding integral equation. The input v is a function defined on J_T.

Remark 3.4. By standard techniques it follows that Z_ε is Lipschitz continuous and continuously differentiable on $\mathrm{C}(\overline{J_T})$. The derivative $Z'_\varepsilon[v;h]$ at v in direction h is given by the unique solution of the integral equation $z(t) = h(t) - \int_0^t \frac{1}{\varepsilon}\Psi''(Z_\varepsilon(v)(s))z(s)\mathrm{d}s$. Moreover, Z_ε is bounded and weakly continuous on $\mathrm{W}^{1,q}(J_T)$ for all $q \in (1,\infty)$. In particular, Z_ε satisfies the properties of $\mathcal{W}$ in Lemma 2.10 (although with a different modulus of continuity than in (2.2)).

Similar to the definition of F in Section 2.4 we denote $(F_\varepsilon(y))(t) := f_\varepsilon(y(t), Z_\varepsilon(Sy)(t))$. Consider the abstract evolution equation

$$\begin{aligned} \dot{y}(t) + (A_p y)(t) &= (F_\varepsilon(y))(t) + u(t) \quad \text{in } X \quad \text{for } t > 0, \\ y(0) &= 0 \in X. \end{aligned} \tag{3.3}$$

Corollary 3.5 (Existence of regularized problem)**.** *Let Assumptions 2.11 and 3.2 hold and let $\varepsilon \in (0, \varepsilon_*]$ be arbitrary. Furthermore, assume $q \in (\frac{1}{1-\alpha}, \infty]$ and set $s = q$ if $q < \infty$ or $s \in (1,\infty)$ arbitrary if $q = \infty$. Then for all $u \in \mathrm{L}^q(J_T;X)$ problem (3.3) 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. Theorem 2.14 neither depends on the concrete choice of $\mathcal{W}$, nor on the exact modulus of continuity of $\mathcal{W}$, since all constants can be adapted, *cf.* Theorems 3.1 and 4.7 of [32]. In particular, $\mathcal{W}$ can be replaced by an operator with appropriate properties ([32], Rem. 4.5), which is Lipschitz continuous and Hadamard directionally differentiable on $\mathrm{C}(\overline{J_T})$ as well as bounded and weakly continuous on $\mathrm{W}^{1,q}(J_T)$ for all $q \in (1,\infty)$. Since all those properties apply for Z_ε (*cf.* Rem. 3.4), Theorem 2.14 holds with $\mathcal{W}$ replaced by Z_ε as required. Existence of unique local solutions of (3.3) is shown by a fixed point argument. Those are extended to $\overline{J_T}$ by linear growth of $F_\varepsilon : \mathrm{C}(\overline{J_T};X^\alpha) \to \mathrm{C}(\overline{J_T};X)$ and a Gronwall argument. Local Lipschitz continuity of G_ε is proved by similar techniques *via* local Lipschitz continuity of F_ε. ☐

In the next step we estimate the norms of the solutions of (3.3) independently of ε by the norm of the source function $u \in \mathrm{L}^q(J_T;X)$. This yields corresponding estimates also for the solutions of (3.1) and (3.2) if u is replaced by $B_i u$.

Lemma 3.6 (Uniform bounds)**.** *Adopt the assumptions and the notation from Corollary 3.5. 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*

$$\|y_\varepsilon^u\|_{Y_{s,0}} \le c(1 + \|u\|_{\mathrm{L}^q(J_T;X)}) \quad \text{and} \quad \|z_\varepsilon^u\|_{\mathrm{C}(\overline{J_T})} \le c(1 + \|u\|_{\mathrm{L}^q(J_T;X)}), \tag{3.4}$$

with $s = q$ if $q < \infty$ and for all $s \in (1,\infty)$ if $q = \infty$. Moreover, there holds

$$0 \le \int_0^T |\dot{z}_\varepsilon^u(s)|^2 \,\mathrm{d}s + \sup_{t \in \overline{J_T}} \frac{1}{\varepsilon}\Psi(z_\varepsilon^u(t)) \le c(1 + \|u\|_{\mathrm{L}^2(J_T;X)})^2. \tag{3.5}$$

Proof. Note first that for $v \in \mathrm{W}^{1,s}(J_T)$ and for $t \in J_T$ we have

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

Moreover, $\Psi'(x)(x - z_0) \ge 0$ for all $x \in \mathbb{R}$ because Ψ is convex and since $\Psi'(z_0) = 0$. We insert $Z_\varepsilon(v)(0) = z_0$ and $\frac{d}{\mathrm{d}s}(Z_\varepsilon(v)) = \dot{v} - \frac{1}{\varepsilon}\Psi'(Z_\varepsilon(v))$ according to Definition 3.3. The triangle inequality and rearranging yields

$$0 \le |Z_\varepsilon(v)(t)| + \frac{1}{\varepsilon}\int_0^t \frac{\Psi'(Z_\varepsilon(v))(Z_\varepsilon(v) - z_0)}{|Z_\varepsilon(v) - z_0|}\mathrm{d}s \le |z_0| + \int_0^t |\dot{v}(s)|\mathrm{d}s.$$

Hence, with $z_\varepsilon^u = Z_\varepsilon(Sy_\varepsilon^u)$ and $v = Sy_\varepsilon^u$ there follows

$$0 \le |z_\varepsilon^u(t)| \le |z_0| + \int_0^t |S\dot{y}_\varepsilon^u(s)| \mathrm{d}s. \tag{3.6}$$

Because the representative $w \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$ of S is contained in $\mathrm{dom}([(A_p+1)^{1-\alpha}]^*)$ by (A3) in Assumption 2.11, we can estimate for all $y \in \mathrm{dom}(A_p)$:

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

For $y = y_\varepsilon^u$, this together with (3.3) and the triangle inequality implies that for a.e. $t \in J_T$

$$|S\dot{y}_\varepsilon^u(t)| \le c_1\|y_\varepsilon^u(t)\|_{X^\alpha} + |Sf_\varepsilon(y_\varepsilon^u(t), z_\varepsilon^u(t))| + |Su(t)|.$$

Consequently, by the linear growth condition on f_ε in $(A3)_\varepsilon$ of Assumption 3.2 we further estimate (3.6) by

$$\begin{aligned}|z_\varepsilon^u(t)| &\le |z_0| + \int_0^t c_1\|y_\varepsilon^u(s)\|_{X^\alpha} + |Sf_\varepsilon(y_\varepsilon^u(s), z_\varepsilon^u(s))| + |Su(s)|\mathrm{d}s\\ &\le |z_0| + \int_0^t (M\|S\|_{X^*} + c_1)\left[\|y_\varepsilon^u(s)\|_{X^\alpha} + |z_\varepsilon^u(s)| + 1\right] + \|S\|_{X^*}\|u(s)\|_X \mathrm{d}s.\end{aligned}$$

Remember that $y_\varepsilon^u(0) = 0$ for any $\varepsilon \in (0, \varepsilon_*]$. Since y_ε^u is the mild solution of (3.3), we can use (2.1) for arbitrary $\gamma \in (0,1)$ and again the linear growth condition on f_ε to obtain

$$\begin{aligned}\|y_\varepsilon^u(t)\|_{X^\alpha} &= \left\|\int_0^t e^{-A_p(t-s)}[f_\varepsilon(y_\varepsilon^u(s), z_\varepsilon^u(s)) + u(s)]\mathrm{d}s\right\|_{X^\alpha}\\ &\le C_\alpha e^{(1-\gamma)T}\int_0^t (t-s)^{-\alpha}[M(\|y_\varepsilon^u(s)\|_{X^\alpha} + |z_\varepsilon^u(s)| + 1) + \|u(s)\|_X]\mathrm{d}s.\end{aligned}$$

Note that $\left(\int_0^t (t-s)^{-\alpha q'}\,\mathrm{d}s\right)^{1/q'} = \left(\frac{t^{1-\alpha q'}}{1-\alpha q'}\right)^{1/q'} = \frac{t^{1/q'-\alpha}}{(1-\alpha q')^{1/q'-\alpha}}$ since $q < \frac{1}{1-\alpha} \Leftrightarrow \frac{1}{q'} - \alpha > 0$. We sum up the estimates for $|z_\varepsilon^u(t)|$ and $\|y_\varepsilon^u(t)\|_{X^\alpha}$ and apply Gronwall's Lemma to arrive at

$$\|y_\varepsilon^u\|_{\mathrm{C}(\overline{J_T};X^\alpha)} \le c_3(1 + \|u\|_{\mathrm{L}^q(J_T;X)}) \quad \text{and} \quad \|z_\varepsilon^u\|_{\mathrm{C}(\overline{J_T})} \le c_3(1 + \|u\|_{\mathrm{L}^q(J_T;X)})$$

for all $q \in (\frac{1}{1-\alpha}, \infty]$ and a constant $c_3 > 0$ which depends on T, q' and α but not on ε and u. By maximal parabolic regularity of A_p, see Remark 2.7, one obtains

$$\|y_\varepsilon^u\|_{Y_{s,0}} \le c_4(1 + \|u\|_{\mathrm{L}^q(J_T;X)})$$

for $s = q$ if $q \in (\frac{1}{1-\alpha}, \infty)$ and for all $s \in (1, \infty)$ if $q = \infty$, again for some $c_4 > 0$ which is independent of ε and u. This shows (3.4). We are left to prove (3.5). Note that $2 > \frac{1}{1-\alpha}$ by (A2) in Assumption 2.11. Because $S \in X^*$, (3.4) yields $\|S\dot{y}^u_\varepsilon\|_{\mathrm{L}^2(J_T)} \le c_5(1+\|u\|_{\mathrm{L}^2(J_T;X)})$ for $c_5 = c_4\|S\|_{X^*}$. We test $\dot{z}^u_\varepsilon$ in Definition 3.3 by $\dot{z}^u_\varepsilon$, integrate over $(0,t)$ and use Young's inequality to compute for $t \in \overline{J_T}$:

$$\begin{aligned}\int_0^t |\dot{z}^u_\varepsilon(s)|^2\,\mathrm{d}s &= \int_0^t S\dot{y}^u_\varepsilon(s)\dot{z}^u_\varepsilon(s)\mathrm{d}s - \frac{1}{\varepsilon}\int_0^t \Psi'(z^u_\varepsilon(s))\dot{z}^u_\varepsilon(s)\mathrm{d}s\\ &\le \frac{1}{2}\int_0^t |\dot{z}^u_\varepsilon(s)|^2\,\mathrm{d}s + \frac{c_5^2}{2}(1+\|u\|_{\mathrm{L}^2(J_T;X)})^2 - \frac{1}{\varepsilon}[\Psi(z^u_\varepsilon(t)) - \Psi(z^u_\varepsilon(0))].\end{aligned}$$

Since $\Psi(z^u_\varepsilon(0)) = 0$ and because $\Psi \ge 0$ it follows

$$0 \le \int_0^T |\dot{z}^u_\varepsilon(s)|^2\,\mathrm{d}s + 2\sup_{t\in\overline{J_T}}\frac{1}{\varepsilon}\Psi(z^u_\varepsilon(t)) \le c_5^2(1+\|u\|_{\mathrm{L}^2(J_T;X)})^2.$$

□

The estimates which we derived in this section are crucial for Section 3.2.

3.2. Dynamics of the regularized states

This section contains ideas from Section 4 of [31] and Section 3.1 of [9]. We prove weak continuity of the solution operator of (3.3) for fixed $\varepsilon \in (0,\varepsilon_*]$. Moreover, we apply Lemma 3.6 to obtain weakly converging subsequences $y_{\varepsilon_k} = y^{u_{\varepsilon_k}}_{\varepsilon_k}$ and $z_{\varepsilon_k} = z^{u_{\varepsilon_k}}_{\varepsilon_k}$ for any weakly converging sequence $u_\varepsilon \rightharpoonup u$ in $\mathrm{L}^q(J_T;X)$, $\varepsilon \to 0$. We show that the limit functions are uniquely determined by $y^u = G(u)$ and $z^u = \mathcal{W}[Sy^u]$. This implies weak convergence of the whole sequences y_ε and z_ε. All results then also hold for the regularized state equations (3.1) and (3.2) and each weakly converging sequence $u_\varepsilon \rightharpoonup u$ in U_i, $i \in \{1,2\}$. In this case, the limit functions $y^{B_iu} = G(B_iu)$ and $z^{B_iu} = \mathcal{W}[Sy^{B_iu}]$ are solutions of (1.2) and (1.3).

Those results are required to prove existence of solutions of the regularized control problems, as defined in Section 3.3, and to show their convergence to an optimal solution of problem (1.1)–(1.3) with $\varepsilon \to 0$.

The following lemma is proved as ([32], Lem. 5.3).

Lemma 3.7. *Let Assumptions 2.11 and 3.2 hold and consider the notation from Corollary 3.5. 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 $y^{u_n}_\varepsilon$ and y^u_ε of (3.3), together with $z^{u_n}_\varepsilon$ and z^u_ε. Then $y^{u_n}_\varepsilon \to y^u_\varepsilon$ with $n \to \infty$ weakly in $Y_{2,0}$ and strongly in $\mathrm{C}(\overline{J_T};X^\alpha)$ and $z^{u_n}_\varepsilon \to z^u_\varepsilon$ 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 $\{y^{u_n}_\varepsilon\}$ 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, $(y^{B_iu_n}_\varepsilon, z^{B_iu_n}_\varepsilon)$ and $(y^{B_iu}_\varepsilon, z^{B_iu}_\varepsilon)$ are solutions of (3.1) and (3.2).*

Furthermore, we have the following convergence result:

Lemma 3.8. *Let Assumptions 2.11 and 3.2 hold and consider the notation from Corollary 3.5. Suppose that $u_\varepsilon \rightharpoonup u$ in $\mathrm{L}^2(J_T;X)$ as $\varepsilon \to 0$. Consider the solutions $y^{u_\varepsilon}_\varepsilon$ of (3.3), together with $z^{u_\varepsilon}_\varepsilon$. Then $y^{u_\varepsilon}_\varepsilon \to y^u$ with $\varepsilon \to 0$ weakly in $Y_{2,0}$ and strongly in $\mathrm{C}(\overline{J_T};X^\alpha)$ and $z^{u_\varepsilon}_\varepsilon \to \mathcal{W}[Sy^u] = z^u$ with $\varepsilon \to 0$ 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 $\{y^{u_\varepsilon}_\varepsilon\}$ 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, $(y^{B_iu_\varepsilon}_\varepsilon, z^{B_iu_\varepsilon}_\varepsilon)$ are solutions of (3.1) and (3.2) and (y^{B_iu}, z^{B_iu}) solves (1.2) and (1.3).*

Proof. The proof combines the proofs of Lemma 3.2 in [9] and Lemma 5.3 in [32]. By Lemma 3.6 we obtain a bound for $y^{u_\varepsilon}_\varepsilon$ in $Y_{2,0}$ and for $z^{u_\varepsilon}_\varepsilon$ in $\mathrm{H}^1(J_T)$ which is independent of $\varepsilon \in (0,\varepsilon_*]$. Hence, there exists a subsequence $\{\varepsilon_k\}$ of the sequence $\{\varepsilon\}$ and functions $\tilde{y} \in Y_{2,0}$ and $\tilde{z} \in \mathrm{H}^1(J_T)$ to which $y^{u_{\varepsilon_k}}_{\varepsilon_k}$ and $z^{u_{\varepsilon_k}}_{\varepsilon_k}$ converge weakly in $Y_{2,0}$ and $\mathrm{H}^1(J_T)$ and strongly in $\mathrm{C}(\overline{J_T};X^\alpha)$ and $\mathrm{C}(\overline{J_T})$ with $k \to \infty$. We abbreviate $y_{\varepsilon_k} := y^{u_{\varepsilon_k}}_{\varepsilon_k}$ and $z_{\varepsilon_k} := z^{u_{\varepsilon_k}}_{\varepsilon_k}$. (3.5) implies that $\Psi(z_{\varepsilon_k}(t)) \to 0$ with $k \to \infty$ for $t \in \overline{J_T}$. By $(A4)_\varepsilon$ in Assumption 3.2 this yields $\tilde{z}(t) \in [a,b]$ for

$t \in \overline{J_T}$. For any $x \in \mathbb{R}$ and $\xi \in [a,b]$ there holds $\Psi'(x)(x-\xi) \geq 0$ because Ψ is convex and since $\Psi'(\xi) = 0$ for $\xi \in [a,b]$. For any $\xi \in [a,b]$ we therefore have

$$\int_0^T (\dot{z}_{\varepsilon_k}(t) - S\dot{y}_{\varepsilon_k}(t))(z_{\varepsilon_k}(t) - \xi)\,\mathrm{d}t = \int_0^T -\frac{1}{\varepsilon}\Psi'(z_{\varepsilon_k}(u_{\varepsilon_k}(t))(z_{\varepsilon_k}(t) - \xi)\,\mathrm{d}t \leq 0.$$

Taking the limit $k \to \infty$ yields $\tilde{z} = \mathcal{W}[S\tilde{y}]$ since $\tilde{z}$ solves (1.4) and (1.5) with $v = S\tilde{y}$. Weak continuity of $\frac{\mathrm{d}}{\mathrm{d}t}$ and A_p implies $\frac{\mathrm{d}}{\mathrm{d}t}y_{\varepsilon_k} + A_p y_{\varepsilon_k} \rightharpoonup \frac{\mathrm{d}}{\mathrm{d}t}\tilde{y} + A_p\tilde{y}$ in $\mathrm{L}^2(J_T;X)$ with $k \to \infty$. For ε_k small enough we estimate with $(A3)_\varepsilon$ in Assumption 3.2:

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

Because the right side converges to zero, we conclude that $F_{\varepsilon_k}[y_{\varepsilon_k}]$ converges to $F[\tilde{y}]$ in $\mathrm{C}(\overline{J_T};X)$ with $k \to \infty$. This together with $\tilde{z} = \mathcal{W}[S\tilde{y}]$ yields $\tilde{y} = G(u)$. Uniqueness of the limit implies convergence of the whole sequence. The statement about strong convergence follows essentially the same way as in Lemma 5.3 of [32]. ☐

3.3. The regularized optimal control problem

In this section, we introduce regularized optimal control problems. It still requires work to obtain adjoint systems for those problems. Nevertheless, we can exploit linearity of the derivatives of the solution operators of (3.1) and (3.2) to derive adjoint systems by a direct approach. This will be done in Section 3.5.

We follow the ideas in Section 3.2 of [9] and Section 4 of [31] in this section. For $i \in \{1,2\}$ consider an optimal control $\overline{u} \in U_i$ of problem (1.1)–(1.3) together with the state $\overline{y} = G(B_i\overline{u})$ and $\overline{z} = \mathcal{W}[S\overline{y}]$. Existence of $\overline{u}$ follows from Theorem 2.15. For $\varepsilon \in (0, \varepsilon_*]$ we introduce the regularized optimal control problem

$$\min_{u \in U_i} J_{\mathrm{reg}}(y, u; \overline{u}) = \min_{u \in U_i} J(y,u) + \frac{1}{2}\|u - \overline{u}\|_{U_i}^2 \tag{3.7}$$

subject to (3.1) and (3.2).

Theorem 3.9 (Convergence of optimal solutions)**.** *Let Assumptions 2.11 and 3.2 hold. For $i \in \{1,2\}$ suppose that $\overline{u} \in U_i$ is an optimal control for problem (1.1)–(1.3). Then for all $\varepsilon \in (0, \varepsilon_*]$ problem (3.1),(3.2),(3.7) 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 3.3), are a solution of the minimization problem (3.7). 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 $Y_{0,2} \hookrightarrow U_1$ is continuous, because $\mathrm{dom}(A_p) \simeq \mathbb{W}^{1,p}_{\Gamma_D}(\Omega) \hookrightarrow [\mathrm{L}^2(\Omega)]^m$. Note also that $\overline{u}$ exists by Theorem 2.15. Existence of optimal controls $\overline{u}_\varepsilon$ for (3.1), (3.2), (3.7) follows essentially the same way as for problem (1.1)–(1.3) by using Lemma 3.7, see also Theorem 2.15. For all $\varepsilon \in (0, \varepsilon_*]$, we deduce from optimality of $(\overline{y}_\varepsilon, \overline{z}_\varepsilon, \overline{u}_\varepsilon)$ for problem (3.1), (3.2), (3.7) and of $(\overline{y}, \overline{z}, \overline{u})$ for problem (1.1)–(1.3) that

$$J(G_\varepsilon(B_i\overline{u}), \overline{u}) = J_{\mathrm{reg}}(G_\varepsilon(B_i\overline{u}), \overline{u}; \overline{u}) \geq J_{\mathrm{reg}}(\overline{y}_\varepsilon, \overline{u}_\varepsilon; \overline{u}) = J(\overline{y}_\varepsilon, \overline{u}_\varepsilon) + \frac{1}{2}\|\overline{u}_\varepsilon - \overline{u}\|_{U_i}^2 \geq J(\overline{y}, \overline{u}). \tag{3.8}$$

Moreover, by (3.4) in Lemma 3.6, $G_\varepsilon(B_i\overline{u}) \in Y_{2,0}$ is uniformly bounded for $\varepsilon \in (0, \varepsilon_*]$ so that $J(G_\varepsilon(B_i\overline{u}), \overline{u}) \leq c$ for some constant $c > 0$. Hence, $c > J_{\mathrm{reg}}(\overline{y}_\varepsilon, \overline{u}_\varepsilon; \overline{u}) = \frac{1}{2}\|\overline{y}_\varepsilon - y_d\|_{U_1}^2 + \frac{\kappa}{2}\|\overline{u}_\varepsilon\|_{U_i}^2 + \frac{1}{2}\|\overline{u}_\varepsilon - \overline{u}\|_{U_i}^2$ and the norms of $\overline{u}_\varepsilon$ in U_i are bounded from above independently of $\varepsilon \in (0, \varepsilon_*]$. Consequently, we can extract a subsequence $\{u_{\varepsilon_k}\}$ which converges weakly in U_i to some $\tilde{u}$ with $k \to \infty$. By Lemma 3.8, $y_{\varepsilon_k} \to G(\tilde{u})$ with $k \to \infty$ weakly in

$Y_{0,2}$ and then also in U_1. Also by Lemma 3.8, $G_\varepsilon(B_i\overline{u}) \to \overline{y}$ with $\varepsilon \to 0$ strongly in $Y_{2,0}$ and then in U_1. J_{reg} is weakly lower semi-continuous. Hence, with (3.8) we obtain

$$J(\overline{y}, \overline{u}) = \lim_{k\to\infty} J(G_{\varepsilon_k}(B_i\overline{u}), \overline{u}) \geq \liminf_{k\to\infty} J_{\text{reg}}(\overline{y}_{\varepsilon_k}, \overline{u}_{\varepsilon_k}; \overline{u}) \geq J(\tilde{y}, \tilde{u}) + \frac{1}{2}\|\tilde{u} - \overline{u}\|_{U_i}^2 \geq J(\overline{y}, \overline{u}).$$

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 3.8. □

3.4. Gâteaux differentiability of the solution operator of the regularized state equation

In this section, we show that G_ε is Gâteaux differentiable for all $\varepsilon \in (0, \varepsilon_*]$.

Lemma 3.10 (Gâteaux differentiability of G_ε)**.** *Let Assumptions 2.11 and 3.2 hold and take the notation from Corollary 3.5. 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 (3.3) 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 $z_\varepsilon^{u,h} = Z'_\varepsilon[Sy_\varepsilon^u; Sy_\varepsilon^{u,h}] \in \mathrm{W}^{1,q}(J_T)$ are the unique solution of*

$$\begin{aligned}
\dot{y}(t) + (A_p y)(t) &= \frac{\partial}{\partial y} f_\varepsilon(y_\varepsilon^u(t), z_\varepsilon^u(t))y(t) + \frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u(t), z_\varepsilon^u(t))z(t) + h(t) \text{ in } X \quad \text{for } t \in J_T,\\
y(0) &= 0 \qquad\qquad \text{in } X, \qquad (3.9)\\
\dot{z}(t) - S\dot{y}(t) &= -\frac{1}{\varepsilon}\Psi''(z_\varepsilon^u(t))z(t) \qquad\qquad \text{for } t \in J_T,\\
z(0) &= 0. \qquad (3.10)
\end{aligned}$$

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 (3.9) with h replaced by $B_i h$ and $z_\varepsilon^{B_i u, B_i h} = Z'_\varepsilon[Sy_\varepsilon^{B_i u}; Sy_\varepsilon^{B_i u, B_i h}]$.*

Proof. As in Corollary 3.5, note that Theorem 2.14 holds with $\mathcal{W}$ replaced by Z_ε, since Z_ε is Lipschitz continuous and Hadamard directionally (even continuously) differentiable on $\mathrm{C}(\overline{J_T})$ as well as bounded and weakly continuous on $\mathrm{W}^{1,q}(J_T)$ for all $q \in (1, \infty)$, *cf.* Remark 3.4. Hadamard directional differentiability of the solution operator G_ε of (3.3) is shown *via* Hadamard directional differentiability and Lipschitz continuity of $F_\varepsilon : \mathrm{C}(\overline{J_T}; X^\alpha) \to \mathrm{L}^q(J_T; X)$. In particular, the derivative $h \mapsto G'_\varepsilon[u; h]$ at $u \in \mathrm{L}^q(J_T; X)$ is Lipschitz continuous from $\mathrm{L}^q(J_T; X)$ to $\mathrm{C}(\overline{J_T}; X^\alpha)$. For $u, h \in \mathrm{L}^q(J_T; X)$, $G'_\varepsilon[u; h]$ is given by the unique solution y of

$$\dot{y}(t) + (A_p y)(t) = F'_\varepsilon[y_\varepsilon^u; y](t) + h(t) \text{ for } t \in J_T, \;\; y(0) = 0,$$

where

$$\begin{aligned}
F'_\varepsilon[y_\varepsilon^u; y](t) &= f'_\varepsilon[(y_\varepsilon^u(t), Z[Sy_\varepsilon^u](t)); (y(t), Z'_\varepsilon[Sy_\varepsilon^u; Sy](t))]\\
&= \frac{\partial}{\partial y} f_\varepsilon(y_\varepsilon^u(t), z_\varepsilon^u(t))y(t) + \frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u(t), z_\varepsilon^u(t))(Z'_\varepsilon[Sy_\varepsilon^u]Sy)(t),
\end{aligned}$$

and with $y_\varepsilon^u = G_\varepsilon(u)$. Since the solution operator $y \mapsto Z'_\varepsilon[Sy_\varepsilon^u]Sy$ of (3.10) is linear, the same applies for the mapping $y \mapsto F'_\varepsilon[y_\varepsilon^u; y] = F'_\varepsilon[y_\varepsilon^u]y$. This implies that the solution operator $h \mapsto G'_\varepsilon[u; h]$ is linear and bounded. Consequently, G_ε is Gâteaux differentiable. To see that $z_\varepsilon^{u,h} \in \mathrm{W}^{1,q}(J_T)$, insert $Sy^{u,h}$ for h in Remark 3.4 and note that the right side is contained in $\mathrm{W}^{1,q}(J_T)$. □

3.5. Adjoint system for the regularized problem

In this section, we derive adjoint systems for the regularized problems (3.1), (3.2), (3.7) with $\varepsilon \in (0, \varepsilon_*]$, see Theorem 3.13. We proceed in a similar way as in Sections 3.3 and 3.5 of [9] and Section 4 of [31]. The following estimates are needed.

Lemma 3.11. *Let Assumptions 2.11 and 3.2 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 $(v, h) \mapsto f'_\varepsilon[(y, z); (v, h)]$ at $(y, z) \in \mathrm{C}(\overline{J_T}; X^\alpha) \times \mathrm{L}^q(J_T)$ is bounded by a constant $K(y) = L(y)(1 + T^{1/q})$, 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 constant in a sufficiently small neighbourhood of y. In particular, for $(v, h) \in \mathrm{C}(\overline{J_T}; X^\alpha) \times \mathrm{L}^q(J_T)$ we can estimate

$$\left\| \frac{\partial}{\partial y} f_\varepsilon(y, z) v \right\|_{\mathrm{L}^q(J_T;X)} + \left\| \frac{\partial}{\partial z} f_\varepsilon(y, z) h \right\|_{\mathrm{L}^q(J_T;X)} \leq K(y)(\|v\|_{\mathrm{C}(\overline{J_T};X^\alpha)} + \|h\|_{\mathrm{L}^q(J_T)}). \tag{3.11}$$

For a.e. $t \in J_T$, there also holds the pointwise estimate

$$\left\| \frac{\partial}{\partial y} f_\varepsilon(y(t), z(t)) v(t) \right\|_X + \left\| \frac{\partial}{\partial z} f_\varepsilon(y(t), z(t)) h(t) \right\|_X \leq K(y)(\|v(t)\|_{X^\alpha} + |h(t)|). \tag{3.12}$$

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. First of all, f_ε is locally Lipschitz continuous and Gâteaux differentiable from the space $\mathrm{C}(\overline{J_T}; X^\alpha) \times \mathrm{L}^q(J_T)$ to $\mathrm{L}^q(J_T; X)$ for all $\varepsilon \in (0, \varepsilon_*]$ and $q \in (\frac{1}{1-\alpha}, \infty)$. This follows from Step 3 in the proof of Theorem 3.1 from [32] and Step 1 in the proof of Theorem 4.7 from [32]. We give a sketch of the proof:

One first applies $(A3)_\varepsilon$ in Assumption 3.2 to show that $(y(\cdot), v) \mapsto f_\varepsilon(y(\cdot), v)$ 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. The proof contains a pointwise estimate of the following form: For $y \in \mathrm{C}(\overline{J_T}; X^\alpha)$ and some neighbourhood $\overline{B_{\mathrm{C}(\overline{J_T};X^\alpha)}(y, \delta)}$ of y there holds

$$\|f_\varepsilon(y_1(t), z_1) - f(y_2(t), z_2)\|_X \leq L(y)(\|y_1(t) - y_2(t)\|_{X^\alpha} + |z_1 - z_2|)$$

for all $y_1, y_2 \in \overline{B_{\mathrm{C}(\overline{J_T};X^\alpha)}(y, \delta)}$, $z_1, z_2 \in \mathbb{R}$ and $t \in \overline{J_T}$ and for some $L(y) > 0$. This local estimate leads to a pointwise estimate of the form

$$\|f_\varepsilon(y_1, z_1)(s) - f_\varepsilon(y_2, z_2)(s)\|_X \leq L(y) \left[\|y_1(s) - y_2(s)\|_{X^\alpha} + |z_1(s) - z_2(s)| \right]$$

for a.e. $s \in J_T$, for any $y_1, y_2 \in \overline{B_{\mathrm{C}(\overline{J_T};X^\alpha)}(y, \delta)}$ and $z_1, z_2 \in \mathrm{L}^q(J_T)$. By Minkowski's inequality, f_ε is locally Lipschitz continuous from $\mathrm{C}(\overline{J_T}; X^\alpha) \times \mathrm{L}^q(J_T)$ to $\mathrm{L}^q(J_T; X)$ with Lipschitz constants of the form $K(y) = L(y)(1 + T^{1/q})$.

In a second step one shows that f_ε is directionally differentiable. 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 $(y, z), (v, h) \in \mathrm{C}(\overline{J_T}; X^\alpha) \times \mathrm{L}^q(J_T)$ follows from $(A3)_\varepsilon$ in Assumption 3.2. Lebesgue's dominated convergence theorem yields directional differentiability of f_ε from the space $\mathrm{C}(\overline{J_T}; X^\alpha) \times \mathrm{L}^q(J_T)$ to

$L^q(J_T;X)$ and the bounds (3.11) and (3.12) for $f'_\varepsilon[(y,z);(\cdot,\cdot)]$. This together with linearity of the derivative implies Gâteaux differentiability of f_ε.

Now for arbitrary $y \in X^\alpha$ with $\|y\|_{X^\alpha} = 1$, we choose the constant function $v \in C(\overline{J}_T;X^\alpha)$, $v(t) = y$ for $t \in \overline{J_T}$ and set $h = 0 \in L^q(J_T)$ in (3.12). This implies that $\frac{\partial}{\partial y} f_\varepsilon(y,z) = \frac{\partial}{\partial y} f_\varepsilon(y(\cdot), z(\cdot))$ is bounded by $K(y)$ in $L^\infty(J_T;\mathcal{L}(X^\alpha,X))$. Then we choose $v = 0 \in C(\overline{J}_T;X^\alpha)$, $h \in L^q(J_T)$, $h(t) = c > 0$ for $t \in \overline{J_T}$ in (3.12) and divide by c on both sides to prove that $\frac{\partial}{\partial z} f_\varepsilon(y,z) = \frac{\partial}{\partial z} f_\varepsilon(y(\cdot), z(\cdot))$ is bounded by $K(y)$ in $L^\infty(J_T;X)$. □

The following lemma provides the main tool to derive adjoint systems for the regularized problems (3.1), (3.2), (3.7). The hardest part in the proof is to find an explicit expression of the adjoint operator $[G'_\varepsilon[u;\cdot]]^* : Y^*_{q',T} \to L^{q'}(J_T;X^*)$ of $G'_\varepsilon[u;\cdot]$ from Lemma 3.10. This comes from the fact that $G'_\varepsilon[u;\cdot]$ is defined as the mapping which assigns to each $h \in L^q(J_T;X)$ the solution $y^{u,h}_\varepsilon \in Y_{q,0}$ of (3.9), which contains the solution $z^{u,h}_\varepsilon$ of (3.10) only implicitly.

Lemma 3.12. *Let Assumptions 2.11 and 3.2 hold and adopt the notation from Lemma 3.10. For $\varepsilon \in (0,\varepsilon_*]$ and any $q \in (\frac{1}{1-\alpha},\infty)$, $h \in L^q(J_T;X)$ and $\eta \in L^{q'}(J_T;[\mathrm{dom}(A_p)]^*)$ there holds*

$$\langle \eta, y^{u,h}_\varepsilon \rangle_{L^q(J_T;\mathrm{dom}(A_p))} = \langle p^\eta_\varepsilon + S q^\eta_\varepsilon, h \rangle_{L^q(J_T;X)},$$

where $p^\eta_\varepsilon \in Y^*_{q',T}$ and $q^\eta_\varepsilon \in L^{q'}(J_T)$ are the unique solution of

$$\begin{aligned} -\dot{p} + A_p^* p &= \left[\frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\right]^* p + S\left[-A_p + \frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\right] q + \eta \text{ for } t \in J_T, \\ p(T) &= 0, \\ -\dot{q} &= \langle p, \frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\rangle_X + S\frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon) q - \frac{1}{\varepsilon}\Psi''(z^u_\varepsilon) q \quad \text{for } t \in J_T, \\ q(T) &= 0, \end{aligned}$$

and where $y^{u,h}_\varepsilon \in Y_{q,0}$ and $z^{u,h}_\varepsilon \in W^{1,q}(J_T)$ are the unique solution of (3.9) and (3.10). Moreover,

$$\|y^{u,h}_\varepsilon\|_{Y_{q,0}} \le C(y^u_\varepsilon)\|h\|_{L^q(J_T;X)} \text{ and } \|z^{u,h}_\varepsilon\|_{C(\overline{J_T})} \le C(y^u_\varepsilon)\|h\|_{L^q(J_T;X)} \tag{3.13}$$

for some constant $C(y^u_\varepsilon) > 0$. $C(y^u_\varepsilon)$ remains constant in a sufficiently small neighbourhood of y^u_ε.

Proof. Let $q \in (\frac{1}{1-\alpha},\infty)$ be arbitrary. Consider the solution operator of

$$\dot{z}(t) = v(t) + \left(S\frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon(t), z^u_\varepsilon(t)) - \frac{1}{\varepsilon}\Psi''(z^u_\varepsilon(t))\right) z(t) \quad \text{for } t \in J_T,\ z(0) = 0,$$

which maps any $v \in L^q(J_T)$ to $z \in W^{1,q}(J_T)$. We denote by $T^u_{z,\varepsilon} : L^q(J_T) \to L^q(J_T)$, $v \mapsto T^u_{z,\varepsilon} v$ the corresponding operator on $L^q(J_T)$.

Consider then the operator $T^u_{y,\varepsilon} := A_p - \frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon) - \frac{\partial}{\partial z} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon) T^u_{z,\varepsilon} S\left(-A_p + \frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\right)$ from $Y_{q,0}$ to $L^q(J_T;X)$. It follows as for the system (3.9) and (3.10) that for each $h \in L^q(J_T;X)$ there exists a unique couple of solutions $(\tilde{y}^{u,h}_\varepsilon, \tilde{z}^{u,h}_\varepsilon)$ in $Y_{q,0} \times L^q(J_T)$ of the system

$$\dot{y}(t) + (T^u_{y,\varepsilon} y)(t) = h(t) \text{ for } t \in J_T,\ y(0) = 0, \tag{3.14}$$

$$z = T^u_{z,\varepsilon} S\left(-A_p + \frac{\partial}{\partial y} f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\right) y. \tag{3.15}$$

This implies that $\left(\frac{\mathrm{d}}{\mathrm{d}t} + T^u_{y,\varepsilon}\right)^{-1}$ is bijective from $\mathrm{L}^q(J_T;X)$ to $Y_{q,0}$. Note the difference between (3.14)–(3.15) and (3.9)–(3.10). We identify $\tilde{z}^{u,h}_\varepsilon \in \mathrm{L}^q(J_T)$ with the corresponding function in $\mathrm{W}^{1,q}(J_T)$ and estimate the norms of $(\tilde{y}^{u,h}_\varepsilon, \tilde{z}^{u,h}_\varepsilon)$. For $t \in \overline{J_T}$ we have

$$|\tilde{z}^{u,h}_\varepsilon(t)| = \int_0^t \frac{\dot{\tilde{z}}^{u,h}_\varepsilon(s)\tilde{z}^{u,h}_\varepsilon(s)}{|\tilde{z}^{u,h}_\varepsilon(s)|}\mathrm{d}s = \int_0^t \frac{-S[(T^u_{y,\varepsilon}\tilde{y}^{u,h}_\varepsilon)(s)]\tilde{z}^{u,h}_\varepsilon(s)}{|\tilde{z}^{u,h}_\varepsilon(s)|}\mathrm{d}s - \frac{1}{\varepsilon}\int_0^t \Psi''(z^u_\varepsilon(s))|\tilde{z}^{u,h}_\varepsilon(s)|\,\mathrm{d}s.$$

With (3.12) in Lemma 3.11 and (A3) in Assumption 2.11 it follows

$$\begin{aligned}
0 &\leq |\tilde{z}^{u,h}_\varepsilon(t)| + \frac{1}{\varepsilon}\int_0^t \Psi''(z^u_\varepsilon(s))|\tilde{z}^{u,h}_\varepsilon(s)|\mathrm{d}s \\
&\leq \int_0^t |SA_p\tilde{y}^{u,h}_\varepsilon(s)| + \left|S\frac{\partial}{\partial y}f_\varepsilon(y^u_\varepsilon(s), z^u_\varepsilon(s))\tilde{y}^{u,h}_\varepsilon(s)\right| + \left|S\frac{\partial}{\partial z}f_\varepsilon(y^u_\varepsilon(s), z^u_\varepsilon(s))\tilde{z}^{u,h}_\varepsilon(s)\right|\mathrm{d}s \\
&\leq (c + \|S\|_{X^*}K(y^u_\varepsilon))\int_0^t \|\tilde{y}^{u,h}_\varepsilon(s)\|_{X^\alpha} + |\tilde{z}^{u,h}_\varepsilon(s)|\mathrm{d}s
\end{aligned}$$

for a constant $c > 0$ which is independent of ε. Note that $\Psi''(z^u_\varepsilon(s)) \geq 0$ because Ψ is convex. Moreover, with (2.1) and again (3.12) in Lemma 3.11 we obtain

$$\begin{aligned}
&\|\tilde{y}^{u,h}_\varepsilon(t)\|_{X^\alpha} \\
&= \left\|\int_0^t e^{-A_p(t-s)}\left[\frac{\partial}{\partial y}f_\varepsilon(y^u_\varepsilon(s), z^u_\varepsilon(s))\tilde{y}^{u,h}_\varepsilon(s) + \frac{\partial}{\partial z}f_\varepsilon(y^u_\varepsilon(s), z^u_\varepsilon(t))\tilde{z}^{u,h}_\varepsilon(s) + h(s)\right]\mathrm{d}s\right\|_{X^\alpha} \\
&\leq C_\alpha(1 + K(y^u_\varepsilon))e^{(1-\gamma)T}\int_0^t (t-s)^{-\alpha}[\|\tilde{y}^{u,h}_\varepsilon(s)\|_{X^\alpha} + |\tilde{z}^{u,h}_\varepsilon(s)| + \|h(s)\|_X]\mathrm{d}s.
\end{aligned}$$

Gronwall's Lemma yields a constant $C_1(y^u_\varepsilon) > 0$ which depends only on $y^u_\varepsilon \in \mathrm{C}(\overline{J_T};X^\alpha)$ such that $\|\tilde{y}^{u,h}_\varepsilon\|_{\mathrm{C}(\overline{J_T};X^\alpha)} \leq C_1(y^u_\varepsilon)\|h\|_{\mathrm{L}^q(J_T;X)}$ and $\|\tilde{z}^{u,h}_\varepsilon\|_{\mathrm{C}(\overline{J_T})} \leq C_1(y^u_\varepsilon)\|h\|_{\mathrm{L}^q(J_T;X)}$ for $q \in (\frac{1}{1-\alpha}, \infty)$. Moreover, there holds $C_1(y^u_\varepsilon) = C_1(y)$ for ε small enough if $\{y^u_\varepsilon\}$ converges to y with $\varepsilon \to 0$. This is the case for the states $\overline{y}_\varepsilon$ in Theorem 3.9. As several times before we apply maximal parabolic regularity of A_p to obtain $\|\tilde{y}^{u,h}_\varepsilon\|_{Y_{q,0}} \leq C_2(y^u_\varepsilon)\|h\|_{\mathrm{L}^q(J_T;X)}$ where $C_2(y^u_\varepsilon) > 0$ has the same dependence on y^u_ε as $C_1(y^u_\varepsilon)$. The inequalities in (3.13) are shown analogously to the estimates which we derived for $(\tilde{y}^{u,h}_\varepsilon, \tilde{z}^{u,h}_\varepsilon)$. We also conclude that there exists a constant $C(y^u_\varepsilon) > 0$ with $\left\|\left(\frac{\mathrm{d}}{\mathrm{d}t} + T^u_{y,\varepsilon}\right)^{-1}\right\|_{\mathcal{L}(\mathrm{L}^q(J_T;X),Y_{q,0})} \leq C(y^u_\varepsilon)$. This proves that $\frac{\mathrm{d}}{\mathrm{d}t} + T^u_{y,\varepsilon} : Y_{q,0} \to \mathrm{L}^q(J_T;X)$ is an isomorphism for $q \in (\frac{1}{1-\alpha}, \infty)$. For ε small enough, also the values $C(y^u_\varepsilon)$ can be chosen independently of ε if $\{y^u_\varepsilon\}$ converges to some y with $\varepsilon \to 0$ as it is the case for the sequence $\{\overline{y}_\varepsilon\}$ in Theorem 3.9. In the proof of Lemma A.5 from [29] it is shown that $-\frac{\mathrm{d}}{\mathrm{d}t} + A^*_p : Y^*_{q',T} \to \mathrm{L}^{q'}(J_T;[\mathrm{dom}(A_p)]^*)$ is an isomorphism if $\frac{\mathrm{d}}{\mathrm{d}t} + A_p : Y_{q,0} \to \mathrm{L}^q(J_T;X)$ is one. The same proof applies for $T^u_{y,\varepsilon}$ and $[T^u_{y,\varepsilon}]^*$, where all involved spaces remain the same. Hence, $-\frac{\mathrm{d}}{\mathrm{d}t} + [T^u_{y,\varepsilon}]^* : Y^*_{q',T} \to \mathrm{L}^{q'}(J_T;[\mathrm{dom}(A_p)]^*)$ is an isomorphism for $q \in (\frac{1}{1-\alpha}, \infty)$. To derive a representation of $[T^u_{y,\varepsilon}]^*$, we collect information about the adjoint mappings of the single components which define $T^u_{y,\varepsilon}$. Lemma 3.11 yields that multiplication with $\frac{\partial}{\partial z}f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)$ is well-defined as a mapping from $\mathrm{L}^q(J_T)$ into $\mathrm{L}^q(J_T;X)$ and $\left[\frac{\partial}{\partial z}f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\right]^* = \langle\cdot, \frac{\partial}{\partial z}f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\rangle_X$. Similarly, $\frac{\partial}{\partial y}f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)$ is a linear continuous mapping from $\mathrm{L}^q(J_T;X^\alpha)$ into $\mathrm{L}^q(J_T;X)$. Moreover, $\left[S\frac{\partial}{\partial y}f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)\right]^*$ is given by multiplication with $S\frac{\partial}{\partial y}f_\varepsilon(y^u_\varepsilon, z^u_\varepsilon)$. The adjoint of $T^u_{z,\varepsilon}$ 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

$$-\dot{q}(t) = v(t) + S\frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u(t), z_\varepsilon^u(t))q(t) - \frac{1}{\varepsilon}\Psi''(z_\varepsilon^u(t))q(t) \quad \text{for } t \in J_T,\ q(T) = 0.$$

S^* and $[SA_p]^*$ are given by multiplication with S and SA_p. Furthermore, $SA_p \in [X^\alpha]^*$ by the assumptions on w in (A3) in Assumption 2.11. All bounds are independent of ε if $\overline{y}_\varepsilon$ and $\overline{z}_\varepsilon$ in Theorem 3.9 are considered and if ε is small enough. We obtain

$$\begin{aligned}
[T_{y,\varepsilon}^u]^* &= A_p^* - \left[\frac{\partial}{\partial y} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u)\right]^* - \left[\frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u) T_{z,\varepsilon}^u S\left(-A_p + \frac{\partial}{\partial y} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u)\right)\right]^* \\
&= A_p^* - \left[\frac{\partial}{\partial y} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u)\right]^* + \left[[SA_p]^* - \left[S\frac{\partial}{\partial y} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u)\right]^*\right][T_{z,\varepsilon}^u]^* \left[\frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u)\right]^* \\
&= A_p^* - \left[\frac{\partial}{\partial y} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u)\right]^* + S\left[A_p - \frac{\partial}{\partial y} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u)\right][T_{z,\varepsilon}^u]^* \langle \cdot, \frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u)\rangle_X.
\end{aligned}$$

Since $-\frac{\mathrm{d}}{\mathrm{d}t} + [T_{y,\varepsilon}^u]^* : Y_{q',T}^* \to \mathrm{L}^{q'}(J_T; [\mathrm{dom}(A_p)]^*)$ is an isomorphism there exists for each $\eta \in \mathrm{L}^{q'}(J_T; [\mathrm{dom}(A_p)]^*)$ a unique function $p_\varepsilon^\eta \in Y_{q',T}^*$ with

$$\left(-\frac{\mathrm{d}}{\mathrm{d}t} + [T_{\varepsilon,y}^u]^*\right) p = \eta.$$

For given $\eta \in \mathrm{L}^{q'}(J_T; [\mathrm{dom}(A_p)]^*)$ let q_ε^η be the representative in $\mathrm{L}^{q'}(J_T)$ of the solution of

$$\begin{aligned}
&-\dot{q}(t) = \langle p_\varepsilon^\eta(t), \frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u(t), z_\varepsilon^u(t))\rangle_X + S\frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u(t), z_\varepsilon^u(t))q(t) - \frac{1}{\varepsilon}\Psi''(z_\varepsilon^u(t))q(t) \text{ for } t \in J_T, \\
&q(T) = 0.
\end{aligned}$$

Let also $(y_\varepsilon^{u,h}, z_\varepsilon^{u,h})$ be the solutions of (3.9) and (3.10) for some given $h \in \mathrm{L}^q(J_T; X)$. Then we obtain with (3.9) and partial integration:

$$\begin{aligned}
&\int_0^T \langle p_\varepsilon^\eta + Sq_\varepsilon^\eta, h\rangle_X \mathrm{d}t \\
&= \int_0^T \langle p_\varepsilon^\eta, \dot{y}_\varepsilon^{u,h} + A_p y_\varepsilon^{u,h} - \frac{\partial}{\partial y} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u) y_\varepsilon^{u,h} - \frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u) z_\varepsilon^{u,h}\rangle_X + \langle Sq_\varepsilon^\eta, h\rangle_X \mathrm{d}t \\
&= \int_0^T \langle -\dot{p}_\varepsilon^\eta + A_p^* p_\varepsilon^\eta - \left[\frac{\partial}{\partial y} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u)\right]^* p_\varepsilon^\eta, y_\varepsilon^{u,h}\rangle_{\mathrm{dom}(A_p)} + \langle Sq_\varepsilon^\eta, h\rangle_X \\
&- \langle p_\varepsilon^\eta, \frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u)\rangle_X z_\varepsilon^{u,h} \mathrm{d}t.
\end{aligned}$$

By definition of q_ε^η the last term on the right side is equal to

$$\int_0^T \left(\dot{q}_\varepsilon^\eta + S\frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u) q_\varepsilon^\eta - \frac{1}{\varepsilon}\Psi''(z_\varepsilon^u) q_\varepsilon^\eta\right) \overline{z}_\varepsilon^{u,h} \mathrm{d}t.$$

Another partial integration together with (3.10) and canceling out some terms yields

$$\int_0^T \langle p_\varepsilon^\eta + S q_\varepsilon^\eta, h\rangle_X \mathrm{d}t = \int_0^T \langle -\dot{p}_\varepsilon^\eta + A_p^* p_\varepsilon^\eta - \left[\frac{\partial}{\partial y} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u)\right]^* p_\varepsilon^\eta, y_\varepsilon^{u,h}\rangle_{\mathrm{dom}(A_p)} + \langle S q_\varepsilon^\eta, h\rangle_X$$
$$- q_\varepsilon^\eta S\left[\left(-A_p + \frac{\partial}{\partial y} f_\varepsilon(y_\varepsilon^u, z_\varepsilon^u)\right) y_\varepsilon^{u,h} + h\right] \mathrm{d}t.$$

By definition of p_ε^η we finally arrive at

$$\int_0^T \langle p_\varepsilon^\eta + S q_\varepsilon^\eta, h\rangle_X \mathrm{d}t = \int_0^T \langle -\dot{p}_\varepsilon^\eta + [T_{y,\varepsilon}^u]^* p_\varepsilon^\eta, y_\varepsilon^{u,h}\rangle_{\mathrm{dom}(A_p)} \mathrm{d}t = \int_0^T \langle \eta, y_\varepsilon^{u,h}\rangle_{\mathrm{dom}(A_p)} \mathrm{d}t.$$

□

We can directly write down an adjoint system for a solution $\overline{u}_\varepsilon$ of problem (3.1), (3.2), (3.7).

Theorem 3.13 (Adjoint system regularized problem)**.** *Adopt the assumptions of Theorem 3.9 and the notation from Lemma 3.12. For $i \in \{1,2\}$ and $\varepsilon \in (0, \varepsilon_*]$ let $\overline{u}_\varepsilon \in U_i$ be an optimal control for problem (3.1), (3.2), (3.7). Then the adjoint variables for $\overline{y}_\varepsilon \in \mathrm{Y}_{2,0}$ and $\overline{z}_\varepsilon \in \mathrm{H}^1(J_T)$ are given by $p_\varepsilon := p_\varepsilon^{\overline{y}_\varepsilon - y_d} \in \mathrm{Y}_{2,T}^*$ and $q_\varepsilon := q_\varepsilon^{\overline{y}_\varepsilon - y_d} \in \mathrm{H}^1(J_T)$. There holds $B_i^*(p_\varepsilon + S q_\varepsilon) = -(\kappa+1)\overline{u}_\varepsilon + \overline{u}$ in $\mathrm{L}^2(J_T; U_i)$ and the following system of evolution equations is satisfied by p_ε and q_ε:*

$$\begin{aligned} -\dot{p}_\varepsilon + A_p^* p_\varepsilon &= \left[\frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{y}_\varepsilon)\right]^* p_\varepsilon + S\left[-A_p + \frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon)\right] q_\varepsilon + \overline{y}_\varepsilon - y_d \quad \text{for } t \in J_T, \\ p_\varepsilon(T) &= 0, \end{aligned} \tag{3.16}$$

$$\begin{aligned} -\dot{q}_\varepsilon &= \langle p_\varepsilon, \frac{\partial}{\partial z} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon)\rangle_X + S \frac{\partial}{\partial z} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon) q_\varepsilon - \frac{1}{\varepsilon}\Psi''(\overline{z}_\varepsilon) q_\varepsilon \qquad \text{for } t \in J_T, \\ q_\varepsilon(T) &= 0. \end{aligned} \tag{3.17}$$

Proof. Note first that we can choose $q = q' = 2$ in Lemma 3.12 since $2 > \frac{1}{1-\alpha} \Leftrightarrow \alpha < \frac{1}{2}$ which is the case by (A2) in Assumption 2.11. Moreover, the expression $\langle \overline{y}_\varepsilon - y_d, y_\varepsilon^{B_i\overline{u}_\varepsilon, B_i h}\rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))} = \int_0^T \int_\Omega (\overline{y}_\varepsilon - y_d) \cdot y_\varepsilon^{B_i\overline{u}_\varepsilon, B_i h} \mathrm{d}x\mathrm{d}t$ is well-defined: With I_p as in Definition 2.5, $y_\varepsilon^{B_i\overline{u}_\varepsilon, B_i h} \in \mathrm{dom}(A_p) = \mathrm{ran}(I_p)$ may be identified with the embedding of $I_p^{-1} y_\varepsilon^{B_i\overline{u}_\varepsilon, B_i h}$ from $\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)$ into $\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega) \simeq X^*$. Note that $p' \le 2 \le p$. Since $(\mathcal{A}_p + I_p)^{-1} \in \mathcal{L}\big(X, \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)\big)$, see Remark 2.7, we can first estimate

$$\begin{aligned} \big\|I_p^{-1} y_\varepsilon^{B_i\overline{u}_\varepsilon, B_i h}(t)\big\|_{\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)} &\le \big\|I_p^{-1} y_\varepsilon^{B_i\overline{u}_\varepsilon, B_i h}(t)\big\|_{\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)} \\ &\le \left\|(\mathcal{A}_p + I_p)^{-1}\right\|_{\mathcal{L}\left(X, \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)\right)} \big\|y_\varepsilon^{B_i\overline{u}_\varepsilon, B_i h}(t)\big\|_{\mathrm{dom}(A_p)} \end{aligned}$$

for a.e. $t \in J_T$ and with the identification of $\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$ with X^* we conclude

$$\begin{aligned} \left|\int_0^T \int_\Omega (\overline{y}_\varepsilon - y_d) \cdot y_\varepsilon^{B_i\overline{u}_\varepsilon, B_i h} \mathrm{d}x\mathrm{d}t\right| &= \left|\int_0^T \langle I_p^{-1} y_\varepsilon^{B_i\overline{u}_\varepsilon, B_i h}, (\overline{y}_\varepsilon - B_1 y_d)\rangle_X \mathrm{d}t\right| \\ \le \|\overline{y}_\varepsilon - B_1 y_d\|_{\mathrm{L}^2(J_T;X)} &\left\|(\mathcal{A}_p + I_p)^{-1}\right\|_{\mathcal{L}\left(X, \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)\right)} \|y_\varepsilon^{B_i\overline{u}_\varepsilon, B_i h}\|_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))}. \end{aligned}$$

The Gâteaux-derivative of $\mathcal{J}_{\mathrm{reg}}(u) := J_{\mathrm{reg}}(G_\varepsilon(B_i u), u; \overline{u}) = J(G_\varepsilon(B_i u), u) + \frac{1}{2}\|u - \overline{u}\|_{U_i}^2$ with respect to u has to be zero at $\overline{u}_\varepsilon$ by optimality. Applying Lemma 3.12 we compute for $h \in U_i$:

$$\begin{aligned}
0 = \mathcal{J}'_{\mathrm{reg}}[\overline{u}_\varepsilon; h] &= \langle \overline{y}_\varepsilon - y_d, y_\varepsilon^{B_i\overline{u}_\varepsilon, B_i h}\rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))} + \kappa\langle \overline{u}_\varepsilon, h\rangle_{U_i} + \langle \overline{u}_\varepsilon - \overline{u}, h\rangle_{U_i} \\
&= \langle (p_\varepsilon + Sq_\varepsilon), B_i h\rangle_{\mathrm{L}^2(J_T;X)} + \langle (\kappa+1)u_\varepsilon - \overline{u}, h\rangle_{U_i} = \langle B_i^*(p_\varepsilon + Sq_\varepsilon) + (\kappa+1)u_\varepsilon - \overline{u}, h\rangle_{U_i}.
\end{aligned}$$

□

3.6. Estimates for the adjoints of the regularized problem

Similar to Section 3.5 of [9] and Lemma 4.14 of [31] we estimate the norms of the adjoint states p_ε and q_ε from Theorem 3.13 independently of ε and of the norms of the optimal controls $\overline{u}_\varepsilon$. In Section 4, we take a sequence $\{\varepsilon\}$ with $\varepsilon \to 0$ and apply those bounds to extract (weakly) converging subsequences of p_ε and q_ε. Those finally yield an adjoint system for problem (1.1)–(1.3), see Theorem 4.12.

Lemma 3.14 (Uniform bounds). *Adopt the assumptions and the notation of Theorem 3.13. 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*

$$0 \le \|q_\varepsilon\|_{\mathrm{C}(\overline{J_T})} + \frac{1}{\varepsilon}\int_0^T \Psi''(\overline{z}_\varepsilon(s))|q_\varepsilon(s)|\mathrm{d}s \le c, \tag{3.18}$$

$$\int_0^T |\dot{q}_\varepsilon(s)|\mathrm{d}s \le c, \tag{3.19}$$

$$\|p_\varepsilon\|_{Y_{2,T}^*} \le c, \tag{3.20}$$

$$\left\| \left[\frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon)\right]^* p_\varepsilon \right\|_{\mathrm{L}^2(J_T;[X^\alpha]^*)} \le c, \tag{3.21}$$

$$\left\| S\frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon) q_\varepsilon \right\|_{\mathrm{L}^2(J_T;[X^\alpha]^*)} \le c, \text{ as well as} \tag{3.22}$$

$$\|SA_p q_\varepsilon\|_{\mathrm{C}(\overline{J_T};[X^\alpha]^*)} \le c. \tag{3.23}$$

Proof. Firstly, Theorem 3.9 yields $\overline{u}_\varepsilon \to \overline{u}$ 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 in the proof of Theorem 3.13 we obtain that $\overline{y}_\varepsilon - y_d$ is bounded in $\mathrm{L}^2(J_T; [\mathrm{dom}(A_p)]^*)$ by $\|\overline{y}_\varepsilon - B_1 y_d\|_{\mathrm{L}^2(J_T;X)} \left\|(\mathcal{A}_p + I_p)^{-1}\right\|_{\mathcal{L}\left(X;\mathbb{W}^{1,p}_{\Gamma_D}(\Omega)\right)} =: c_0$. This constant can be estimated independently of ε because $\{\overline{y}_\varepsilon\}$ is uniformly bounded in $\mathrm{C}(\overline{J_T}; X)$. For any $\xi \in \mathrm{L}^2(J_T; X)$, Lemma 3.12 yields

$$\langle p_\varepsilon + Sq_\varepsilon, \xi\rangle_{\mathrm{L}^2(J_T;X)} = \langle \overline{y}_\varepsilon - y_d, y_\varepsilon^{B_i\overline{u}_\varepsilon,\xi}\rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))} \le c_0 C(\overline{y}_\varepsilon)\|\xi\|_{\mathrm{L}^2(J_T;X)}.$$

Because $\overline{y}_\varepsilon \to \overline{y}$ in $\mathrm{C}(\overline{J_T}; X^\alpha)$ we can find some $\varepsilon_0 > 0$ such that $C(\overline{y}_\varepsilon) = C(\overline{y})$ for all $\varepsilon \in (0, \varepsilon_0)$. From reflexivity of $\mathrm{L}^2(J_T; X)$ we conclude

$$\|p_\varepsilon + Sq_\varepsilon\|_{\mathrm{L}^2(J_T;X^*)} \le c_0 C(\overline{y}) =: c_1 \tag{3.24}$$

for all $\varepsilon \in (0, \varepsilon_0)$. We continue with estimates for q_ε. We test (3.17) with $q_\varepsilon/|q_\varepsilon|$, integrate from any $t \in J_T$ to T and apply (3.11) from Lemma 3.11 and (3.24) to obtain

$$\begin{aligned}
|q_\varepsilon(t)| + \frac{1}{\varepsilon}\int_t^T \Psi''(\overline{z}_\varepsilon(s))|q_\varepsilon(s)|\mathrm{d}s &= \int_t^T \langle p_\varepsilon(s) + Sq_\varepsilon(s), \frac{\partial}{\partial z} f_\varepsilon(\overline{y}_\varepsilon(s), \overline{z}_\varepsilon(s))\rangle_X \frac{q_\varepsilon(s)}{|q_\varepsilon(s)|}\,\mathrm{d}s \\
&\le c_1 \left\|\frac{\partial}{\partial z} f_\varepsilon(y_\varepsilon, \overline{z}_\varepsilon)\right\|_{\mathrm{L}^2(J_T;X)} \le c_1 K(\overline{y}_\varepsilon).
\end{aligned} \tag{3.25}$$

W.l.o.g. for the same ε_0 as before there holds $c_1 K(\overline{y}_\varepsilon) = c_1 K(\overline{y}) =: c_2$ for all $\varepsilon \in (0, \varepsilon_0)$. Note that $\Psi''(\overline{z}_\varepsilon) \geq 0$ by convexity of Ψ. This yields

$$0 \leq \|q_\varepsilon\|_{\mathrm{C}(\overline{J_T})} + \frac{1}{\varepsilon}\int_0^T \Psi''(\overline{z}_\varepsilon(s))|q_\varepsilon(s)|\mathrm{d}s \leq c_2 \tag{3.26}$$

for all $\varepsilon \in (0, \varepsilon_0)$. We conclude $Sq_\varepsilon \in \mathrm{L}^2(J_T; X^*)$ and then by (3.24) also $p_\varepsilon \in \mathrm{L}^2(J_T; X^*)$, both with a norm which is independent of $\varepsilon \in (0, \varepsilon_0)$. We continue by estimating

$$\int_0^T |\dot{q}_\varepsilon(s)|\mathrm{d}s \leq \int_0^T |\langle p_\varepsilon(s) + Sq_\varepsilon(s), \frac{\partial}{\partial z} f_\varepsilon(\overline{y}_\varepsilon(s), \overline{z}_\varepsilon(s))\rangle_X|\mathrm{d}s + \frac{1}{\varepsilon}\int_0^T \Psi''(\overline{z}_\varepsilon(s))|q_\varepsilon(s)|\mathrm{d}s.$$

Because of (3.25) the right side is bounded by $2c_2$ so that $\int_0^T |\dot{q}_\varepsilon(s)|\mathrm{d}s \leq 2c_2 =: c_3$ for $\varepsilon \in (0, \varepsilon_0)$.

To proceed, we exploit boundedness of the mapping $\left(-\frac{\mathrm{d}}{\mathrm{d}t} + A_p^*\right)^{-1} : \mathrm{L}^2(J_T; [\mathrm{dom}(A_p)]^*) \to Y_{2,T}^*$ and (3.16) to obtain

$$\begin{aligned}
\|p_\varepsilon\|_{Y_{2,T}^*} \leq{}& \left\|\left(-\frac{\mathrm{d}}{\mathrm{d}t} + A_p^*\right)^{-1}\right\|_{\mathcal{L}(\mathrm{L}^2(J_T;[\mathrm{dom}(A_p)]^*), Y_{2,T}^*)} \\
& \left\|\left[\frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon)\right]^* p_\varepsilon + S\left[-A_p + \frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon)\right] q_\varepsilon + \overline{y}_\varepsilon - y_d\right\|_{\mathrm{L}^2(J_T;[\mathrm{dom}(A_p)]^*)} \\
\leq{}& \left\|\left(-\frac{\mathrm{d}}{\mathrm{d}t} + A_p^*\right)^{-1}\right\|_{\mathcal{L}(\mathrm{L}^2(J_T;[\mathrm{dom}(A_p)]^*), Y_{2,T}^*)} \\
& \Bigg(\left\|\frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon)\right\|_{\mathcal{L}(\mathrm{L}^2(J_T;X^\alpha), \mathrm{L}^2(J_T;X))} \|p_\varepsilon + Sq_\varepsilon\|_{\mathrm{L}^2(J_T;X^*)} \\
& + \|SA_p\|_{[X^\alpha]^*}\|q_\varepsilon\|_{\mathrm{C}(\overline{J_T})} + \|\overline{y}_\varepsilon - y_d\|_{\mathrm{L}^2(J_T;[\mathrm{dom}(A_p)]^*)}\Bigg).
\end{aligned}$$

(3.11) from Lemma 3.11, (3.24), (3.26) and the bound $\|\overline{y}_\varepsilon - y_d\|_{\mathrm{L}^2(J_T;[\mathrm{dom}(A_p)]^*)} \leq c_0$ yield

$$\|p_\varepsilon\|_{Y_{2,T}^*} \leq \left\|\left(-\frac{\mathrm{d}}{\mathrm{d}t} + A_p^*\right)^{-1}\right\|_{\mathcal{L}(\mathrm{L}^2(J_T;[\mathrm{dom}(A_p)]^*), Y_{2,T}^*)} \left(c_1 K(\overline{y}) + \|SA_p\|_{[X^\alpha]^*} c_2 + c_0\right) =: c_4$$

for $\varepsilon \in (0, \varepsilon_0)$. In a similar way one obtains (3.21)–(3.23) from the estimates

$$\begin{aligned}
\left\|\left[\frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon)\right]^* p_\varepsilon\right\|_{\mathrm{L}^2(J_T;[X^\alpha]^*)} &\leq \left\|\frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon)\right\|_{\mathcal{L}(\mathrm{L}^2(J_T;X^\alpha), \mathrm{L}^2(J_T;X))} \|p_\varepsilon\|_{\mathrm{L}^2(J_T;X^*)}, \\
\left\|S\frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon) q_\varepsilon\right\|_{\mathrm{L}^2(J_T;[X^\alpha]^*)} &\leq \left\|\frac{\partial}{\partial y} f_\varepsilon(\overline{y}_\varepsilon, \overline{z}_\varepsilon)\right\|_{\mathcal{L}(\mathrm{L}^2(J_T;X^\alpha), \mathrm{L}^2(J_T;X))} \|Sq_\varepsilon\|_{\mathrm{L}^2(J_T;X^*)}
\end{aligned}$$

and

$$\|SA_p q_\varepsilon\|_{\mathrm{C}(\overline{J_T};[X^\alpha]^*)} \leq \|SA_p\|_{[X^\alpha]^*}\|q_\varepsilon\|_{\mathrm{C}(\overline{J_T})}.$$

□

4. Adjoint system and optimality conditions for the optimal control problem

As in Section 4 of [9] and Theorem 4.15 of [31] we are interested in the limit $\varepsilon \to 0$ in Theorem 3.13 to obtain an adjoint system for problem (1.1)–(1.3). In Sections 4.1–4.3 we study the general case with spatially distributed or boundary controls, *i.e.* $i \in \{1,2\}$. Particularly, in Section 4.1 we derive an adjoint system (p,q) for problem (1.1)–(1.3) for the optimal control $\overline{u}$ from Theorem 3.9, see Lemma 4.1. Moreover, we gather information about the continuity properties of q. Section 4.2 contains the optimality conditions for problem (1.1)–(1.3) for the optimal control $\overline{u}$ in terms of the pair p and q, see Lemma 4.11. In Section 4.3 we summarize the results from Sections 4.1–4.2 in Theorem 4.12. Afterwards, we consider the particular case when f is continuously differentiable. In Corollary 4.13 we improve the optimality condition (4.7) from Theorem 4.12 for this instance. Both optimality conditions (4.7) and (4.13) are restricted to test functions $y^{B_i\overline{u},B_ih}$ with $h \in U_i$, $i \in \{1,2\}$.

In Section 4.4 we focus on the setting when the controls act inside of Ω, *i.e.* on $i=1$. In Corollary 4.14 we improve the optimality conditions from Theorem 4.12 as well as those from Corollary 4.13 by extending inequalities (4.7) and (4.13) to any test function of the form $v\varphi$ with $v \in \mathrm{dom}(A_p)$, $Sv > 0$ and $\varphi \in \mathrm{C}_0^\infty(J_T)$. Dividing the corresponding inequality by Sv yields, at least in (4.13), an optimality condition with arbitrary test functions $\varphi \in \mathrm{C}_0^\infty(J_T)$. For $i=1$ we also prove uniqueness of p and q if f is continuously differentiable, see Corollary 4.15.

4.1. Adjoint system for distributed or boundary controls

In this section, we derive an adjoint system (p,q) for problem (1.1)–(1.3) and collect regularity properties of p and q. The evolution equation of p can be derived pretty much straight forward as the limit equation of (3.16) for $\varepsilon \to 0$, see Lemma 4.1. This is not possible for q. The reason is that in Lemma 3.14 we could bound the norm of $\dot{q}_\varepsilon$ independently of ε only in $\mathrm{L}^1(J_T)$. As a remedy we split the interval J_T into the set I_0 of times t at which the limit $\overline{z}(t)$ is contained in the open interval (a,b) and the rest I_∂ where $\overline{z}(t) \in \{a,b\}$. It turns out, that the evolution of q in I_0 can be described in form of an evolution equation, see Lemma 4.3. As for I_∂, we have to pass to weak-$*$ convergence of q_ε and consider the limit $\mathrm{d}\mu$ of $\frac{1}{\varepsilon}\Psi''(\overline{z}_\varepsilon)q_\varepsilon$ in $\mathrm{C}(\overline{J_T})^*$. Driving ε to zero yields an equality for $\mathrm{d}q$ in the sense of measures on I_∂, see Lemma 4.5. The abstract measure $\mathrm{d}\mu$, having support in I_∂, remains part of this evolution equation. It also appears in the optimality conditions for problem (1.1)–(1.3) in (4.7). In order to complete the description of q by analyzing the measure $\mathrm{d}\mu$, we will introduce a regularity Assumption 4.7 on $S\overline{y}(t)$ for $t \in I_\partial$. With this assumption, we can characterize $\mathrm{d}\mu$ in a subset of I_∂. This allows us to characterize q in open subintervals of I_∂ and we can prove continuity of q at so-called $(0,\partial)$-switching times, see Lemma 4.9. In Remark 4.10 we generalize Lemma 4.9 to the setting when Assumption 4.7 is not satisfied.

Lemma 4.1 (Adjoint system in the limit)**.** *Adopt the assumptions and the notation of Theorem 3.13. 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 3.9. 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 $\lambda_1, \lambda_2 \in \mathrm{L}^2(J_T;[X^\alpha]^*)$ such that as $k \to \infty$, $p_{\varepsilon_k} \rightharpoonup p$ in $Y^*_{2,T}$ and*

$$\left[\frac{\partial}{\partial y} f_{\varepsilon_k}(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k})\right]^* p_{\varepsilon_k} \rightharpoonup \lambda_1 \quad \text{in } \mathrm{L}^2(J_T;[X^\alpha]^*),$$

$$S\frac{\partial}{\partial y} f_{\varepsilon_k}(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k}) q_{\varepsilon_k} \rightharpoonup \lambda_2 \quad \text{in } \mathrm{L}^2(J_T;[X^\alpha]^*).$$

Moreover, there exists a function q which has bounded variation, i.e. *$q \in \mathrm{BV}(J_T)$, such that q_{ε_k} converges pointwise to q with $k \to \infty$. There holds $\mathrm{Var}(q) \leq \liminf_{\varepsilon_k \to 0} \mathrm{Var}(q_{\varepsilon_k})$. Alternatively, $\dot{q}_{\varepsilon_k} \to \mathrm{d}q$ weak-$*$ in $\mathrm{C}(\overline{J_T})^*$ with $k \to \infty$ for some signed regular Borel measure $\mathrm{d}q \in \mathrm{C}(\overline{J_T})^*$. The relation between q and $\mathrm{d}q$ is given by $q(t-) - q(s+) = \mathrm{d}q((s,t))$ and $q(t+) - q(s-) = \mathrm{d}q([s,t])$ for $[s,t] \subset \overline{J_T}$.*

The function p solves the evolution equation

$$-\dot{p} + A_p^* = \lambda_1 + \lambda_2 - SA_p q + \overline{y} - y_d \quad \text{for } t \in J_T,\ p(T) = 0. \tag{4.1}$$

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,

$$B_i^*(p + Sq) = -\kappa \overline{u} \quad \text{in } U_i. \tag{4.2}$$

Proof. Theorem 3.9 implies $u_\varepsilon \to \overline{u}$ 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 $\varepsilon \to 0$. By (3.20), (3.21) and (3.22) in Lemma 3.14, reflexivity of all spaces yields a subsequence $\{\varepsilon_k\}$ and some functions p, λ_1 and λ_2 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 $k \to \infty$. The condition $p(T) = 0$ is included in the definition of the space $Y_{2,T}^*$. From (3.19) we conclude that q_ε has bounded variation, *i.e.* $q_\varepsilon \in \mathrm{BV}(J_T)$, with a norm which is bounded independently of ε. This 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, $\dot{q}_{\varepsilon_k} \to \mathrm{d}q$ weak-* in $\mathrm{C}(\overline{J_T})^*$ with $k \to \infty$ for some signed regular Borel measure $\mathrm{d}q \in \mathrm{C}(\overline{J_T})^*$ and the relation between q and $\mathrm{d}q$ is given by $q(t-) - q(s+) = \mathrm{d}q((s,t))$ and $q(t+) - q(s-) = \mathrm{d}q([s,t])$ for $[s,t] \subset \overline{J_T}$ ([9], Sect. 4). We apply weak continuity of $-\frac{\mathrm{d}}{\mathrm{d}t} + A_p^*$ from $Y_{2,T}^*$ to $\mathrm{L}^2(J_T; [\mathrm{dom}(A_p)]^*)$ to conclude

$$\begin{aligned} 0 &= -\dot{p}_{\varepsilon_k} + A_p^* p_{\varepsilon_k} - \left[\frac{\partial}{\partial y} f_{\varepsilon_k}(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k})\right]^* p_{\varepsilon_k} + SA_p q_{\varepsilon_k} - S\frac{\partial}{\partial y} f_\varepsilon(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k}) q_{\varepsilon_k} - (\overline{y}_{\varepsilon_k} - y_d) \\ &\rightharpoonup -\dot{p} + A_p^* - \lambda_1 - \lambda_2 + SA_p q - (\overline{y} - y_d) \end{aligned}$$

in $\mathrm{L}^2(J_T; [\mathrm{dom}(A_p)]^*)$ with $k \to \infty$. Consequently, $p \in Y_{2,T}^*$ solves equation (4.1). Note that we can set $f_\varepsilon \equiv f$ if f is continuously differentiable from $X^\alpha \times \mathbb{R}$ into X and in this case $\lambda_1 = \left[\frac{\partial}{\partial y} f(\overline{y}, \overline{z})\right]^* p$ and $\lambda_2 = S\frac{\partial}{\partial y} f(\overline{y}, \overline{z}) q$. Moreover,

$$0 = B_i^*(p_{\varepsilon_k} + Sq_{\varepsilon_k}) + (\kappa + 1) u_{\varepsilon_k} - \overline{u} \rightharpoonup B_i^*(p + Sq) + \kappa \overline{u}$$

in U_i with $k \to \infty$ since B_i^* is weakly continuous . This implies (4.2). ☐

To obtain information about q from Lemma 4.1 we continue similar as in Section 4 of [9].

Definition 4.2 (Partition of J_T)**.** Consider $\overline{z}$ from Theorem 3.9. We split the interval $\overline{J_T}$ 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 $I_\partial^b := \{t \in \overline{J_T} : \overline{z}(t) = b\}$.

Note that I_0 is open because $\overline{z}$ is continuous.

Lemma 4.3 (q in I_0)**.** *Adopt the assumptions and the notation of Lemma 4.1 and consider the subdivision of $\overline{J_T}$ from Definition 4.2. For any interval $(c,d) \subset I_0$ the limit q in Lemma 4.1 belongs to $\mathrm{H}^1(c,d)$ and there exists some $\nu \in \mathrm{L}^2(J_T)$ such that $-\dot{q} = \nu$ in $\mathrm{L}^2(c,d)$. If f is continuously differentiable from $X^\alpha \times \mathbb{R}$ into X then $\nu = \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X$.*

Proof. By Theorem 3.9, $\overline{z}_\varepsilon \to \overline{z}$ uniformly in $\overline{J_T}$. Let $(c,d) \subset I_0$ and choose $[s,t] \subset (c,d)$ arbitrary. $(A4)_\varepsilon$ in Assumption 3.2 implies that (w.l.o.g for $\varepsilon_0 > 0$ from Lemma 3.14) $\Psi''(\overline{z}_\varepsilon) \equiv 0$ on $[s,t]$ for all $\varepsilon \in (0, \varepsilon_0)$. For

$\varepsilon \in (0, \varepsilon_0)$ we integrate from (s,t) in (3.17) in Theorem 3.13 and obtain

$$q_\varepsilon(t) - q_\varepsilon(s) = \int_s^t -\langle p_\varepsilon(s) + Sq_\varepsilon(s), \frac{\partial}{\partial z} f_\varepsilon(\overline{y}_\varepsilon(s), \overline{z}_\varepsilon(s))\rangle_X \mathrm{d}s.$$

Consider $\{\varepsilon_k\}$ from Lemma 4.1. Lemma 3.14 together with Lemma 3.11 imply 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, we obtain a subsequence of $\{\varepsilon_k\}$ (still denoted by $\{\varepsilon_k\}$) and functions $\nu_1, \nu_2 \in \mathrm{L}^2(J_T)$, such that $\langle p_{\varepsilon_k}, \frac{\partial}{\partial z} f_{\varepsilon_k}(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k})\rangle_X \rightharpoonup \nu_1$ and $\langle Sq_{\varepsilon_k}, \frac{\partial}{\partial z} f_{\varepsilon_k}(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k})\rangle_X \rightharpoonup \nu_2$ in $\mathrm{L}^2(J_T)$ with $k \to \infty$. If f is continuously differentiable from $X^\alpha \times \mathbb{R}$ into X we can set $f_\varepsilon \equiv f$ and get $\nu_1 = \langle p, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X$ and $\nu_2 = \langle Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X$. In the general case we obtain

$$q_{\varepsilon_k}(t) - q_{\varepsilon_k}(s) = \int_0^T -\langle p_{\varepsilon_k} + Sq_{\varepsilon_k}, \frac{\partial}{\partial z} f_{\varepsilon_k}(\overline{y}_{\varepsilon_k}, \overline{z}_{\varepsilon_k})\rangle_X \chi_{[s,t]} \mathrm{d}s \to \int_s^t -\nu_1 - \nu_2 \mathrm{d}s =: \int_s^t -\nu \mathrm{d}s$$

with $k \to \infty$. So the weak derivative of q exists in $\mathrm{L}^2(c,d)$ and is given by $-\nu$. □

Our next goal is to understand the behaviour of q in I_∂.

Lemma 4.4 (q in I_∂: Relation to $\mathcal{P}(S\overline{y})$)**.** *Adopt the assumptions and the notation of Lemma 4.1 and consider the subdivision of $\overline{J_T}$ from Definition 4.2. With $\mathcal{P} = \mathrm{Id} - \mathcal{W}$, cf. Lemma 2.10, there holds $\left[\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}[S\overline{y}](t)\right] q(t) = 0$ for a.e. $t \in I_\partial$.*

Proof. Consider Ψ from Definition 3.1 and c and ε_0 from Lemma 3.14. By Theorem 3.9, $\overline{z}_\varepsilon \to z$ uniformly so that $\overline{z}_\varepsilon(t) \to b$ for $t \in I_\partial^b$ and $\overline{z}_\varepsilon(t) \to a$ for $t \in I_\partial^a$ with $\varepsilon \to 0$. Hence, there exists some $\varepsilon_1 \in (0, \varepsilon_0]$ such that

$$a < \overline{z}_\varepsilon(t) < b + 1 \quad \text{for } t \in I_\partial^b \tag{4.3}$$

for all $\varepsilon \in (0, \varepsilon_1)$. Remember $\Psi_1(x) = (x-b)^3(4+b-x)$ and $\Psi \equiv 0$ on $[a,b]$. For $\varepsilon \in (0, \varepsilon_1)$ and $t \in I_\partial^b$ we obtain

$$\Psi'(\overline{z}_\varepsilon(t)) = \Psi_1'(\overline{z}_\varepsilon(t))\chi_{\{b<\overline{z}_\varepsilon\le b+2\}}(t) = 4(3 - (\overline{z}_\varepsilon(t) - b))(\overline{z}_\varepsilon(t) - b)^2 \chi_{\{b<\overline{z}_\varepsilon\le b+2\}}(t), \tag{4.4}$$
$$\Psi''(\overline{z}_\varepsilon(t)) = 12(\overline{z}_\varepsilon(t) - b)[2 - (\overline{z}_\varepsilon(t) - b)]\chi_{\{b<\overline{z}_\varepsilon\le b+2\}}(t). \tag{4.5}$$

We apply estimate (3.18) from Lemma 3.14 together with (4.3) and (4.5) to conclude

$$\begin{aligned} c &\ge \frac{1}{\varepsilon}\int_0^T \Psi''(\overline{z}_\varepsilon(s))|q_\varepsilon(s)|\mathrm{d}s \ge \frac{1}{\varepsilon}\int_{I_\partial^b} \Psi''(\overline{z}_\varepsilon(s))|q_\varepsilon(s)|\mathrm{d}s \\ &= \frac{1}{\varepsilon}\int_{I_\partial^b} 12(\overline{z}_\varepsilon(t) - b)[2 - (\overline{z}_\varepsilon(t) - b)]\chi_{\{b<\overline{z}_\varepsilon\le b+2\}}|q_\varepsilon(s)|\mathrm{d}s \\ &\ge \frac{1}{\varepsilon}\int_{I_\partial^b} 12(\overline{z}_\varepsilon(t) - b)\chi_{\{b<\overline{z}_\varepsilon\le b+2\}}|q_\varepsilon(s)|\mathrm{d}s \end{aligned} \tag{4.6}$$

for all $\varepsilon \in (0, \varepsilon_1)$. We apply the convergence results from Theorem 3.9 in (3.2) and the representation $\mathcal{W} + \mathcal{P} = \mathrm{Id}$ from Lemma 2.10 to obtain the weak convergence

$$\frac{1}{\varepsilon}\Psi'(\overline{z}_\varepsilon) = S\dot{\overline{y}}_\varepsilon - \dot{\overline{z}}_\varepsilon \rightharpoonup S\dot{\overline{y}} - \dot{\overline{z}} = \frac{\mathrm{d}}{\mathrm{d}t}(S\overline{y} - \mathcal{W}[S\overline{y}]) = \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}[S\overline{y}]$$

in $\mathrm{L}^2(J_T)$ with $\varepsilon \to 0$. Furthermore, by Lemma 4.1, $|q_{\varepsilon_k}| \to |q|$ strongly in $\mathrm{L}^2(J_T)$ with $k \to \infty$ and $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}[S\overline{y}] = \left|\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}[S\overline{y}]\right|$ a.e. in I_∂^b by definition of I_∂^b. This together with (4.4) and (4.6) yields

$$\begin{aligned}
0 &\leq \int_{I_\partial^b} \left|\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}[S\overline{y}]\right| |q(s)|\mathrm{d}s = \lim_{k\to\infty} \frac{1}{\varepsilon_k} \int_{I_\partial^b} \Psi'(\overline{z}_{\varepsilon_k}(s))|q_{\varepsilon_k}(s)|\mathrm{d}s \\
&= \lim_{k\to\infty} \frac{1}{\varepsilon_k} \int_{I_\partial^b} 4(3-(\overline{z}_\varepsilon(t)-b))(\overline{z}_\varepsilon(t)-b)^2 \chi_{\{b<\overline{z}_{\varepsilon_k}\leq b+2\}}|q_{\varepsilon_k}(s)|\mathrm{d}s \\
&\leq \lim_{k\to\infty} \frac{12}{\varepsilon_k} \int_{I_\partial^b} (\overline{z}_{\varepsilon_k}(s)-b)^2 \chi_{\{b<\overline{z}_{\varepsilon_k}\leq b+2\}}|q_{\varepsilon_k}(s)|\mathrm{d}s \leq c \lim_{k\to\infty} \sup_{s\in I_\partial^b} (\overline{z}_{\varepsilon_k}(s)-b) = 0.
\end{aligned}$$

Similar estimates for I_∂^a and the decomposition $I_\partial = I_\partial^a \cup I_\partial^b$ prove the statement. ☐

Next, we pass to the limit in (3.17) to obtain the following result:

Lemma 4.5 (q in I_∂: Relation to $\mathrm{d}\mu$)**.** *Adopt the assumptions and the notation of Lemma 4.1 and let ν be as in Lemma 4.3. Consider the subdivision of $\overline{J_T}$ from Definition 4.2. We denote $\mathrm{d}\mu_\varepsilon := \frac{1}{\varepsilon}\Psi''(\overline{z}_\varepsilon)q_\varepsilon$. There exists a measure $\mathrm{d}\mu \in \mathrm{C}(\overline{J_T})^*$, such that a subsequence $\{\mathrm{d}\mu_{\varepsilon_k}\}$ (w.l.o.g we may consider $\{\varepsilon_k\}$ from Lemma 4.1) converges weak-* to $\mathrm{d}\mu$ in $\mathrm{C}(\overline{J_T})^*$ with $k \to \infty$. The support of $\mathrm{d}\mu$ is contained in I_∂. For any $\varphi \in \mathrm{C}(\overline{J_T})$ there holds*

$$\int_0^T -\varphi(t)\mathrm{d}q(t) + \int_{I_\partial} \varphi(t)\mathrm{d}\mu(t) = \int_0^T \varphi(t)\nu(t)\mathrm{d}t.$$

This implies $\mathrm{d}\mu = \mathrm{d}q + \nu\mathrm{d}t$ as measures on I_∂.

Proof. By (3.18) in Lemma 3.14 the functions $\mathrm{d}\mu_\varepsilon$ are bounded in $\mathrm{L}^1(J_T)$ independently of ε for all $\varepsilon \in (0,\varepsilon_0)$. Consequently, a subsequence of $\{\mathrm{d}\mu_\varepsilon\}$ converges weak-* in $\mathrm{C}(\overline{J_T})^*$ to some measure $\mathrm{d}\mu$. By $(A4)_\varepsilon$ in Assumption 3.2 and the uniform convergence of $\overline{z}_\varepsilon$ to $\overline{z}$ there holds $\varphi\,\frac{1}{\varepsilon}\Psi''(\overline{z}_\varepsilon)q_\varepsilon \equiv 0$ as soon as ε is small enough, if $\varphi \in \mathrm{C}(\overline{J_T})$ has compact support in I_0. Therefore, the support of $\mathrm{d}\mu$ is contained in I_∂ ([9], p. 343). The other statements are shown similar as ([9], Lems. 4.6 and 4.7). ☐

There also holds:

Lemma 4.6 (Discontinuity properties of q)**.** *Adopt the assumptions and notation of Lemma 4.1. The absolute value of q can only jump downwards in reverse time. Consequently, for any $t \in \overline{J_T}$ there holds $|q(t-)| \leq |q(t+)|$ and $q(T-) = q(T) = 0$. Moreover, q is right continuous in $[0,T)$ and left continuous at T.*

Proof. From Lemma 4.1 we conclude that q_{ε_k} converges to q in $\mathrm{L}^1(J_T)$ and that $\mathrm{d}q_{\varepsilon_k} = \dot{q}_{\varepsilon_k}\mathrm{d}t$ converges to $\mathrm{d}q$ weak-* in $\mathrm{C}(\overline{J_T})^*$. From Chapter XII.7 of [40] it follows that q has bounded variation and that the limit is right continuous in $[0,T)$ and left continuous at T. The rest of the statements are shown just as ([9], Lem. 4.4). ☐

The unknown measure $\mathrm{d}\mu$ has support in I_∂ so that we only know the behaviour of the sum $-\mathrm{d}q + \mathrm{d}\mu$ in $\mathrm{C}(\overline{J_T})^*$ but not that of $\mathrm{d}q$ alone. In order to analyze q also in I_∂ we make the following regularity assumption, *cf.* ([9]

Assumption 4.7 (Regularity assumption)**.** Let $\overline{y}$ be as in Theorem 3.9 and consider the subdivision of $\overline{J_T}$ from Definition 4.2. We suppose that the function $\mathcal{P}[S\overline{y}]$ satisfies $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}[S\overline{y}] \neq 0$ 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.

In order to analyze the behaviour of q and $\mathrm{d}q$ in $\overline{I_0} \cap I_\partial$ we introduce the following categories of times as in [9]:

Definition 4.8 (Switching times)**.** Consider the subdivision of $\overline{J_T}$ from Definition 4.2. We call a time t a $(0,\partial)$-switching time if $t \in \overline{I_0} \cap I_\partial$ and if there exists 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 $(t, t+\varepsilon) \subset I_0$.

Lemma 4.9 (q at switching times)**.** *Adopt the assumptions and the notation of Lemma 4.1. If t is a $(0,\partial)$-switching time in the sense of Definition 4.8 and if Assumption 4.7 holds then there exits a constant $\varepsilon > 0$ 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 $q \equiv 0$ in $[c,d)$.*

Proof. Let $(c,d) \subset I_\partial$ be arbitrary and suppose that Assumption 4.7 holds. Then Lemma 4.4 implies $q(t) = 0$ for a.e. $t \in (c,d)$. By Lemma 4.6, q is right continuous in $[0,T)$ so that $q \equiv 0$ in $[c,d)$. Consequently, for every subinterval $[\beta,\gamma] \subset (c,d)$ we have $0 = q(\gamma-) - q(\beta+) = \mathrm{d}q((\beta,\gamma))$ so that $\mathrm{d}q = 0$ as a measure on (c,d). Again by Lemma 4.6 the absolute value of q can only jump downwards in reverse time. By Lemma 4.3, $q \in \mathrm{H}^1(e,c)$ for any interval $(e,c) \subset I_0$. Consequently, whenever an interval $(e,c) \subset I_0$ is followed by an interval $[c,d] \subset I_\partial$, then q is absolutely continuous on $[e,d)$.

Let t be a $(0,\partial)$-switching and consider $\varepsilon > 0$ such that $(t-\varepsilon,t) \subset I_0$ and $[t,t+\varepsilon) \subset I_\partial$. Then setting $e = t-\varepsilon$, $c = t$ and $d = t+\varepsilon$ proves the rest of the lemma.

□

Remark 4.10. In the setting of Lemma 4.9 one can prove even more about the continuity properties of q if f is continuously differentiable and for $i = 1$, even in absence of Assumption 4.7:

– Note first that when $t \in I_\partial$ is a $(\partial,0)$-switching time then q might jump at t no matter if Assumption 4.7 holds or not. If it q is continuous at t then under Assumption 4.7 there holds $q(t) = 0$. It is also possible to prove that q may only jump up at t if $\int_t^{t^-} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y},\overline{z})\rangle_X \,\mathrm{d}s > 0$, where either $t^- = t^-(t) \in (t,T] \cap I_\partial^a$ is (essentially) the first time in (t,T) for which there exists some $\varepsilon > 0$ such that $S\dot{\overline{y}} < 0$ a.e. in $(t^-, t^- + \varepsilon)$, or $t^- = T$. It can further be shown that the size of the jump is bounded by $\int_t^{t^-} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y},\overline{z})\rangle_X \,\mathrm{d}s$. Analogously, one can prove that q may only jump down at t if $\int_t^{t^+} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y},\overline{z})\rangle_X \,\mathrm{d}s < 0$, where either $t^+ = t^+(t) \in (t,T] \cap I_\partial^b$ is (essentially) the first time in (t,T) for which there exists some $\varepsilon > 0$ such that $S\dot{\overline{y}} > 0$ a.e. in $(t^+, t^+\varepsilon)$, or $t^+ = T$. In this case the size of the jump is bounded by $-\int_t^{t^+} \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y},\overline{z})\rangle_X \,\mathrm{d}s$.
– Other categories of times can be considered. Those include isolated times in I_0 or subintervals of I_∂ in which $S\dot{\overline{y}} = 0$ a.e. The latter can only occur if Assumption 4.7 does not hold. Also for those categories one can show sign conditions for $\mathrm{d}q$ and $\mathrm{d}\mu$ and upper bounds for jumps.

The proof of these continuity properties is very technical and exceeds the scope of this work. The results will be published in the dissertation of the author.

4.2. Optimality conditions for distributed or boundary controls

We derive optimality conditions for problem (1.1)–(1.3) for the optimal control $\overline{u}$ from Theorem 3.9 in terms of the pair p and q from Lemma 4.1. We can not expect a pointwise maximum condition as in Section 5 from [31] since the hysteresis and its derivative, and then also $F'[\overline{y},\cdot]$ in Theorem 2.14, act non-local in time. This implies that if for some direction $\zeta \in \mathrm{C}(\overline{J_T}; X^\alpha)$ and some set $I \subset J_T$ of positive measure $F'[y;\zeta](\tau) = f'[(y(\tau), \mathcal{W}[Sy](\tau)); (\zeta(\tau), \mathcal{W}'[Sy; S\zeta](\tau))] \neq 0$ for all $\tau \in I$, then the values of the derivative in I might have an influence on its value at any t with $\max\{\tau \in I\} < t \leq T$. That is, we can only expect an optimality condition for problem (1.1)–(1.3) which includes integration at least over a part of the time interval J_T. Nevertheless, we follow the steps in Section 5 from [31] as long as possible. The optimality condition for $i \in \{1,2\}$ is derived in Lemma 4.11 and improved in Corollary 4.13 for the case when f is continuously differentiable. We can even further improve this condition for the case when the controls act inside of Ω, *i.e.* for $i = 1$. Also in this case we cannot expect to obtain an inequality without integration in time. But since the range of B_1 is dense in

X, we are able to derive a condition without variation in space. The result can be found in Corollary 4.14 in Section 4.4.1. For $i = 1$ we are also able to prove uniqueness of p, q and $\mathrm{d}\mu$ if f is continuously differentiable, see Corollary 4.15 in Section 4.4.2.

Because the range of B_2 is not dense in X, we treat the general case $i \in \{1, 2\}$ first.

Lemma 4.11 (Optimality condition)**.** *Adopt the assumptions and the notation of Lemma 4.1 and let ν be as in Lemma 4.3. 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 Thm. 2.14), there holds the optimality condition*

$$\begin{aligned}
&\int_0^T \langle \lambda_1 + \lambda_2 + S\nu, y^{B_i\overline{u},B_ih}\rangle_{\mathrm{dom}(A_p)}\mathrm{d}t \\
&\le \int_{I_\partial} Sy^{B_i\overline{u},B_ih}\mathrm{d}\mu + \int_0^T \langle p + Sq, F'[\overline{y}; y^{B_i\overline{u},B_ih}]\rangle_X \mathrm{d}t.
\end{aligned} \tag{4.7}$$

Proof. Since $\overline{u}$ is an optimal control, the directional derivative of the reduced cost function $\mathcal{J}$ has to be greater or equal than zero in each direction. With $y^{B_i\overline{u},B_ih} = G'[B_i\overline{u}; B_ih]$ this implies for any $h \in U_i$ that

$$0 \le \mathcal{J}'[\overline{u}; h] = \langle \overline{y} - y_d, y^{B_i\overline{u},B_ih}\rangle_{\mathrm{L}^2(J_T;\mathrm{dom}(A_p))} + \kappa\langle \overline{u}, h\rangle_{U_i}. \tag{4.8}$$

The function $y^{B_i\overline{u},B_ih}$ solves the evolution equation (2.4) in Theorem 2.14 with y replaced by $\overline{y}$ and h replaced by B_ih. We test this equation with $p + Sq$, integrate over time and apply (4.2) to compute

$$\begin{aligned}
&\int_0^T \langle p + Sq, \dot{y}^{B_i\overline{u},B_ih} + A_py^{B_i\overline{u},B_ih}\rangle_X \mathrm{d}t - \int_0^T \langle p + Sq, F'[\overline{y}; y^{B_i\overline{u},B_ih}]\rangle_X \mathrm{d}t \\
&= \int_0^T \langle p + Sq, B_ih\rangle_X \mathrm{d}t = -\kappa\langle \overline{u}, h\rangle_{U_i}.
\end{aligned} \tag{4.9}$$

We integrate the first term on the left side of (4.9) by parts, insert (4.1) from Lemma 4.1 and use the representation of $\mathrm{d}q$ from Lemma 4.5 to observe

$$\begin{aligned}
&\int_0^T \langle p + Sq, \dot{y}^{B_i\overline{u},B_ih} + A_py^{B_i\overline{u},B_ih}\rangle_X \mathrm{d}t \\
&= \int_0^T \langle \lambda_1 + \lambda_2 - SA_pq + \overline{y} - y_d, y^{B_i\overline{u},B_ih}\rangle_{\mathrm{dom}(A_p)}\mathrm{d}t \\
&\quad - \int_0^T Sy^{B_i\overline{u},B_ih}\mathrm{d}q + \int_0^T \langle SA_pq, y^{B_i\overline{u},B_ih}\rangle_{\mathrm{dom}(A_p)}\mathrm{d}t \\
&= \int_0^T \langle \lambda_1 + \lambda_2 + \overline{y} - y_d, y^{B_i\overline{u},B_ih}\rangle_{\mathrm{dom}(A_p)}\mathrm{d}t - \int_0^T Sy^{B_i\overline{u},B_ih}\mathrm{d}q \\
&= \int_0^T \langle \lambda_1 + \lambda_2 + \overline{y} - y_d, y^{B_i\overline{u},B_ih}\rangle_{\mathrm{dom}(A_p)}\mathrm{d}t - \int_{I_\partial} Sy^{B_i\overline{u},B_ih}\mathrm{d}\mu + \int_0^T \nu Sy^{B_i\overline{u},B_ih}\mathrm{d}t.
\end{aligned} \tag{4.10}$$

We insert (4.9) into (4.8) and apply (4.10) to obtain

$$
\begin{aligned}
0 &\le \int_0^T \langle \overline{y} - y_d, y^{B_i\overline{u},B_ih}\rangle_{\mathrm{dom}(A_p)} \mathrm{d}t + \kappa\langle \overline{u}, h\rangle_{U_i} \\
&= -\int_0^T \langle \lambda_1 + \lambda_2, y^{B_i\overline{u},B_ih}\rangle_{\mathrm{dom}(A_p)}\mathrm{d}t + \int_{I_\partial} Sy^{B_i\overline{u},B_ih}\mathrm{d}\mu - \int_0^T \nu S y^{B_i\overline{u},B_ih}\mathrm{d}t \\
&\quad + \int_0^T \langle p + Sq, F'[\overline{y}; y^{B_i\overline{u},B_ih}]\rangle_X \mathrm{d}t.
\end{aligned}
$$

□

4.3. Summary: adjoint system and optimality conditions for distributed or boundary controls

We summarize our results for the general control problem with $i \in \{1, 2\}$.

Theorem 4.12 (Adjoint system and optimality condition)**.** *Let Assumptions 2.11 and 3.2 hold. For $i \in \{1,2\}$ suppose that $\overline{u} \in U_i$ is an optimal control for problem (1.1)–(1.3) 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 4.2. Then there exist adjoint states $p \in Y^*_{2,T}$ and $q \in \mathrm{BV}(J_T)$ of the following kind: There holds $B_i^*(p + Sq) = -\kappa\overline{u}$ in U_i. For some functions $\lambda_1, \lambda_2 \in \mathrm{L}^2(J_T; [X^\alpha]^*)$ we have*

$$
-\dot{p} + A_p^* p = \lambda_1 + \lambda_2 - SA_p q + \overline{y} - y_d \quad \textit{for } t \in J_T,\ p(T) = 0.
$$

q is right continuous in J_T, left continuous at T and absolutely continuous in I_0. There exists $\nu \in \mathrm{L}^2(J_T)$ such that q solves $-\dot{q} = \nu$ in every open subinterval of I_0. $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}[S\overline{y}](t)q(t) = 0$ for a.e. $t \in I_\partial$ and there exists a measure $\mathrm{d}\mu \in \mathrm{C}(\overline{J_T})^*$ with support in I_∂ such that $\mathrm{d}\mu = \mathrm{d}q + \nu\mathrm{d}t$ as measures on I_∂. For all $h \in U_i$ and $y^{B_i\overline{u},B_ih} = G'[B_i\overline{u}; B_ih]$ (see Theorem 2.14) there holds the optimality condition

$$
\begin{aligned}
&\int_0^T \langle \lambda_1 + \lambda_2 + S\nu, y^{B_i\overline{u},B_ih}\rangle_{\mathrm{dom}(A_p)}\mathrm{d}t \\
&\le \int_{I_\partial} Sy^{B_i\overline{u},B_ih}\mathrm{d}\mu + \int_0^T \langle p + Sq, F'[\overline{y}; y^{B_i\overline{u},B_ih}]\rangle_X \mathrm{d}t,
\end{aligned} \tag{4.11}
$$

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 $|q(t-)| \le |q(t+)|$ for all $t \in \overline{J_T}$. If the regularity Assumption 4.7 is valid then q is continuous at every $(0, \partial)$-switching time t (see Def. 4.8) with $q(t) = 0$. In this case, $q \equiv 0$ on $[c, d)$ holds for every open interval $(c, d) \subset I_\partial$.

We can improve the results of Theorem 4.12 if f is continuously differentiable:

Corollary 4.13 (Adjoint system and optimality condition for regular f)**.** *Let Assumptions 2.11 and 3.2 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 (1.1)–(1.3) 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 4.2. Then there exist adjoint states $p \in Y^*_{2,T}$ and $q \in \mathrm{BV}(J_T)$ of the following kind: There holds $B_i^*(p + Sq) = -\kappa\overline{u}$ in U_i. We have*

$$
-\dot{p} + A_p^* p = \left[\frac{\partial}{\partial y} f(\overline{y}, \overline{z})\right]^* (p + Sq) - SA_p q + \overline{y} - y_d \quad \textit{for } t \in J_T,\ p(T) = 0. \tag{4.12}
$$

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. $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}[S\overline{y}](t)q(t) = 0$ for a.e. $t \in I_\partial$ and there exists a measure $\mathrm{d}\mu \in \mathrm{C}(\overline{J_T})^*$ with support in I_∂ such that $\mathrm{d}\mu = \mathrm{d}q + \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X \mathrm{d}t$ as measures on I_∂. For all $h \in U_i$, $y^{B_i\overline{u},B_ih} = G'[B_i\overline{u}; B_ih]$ (see Thm. 2.14) and $\mathcal{P} = \mathrm{Id} - \mathcal{W}$ (see Lemma 2.10) there holds the optimality condition

$$\int_0^T \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X \mathcal{P}'[S\overline{y}; Sy^{B_i\overline{u},B_ih}]\mathrm{d}t \leq \int_{I_\partial} Sy^{B_i\overline{u},B_ih}\mathrm{d}\mu. \tag{4.13}$$

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 4.7 is valid then q is continuous at every $(0, \partial)$-switching time t (see Def. 4.8) with $q(t) = 0$. In this case, $q \equiv 0$ on $[c, d)$ holds for every open interval $(c, d) \subset I_\partial$.

Proof. If f is continuously differentiable then according to Lemma 4.1 and Lemma 4.3 we can replace $\lambda_1 = \left[\frac{\partial}{\partial y} f(\overline{y}, \overline{z})\right]^* p$, $\lambda_2 = S\frac{\partial}{\partial y} f(\overline{y}, \overline{z})q$ and $\nu = \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X$ in Theorem 4.12. This yields all statements except for the optimality condition. (4.7) takes the form

$$\begin{aligned}
&\int_0^T \langle \left[\frac{\partial}{\partial y} f(\overline{y}, \overline{z})\right]^* (p + Sq), y^{B_i\overline{u},B_ih}\rangle_{\mathrm{dom}(A_p)} + \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X Sy^{B_i\overline{u},B_ih}\mathrm{d}t \\
&\leq \int_{I_\partial} Sy^{B_i\overline{u},B_ih}\mathrm{d}\mu + \int_0^T \langle p + Sq, \frac{\partial}{\partial y} f(\overline{y}, \overline{z})y^{B_i\overline{u},B_ih} + \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\mathcal{W}'[S\overline{y}; Sy^{B_i\overline{u},B_ih}]\rangle_X \mathrm{d}t.
\end{aligned}$$

Because $\mathcal{P} = \mathrm{Id} - \mathcal{W}$ (see Lem. 2.10) we have $Sy^{B_i\overline{u},B_ih} - \mathcal{W}'[S\overline{y}; Sy^{B_i\overline{u},B_ih}] = \mathcal{P}'[S\overline{y}; Sy^{B_i\overline{u},B_ih}]$. This yields the optimality condition (4.13). □

4.4. Improved optimality conditions and uniqueness for distributed controls

We want to replace $y^{B_i\overline{u},B_ih}$ in (4.11) and (4.13) by an arbitrary function of an appropriate space. This would certainly improve the optimality conditions in Theorem 4.12 and Corollary 4.13. It is not possible in the general case $i \in \{1, 2\}$ without density of the range of B_i. Therefore, we restrict ourselves to problem (1.1)–(1.3) with distributed controls $u \in U_1$ in this section. Suppose that p in (A1) in Assumption 2.11 is chosen close to two such that $\frac{1}{2} > 1 - \frac{1}{p} - \frac{1}{d}$. Then $2 < \frac{dp'}{d-p'}$ and by Remark 2.7 from [32] we have the compact embedding $\mathbb{W}^{1,p'}_{\Gamma_D}(\Omega) \hookrightarrow [\mathrm{L}^2(\Omega)]^m$ which is also one-to-one. That is, B_1 has dense range. In Corollary 4.14 in Section 4.4.1 we improve the optimality conditions from Theorem 4.12 and Corollary 4.13 for this case. For $i = 1$ we also prove uniqueness of p, q and $\mathrm{d}\mu$ if f is continuously differentiable, see Corollary 4.15 in Section 4.4.2.

4.4.1. Improved optimality conditions

We improve the optimality conditions (4.11) and (4.13).

Corollary 4.14 (Optimality condition for distributed controls)**.** *Let Assumptions 2.11 and 3.2 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 (1.1)–(1.3) 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 (4.11) in Theorem 4.12 there holds*

$$\begin{aligned}
&\int_0^T \langle \lambda_1 + \lambda_2, \frac{v}{Sv}\varphi\rangle_{\mathrm{dom}(A_p)} + \nu\varphi\mathrm{d}t \\
&\leq \int_{I_\partial} \varphi\mathrm{d}\mu + \int_0^T \langle p + Sq, f'[(\overline{y}, \overline{z}); ((v/Sv)\varphi, \mathcal{W}'[S\overline{y}; \varphi])]\rangle_X \mathrm{d}t.
\end{aligned}$$

If f is continuously differentiable then in addition to (4.13) in Corollary 4.13 there holds

$$\int_0^T \langle p + Sq, \frac{\partial}{\partial z} f(\overline{y}, \overline{z})\rangle_X \mathcal{P}'[S\overline{y}; \varphi] \mathrm{d}t \leq \int_{I_\partial} \varphi \,\mathrm{d}\mu \quad \textit{for all } \varphi \in \mathrm{C}_0^\infty(J_T).$$

Proof. Since B_1 has dense range one proves just as in Lemma 5.2 of [31] that the set $\{y^{B_1\overline{u}, B_1 h} : h \in U_1\}$ is dense in $Y_{2,0}$. Unfortunately, we can not continue as in Theorem 5.3 of [31] to derive a pointwise optimality condition. The reason is that for any $\zeta \in \mathrm{C}(\overline{J_T}; X^\alpha)$ the function $\mathcal{W}'[S\overline{y}; S\zeta]$ is non-local in time. Nevertheless, we can still make use of the fact that $\mathcal{W}'[S\overline{y}; \cdot]$ and f' are positive homogeneous. For arbitrary given $\eta \in Y_{2,0}$, we choose a sequence in $\{y^{B_1\overline{u}, B_1 h} : h \in U_1\}$ which converges to η. We pass to the limit in (4.11) and obtain

$$\int_0^T \langle \lambda_1 + \lambda_2 + S\nu, \eta\rangle_{\mathrm{dom}(A_p)} \mathrm{d}t \leq \int_{I_\partial} S\eta \mathrm{d}\mu + \int_0^T \langle p + Sq, F'[\overline{y}; \eta]\rangle_X \mathrm{d}t,$$

where $F'[\overline{y}; \eta](t) = f'[(\overline{y}(t), \mathcal{W}[S\overline{y}](t)); (\eta(t), \mathcal{W}'[S\overline{y}; S\eta](t))]$. Let $v \in \mathrm{dom}(A_p)$ with $Sv > 0$ be given. Furthermore, let $\varphi \in \mathrm{C}_0^\infty(J_T)$ be arbitrary. Then $v\varphi \in Y_{2,0}$ and $\mathcal{W}'[S\overline{y}; S(v\varphi)] = Sv\mathcal{W}'[S\overline{y}; \varphi]$. Setting $\eta = \varphi v$ and rearranging yields

$$\begin{aligned}
&\int_0^T \langle \lambda_1 + \lambda_2, \varphi v\rangle_{\mathrm{dom}(A_p)} + \varphi\nu S v \mathrm{d}t \\
&\leq \int_{I_\partial} \varphi S v \mathrm{d}\mu + \int_0^T \langle p + Sq, f'[(\overline{y}, \overline{z}); (v\varphi, Sv\mathcal{W}'[S\overline{y}; \varphi])]\rangle_X \mathrm{d}t.
\end{aligned}$$

Dividing both sides by Sv proves the first statement. The second inequality is shown analogously. ☐

4.4.2. Uniqueness of the adjoint variables

If f is continuously differentiable we can also show uniqueness of the adjoint couple.

Corollary 4.15 (Unique adjoint system for distributed controls)**.** *Let Assumptions 2.11 and 3.2 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 (1.1)–(1.3) 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 4.13 the adjoint couple $p \in Y_{2,T}^*$ and $q \in \mathrm{BV}(J_T)$ together with the measure $\mathrm{d}\mu$ in $\mathrm{C}(\overline{J_T})^*$ are unique.*

Proof. Because B_1 has dense range we have $\ker(B_1^*) = \overline{\mathrm{ran}(B_1)}^\perp = \{0\}$. Therefore, Corollary 4.13 implies

$$p + Sq = -\kappa(B_1^*)^{-1}\overline{u} \quad \text{in } X^* \text{ a.e. in } J_T, \tag{4.14}$$

cf. Theorem 4.15 from [31]. Suppose that two adjoint couples $(p_1, q_1), (p_2, q_2)$ exist which satisfy the conditions of Corollary 4.13. Let $\zeta \in \mathrm{L}^2(J_T; \mathrm{dom}(A_p))$ be arbitrary. Then by (4.12) and (4.14) there holds

$$\begin{aligned}
&\langle \dot{p}_2 - \dot{p}_1, \zeta\rangle_{\mathrm{L}^2(J_T; \mathrm{dom}(A_p))} \\
&= \langle \left[\frac{\partial}{\partial y} f(\overline{y}, \overline{z})\right]^* (p_1 + Sq_1 - (p_2 + Sq_2)) - A_p^*(p_2 - p_1) - SA_p(q_2 - q_1), \zeta\rangle_{\mathrm{L}^2(J_T; \mathrm{dom}(A_p))} \\
&= \langle p_1 + Sq_1 - (p_2 + Sq_2), \frac{\partial}{\partial y} f(\overline{y}, \overline{z})\zeta\rangle_{\mathrm{L}^2(J_T; X)} - \langle p_2 + Sq_2 - (p_1 + Sq_1), A_p\zeta\rangle_{\mathrm{L}^2(J_T; X)} = 0.
\end{aligned}$$

This implies $\dot p_2 = \dot p_1$ in $\mathrm{L}^2(J_T; [\mathrm{dom}(A_p)]^*)$. Together with $p_1(T) = p_2(T) = 0 \in [\mathrm{dom}(A_p)]^*$ we obtain $p_1 = p_2$ in $\mathrm{L}^2(J_T; [\mathrm{dom}(A_p)]^*)$. Since the embedding $\mathrm{dom}(A_p) \hookrightarrow X$ is dense, the embedding of X^* into $[\mathrm{dom}(A_p)]^*$ is one-to-one and $p_1 = p_2$ also in $\mathrm{L}^2(J_T; X^*)$ and then in $Y^*_{2,T}$. Let $v \in \mathrm{dom}(A_p)$ be given with $Sv > 0$. We already know $p_1 = p_2$ so that $S(q_1 - q_2) = 0$ in X^* a.e. in J_T because of (4.14). Hence,

$$q_1 - q_2 = \frac{(q_1 - q_2)Sv}{Sv} = \frac{\langle S(q_1 - q_2), v\rangle_X}{Sv} = 0 \quad \text{in } X^* \text{ a.e. in } J_T,$$

so that $q_1 = q_2$ in $\mathrm{L}^1(J_T)$. This way we obtain

$$\int_{[0,T]} |\mathrm{d}q_1 - \mathrm{d}q_2| = \sup \left\{ \int_0^T (q_1 - q_2)\dot\varphi \mathrm{d}t : \varphi \in \mathrm{C}^1_0(J_T),\ |\varphi| \le 1 \right\} = 0,$$

which implies $\mathrm{d}q_1 - \mathrm{d}q_2 = 0$ as measures on $\overline{J_T}$ according to page XII.7 of [40]. This yields $q_1 = q_2 \in \mathrm{BV}(0,T)$. From Corollary 4.13 we conclude $\mathrm{d}\mu_1 = \mathrm{d}\mu_2$ and the proof is complete. ☐

5. Higher regularity of the solutions of the optimal control problem

In this section we improve the regularity of the optimal control $\overline{u} \in U_i$, $i \in \{1,2\}$, and then also of the optimal state $\overline{y} = G(B_i\overline{u})$ and $\overline{z} = \mathcal{W}[S\overline{y}]$. We denote $\tilde U_1 := [\mathrm{L}^2(\Omega)]^m$ and $\tilde U_2 := \prod_{j=1}^m \mathrm{L}^2(\Gamma_{N_j}, \mathcal{H}_{d-1})$. We want to exploit the equation $B_i^*(p + Sq) = -\kappa\overline{u}$ in $[\tilde U_i]^*$ a.e. in J_T which follows from Theorem 4.12. In order to apply the time-regularity of $p + Sq$ we need to enforce the conditions on B_i.

Assumption 5.1. For $i \in \{1,2\}$, the operator $B_i : \tilde U_i \to X$ in (N5) 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 5.2 (Higher regularity)**.** *In the setting of Theorem 4.12 let Assumption 5.1 hold for some $\gamma \in (0,1]$.*

If $\gamma > \frac12$, 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 $d < p$, which is the case when $d = 2$ and $p > 2$ in (A1) in Assumption 2.11, 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 $\gamma \le \frac12$, 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 $\theta \in (0, \frac12 + \gamma)$. If $\gamma > \frac{d}{2p}$, with d and p from (A1) in Assumption 2.11, 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. First note that for $0 \le \beta \le \gamma \le 1$ we have the compact and dense embeddings $X^\gamma \hookrightarrow X^\beta \hookrightarrow X$ ([25], Thm. 1.4.8). This implies $X^* \hookrightarrow [X^\gamma]^* \hookrightarrow [X^\beta]^*$. By standard results on complex interpolation and Remark 2.7 there holds

$$[[\mathrm{dom}(A)]^*, X^*]_{1-\gamma} = [X^*, [\mathrm{dom}(A_p)]^*]_\gamma = [X, \mathrm{dom}(A_p)]^*_\gamma \simeq [X^\gamma]^*. \tag{5.1}$$

– We prove statement when Assumption 5.1 is fulfilled with $\gamma > \frac12$:
Since $1 - \gamma < \frac12$ we obtain (as in Rem. 2.8) an embedding

$$Y^*_{2,T} \subset \mathrm{H}^1(J_T; [\mathrm{dom}(A)]^*) \cap \mathrm{L}^2(J_T; X^*) \hookrightarrow \mathrm{C}(\overline{J_T}; [[\mathrm{dom}(A)]^*, X^*]_{1-\gamma}).$$

Therefore, by (5.1) the regularity $p \in Y^*_{2,T}$ in Theorem 4.12 implies that we can identify the function $p \in \mathrm{L}^2(J_T; X^*)$ and the representative $\tilde p$ of p in $\mathrm{C}(\overline{J_T}; [X^\gamma]^*)$. This allows the identification of $B_i^* p \in \mathrm{L}^2(J_T; [\tilde U_i]^*)$ and $\tilde B_i^* \tilde p \in \mathrm{C}(\overline{J_T}; [\tilde U_i]^*)$. We also have $Sq \in \mathrm{L}^\infty(J_T; X^*)$ since $q \in \mathrm{BV}(J_T)$ by Theorem 4.12 and because

$S \in X^*$ by (A3) in Assumption 2.11. That is, $B_i^* Sq \in \mathrm{L}^\infty(J_T; [\tilde{U}_i]^*)$. Again with Theorem 4.12 and we arrive at

$$\tilde{B}_i^* \tilde{p} + B_i^* Sq = B_i^*(p + Sq) = -\kappa \overline{u} \quad \text{in } [\tilde{U}_i]^*, \text{ a.e. in } J_T.$$

Both functions on the left side are contained in $\mathrm{L}^\infty(J_T; [\tilde{U}_i]^*)$. We identify $[\tilde{U}_i]^*$ with $\tilde{U}_i$, so that $\overline{u} \in \mathrm{L}^\infty(J_T; \tilde{U}_i)$. Now we use the higher regularity of $\overline{u}$ to prove a better regularity also for $\overline{y}$. Since $\overline{u} \in \mathrm{L}^\infty(J_T; \tilde{U}_i)$, Theorem 2.14 yields $\overline{y} \in Y_{s,0}$ for arbitrary $s \in (1, \infty)$. From Remarks 2.7 and 2.8 we obtain $\overline{y} \in \mathrm{C}(\overline{J_T}; X^\theta)$ for arbitrary $\theta \in [0, 1)$. According to Theorem 3.3 from [37], X^θ is a subset of $[\mathrm{L}^\infty(\Omega)]^m$ if $\theta > \frac{1}{2}(1 + \frac{d}{p})$. By Remark 2.7 we are guaranteed that we can choose $p > 2$. So at least if $d = 2$ there is some $\theta \in (0, 1)$ with $\theta > \frac{1}{2}(1 + \frac{d}{p})$ and therefore $\overline{y} \in \mathrm{C}(\overline{J_T}; [\mathrm{L}^\infty(\Omega)]^m)$. If $d = 2$ and $p > 2$ and if Ω is regular enough, for example a Lipschitz domain, then by Theorem 4.5 of [16] the state $\overline{y}$ is even Hölder continuous in time and space.

– We prove the statement for the case when Assumption 5.1 is fulfilled with $\gamma \le \frac{1}{2}$:
From Theorem 3 and (22) of [1] it follows

$$Y_{2,T}^* \subset \mathrm{H}^1(J_T; [\mathrm{dom}(A)]^*) \cap \mathrm{L}^2(J_T; X^*) \hookrightarrow \mathrm{L}^{\frac{2}{1-2s}}(J_T; [[\mathrm{dom}(A)]^*, X^*]_{1-\gamma})$$

for arbitrary $s \in (0, \gamma)$. Hence, the regularity $p \in Y_{2,T}^*$ in Theorem 4.12 together with (5.1) implies that we can identify $p \in \mathrm{L}^2(J_T; X^*)$ and the representative $\tilde{p}$ of p in $\mathrm{L}^{\frac{2}{1-2s}}(J_T; [X^\gamma]^*)$ and then $B_i^* p \in \mathrm{L}^2(J_T; [\tilde{U}_i]^*)$ and $\tilde{B}_i^* \tilde{p} \in \mathrm{L}^{\frac{2}{1-2s}}(J_T; [\tilde{U}_i]^*)$. We proceed as for the case $\gamma > \frac{1}{2}$ to prove $\overline{u} \in \mathrm{L}^{\frac{2}{1-2s}}(J_T; \tilde{U}_i)$ for arbitrary $s \in (0, \gamma)$. Theorem 2.14 yields $\overline{y} \in Y_{2/(1-2s),0}$ for arbitrary $s \in (0, \gamma)$ and from Remarks 2.7 and 2.8 it follows $\overline{y} \in \mathrm{C}(\overline{J_T}; X^\theta)$ for arbitrary $\theta \in \big[0, 1 - \left(\frac{2}{1-2s}\right)^{-1}\big) = [0, \frac{1}{2} + s)$. Because $s \in (0, \gamma)$ is arbitrary, this holds for all $\theta \in [0, \frac{1}{2} + \gamma)$. The remaining statements are shown just as for $\gamma > \frac{1}{2}$. □

Remark 5.3. For example, take $d = 2$ and $p > 2$ in (A1) in Assumption 2.11 and adopt the assumptions and the notation from Theorem 4.12. By Remark 2.7 of [32] we have the compact embedding

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

because $p' > 1$ and then $2 < \frac{dp'}{d-p'} = \frac{2p'}{2-p'}$. Therefore we can interpret functions $u \in \tilde{U}_1$ as elements $\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega)$ by the assignment $\int_\Omega u \cdot v \,\mathrm{d}x$, $\forall v \in \mathbb{W}^{1,p'}_{\Gamma_D}(\Omega)$. We slightly reinforce Assumption 2.11. Suppose that Assumption 2.2 from [21] holds for Ω and for all Γ_{D_j}, $j \in \{1, \dots, m\}$. This essentially means that Assumption 2.2 holds for all $x \in \partial\Omega$ and that the functional determinant of each bi-Lipschitz transformation ϕ_x is constant a.e. For example, this is the case if Ω is a Lipschitz domain ([21], Rem. 2.3). The rest of Assumption 2.11 remains the same. With this assumption one has

$$[\mathbb{W}^{-1,p_1}_{\Gamma_D}(\Omega), [\mathrm{L}^2(\Omega)]^m]_\theta = \mathbb{W}^{-\theta,p}_{\Gamma_D}(\Omega)$$

for $\theta \in (0,1)\backslash\{\frac{1}{p'}\}$, $\frac{1}{p} = \frac{1-\theta}{2} + \frac{1}{p_1}$ ([21], Thm. 3.1). This way we obtain an embedding $\tilde{U}_1 \hookrightarrow \mathbb{W}^{-\theta,p}_{\Gamma_D}(\Omega)$. Furthermore, we have

$$\mathbb{W}^{-\theta,p}_{\Gamma_D}(\Omega) = [\mathbb{W}^{-1,p}_{\Gamma_D}(\Omega), \mathbb{W}^{1,p}_{\Gamma_D}(\Omega)]_\gamma \simeq X^\gamma$$

for $-\theta = -1 + 2\gamma$ by ([21], Thm. 3.5) and Remark 2.7. For any $\gamma \in (0, \frac{1}{2})\backslash\{\frac{1}{2p}\}$ there holds $\theta \in (0,1)\backslash\{\frac{1}{p'}\}$ for $\theta = 1 - 2\gamma$ so that we obtain an embedding $\tilde{U}_1 \hookrightarrow X^\gamma$. Therefore, Assumption 5.1 is fulfilled for B_1 with any $\gamma \in (0, \frac{1}{2})$. By Theorem 5.2 it follows $\overline{u} \in \mathrm{L}^{\frac{2}{1-2s}}(J_T; \tilde{U}_1)$, $\overline{y} \in Y_{1/(1-2s),0}$ and $\overline{z} \in \mathrm{W}^{1,\frac{2}{1-2s}}(J_T)$ for arbitrary $s \in (0, \gamma)$.

Since $d = 2$ and $p > 2$ we can choose $\gamma \in (0, \frac{1}{2})$ such that $\gamma > \frac{d}{2p}$. Theorem 5.2 yields $\overline{y} \in \mathrm{C}(\overline{J_T}; [\mathrm{L}^\infty(\Omega)]^m)$. If Ω is a Lipschitz domain, then $\overline{y}$ is Hölder continuous in time and space.

References

[1] H. Amann, Nonautonomous parabolic equations involving measures. *J. Math. Sci.* **130** (2005) 4780–4802.

[2] M. Arnold, N. Begun, P. Gurevich, E. Kwame, H. Lamba and D. Rachinskii, Dynamics of discrete time systems with a hysteresis stop operator. *SIAM J. Appl. Dyn. Syst.* **16** (2017) 91–119.

[3] P. Auscher, N. Badr, R. Haller-Dintelmann and J. Rehberg, The square root problem for second-order, divergence form operators with mixed boundary conditions on L^p. *J. Evol. Equ.* **15** (2014) 165–208.

[4] W. Barthel, C. John and F. Tröltzsch, Optimal boundary control of a system of reaction diffusion equations. *ZAMM – J. Appl. Math. Mech./Z. Angew. Math. Mech.* **90** (2010) 966–982.

[5] J.F. Bonnans and E. Casas, On the choice of the function spaces for some state-constrained control problems. *Numer. Funct. Anal. Optim.* **7** (1985) 333–348.

[6] M. Brokate, Optimale Steuerung von gewöhnlichen Differentialgleichungen mit Nichtlinearitäten vom Hysteresis-Typ. *Methoden und Verfahren der mathematischen Physik*. P. Lang (1987).

[7] M. Brokate, Optimal control of ode systems with hysteresis nonlinearities, in Trends in Mathematical Optimization. Springer (1988) 25–41.

[8] M. Brokate, Optimal control of systems described by ordinary differential equations with nonlinear characteristics of the hysteresis type. *Autom. Remote Control* **52** (1991) 1639–1681.

[9] M. Brokate and P. Krejčì, Optimal control of ode systems involving a rate independent variational inequality. *Discrete Contin. Dynam. Syst.* **18** (2013) 331–348.

[10] M. Brokate and P. Krejčì, Weak differentiability of scalar hysteresis operators. *Discrete Contin. Dynam. Syst.* **35** (2015) 2405–2421.

[11] C.M. Carracedo and M. Sanz Alix The Theory of Fractional Powers of Operators. *North-Holland Mathematics Studies*. Elsevier (2001).

[12] E. Casas, Pontryagin's principle for state-constrained boundary control problems of semilinear parabolic equations. *SIAM J. Control Optim.* **35** (1997) 1297–1327.

[13] C. Castaing, M. Monteiro Marques and P.R. de Fitte, Some problems in optimal control governed by the sweeping process. *J. Nonlinear Convex Anal.* **15** (2014) 1043–1070.

[14] C. Cavaterra and F. Colombo, Automatic control problems for reaction-diffusion systems. *J. Evol. Equ.* **2** (2002) 241–273.

[15] G. Colombo, R. Henrion, N.D. Hoang and B.S. Mordukhovich, Optimal control of the sweeping process. *Dyn. Contin. Discrete Impuls. Syst. Ser. B: Appl. Algorithms* **19** (2012) 117–159.

[16] K. Disser, A.F.M. ter Elst and J. Rehberg, Hölder Estimates for Parabolic Operators on Domains With Rough Boundary. Preprint arXiv:1503.07035 (2015).

[17] G. Dudziuk and M. Niezgódka, Closed-loop control of a reaction-diffusion system. *Adv. Math. Sci. Appl.* **21** (2011) 383–402.

[18] M. Eleuteri and L. Lussardi, Thermal control of a rate-independent model for permanent inelastic effects in shape memory materials. *Evol. Equ. Control Theory* **3** (2014) 411–427.

[19] M. Eleuteri, L. Lussardi and U. Stefanelli, Thermal control of the Souza-Auricchio model for shape memory alloys. *Discrete Cont. Dynam. Syst. Ser. S* **6** (2013) 369–386.

[20] C. Giovanni, R. Henrion, D. Hoang Nguyen and B.S. Mordukhovich, Optimal control of the sweeping process over polyhedral controlled sets. *J. Differ. Equ.* **260** (2016) 3397–3447.

[21] J.A. Griepentrog, K. Groeger, H.C. Kaiser and J. Rehberg, Interpolation for function spaces related to mixed boundary value problems. *Math. Nachr.* **241** (2002) 110–120.

[22] R. Griesse, *Parametric Sensitivity Analysis for Control-Constrained Optimal Control Problems Governed by Systems of Parabolic Partial Differential Equations*. Ph.D. thesis, Universität Bayreuth, Fakultät für Mathematik und Physik (2003).

[23] 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) 127–149.

[24] R. Haller-Dintelmann, A. Jonsson, D. Knees and J. Rehberg, Elliptic and parabolic regularity for second-order divergence operators with mixed boundary conditions. *Math. Methods Appl. Sci.* **39** (2016) 5007–5026.

[25] D. Henry, Geometric Theory of Semilinear Parabolic Equations. *Lecture Notes in Mathematics*. Springer (1981).

[26] R. Herzog, C. Meyer and G. Wachsmuth, C-stationarity for optimal control of static plasticity with linear kinematic hardening. *SIAM J. Control Optim.* **50** (2012) 3052–3082.

[27] R. Herzog, C. Meyer and G. Wachsmuth, B- and strong stationarity for optimal control of static plasticity with hardening. *SIAM J. Optim.* **23** (2013) 321–352.

[28] R. Herzog, C. Meyer and G. Wachsmuth, Optimal Control of Elastoplastic Processes: Analysis, Algorithms, Numerical Analysis and Applications. Springer International Publishing (2014) 27–41.

[29] R. Herzog, C. Meyer and A. Stötzner, Existence of solutions of a thermoviscoplastic model and associated optimal control problems. *Nonlinear Anal.: Real World Appl.* **35** (2017) 75–101.

[30] D. Hömberg, K. Krumbiegel and J. Rehberg, Optimal control of a parabolic equation with dynamic boundary condition. *Appl. Math. Optim.* **67** (2013) 3–31.

[31] C. Meyer and L. Susu, Optimal control of nonsmooth, semilinear parabolic equations. *SIAM J. Control Optim.* **55** (2017) 2206–2234.

[32] C. Münch, Global existence and Hadamard differentiability of hysteresis reaction–diffusion systems. *J. Evol. Equ.* (2017).

[33] A. Pazy, Semigroups of Linear Operators and Applications to Partial Differential Equations. *Applied Mathematical Sciences*. Springer (1983).

[34] J.P. Raymond and H. Zidani, Pontryagin's principle for state-constrained control problems governed by parabolic equations with unbounded controls. *SIAM J. Control Optim.* **36** (1998) 1853–1879.

[35] F. Rindler, Optimal control for nonconvex rate-independent evolution processes. *SIAM J. Control Optim.* **47** (2008) 2773–2794.

[36] F. Rindler, Approximation of rate-independent optimal control problems. *SIAM J. Numer. Anal.* **47** (2009) 3884–3909.

[37] A.F.M. Ter Elst and J. Rehberg, L^∞-estimates for divergence operators on bad domains. *Anal. Appl.* **10** (2012) 207–214.

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

[39] S. Ulisse, D. Wachsmuth and G. Wachsmuth, Optimal control of a rate-independent evolution equation via viscous regularization. *Discrete Contin. Dyn. Syst. S* **10** (2017) 1467–1485.

[40] A. Visintin, Differential Models of Hysteresis, Vol. 111. Springer Science & Business Media (2013).

[41] G. Wachsmuth, Optimal control of quasi-static plasticity with linear kinematic hardening, part I: Existence and discretization in time. *SIAM J. Control Optim.* **50** (2012) 2836–2861.

[42] G. Wachsmuth, Optimal control of quasistatic plasticity with linear kinematic hardening II: Regularization and differentiability. *Z. Anal. Anwend.* **34** (2015) 391–418.

[43] G. Wachsmuth, Optimal control of quasistatic plasticity with linear kinematic hardening III: Optimality conditions. *Z. Anal. Anwend.* **35** (2016) 81–118.