OPTIMAL CONTROL OF NONSMOOTH, SEMILINEAR PARABOLIC EQUATIONS

C. MEYER* AND L. M. SUSU*

Abstract. This paper is concerned with an optimal control problem governed by a semilinear, nonsmooth operator differential equation. The nonlinearity is locally Lipschitz-continuous and directionally differentiable, but not Gâteaux-differentiable. Two types of necessary optimality conditions are derived, the first one by means of regularization, the second one by using the directional differentiability of the control-to-state mapping. The paper ends with the application of the general results to a semilinear heat equation involving the max-function.

Key words. Optimal control of PDEs, nonsmooth optimization, strong stationarity

1. Introduction. This paper is concerned with an optimal control problem governed by a semilinear operator differential equation. The essential feature of the problem under consideration is that the non-linearity in the state equation is only Lipschitz continuous and not necessarily Gâteaux-differentiable. Therefore, the standard adjoint calculus for the derivation of qualified optimality conditions is not applicable in our situation.

We present two alternative strategies to overcome this issue. The first one is to regularize the state equation in order to obtain a differentiable control-to-state mapping, which allows to derive Karush-Kuhn-Tucker (KKT) conditions by standard arguments. A passage to the limit then yields an optimality system, which turns out to be rather weak. A sharper result is obtained by employing the directional differentiability of the original unregularized control-to-state map. The optimality system obtained in this way is equivalent to the classical purely primal optimality condition, saying that the directional derivative of the objective in feasible directions is non-negative.

A similar situation, where various optimality conditions of different strength are known, arise in mathematical programs with equilibrium constraints (MPECs). In this context, the purely primal conditions are called Bouligand (B) stationarity. The most rigorous stationarity concept is strong stationarity. Roughly speaking, the strong stationarity conditions involve an optimality system, which is equivalent to B-stationarity. Other stationarity concepts such as Clarke (C) and Mordukhovich (M) stationarity are less rigorous compared to strong stationarity. The weakest concept is weak stationarity, which involves the existence of multipliers but no sign conditions for the multipliers at all. For a detailed overview we refer to [27] for the finite dimensional and [18] for the infinite dimensional case. If one adopts this scheme of stationarity conditions for our problem, the optimality conditions derived via regularization can be seen as weak stationarity, whereas the optimality system established by means of the directional derivative can be interpreted as strong stationarity.

Let us put our work into perspective. To keep the depiction concise we concentrate on parabolic problems. While there is a plenty of contributions in the field of optimal control of smooth semilinear parabolic equation, see e.g. [32] and the references therein, less papers are dealing with non-smooth equations. Most of the contributions in this field focus on variational inequalities of the first kind such as the parabolic obstacle problem. We only refer to [3, 9, 10, 15, 19, 20], where different regularization and

*Faculty of Mathematics, Technische Universität Dortmund, Dortmund-Germany.

relaxation schemes are used to smooth the problem. The optimality systems derived by regularization and relaxation are of intermediate strength such as C stationarity. For optimal control of the parabolic obstacle problem a strong stationarity system can be found in [24], but no rigorous proof is given there. Much less is known concerning parabolic equations involving more regular nonlinearities, which are not only maximal monotone operators but single-valued, Lipschitz continuous functions as in our case. To the best of our knowledge there are only contributions dealing with optimal control of ODEs in this case, see e.g. [6, 22] and [7, Chapter 5]. Our main result is the strong stationarity condition in Section 5. As we will exemplarily demonstrate by a comparison with the results of [7, Chapter 5], these conditions are more rigorous than the optimality conditions known so far for optimal control of non-smooth ODEs. This is due to the special structure of our problem, which is employed by the analysis in Section 5.

The paper is organized as follows: After a short introduction of our notation, we lay the foundations for our analysis in Section 2. We prove existence and uniqueness of the state equation in suitable spaces, which allows to define the control-to-state map. Section 3 is then devoted to a further investigation of the control-to-state map. We show that this mapping is directionally differentiable, which is the basis for our main result, the strong stationarity conditions in Section 5. In Section 4 we pursue the regularization approach to derive necessary optimality conditions. As a result weak stationarity conditions are established. The use of this approach is twofold. First it guarantees an improved regularity of locally optimal controls, which is required for the proof of strong stationarity. Secondly, weak stationarity enables a comparison to the strong stationarity conditions, which demonstrates that the latter ones are comparatively rigorous. After having established strong stationarity in Section 5, we address a specific example in Section 6. While Sections 2–5 deal with a general operator differential equation, Section 6 is concerned with the application of the general results to a semilinear heat equation involving the max-function. By employing the special structure of the max-function, the optimality conditions can further be sharpened.

Notation. Throughout the paper, c denotes a generic positive constant. If X and Y are two linear normed spaces, the space of linear and bounded operators from X to Y is denoted by $\mathcal{L}(X, Y)$. The open ball in X around $x \in X$ with radius R is denoted by $B_X(x, R)$. The dual of a linear normed space X will be denoted by X^*. For the dual pairing between X and X^* we write $\langle ., . \rangle_X$. If X is compactly embedded in Y, we write $X \hookrightarrow\hookrightarrow Y$, and $X \overset{d}{\hookrightarrow} Y$ means that X is dense in Y.

2. Standing Assumptions and Preliminaries.

Our optimal control problem reads as follows:

$$\left.\begin{aligned} \min \quad & J(y, u) \\ \text{s.t.} \quad & \dot{y}(t) + A\,y(t) + f(y(t)) = B\,u(t) \quad \text{in }]0, T[\\ & y(0) = 0. \end{aligned}\right\} \tag{P}$$

For the quantities in (P) we require the following:

Assumption 2.1.

1. *$T > 0$ is a given fixed final time.*
2. *X and U are real reflexive and separable Banach spaces. U is equipped with a norm such that U and U^* become locally uniformly convex spaces.*

3. *$A : X \to X$ is a linear, unbounded, and closed operator. Its domain of definition*

$$\mathcal{D} = \{x \in X : \|Ax\|_X < \infty\}$$

is densely and compactly embedded in X. Moreover, A is the infinitesimal generator of an analytic semigroup e^{-tA} on X. In addition, $0 \notin \sigma(A)$, where $\sigma(A)$ denotes the spectrum of A.

4. *For the nonlinearity we require $f : Y \to X$, where Y is a real reflexive seperable Banach space. Moreover, there exists a real number $\theta \in\,]0,1[$ such that $(X,\mathcal{D})_{\theta,\infty} \hookrightarrow Y$, where $(X,\mathcal{D})_{\theta,\infty}$ denotes the real interpolation space, see e.g. [31]. Furthermore, f is assumed to be Lipschitz continuous on bounded sets, i.e., for every $M > 0$, there exists $L(M) > 0$ so that*

$$\|f(x) - f(y)\|_X \leq L(M)\, \|x - y\|_Y \quad \forall\, x, y \in \overline{B_Y(0,M)} \tag{2.1}$$

and satisfies the following growth condition

$$\|f(y)\|_X \leq K\big(1 + \|y\|_Y\big) \quad \forall\, y \in Y \tag{2.2}$$

with a constant $K > 0$.

5. *Moreover, f is assumed to be directionally differentiable, i.e.,*

$$\Big\|\frac{f(x + \tau\, h) - f(x)}{\tau} - f'(x;h)\Big\|_X \overset{\tau \searrow 0}{\longrightarrow} 0 \quad \forall\, x, h \in Y, \tag{2.3}$$

and, similarly to f itself, its directional derivative is supposed to satisfy

$$\begin{aligned} &\forall\, M > 0 \;\; \exists\, Q(M) > 0 \text{ such that} \\ &\quad \|f'(y;h) - f'(y;s)\|_X \leq Q(M)\, \|h - s\|_Y \;\forall\, h, s \in Y,\, y \in \overline{B_Y(0,M)}, \end{aligned} \tag{2.4}$$

$$\begin{aligned} &\forall\, M > 0 \;\; \exists\, K(M) > 0 \text{ such that} \\ &\qquad \|f'(y;h)\|_X \leq K(M)\, \|h\|_Y \quad \forall\, h \in Y,\, y \in \overline{B_Y(0,M)}. \end{aligned} \tag{2.5}$$

6. *The operator $B : U \to X$ is linear and bounded. Its range is dense in X.*
7. *Let*

$$r > \frac{1}{1 - \theta} \tag{2.6}$$

with $\theta \in]0,1[$ from Assumption 2.1.4. Then the objective $J : L^r(]0,T[;\mathcal{D}) \times L^r(]0,T[;U) \to \mathbb{R}$ is convex and continuously Fréchet-differentiable w.r.t. both variables.

Remark 2.2. *We point out that one always find a norm such that U and U^* become locally uniformly convex spaces, since U is assumed to be reflexive, cf. [29, Theorem 4.7.12]. If U is a Hilbert space, then one can simply take the natural norm induced by the scalar product, see [29, Example 4.7.7].*

Remark 2.3. *With a little abuse of notation the Nemystkii-operator associated with f, considered with different ranges, will be denoted by the same symbol.*

We start the discussion of (P) with a global existence and uniqueness result for the state equation in Proposition 2.5 below. Although such a result is not surprising in view of the Lipschitz and growth conditions on f, we give a detailed proof in Appendix

A, as, for the best of our knowledge, there is no suitable reference for a global existence result for our particular equation.

Since A is closed, its domain $\mathcal{D}$ equipped with the graph norm forms a Banach space. Moreover, since $0 \notin \sigma(A)$, the graph norm is equivalent to

$$\|x\|_{\mathcal{D}} := \|Ax\|_X,$$

which is the norm we use for $\mathcal{D}$ in the sequel. The following result will be useful for various parts of the paper.

Lemma 2.4. *There holds*

$$\int_0^T \|e^{-tA}\|_{\mathcal{L}(X,Y)}^\alpha \, dt < \infty \quad \forall\, \alpha \in [0, \theta^{-1}[.$$

Proof. Using Assumption 2.1.4, [31, Theorem 1.15.2], and [26, Theorem 2.6.13] we estimate

$$\begin{aligned} \|e^{-tA}\|_{\mathcal{L}(X,Y)} &\leq c\, \|e^{-tA}\|_{\mathcal{L}(X,(X,\mathcal{D})_{\theta,\infty})} \\ &\leq c\, \|e^{-tA}\|_{\mathcal{L}(X,\mathcal{D}(A^\theta))} = c\, \|A^\theta e^{-tA}\|_{\mathcal{L}(X,X)} \leq c\, t^{-\theta}. \end{aligned} \tag{2.7}$$

Herein, $\mathcal{D}(A^\theta)$ denotes the domain of definition of A^θ. Since $\alpha < 1/\theta$, the right hand side in (2.7) is an element of $L^\alpha(0,T)$ giving the claim. □

Proposition 2.5. *For every $u \in L^r(]0,T[;U)$ there exists a unique mild solution $y \in C([0,T];Y)$ of*

$$\begin{aligned} \dot{y}(t) + A\, y(t) + f(y(t)) = B\, u(t) \quad \textit{in }]0,T[\\ y(0) = 0, \end{aligned} \tag{2.8}$$

which satisfies the following integral equation

$$y(t) = \int_0^t e^{-(t-s)A}\big(Bu(s) - f(y(s))\big)\, ds. \tag{2.9}$$

The associated solution operator $S : L^r(]0,T[;U) \ni u \mapsto y \in C([0,T];Y)$ is locally Lipschitz continuous in the following sense: For every $R > 0$ there exists a constant $\mathcal{L}(R) > 0$ such that

$$\|S(u_1) - S(u_2)\|_{C([0,T];Y)} \leq \mathcal{L}(R) \|u_1 - u_2\|_{L^r(]0,T[;U)} \tag{2.10}$$

for all $u_1, u_2 \in \overline{B_{L^r(]0,T[;U)}(0,R)}$.

The proof of Proposition 2.5 is based on classical results of semi-group theory, in particular Lemma 2.4. As mentioned above, we did not found a reference suited for (2.8) so that we added the proof in Appendix A for convenience of the reader. An inspection of part (ii) of the proof of Proposition 2.5, in particular the arguments leading to (A.1), immediately shows the following result, which will be frequently used in the sequel:

Corollary 2.6. *There is a constant $C > 0$ such that, for all $u \in L^r(]0,T[;U)$,*

$$\|S(u)\|_{C([0,T];Y)} \leq C\big(1 + \|u\|_{L^r(]0,T[;U)}\big)$$

holds true.

DEFINITION 2.7. *The operator A is said to satisfy maximal parabolic $L^r(]0,T[;X)$-regularity, $r \in]1,\infty[$, iff, for every $g \in L^r(]0,T[;X)$, the equation $\dot{w} + Aw = g$ admits a unique solution $w \in W_0^{1,r}(]0,T[;X) \cap L^r(]0,T[;\mathcal{D})$. In the following we abbreviate $\mathbb{W}_0^r(\mathcal{D},X) := W_0^{1,r}(]0,T[;X) \cap L^r(]0,T[;\mathcal{D})$. Moreover, we sometimes just say maximal parabolic regularity, i.e., we drop $L^r(]0,T[;X)$, if the context is clear.*

PROPOSITION 2.8. *If A satisfies maximal parabolic $L^r(]0,T[;X)$-regularity, then, for every $u \in L^r(]0,T[;U)$, there exists a unique solution $y \in \mathbb{W}_0^r(\mathcal{D},X)$ of (2.8). The solution operator $S : L^r([0,T];U) \to \mathbb{W}_0^r(\mathcal{D},X)$ is locally Lipschitz continuous in the sense of Proposition 2.5.*

Proof. We apply a standard boot strapping argument. Let $y \in C([0,T];Y)$ denote the solution of (2.8) according to Proposition 2.5. Consider the following auxiliary linear equation

$$\dot{w} + Aw = B\,u - f(y), \quad w(0) = 0. \tag{2.11}$$

Because of $y \in C([0,T];Y)$ we have $f(y) \in L^\infty(]0,T[;X)$ by (2.2). According to maximal parabolic regularity (2.11) admits a unique solution $w \in \mathbb{W}_0^r(\mathcal{D},X)$. This solution is given by

$$w(t) = \int_0^t e^{-(t-s)A}\big(Bu(s) - f(y(s))\big)\,ds,$$

which in view of (2.9) implies $y = w \in \mathbb{W}_0^r(\mathcal{D},X)$.

To prove the local Lipschitz continuity, let $R > 0$ be arbitrary and consider two functions $u_1, u_2 \in L^r(]0,T[;U)$ with $\|u_i\|_{L^r(]0,T[;U)} \le R$. Then, due to Corollary 2.6, there holds

$$\|y_i(t)\|_Y \le C(1+R) =: M, \quad i = 1,2.$$

Thanks to the open mapping theorem, $\partial_t + A : \mathbb{W}_0^r(\mathcal{D},X) \to L^r(]0,T[;X)$ is continuously invertible, if A satisfies maximal parabolic regularity. Therefore one arrives at

$$\begin{aligned}
\|y_1 - y_2\|_{\mathbb{W}_0^r(\mathcal{D},X)} &\le \|(\partial_t + A)^{-1}\|_{\mathcal{L}(L^r(]0,T;X[);\mathbb{W}_0^r(\mathcal{D},X))} \\
&\qquad \big(\|B(u_1-u_2)\|_{L^r(]0,T[;X)} + \|f(y_1) - f(y_2)\|_{L^r(]0,T[;X)}\big) \\
&\le c\big(\|B\|_{\mathcal{L}(U,X)}\|u_1 - u_2\|_{L^r(]0,T[;U)} + L(M)\|y_1 - y_2\|_{L^r(]0,T[;Y)}\big) \\
&\le c\,\big(\|B\|_{\mathcal{L}(U,X)} + L(M)\,\mathcal{L}(R)\big)\|u_1 - u_2\|_{L^r(]0,T[;U)},
\end{aligned}$$

where we used (2.1) and (2.10). ◻

REMARK 2.9. *Thanks to (2.6), we have $1 - 1/r > \theta$ and therefore, by [1, Ch. III, Thm. 4.10.2], it holds*

$$\mathbb{W}_0^r(\mathcal{D},X) \hookrightarrow C([0,T];(X,\mathcal{D})_{1-1/r,r}) \hookrightarrow C([0,T];(X,\mathcal{D})_{\theta,\infty}) \hookrightarrow C([0,T];Y).$$

Thus the assertion of Proposition 2.8 is indeed sharper than the one of Proposition 2.5.

LEMMA 2.10. *Let A satisfy maximal parabolic regularity. Then $S : L^r(]0,T[;U) \to \mathbb{W}_0^r(\mathcal{D},X)$ is weakly continuous.*

Proof. Let $u_n \rightharpoonup u$ in $L^r(]0,T[;U)$ and set $y_n := S(u_n)$ and $y = S(u)$. Then, due to Corollary 2.6 and (2.2), there exists a constant $C > 0$ such that

$$\begin{aligned}
&\|y_n\|_{\mathbb{W}_0^r(\mathcal{D},X)} \\
&\quad \leq \|(\partial_t + A)^{-1}\|_{\mathcal{L}(L^r(]0,T;X[);\mathbb{W}_0^r(\mathcal{D},X))}\big(\|Bu_n\|_{L^r(]0,T[;X)} + \|f(y_n)\|_{L^r(]0,T[;X)}\big) \\
&\quad \leq C\big(1 + \|u_n\|_{L^r(]0,T[;U)}\big) \quad \forall\, n \in \mathbb{N}.
\end{aligned}$$

Therefore there exists a further subsequence, for simplicity denoted by the same symbol, so that

$$y_n \rightharpoonup y \quad \text{in } \mathbb{W}_0^r(\mathcal{D},X) \text{ for } n \to \infty.$$

Note that $\mathbb{W}_0^r(\mathcal{D},X)$ is reflexive, as X and $\mathcal{D}$ are assumed to be so, see [11, Thm. I.5.13]. Since ∂_t, A, and B are weakly continuous, we immediately obtain

$$\dot{y}_n + A\,y_n \rightharpoonup \dot{y} + A\,y, \quad Bu_n \rightharpoonup Bu \quad \text{in } L^r(]0,T[;X).$$

Because of $\mathbb{W}_0^r(\mathcal{D},X) \hookrightarrow C([0,T];Y)$, cf. Remark 2.9, the sequence $\{y_n\}$ is bounded in $C([0,T];Y)$ by a constant $M > 0$ so that (2.1) yields

$$\|f(y_n) - f(y)\|_{L^r(]0,T[;X)} \leq L(M)\|y_n - y\|_{L^r(]0,T[;Y)}.$$

Since the embedding $\mathbb{W}_0^r(\mathcal{D},X) \hookrightarrow\hookrightarrow L^r(]0,T[;Y)$ is compact because of $\mathcal{D} \hookrightarrow\hookrightarrow X$, cf. Assumption 2.1.3, (2.6), and [1, Thm. II.4.10.2], we deduce $f(y_n) \to f(y)$ in $L^r(]0,T[;X)$. This allows to pass to the limit in (2.8), and its unique solvability yields $y = S(u)$. By a known argument the uniqueness of the weak limit gives the weak convergence of the whole sequence. □

PROPOSITION 2.11. *In addition to Assumption 2.1 and the maximal parabolic regularity of A, suppose that $L^r(]0,T[;U) \ni u \mapsto J(S(u),u) \in \mathbb{R}$ is radially unbounded. Then there exists at least one (not necessarily unique) solution to the optimal control problem* (P).

Proof. The arguments are standard. Every minimizing sequence is bounded due to the radial unboundedness. As $L^r(]0,T[;U)$ is reflexive by assumption, we can therefore select a weakly converging subsequence. By Lemma 2.10, the sequence of associated states also converges weakly in $\mathbb{W}_0^r(\mathcal{D},X)$. As the objective is continuous and convex by assumption, thus weakly lower semicontinuous, we can pass to the limit in the objective, giving optimality of the weak limit. □

Note that the additional assumption of radial unboundedness is not necessary for the rest of the paper, as we are investigating the characterization of local minima. This is why we do not impose this additional hypothesis in our standing assumptions.

3. Directional Differentiability.

This section is devoted to the directional differentiability of the solution operator S. In Section 5 this result will be used to derive a comparatively rigorous optimality condition. We start the discussion with a differentiability result of the nonlinearity in (2.8) in abstract function spaces.

LEMMA 3.1. *Under Assumption 2.1.4 and 5, the function f is directionally differentiable from $C([0,T];Y)$ to $L^\beta(]0,T[;X)$ for every $\beta < \infty$. If the Lipschitz and boundedness conditions in (2.1) and (2.5) are satisfied globally, i.e., with constants L and K independent of M, then f is directionally differentiable from $L^\beta(]0,T[;Y)$ to $L^\beta(]0,T[;X)$ for every $\beta < \infty$.*

Proof. Let $y, v \in C([0,T];Y)$ be given. Thanks to (2.3) it holds

$$A_\tau(t) := \Big\| \frac{f(y(t) + \tau\, v(t)) - f(y(t))}{\tau} - f'(y(t); v(t)) \Big\|_X \overset{\tau\searrow 0}{\longrightarrow} 0 \quad \text{f.a.a. } t \in]0,T[. \tag{3.1}$$

In view of (2.1) and (2.5) we find

$$|A_\tau(t)| \le \big(L(M) + K(M)\big) \|v(t)\|_Y \quad \text{f.a.a. } t \in]0,T[, \tag{3.2}$$

with $M := \|y\|_{C([0,T];Y)} + \|v\|_{C([0,T];Y)}$. Thus, from Lebesgue's dominated convergence theorem, we deduce

$$A_\tau \overset{\tau\searrow 0}{\longrightarrow} 0 \quad \text{in } L^\beta(]0,T[) \quad \text{for all } \beta < \infty,$$

which shows the first claim.

If (2.1) and (2.5) are satisfied globally, then (3.2) implies $|A_\tau(t)| \le (L+K)\,\|v(t)\|_Y$ for a.a. $t \in]0,T[$ and all $y, v \in L^\beta(]0,T[;Y)$, and again Lebesgue's dominated convergence theorem yields the second assertion. □

THEOREM 3.2. *The solution operator $S : L^r(]0,T[;U) \to C([0,T];Y)$ is directionally differentiable and its directional derivative $\eta = S'(u;h)$ at $u \in L^r(]0,T[;U)$ in direction $h \in L^r(]0,T[;U)$ is given by the mild solution of*

$$\begin{aligned} \dot\eta(t) + A\eta(t) + f'(y(t);\eta(t)) &= B\,h(t) \quad \textit{in } [0,T] \\ \eta(0) &= 0. \end{aligned} \tag{3.3}$$

with $y = S(u)$.

If A satisfies maximal $L^r(]0,T[;X)$-regularity, then S is directionally differentiable from $L^r(]0,T[;U)$ to $\mathbb{W}^r_0(\mathcal{D},X)$.

Proof. We first shortly address the existence of unique solutions to (3.3). For this purpose, let $u, h \in L^r(]0,T[;U)$ be arbitrary and set $y = S(u)$. By Corollary 2.6, (2.4), and (2.5) we have

$$\exists\, Q_y > 0 \text{ such that } \|f'(y;\eta_1) - f'(y;\eta_2)\|_X \le Q_y\,\|\eta_1 - \eta_2\|_Y \;\forall\, \eta_1, \eta_2 \in Y, \tag{3.4}$$

$$\exists\, K_y > 0 \text{ such that } \|f'(y;\eta)\|_X \le K_y \|\eta\|_Y \quad \forall\, \eta \in Y, \tag{3.5}$$

i.e., $\eta \mapsto f(y;\eta)$ satisfies the same Lipschitz and growth conditions as stated for $y \mapsto f(y)$ in Assumption 2.1.4. Therefore the same arguments apply as in the proofs of Proposition 2.5 and 2.8, and we obtain a unique solution $\eta \in C([0,T];Y)$, which fulfills $\eta \in \mathbb{W}^r_0(\mathcal{D},X)$, provided that A satisfies maximal parabolic regularity.

The associated integral equation reads

$$\eta(t) = \int_0^t e^{-(t-s)A}\big(Bh(s) - f'(y(s);\eta(s))\big)\,ds. \tag{3.6}$$

Subtracting this equation from the ones for $y = S(u)$ and $y^\tau = S(u + \tau\, h)$, $\tau \in [0,1]$, yields

$$\Big(\frac{y^\tau - y}{\tau} - \eta\Big)(t) = \int_0^t e^{-(t-s)A}\Big(\frac{f(y^\tau(s)) - f(y(s))}{\tau} - f'(y(s);\eta(s))\Big)\,ds.$$

Consequently, one obtains

$$
\begin{aligned}
&\Big\|\Big(\frac{y^\tau - y}{\tau} - \eta\Big)(t)\Big\|_Y \\
&\le \int_0^t \|e^{-(t-s)A}\|_{L(X,Y)} \Big(\overbrace{\Big\|\frac{f\big((y+\tau\eta)(s)\big) - f(y(s))}{\tau} - f'(y(s);\eta(s))\Big\|_X}^{=:A_\tau(s)} \\
&\qquad + \underbrace{\Big\|\frac{f(y^\tau(s)) - f\big((y+\tau\eta)(s)\big)}{\tau}\Big\|_X}_{=:B_\tau(s)}\Big) ds.
\end{aligned}
\tag{3.7}
$$

According to Lemma 3.1 there holds

$$
A_\tau \overset{\tau\searrow 0}{\longrightarrow} 0 \quad \text{in } L^\beta(]0,T[) \quad \text{for all } \beta < \infty. \tag{3.8}
$$

To estimate B_τ, first observe that Corollary 2.6 gives

$$
\begin{aligned}
\|y^\tau\|_{C(]0,T[;Y)} &\le C(1 + \|u + \tau h\|_{L^r(]0,T[;U)}) \\
&\le C(1 + \|u\|_{L^r(]0,T[;U)} + \|h\|_{L^r(]0,T[;U)}) =: \rho.
\end{aligned}
$$

By setting

$$
M := \max\big\{\rho, \|y\|_{C([0,T];Y)} + \|\eta\|_{C([0,T];Y)}\big\},
$$

(2.1) gives

$$
B_\tau(t) \le L(M) \Big\|\Big(\frac{y^\tau - y}{\tau} - \eta\Big)(t)\Big\|_Y \tag{3.9}
$$

for all $t \in [0,T]$. Together with (3.7), Gronwall's inequality, and Lemma 2.4, it follows that

$$
\begin{aligned}
&\Big\|\Big(\frac{y^\tau - y}{\tau} - \eta\Big)(t)\Big\|_Y \\
&\le \exp\Big(L(M)\int_0^t \|e^{-(t-s)A}\|_{L(X,Y)} ds\Big) \int_0^T \|e^{-(t-s)A}\|_{L(X,Y)} A_\tau(s)\, ds \\
&\le c \int_0^T \|e^{-sA}\|^{\beta'}_{L(X,Y)} ds^{1/\beta'} \, \|A_\tau\|_{L^\beta(]0,T[)},
\end{aligned}
$$

with $\beta > 1/(1-\theta)$ and θ from Assumption 2.1.4. Therefore, we conclude

$$
\Big\|\frac{y^\tau - y}{\tau} - \eta\Big\|_{C([0,T];Y)} \le c\|A_\tau\|_{L^\beta(]0,T[)} \overset{\tau\searrow 0}{\longrightarrow} 0 \tag{3.10}
$$

giving the first assertion.

If A satisfies maximal parabolic regularity, then

$$
\begin{aligned}
&\Big\|\frac{y^\tau - y}{\tau} - \eta\Big\|_{\mathbb{W}_0^r(\mathcal{D},X)} \\
&\le \|(\partial_t + A)^{-1}\|_{L(L^r(]0,T[;X),\mathbb{W}_0^r(\mathcal{D},X))} \Big\|\frac{f(y^\tau) - f(y)}{\tau} - f'(y;\eta)\Big\|_{L^r(]0,T[;X)} \\
&\le c\Big(\int_0^T |A_\tau(t)|^r \, dt^{1/r} + \int_0^T |B_\tau(t)|^r \, dt^{1/r}\Big) \\
&\le c\Big(\|A_\tau\|_{L^r(]0,T[)} + L(M)\Big\|\frac{y^\tau - y}{\tau} - \eta\Big\|_{L^r(]0,T[;Y)}\Big) \overset{\tau\searrow 0}{\longrightarrow} 0,
\end{aligned}
$$

where we used (3.9), (3.8), and (3.10). □

Remark 3.3. *Since S is directionally differentiable and locally Lipschitz continuous, it is Hadamard directionally differentiable, see e.g. [4, Proposition 2.49].*

Next we turn to the linearized equation with right-hand sides in $L^r(]0,T[;X)$, i.e.

$$\begin{aligned} \dot\eta(t) + A\eta(t) + f'(y(t);\eta(t)) = \xi(t) \quad \text{in } [0,T] \\ \eta(0) = 0 \end{aligned} \tag{3.11}$$

with $\xi \in L^r(]0,T[;X)$ and $y = S(u)$. In view of (3.5) and (3.4), a straight forward adaptation of the proof of Proposition 2.5 shows that this equation admits a unique solution η for every $\xi \in L^r(]0,T[;X)$. Moreover, if A satisfies maximal parabolic regularity, then we can argue as in the proof of Proposition 2.8 to show that $\eta \in \mathbb{W}_0^r(\mathcal{D},X)$. We denote the associated solution operator by $S^u : L^r(]0,T[;X) \to \mathbb{W}_0^r(\mathcal{D},X)$. Moreover, similarly to part (iii) of the proof of Proposition 2.5 one shows that S^u is globally Lipschitz continuous. To see this, let $\xi_1, \xi_2 \in L^r(]0,T[;X)$ and $u \in \overline{B_{L^r(]0,T[;U)}(0,R)}$ be given. Then Corollary 2.6 implies for $y = S(u)$ that

$$\|y\|_{C([0,T];Y)} \le C(1 + \|u\|_{L^r(]0,T[;U)}) \le C(1+R) =: M_R,$$

and the integral equation associated with (3.11) in combination with (2.4) results in

$$\begin{aligned} \|\eta_1(t) - \eta_2(t)\|_Y \le \int_0^T \|e^{-sA}\|_{\mathcal{L}(X,Y)}^{r'}\, ds^{1/r'}\, \|\xi_1 - \xi_2\|_{L^r(]0,T[;X)} \\ + Q(M_R) \int_0^t \|e^{-(t-s)A}\|_{\mathcal{L}(X,Y))} \|\eta_1(s) - \eta_2(s)\|_Y\, ds, \end{aligned}$$

cf. also (A.4). Then Gronwall's inequality and Lemma 2.4 yields

$$\|\eta_1(t) - \eta_2(t)\|_Y \le c\, Q_R\, \|\xi_1 - \xi_2\|_{L^r(]0,T[;X)}$$

with

$$Q_R := c \exp\Big(Q(M_R) \int_0^T \|e^{-tA}\|_{\mathcal{L}(X,Y))} dt\Big).$$

The estimate w.r.t. the $\mathbb{W}_0^r(\mathcal{D},X)$ follows completely analogously to the end of the proof of Proposition 2.8. The resulting estimate reads

$$\|\eta_1 - \eta_2\|_{\mathbb{W}_0^r(\mathcal{D},X)} \le c\,(1 + Q(M_R)\, Q_R) \|\xi_1 - \xi_2\|_{L^r(]0,T[;X)}.$$

Note that the Lipschitz constant does only depend on R. We collect these results in the following

Lemma 3.4. *Let A satisfy maximal parabolic regularity and $u \in L^r(]0,T[;U)$ be arbitrary with assiciated state $y = S(u)$. Then, for every $\xi \in L^r(]0,T[;X)$, there exists a unique solution $\eta \in \mathbb{W}_0^r(\mathcal{D},X)$ of (3.11). The associated solution operator S^u is globally Lipschitz continuous in the following sense: For every $R > 0$ there exists a constant $\mathcal{Q}(R) > 0$ so that for every $u \in L^r(]0,T[;U)$ with $\|u\|_{L^r(]0,T[;U)} \le R$ the following estimate*

$$\|S^u(\xi_1) - S^u(\xi_2)\|_{\mathbb{W}_0^r(\mathcal{D},X)} \le \mathcal{Q}(R)\, \|\xi_1 - \xi_2\|_{L^r(]0,T[;X)}\ \forall \xi_1, \xi_2 \in L^r(]0,T[;X). \tag{3.12}$$

holds true.

4. Regularization. In the following section, we regularize the nonlinearity f in (2.8) in order to obtain a Gâteaux-differentiable mapping, which enables us in turn to derive first-order optimality conditions in a standard way by using an adjoint calculus. Afterwards a limit analysis for vanishing regularization gives an optimality system for the original non-smooth problem, which is of *weak stationary type*, see Theorem 4.15 below. For this purpose we have to require several additional assumptions listed in the following.

Assumption 4.1 (Maximal parabolic regularity). *For the rest of the paper we assume that A satisfies maximal parabolic $L^r(]0,T[;X)$-regularity.*

Assumption 4.2 (Regularization). *For every $\varepsilon > 0$ there exists a function $f_\varepsilon : Y \to X$ such that*

1. *For all $y \in Y$ it holds*

$$\|f(y) - f_\varepsilon(y)\|_X \to 0 \quad \text{as } \varepsilon \to 0.$$

2. *f_ε satisfies the same hypotheses as f in Assumption 2.1.4, i.e.*

$$\begin{aligned} &\forall M > 0\, \exists L(M) > 0 : \\ &\qquad \|f_\varepsilon(x) - f_\varepsilon(y)\|_X \le L(M)\, \|x - y\|_Y \quad \forall x, y \in \overline{B_Y(0, M)}, \end{aligned} \tag{4.1}$$

$$\exists K > 0 : \quad \|f_\varepsilon(y)\|_X \le K\big(1 + \|y\|_Y\big) \quad \forall y \in Y, \tag{4.2}$$

where $L(.)$ and K are independent of ε.

3. *$f_\varepsilon : Y \to X$ is Gâteaux differentiable. Similarly to Assumption 2.1.5, its derivative is assumed to satisfy*

$$\begin{aligned} &\forall M > 0 \;\; \exists K(M) > 0 \text{ such that} \\ &\qquad \|f_\varepsilon'(y)\|_{\mathcal{L}(Y,X)} \le K(M) \quad \forall y \in \overline{B_Y(0, M)}. \end{aligned} \tag{4.3}$$

with $K(.)$ independent of ε.

Remark 4.3. *Note that, due to the linearity of the derivative, (4.3) is equivalent to a Lipschitz and boundedness condition analogous to (2.5) and (2.4).*

With a little abuse of notation the Lipschitz and boundedness constants in Assumption 4.2.2 and 4.2.3 are denoted by the same symbols as in Assumption 2.1, as all these constants do not depend on ε. The same holds true for the Lipschitz and boundedness constants in the sequel of this section.

Consider now the following regularized counterpart to (2.8)

$$\begin{aligned} \dot{y}(t) + A\, y(t) + f_\varepsilon(y(t)) &= B\, u(t) \quad \text{in }]0, T[\\ y(0) &= 0. \end{aligned} \tag{4.4}$$

Since f_ε satisfies the same conditions as f by Assumption 4.2.2, we immediately obtain the following

Lemma 4.4. *For every $u \in L^r(]0,T[;U)$ the equation (4.4) admits a unique solution $y \in \mathbb{W}_0^r(\mathcal{D}, X)$. The associated solution operator $S_\varepsilon : L^r(]0,T[;U) \ni u \mapsto y \in \mathbb{W}_0^r(\mathcal{D}, X)$ is locally Lipschitz continuous in the sense of Proposition 2.5 and 2.8, respectively, i.e., for every $R > 0$ there exists a constant $\mathcal{L}(R)$ so that*

$$\|S_\varepsilon(u_1) - S_\varepsilon(u_2)\|_{\mathbb{W}_0^r(\mathcal{D},X)} \le \mathcal{L}(R)\, \|u_1 - u_2\|_{L^r(]0,T[;U)} \quad \forall u_1, u_2 \in \overline{B_{L^r(]0,T[;U)}(0, R)}.$$

The Lipschitz constant $\mathcal{L}(.)$ does not depend on ε, since the constants in Assumption 4.2.2 do so.

In the sequel we will frequently use the following estimate

$$\|S_\varepsilon(u)\|_{C([0,T];Y)} \leq C\big(1 + \|u\|_{L^r(]0,T[;U)}\big), \tag{4.5}$$

with a constant $C > 0$, which does not depend on ε, as K in (4.2) does not. This estimate can be shown completely analogously to (A.1), and for simplicity we denote the constants by the same symbol.

LEMMA 4.5. *The regularized solution mapping $S_\varepsilon : L^r(]0,T[;U) \to \mathbb{W}_0^r(\mathcal{D}, X)$ is Gâteaux differentiable. Its derivative at $u \in L^r(]0,T[;U)$ in direction $h \in L^r(]0,T[;U)$ is given as the unique solution of*

$$\begin{aligned} \dot\eta(t) + A\eta(t) + f_\varepsilon'(y(t))\eta(t) &= B\,h(t) \quad \text{in }]0,T[\\ \eta(0) &= 0. \end{aligned} \tag{4.6}$$

Proof. The fact that $S_\varepsilon : L^r(]0,T[;U) \to \mathbb{W}_0^r(\mathcal{D}, X)$ is directionally differentiable follows with exactly the same arguments as in Theorem 3.2. The linearity of the derivative follows from Assumption 4.2.3. □

Based on Assumption 4.2.3 the following lemma can be proven in exactly the same way as Lemma 3.4. Note that, in case of a linear operator, Lipschitz continuity is equivalent to boundedness.

LEMMA 4.6. *For every $u \in L^r(]0,T[;U)$ and every $\xi \in L^r(]0,T[;X)$ the equation*

$$\begin{aligned} \dot\eta(t) + A\eta(t) + f_\varepsilon'(y(t))\eta(t) &= \xi(t) \quad \text{in }]0,T[\\ \eta(0) &= 0 \end{aligned} \tag{4.7}$$

admits a unique solution $\eta \in \mathbb{W}_0^r(\mathcal{D}, X)$. The associated linear solution operator is denoted by S_ε^u. Moreover, for every $R > 0$ there exists a constant $\mathcal{K}(R)$ such that

$$\|S_\varepsilon^u\|_{\mathcal{L}(L^r(]0,T[;X),\mathbb{W}_0^r(\mathcal{D},X))} \leq \mathcal{K}(R) \quad \forall\, u \in \overline{B_{L^r(]0,T[;U)}(0,R)}. \tag{4.8}$$

In view of Assumption 4.2.3 the constant $\mathcal{K}(.)$ does not depend on ε.

LEMMA 4.7 (Convergence of the regularization). *Let $u \in L^r(]0,T[;U)$ be arbitrary. Then*

$$S_\varepsilon(u) \to S(u) \quad \text{in } \mathbb{W}_0^r(\mathcal{D}, X), \ \text{as } \varepsilon \searrow 0. \tag{4.9}$$

Proof. The proof is similar to the one of Theorem 3.2. From the integral equation we get

$$y(t) - y_\varepsilon(t) = \int_0^t e^{-(t-s)A}\big(f(y(s)) - f_\varepsilon(y_\varepsilon(s))\big)\, ds. \tag{4.10}$$

In view of Corollary 2.6 and (4.5) one has that

$$M := \max\{\|y\|_{C([0,T];Y)}, \|y_\varepsilon\|_{C([0,T];Y)}\}$$

is finite and independent of ε. Therefore, (4.10) together with the triangle inequality and (4.1) leads to

$$\begin{aligned}&\|y(t)-y_\varepsilon(t)\|_Y\\&\quad\leq\int_0^t\|e^{-(t-s)A}\|_{\mathcal{L}(X,Y)}\Big(L(M)\|y(s)-y_\varepsilon(s)\|_Y+\underbrace{\|f(y(s))-f_\varepsilon(y(s)\|_X}_{=:A_\varepsilon(s)}\Big)\,ds.\end{aligned}$$

Gronwall's inequality and Lemma 2.4 then imply

$$\begin{aligned}&\|y(t)-y_\varepsilon(t)\|_Y\\&\quad\leq\exp\Big(L(M)\int_0^t\|e^{-(t-s)A}\|_{L(X,Y)}ds\Big)\int_0^T\|e^{-(t-s)A}\|_{L(X,Y)}A_\varepsilon(s)\,ds\\&\quad\leq c\int_0^T\|e^{-sA}\|_{L(X,Y)}^{\beta'}ds^{1/\beta'}\,\|A_\varepsilon\|_{L^\beta(]0,T[)}\end{aligned}$$

with $\beta\geq r>1/(1-\theta)$. Thanks to (2.2), (4.2), and Corollary 2.6, it holds

$$\begin{aligned}A_\varepsilon(t)&\to 0 &&\text{f.a.a. } t\in]0,T[\\|A_\varepsilon(t)|&\leq c(1+\|y(t)\|_Y)\leq c(1+\|u\|_{L^r(]0,T[;U)}) &&\text{f.a.a. } t\in]0,T[.\end{aligned}$$

and therefore, by Lebesgue's dominated convergence theorem,

$$A_\varepsilon=\|f(y(.))-f_\varepsilon(y(.))\|_X\to 0\quad\text{ in } L^\beta(]0,T[)\text{ as }\varepsilon\searrow 0. \tag{4.11}$$

Therefore we arrive at

$$\|y-y_\varepsilon\|_{C([0,T];Y)}\to 0. \tag{4.12}$$

Using the fact that $\partial_t+A:\mathbb{W}_0^r(\mathcal{D},X)\to L^r(]0,T[;X)$ is continuously invertible by Assumption 4.1, one obtains

$$\begin{aligned}\|y-y_\varepsilon\|_{\mathbb{W}_0^r(\mathcal{D},X)}&\leq\|(\partial_t+A)^{-1}\|_{\mathcal{L}(L^r(]0,T;X[);\mathbb{W}_0^r(\mathcal{D},X))}\|f(y)-f_\varepsilon(y_\varepsilon)\|_{L^r(]0,T[;X)}\\&\leq c\left(L(M)\|y_\varepsilon-y\|_{L^r(]0,T[;Y)}+\|A_\varepsilon\|_{L^r(]0,T[)}\right)\to 0,\end{aligned}$$

where we used (4.11) and (4.12). □

Corollary 4.8. *If $u_\varepsilon\to u$ in $L^r(]0,T[;U)$, then $S_\varepsilon(u_\varepsilon)\to S(u)$ in $\mathbb{W}_0^r(\mathcal{D},X)$, as $\varepsilon\searrow 0$.*

Proof. The triangle inequality yields

$$\|S_\varepsilon(u_\varepsilon)-S(u)\|_{\mathbb{W}_0^r(\mathcal{D},X)}\leq\|S_\varepsilon(u_\varepsilon)-S_\varepsilon(u)\|_{\mathbb{W}_0^r(\mathcal{D},X)}+\|S_\varepsilon(u)-S(u)\|_{\mathbb{W}_0^r(\mathcal{D},X)}.$$

While the second addend converges to zero thanks to Lemma 4.7, the first one can be estimated with the help of the local Lipschitz continuity of S_ε, see Lemma 4.4. Note that $\|u_\varepsilon\|_{L^r(]0,T[;U)}$ is uniformly bounded due to convergence. □

Lemma 4.9. *If $u_\varepsilon\rightharpoonup u$ in $L^r(]0,T[;U)$, then $S_\varepsilon(u_\varepsilon)\rightharpoonup S(u)$ in $\mathbb{W}_0^r(\mathcal{D},X)$, as $\varepsilon\searrow 0$.*

Proof. The proof follows the lines of the proof of Lemma 2.10. Thanks to (4.2) and (4.5), one shows completely analogously that the sequence $\{y_\varepsilon\}$ with $y_\varepsilon:=S_\varepsilon(u_\varepsilon)$ is bounded in $\mathbb{W}_0^r(\mathcal{D},X)$. Note in this context that the constant K in Assumption 4.2.2 does not depend on ε. Therefore there is a weakly converging subsequence, denoted by the same symbol, i.e., $y_\varepsilon \rightharpoonup y$ in $\mathbb{W}_0^r(\mathcal{D}, X)$. As in the proof of Lemma 2.10, we can pass to the limit in the linear parts of (4.4) by weak continuity. For the nonlinear part, we obtain

$$\|f_\varepsilon(y_\varepsilon) - f(y)\|_{L^r(]0,T[;X)} \le \|f_\varepsilon(y_\varepsilon) - f_\varepsilon(y)\|_{L^r(]0,T[;X)} + \|f_\varepsilon(y) - f(y)\|_{L^r(]0,T[;X)}$$

While the convergence of the second addend follows from (4.11), the first part converges to zero due to the compact embedding $\mathbb{W}_0^r(\mathcal{D}, X) \hookrightarrow\hookrightarrow L^r(]0,T[;Y)$ and the local Lipschitz continuity of f_ε in (4.1) with a constant independent of ε, see also the argument at the end of the proof of Lemma 2.10. Therefore the weak limit satisfies (2.8), and the uniqueness of the weak limit gives the convergence of the whole sequence. □

For the rest of this section let $\bar u \in L^r(]0,T[;U)$ an arbitrary local minimizer of (P) and consider the following regularization of (P):

$$\left.\begin{aligned} \min\quad & J(y,u) + \frac{1}{r}\|u - \bar u\|^r_{L^r(]0,T[;U)} \\ \text{s.t.}\quad & \dot y(t) + A\,y(t) + f_\varepsilon(y(t)) = B\,u(t) \quad \text{in }]0,T[\\ & y(0) = 0. \end{aligned}\right\} \tag{P$_\varepsilon$}$$

In order to derive necessary optimality conditions for (P$_\varepsilon$) we need the following

Lemma 4.10 (Adjoint equation). *Let $u \in L^r(]0,T[;U)$ be arbitrary. Then for every $\nu \in L^{r'}(]0,T[;\mathcal{D}^*)$ there holds $S_\varepsilon'(u)^*\nu = B^* p_\varepsilon$, where*

$$p_\varepsilon \in \mathbb{W}_T^{r'}(X^*, \mathcal{D}^*) := \{v \in W^{1,r'}(]0,T[;\mathcal{D}^*) \cap L^{r'}(]0,T[;X^*) : v(T) = 0\}$$

is the solution of the following linear equation

$$-\dot p_\varepsilon + A^* p_\varepsilon + [f_\varepsilon'(y_\varepsilon)]^* p_\varepsilon = \nu, \quad p_\varepsilon(T) = 0. \tag{4.13}$$

Proof. Since S_ε is Gâteaux-differentiable, the assertion follows by standard adjoint calculus. First note that

$$S_\varepsilon'(u) = S_\varepsilon^u\, B$$

with S_ε^u as defined in Lemma 4.6, i.e.,

$$S_\varepsilon^u = [\partial_t + A + f_\varepsilon'(y_\varepsilon)]^{-1} \in \mathcal{L}(L^r(]0,T[;X), \mathbb{W}_0^r(\mathcal{D}, X)).$$

with $y_\varepsilon = S_\varepsilon(u)$. This means that the $A + f_\varepsilon'(y_\varepsilon)$ satisfies maximal parabolic $L^r(]0,T[;X)$-regularity and, by [16, Lemma 36], the adjoint $A^* + f_\varepsilon'(y_\varepsilon)^*$ thus satisfies maximal parabolic $L^{r'}(]0,T[;\mathcal{D}^*)$-regularity, giving in turn that, for every $\nu \in L^{r'}(]0,T[;\mathcal{D}^*)$, there is a unique solution in $\mathbb{W}_T^{r'}(X^*, \mathcal{D}^*)$ of (4.13). Now let $h \in L^r(]0,T[;U)$ be arbitrary and denote by η the solution of (4.6), i.e., $\eta = S_\varepsilon'(u)h$. By applying the formula of integration by parts from [2, Proposition 5.1] in combination with the initial and final time conditions in (4.6) and (4.13), respectively, one shows that

$$\begin{aligned} \langle B^* p_\varepsilon, h\rangle_{L^r(]0,T[;U)} &= \int_0^T \langle p_\varepsilon, B\,h\rangle_X\, dt \\ &= \int_0^T \langle p_\varepsilon, \dot\eta + A\eta + f_\varepsilon'(y_\varepsilon)\eta\rangle_X\, dt \\ &= \int_0^T \langle -\dot p_\varepsilon + A^* p_\varepsilon + f_\varepsilon'(y_\varepsilon)^* p_\varepsilon, \eta\rangle_{\mathcal{D}}\, dt \\ &= \langle \nu, S_\varepsilon'(u)h\rangle_{L^r(]0,T[;\mathcal{D})}, \end{aligned} \tag{4.14}$$

cf. also [16, Appendix A.2]. Since h was arbitrary, this proves the claim. □

Definition 4.11. *Let $\varepsilon > 0$ be given. We define the reduced cost functionals of* (P) *and* (P_ε)*, respectively, by*

$$j : L^r(]0,T[;U) \ni u \mapsto J(S(u),u) \in \mathbb{R}, \tag{4.15}$$

$$j_\varepsilon : L^r(]0,T[;U) \ni u \mapsto J(S_\varepsilon(u),u) + \frac{1}{r}\|u-\bar{u}\|^r_{L^r(]0,T[;U)} \in \mathbb{R}. \tag{4.16}$$

Proposition 4.12 (Optimality system for the regularized problem). *Every local solution u_ε of* (P_ε) *fulfills together with the state $y_\varepsilon \in \mathbb{W}^r_0(\mathcal{D},X)$ and the adjoint state $p_\varepsilon \in \mathbb{W}^{r'}_T(X^*,\mathcal{D}^*)$ the following optimality system*

$$\dot{y}_\varepsilon + A\,y_\varepsilon + f_\varepsilon(y_\varepsilon) = Bu_\varepsilon, \quad y_\varepsilon(0) = 0 \tag{4.17a}$$

$$-\dot{p}_\varepsilon + A^*\,p_\varepsilon + [f'_\varepsilon(y_\varepsilon)]^* p_\varepsilon = \partial_y J(y_\varepsilon,u_\varepsilon), \quad p_\varepsilon(T) = 0, \tag{4.17b}$$

$$B^* p_\varepsilon + \partial_u J(y_\varepsilon,u_\varepsilon) + F'(u_\varepsilon) = 0 \tag{4.17c}$$

Proof. In view of Lemma 4.10 the arguments are standard. First note that u_ε is a local solution of the free optimization problem

$$\min_{u\in L^r(]0,T[;U)} j_\varepsilon(u).$$

By the chain rule and Lemma C.1 in Appendix C, j_ε is Gâteaux-differentiable and therefore

$$0 = j'_\varepsilon(u_\varepsilon) = [S'_\varepsilon(u_\varepsilon)]^* \partial_y J(y_\varepsilon,u_\varepsilon) + \partial_u J(y_\varepsilon,u_\varepsilon) + F'(u_\varepsilon) \quad \text{in } L^{r'}(]0,T[;U^*).$$

Note that the assumptions of Lemma C.1 are fulfilled due to Assumption 2.1.2. Lemma 4.10 then gives the result. □

Proposition 4.13 (Convergence of the minimizers). *Let $\bar{u} \in L^r(]0,T[;U)$ be a local minimizer of* (P)*. Then there exists a sequence $\{u_\varepsilon\}$ of local minimizers of* (P_ε) *such that*

$$u_\varepsilon \to \bar{u} \quad \text{in } L^r(]0,T[;U) \text{ as } \varepsilon \searrow 0. \tag{4.18}$$

Moreover

$$S_\varepsilon(u_\varepsilon) \to S(\bar{u}) \text{ in } \mathbb{W}^r_0(\mathcal{D},X). \tag{4.19}$$

Proof. The arguments are standard and combine a technique used for instance in [25] with a localization argument of [5]. For convenience of the reader we recall the arguments.

Let $B(\bar{u},\rho) := \overline{B_{L^r(]0,T[;U)}(\bar{u},\rho)}$ be the ball of local optimality of $\bar{u}$, i.e.,

$$j(\bar{u}) \le j(u) \quad \forall\, u \in B(\bar{u},\rho). \tag{4.20}$$

Then we consider the following auxiliary optimal control problem:

$$\left.\begin{array}{ll} \min & j_\varepsilon(u) \\ \text{s.t.} & u \in B(\bar{u},\rho), \end{array}\right\} \tag{P^ρ_ε}$$

which coincides with (P_ε) except for the additional constraints on u. The existence of global minimizers for ($\mathrm{P}_\varepsilon^\rho$) can be shown completely analogously to the proof of Proposition 2.11 by standard arguments. Note that, this time, the radial unboundedness of J is not needed, since the additional constraint $u \in B(\bar{u}, \rho)$ ensures the boundedness of minimizing sequences. Note further that $B(\bar{u}, \rho)$ is convex and closed, thus weakly closed, which gives the feasibility of the weak limit. In the sequel we denote by u_ε a global minimizer of ($\mathrm{P}_\varepsilon^\rho$).

Now let $\varepsilon \searrow 0$. Then, due to $\{u_\varepsilon\} \subset B(\bar{u}, \rho)$, and we can select a weakly convergent subsequence, which we denote by the same symbol, i.e.,

$$u_\varepsilon \rightharpoonup \tilde{u} \quad \text{in } L^r(]0, T[; U). \tag{4.21}$$

As $\bar{u}$ is clearly feasible for ($\mathrm{P}_\varepsilon^\rho$), one obtains by applying Lemma 4.7 that

$$J(S_\varepsilon(u_\varepsilon), u_\varepsilon) \le j_\varepsilon(u_\varepsilon) \le j_\varepsilon(\bar{u}) = J(S_\varepsilon(\bar{u}), \bar{u}) \to J(S(\bar{u}), \bar{u}) = j(\bar{u}). \tag{4.22}$$

Thus we arrive at

$$\begin{aligned} j(\bar{u}) \ge \limsup_{\varepsilon \to 0} j_\varepsilon(u_\varepsilon) &\ge \liminf_{\varepsilon \to 0} j_\varepsilon(u_\varepsilon) \\ &= \liminf_{\varepsilon \to 0} \Big(J(S_\varepsilon(u_\varepsilon), u_\varepsilon) + \frac{1}{r}\|u_\varepsilon - \bar{u}\|^r_{L^r(]0,T[;U)} \Big) \\ &\ge J(S(\tilde{u}), \tilde{u}) + \frac{1}{r}\|\tilde{u} - \bar{u}\|^r_{L^r(]0,T[;U)} \\ &\ge j(\bar{u}) + \frac{1}{r}\|\tilde{u} - \bar{u}\|^r_{L^r(]0,T[;U)} \end{aligned} \tag{4.23}$$

The last two inequalities follow from the weak lower semicontinuity of the norm and of J (by convexity), Lemma 4.9, and (4.20), respectively. Note that $\tilde{u} \in B(\bar{u}, \rho)$, as $B(\bar{u}, \rho)$ is weakly closed. The first thing to note in (4.23) is that $\tilde{u} = \bar{u}$.

Furthermore, by applying the same arguments with the very left side of (4.22), it follows

$$j(\bar{u}) \ge \limsup_{\varepsilon \to 0} J(S_\varepsilon(u_\varepsilon), u_\varepsilon) \ge \liminf_{\varepsilon \to 0} J(S_\varepsilon(u_\varepsilon), u_\varepsilon) \ge J(S(\tilde{u}), \tilde{u}) \ge j(\bar{u}).$$

Together with $j_\varepsilon(u_\varepsilon) \to j(\bar{u})$, which follows from (4.23), this yields

$$\frac{1}{r}\,\|u_\varepsilon - \bar{u}\|^r_{L^r(]0,T[;U)} = j_\varepsilon(u_\varepsilon) - J(S_\varepsilon(u_\varepsilon), u_\varepsilon) \to 0, \tag{4.24}$$

which shows (4.18). The convergence of the states in (4.19) is then an immediate consequence of Corollary 4.8.

It remains to show that u_ε is a local minimizer of (P_ε) for sufficiently small $\varepsilon > 0$. To this end, let $u \in B_{L^r(]0,T[;U)}(u_\varepsilon, \rho/2)$ be arbitrary. Then, for sufficiently small $\varepsilon > 0$, (4.18) leads to

$$\|u - \bar{u}\|_{L^r(]0,T[;U)} \le \|u_\varepsilon - \bar{u}\|_{L^r(]0,T[;U)} + \|u - u_\varepsilon\|_{L^r(]0,T[;U)} \le \frac{\rho}{2} + \frac{\rho}{2} = \rho.$$

This yields $u \in B(\bar{u}, \rho)$ and the global optimality of u_ε for ($\mathrm{P}_\varepsilon^\rho$) implies $j_\varepsilon(u_\varepsilon) \le j_\varepsilon(u)$. Since $u \in B_{L^r(]0,T[;U)}(u_\varepsilon, \rho/2)$ was arbitrary this gives the claim. □

Lemma 4.14 (Boundedness of dual variables). *Let $\{p_\varepsilon\}$ be sequence of adjoint states given by (4.27c) associated with the sequence of local solutions $\{u_\varepsilon\}$ in Proposition 4.13. Define*

$$\lambda_\varepsilon := [f_\varepsilon'(y_\varepsilon)]^* p_\varepsilon. \tag{4.25}$$

Then there exist constants $C_1, C_2 > 0$, independent of ε, such that

$$\|p_\varepsilon\|_{\mathbb{W}_T^{r'}(X^*,\mathcal{D}^*)} \le C_1, \quad \|\lambda_\varepsilon\|_{L^{r'}(]0,T[;Y^*)} \le C_2$$

for all $\varepsilon > 0$.

Proof. We first address the boundedness of $\{p_\varepsilon\}$ in $\mathbb{W}_T^{r'}(X^*,\mathcal{D}^*)$. Since u_ε and y_ε converge to $\bar{u}$ and $\bar{y}$ in $L^r(]0,T[;U)$ and $\mathbb{W}_0^r(\mathcal{D},X)$, respectively, the continuous Fréchet-differentiability of J implies that $J'(y_\varepsilon, u_\varepsilon)$ converges and is therefore bounded, i.e.,

$$\|J'(y_\varepsilon,u_\varepsilon)\|_{(L^r(]0,T[;\mathcal{D}))^*\times(L^r(]0,T[;U))^*} \le c < \infty \quad \forall\,\varepsilon > 0. \tag{4.26}$$

With regard to the adjoint equation in (4.27c) an integration by parts, completely analogously to (4.14) yields

$$\begin{aligned}\langle p_\varepsilon,\xi\rangle_{L^r(]0,T[;X)} &= \langle \partial_y J(y_\varepsilon,u_\varepsilon), S_\varepsilon^{u_\varepsilon}\,\xi\rangle_{L^r(]0,T[;\mathcal{D})}\\ &\le \|\partial_y J(y_\varepsilon,u_\varepsilon)\|_{(L^r(]0,T[;\mathcal{D}))^*}\|S_\varepsilon^{u_\varepsilon}\|_{\mathcal{L}(L^r(]0,T[;X),\mathbb{W}_0^r(\mathcal{D},X))}\|\xi\|_{L^r(]0,T[;X)}\\ &\le c\,\mathcal{K}(R)\,\|\xi\|_{L^r(]0,T[;X)},\end{aligned}$$

where we used (4.26), (4.8), and $\|u_\varepsilon\|_{L^r(]0,T[;U)} \le R$ for the last estimate. Since X is reflexive and separable and thus $(L^r(]0,T[;X))^* = L^{r'}(]0,T[;X^*)$, we obtain

$$\|p_\varepsilon\|_{L^{r'}(]0,T[;X^*)} = \sup_{\xi\neq 0}\frac{\langle p_\varepsilon,\xi\rangle_{L^r(]0,T[;X)}}{\|\xi\|_{L^r(]0,T[;X)}} \le c.$$

The maximal parabolic $L^{r'}(]0,T[;\mathcal{D}^*)$-regularity of A^*, cf. [16, Lemma 36] and Assumption 4.2.3 together with the boundedness of $\{y_\varepsilon\}$ in $\mathbb{W}_0^r(\mathcal{D},X) \hookrightarrow C([0,T];Y)$ then yield

$$\begin{aligned}&\|p_\varepsilon\|_{\mathbb{W}_T^{r'}(X^*,\mathcal{D}^*)}\\ &\quad\le \|(-\partial_t + A^*)^{-1}\|_{\mathcal{L}(L^{r'}(]0,T[;\mathcal{D}^*),\mathbb{W}_T^{r'}(X^*,\mathcal{D}^*))}\|\partial_y J(y_\varepsilon,u_\varepsilon) - [f_\varepsilon'(y_\varepsilon)]^* p_\varepsilon\|_{L^{r'}(]0,T[;\mathcal{D}^*)}\\ &\quad\le c\Big(\|\partial_y J(y_\varepsilon,u_\varepsilon)\|_{L^{r'}(]0,T[;\mathcal{D}^*)}\\ &\qquad\quad + \|[f_\varepsilon'(y_\varepsilon)]\|_{\mathcal{L}(L^r(]0,T[;Y),L^r(]0,T[;X))}\|p_\varepsilon\|_{L^{r'}(]0,T[;X^*)}\Big) \le c \quad \forall\,\varepsilon > 0.\end{aligned}$$

For λ_ε as defined in (4.25) one similarly obtains

$$\begin{aligned}\|\lambda_\varepsilon\|_{L^{r'}(]0,T[;Y^*)} &= \|[f_\varepsilon'(y_\varepsilon)]^* p_\varepsilon\|_{L^{r'}(]0,T[;Y^*)}\\ &\le \|[f_\varepsilon'(y_\varepsilon)]\|_{\mathcal{L}(L^r(]0,T[;Y),L^r(]0,T[;X))}\|p_\varepsilon\|_{L^{r'}(]0,T[;X^*)} \le c < \infty\;\forall\,\varepsilon > 0,\end{aligned}$$

which finally proves the assertion. □

Theorem 4.15 (Optimality system after passing to the limit). *Let $\bar{u}$ be a local minimizer of (P) with associated state $\bar{y} = S(\bar{u}) \in \mathbb{W}_0^r(\mathcal{D},X)$. Then there exist unique adjoint state $p \in \mathbb{W}_T^{r'}(X^*,\mathcal{D}^*)$ and $\lambda \in L^{r'}(]0,T[;Y^*)$ such that*

$$\dot{\bar{y}} + A\,\bar{y} + f(\bar{y}) = B\bar{u}, \quad \bar{y}(0) = 0 \tag{4.27a}$$

$$-\dot{p} + A^*\,p + \lambda = \partial_y J(\bar{y},\bar{u}), \quad p(T) = 0 \tag{4.27b}$$

$$B^* p + \partial_u J(\bar{y},\bar{u}) = 0. \tag{4.27c}$$

Proof. Let $\{u_\varepsilon\}$ be the sequence from Proposition 4.13. Thanks to Lemma 4.14 and the reflexivity of $\mathbb{W}_T^{r'}(X^*, \mathcal{D}^*)$, cf. , see [11, Thm. I.5.13], and $L^{r'}(]0,T[;Y^*)$, there exist p and λ such that

$$p_\varepsilon \rightharpoonup p \text{ in } \mathbb{W}_T^{r'}(X^*, \mathcal{D}^*), \quad \lambda_\varepsilon \rightharpoonup \lambda \text{ in } L^{r'}(]0,T[;Y^*). \tag{4.28}$$

Therefore, in view of Lemma 4.14 and the weak continuity of $-\partial_t + A^*$ from $\mathbb{W}_T^{r'}(X^*, \mathcal{D}^*)$ to $L^{r'}(]0,T[;\mathcal{D}^*)$, passing to the limit in (4.17b) leads to

$$\begin{aligned} 0 &= -\dot{p}_\varepsilon + A^* p_\varepsilon + [f_\varepsilon'(y_\varepsilon)]^* p_\varepsilon \\ &= -\dot{p}_\varepsilon + A^* p_\varepsilon + \lambda_\varepsilon \rightharpoonup -\dot{p} + A^* p + \lambda \quad \text{in } L^{r'}(]0,T[;\mathcal{D}^*). \end{aligned}$$

In view of $p \in \mathbb{W}_T^{r'}(X^*, \mathcal{D}^*)$, which implies $p(T) = 0$, this is just (4.27b).

Next we pass to the limit in (4.17c). From the weak convergence of $\{p_\varepsilon\}$ together with the weak continuity of the linear operator B^*, Lemma C.1, the continuous differentiability of J, and the strong convergence of $\{(y_\varepsilon, u_\varepsilon)\}$, it follows

$$B^* p_\varepsilon + \partial_u J(y_\varepsilon, u_\varepsilon) + F'(u_\varepsilon) \rightharpoonup B^* p + \partial_u J(\bar{y}, \bar{u}) \quad \text{in } L^{r'}(]0,T[;U^*) \tag{4.29}$$

so that (4.17c) gives (4.27c).

It now remains to show the uniqueness of p. In view of (4.27b) this will also imply the uniqueness of λ. Since $\operatorname{Rg}(B)$ is dense in X by Assumption 2.1.6, we obtain

$$\ker(B^*) = \operatorname{Rg}(B)^\perp = \{0\}$$

so that (4.27c), i.e.,

$$B^* p(t) = -\partial_u J(\bar{y}, \bar{u})(t) \quad \text{in } U^*, \text{ f.a.a. } t \in]0,T[, \tag{4.30}$$

gives the uniqueness of p. □

Remark 4.16. *Since the optimality system (4.27) does not contain any sign condition, neither on p nor on λ, it can be seen as a* weak stationarity condition. *If a concrete instance for the nonlinearity f is given, then it may be possible to derive additional conditions when passing to the limit in the regularized optimality system (4.17). We will demonstrate this in Section 6 below.*

Note in this context that, in general, one cannot expect any regularization of f to have better approximation properties than Assumption 4.2.1, in the sense that also the directional derivative is approximated, even not in a weak sense. To see this, assume that for some $y \in Y$, we have

$$\lim_{\varepsilon \searrow 0} \langle \xi, f_\varepsilon'(y) v \rangle_X = \langle \xi, f'(y; v) \rangle_X \quad \forall\, v \in Y,\, \xi \in X^*$$

Then one easily deduces that $f'(y;.)$ is linear, and thus f is Gâteaux-differeniable at y. It is therefore in general not possible to derive a precise characterization of the weak limit λ in terms of the directional derivative of f by passing to the limit in the regularized optimality system. Under further assumptions on f_ε however, sign conditions for p and λ can be shown. This is for instance the case, if, for all $\varepsilon > 0$, f_ε is the derivative of a convex potential as in case of the max*-function in Section 6 below.*

5. Strong Stationarity. In this section we derive an optimality system, which employs the particular structure of the optimal control problem under consideration. The underlying analysis is based on the directional differentiability that was established in Section 3. It turns out that the optimality system derived in this way is significantly sharper compared to the one obtained via regularization in Section 4. For convenience let us recall the optimal control problem:

$$\left.\begin{aligned} \min \quad & J(y,u) \\ \text{s.t.} \quad & \dot y(t) + A\,y(t) + f(y(t)) = B\,u(t) \quad \text{in }]0,T[\\ & y(0) = 0. \end{aligned}\right\} \tag{P}$$

Throughout this section we again assume that $\bar u \in L^r(]0,T[;U)$ is locally optimal for (P) with associated state $\bar y = S(\bar u)$. As in Section 4 we suppose that A satisfies maximal parabolic $L^r(]0,T[;X)$-regularity. We start our analysis with a purely primal optimality condition, which is an immediate consequence of the directional differentiability of the control-to-state mappint S. In accordance with the notion known for MPECs, we call this optimality condition *Bouligand (B) stationarity.*

Lemma 5.1 (B-stationarity). *If $\bar u \in L^r(]0,T[;U)$ is locally optimal for (P), then there holds*

$$\partial_y J(\bar y,\bar u)S'(\bar u;h) + \partial_u J(\bar y,\bar u)h \geq 0 \quad \forall\, h \in L^r(]0,T[;U). \tag{5.1}$$

Proof. According to Theorem 3.2 and [17, Lemma 3.9] the composite mapping $L^r(]0,T[;U) \ni u \mapsto J(S(u),u) \in \mathbb{R}$ is directionally differentiable with directional derivative $\partial_y J(\bar y,\bar u)S'(\bar u;h) + \partial_u J(\bar y,\bar u)h$. The result then follows immediately from the local optimality of $\bar u$. ◻

Lemma 5.2. *The set $\{S'(\bar u;h) : h \in L^r(]0,T[;U)\}$ is dense in $\mathbb{W}_0^r(\mathcal{D},X)$.*

Proof. Let $\eta \in \mathbb{W}_0^r(\mathcal{D},X)$ be arbitrary and define

$$\xi := \dot\eta(t) + A\eta(t) + f'(\bar y(t);\eta(t)) \in L^r(]0,T[;X).$$

With regard to Assumption 2.1.6 there exists a sequence $\{h_n\} \subset L^r(]0,T[;U)$ such that

$$B\,h_n \to \xi \quad \text{in } L^r(]0,T[;X).$$

This follows from Lemma B.1, which implies that $L^r(]0,T[;Rg(B)) \overset{d}{\hookrightarrow} L^r(]0,T[;X)$. The global Lipschitz continuity according to Lemma 3.4 gives

$$S'(\bar u;h_n) = S^{\bar u}(B\,h_n) \to S^{\bar u}(\xi) = \eta \quad \text{in } \mathbb{W}_0^r(\mathcal{D},X).$$

This completes the proof. ◻

We are now in the position to state our main result:

Theorem 5.3 (Strong stationarity). *Let $\bar u \in L^r(]0,T[;U)$ be locally optimal with associated state $\bar y = S(\bar u)$. Then there exists a unique adjoint state*

$$p \in \mathbb{W}_T^{r'}(X^*,\mathcal{D}^*) = \{v \in W^{1,r'}(]0,T[;\mathcal{D}^*) \cap L^{r'}(]0,T[;X^*) : v(T) = 0\}$$

and a unique multiplier $\lambda \in L^{r'}(]0,T[;Y^*)$ *such that*

$$-\dot{p} + A^* p + \lambda = \partial_y J(\bar{y}, \bar{u}), \quad p(T) = 0 \tag{5.2a}$$

$$\langle \lambda(t), v\rangle_{\mathcal{D}} \geq \langle p(t), f'(\bar{y}(t); v)\rangle_X \quad \forall\, v \in \mathcal{D},\ \textit{f.a.a. } t \in]0,T[\tag{5.2b}$$

$$B^* p(t) + \partial_u J(\bar{y}, \bar{u})(t) = 0 \quad \textit{a.e. in }]0,T[. \tag{5.2c}$$

Proof. From Theorem 4.15 we know that there is a unique function $p \in \mathbb{W}^{r'}_T(X^*, \mathcal{D}^*)$ such that

$$B^* p(t) + \partial_u J(\bar{y}, \bar{u})(t) = 0 \quad \text{a.e. in }]0,T[. \tag{5.3}$$

In the following let $h \in L^r(]0,T[;U)$ be arbitrary. As $p \in \mathbb{W}^{r'}_T(X^*, \mathcal{D}^*) \hookrightarrow L^{r'}(]0,T[;X^*)$, we can test (3.3) with p to obtain

$$\int_0^T \langle p(t), \dot{\eta}(t) + A\eta(t)\rangle_X \, dt + \int_0^T \langle p(t), f'(\bar{y}(t); \eta(t))\rangle_X \, dt \\ = \int_0^T \langle p(t), B\, h(t)\rangle_X \, dt.$$

Using the integration by parts formula in [2, Proposition 5.1], we deduce from this that

$$\int_0^T \langle -\dot{p}(t) + A^* p(t), \eta(t)\rangle_{\mathcal{D}} \, dt + \int_0^T \langle p(t), f'(\bar{y}(t); \eta(t))\rangle_X \, dt \\ = \int_0^T \langle B^* p(t), h(t)\rangle_U \, dt = -\int_0^T \langle \partial_u J(\bar{y}, \bar{u})(t), h(t)\rangle_U \, dt,$$

where we used (5.3) for the last equation. Inserting this into (5.1) gives

$$\int_0^T \langle \partial_y J(\bar{y}, \bar{u})(t), \eta(t)\rangle_{\mathcal{D}} \, dt - \int_0^T \langle -\dot{p}(t) + A^* p(t), \eta(t)\rangle_{\mathcal{D}} \, dt \\ \geq \int_0^T \langle p(t), f'(\bar{y}(t); \eta(t))\rangle_X \, dt. \tag{5.4}$$

Next let us define

$$\lambda := \dot{p} - A^* p + \partial_y J(\bar{y}, \bar{u}) \in L^{r'}(]0,T[;\mathcal{D}^*) \tag{5.5}$$

so that (5.4) reads

$$\int_0^T \langle \lambda(t), \eta(t)\rangle_{\mathcal{D}} \, dt \geq \int_0^T \langle p(t), f'(\bar{y}(t); \eta(t))\rangle_X \, dt. \tag{5.6}$$

Now let $\zeta \in \mathbb{W}^r_0(\mathcal{D}, X)$ be arbitrary. According to Lemma 5.2 there is a sequence $\{h_n\} \subset L^r(]0,T[;U)$ such that $S'(\bar{u}; h_n) \to \zeta$ in $\mathbb{W}^r_0(\mathcal{D}, X)$. Due to $\mathbb{W}^r_0(\mathcal{D}, X) \hookrightarrow C([0,T];Y)$, see Remark 2.9, it follows that $S'(\bar{u}; h_n) \to \zeta$ in $C([0,T];Y)$ such that (2.4) gives

$$f'(\bar{y}; S'(\bar{u}; h_n)) \to f'(\bar{y}; \zeta) \quad \text{in } L^r(]0,T[;X).$$

Thus we are allowed to pass to the limit in (5.6) to obtain

$$\int_0^T \langle \lambda(t), \zeta(t)\rangle_{\mathcal{D}}\, dt \geq \int_0^T \langle p(t), f'(\bar y(t); \zeta(t))\rangle_X \, dt \quad \forall\, \zeta \in \mathbb{W}^r_0(\mathcal{D}, X). \tag{5.7}$$

Now, let $v \in \mathcal{D}$ and $\varphi \in C_0^\infty(]0,T[)$ with $\varphi \geq 0$ be arbitary. Then $\varphi\, v \in \mathbb{W}^r_0(\mathcal{D}, X)$ so that it can be chosen as test function in (5.7). Since the directional derivative $\eta \mapsto f'(\bar y; \eta)$ is positively homogeneous, this yields

$$\int_0^T \langle \lambda(t), v\rangle_{\mathcal{D}}\, \varphi(t)\, dt \geq \int_0^T \langle p(t), f'(\bar y(t); v)\rangle_X \, \varphi(t)\, dt \quad \forall\, v \in \mathcal{D},\ \varphi \in C_0^\infty(]0,T[).$$

The fundamental lemma of the calculus of variations then gives (5.2b). Together with (5.3) and (5.5) this is the claimed optimality system.

It remains to prove the improved regularity of λ. By (5.2b) and (2.5) we find for all $v \in L^r(]0,T[; \mathcal{D})$ that

$$\begin{aligned}
\int_0^T \langle \lambda(t), v(t)\rangle_{\mathcal{D}}\, dt &\leq \int_0^T \big|\langle p(t), f'(\bar y(t); -v(t))\rangle_X\big|\, dt \\
&\leq K_y \int_0^T \|v(t)\|_Y\, \|p(t)\|_{X^*}\, dt \\
&\leq c\, \|p(t)\|_{L^{r'}(]0,T[;X^*)}\, \|v\|_{L^r(]0,T[;Y)}.
\end{aligned}$$

Therefore, by the Hahn-Banach theorem, λ can be extended to a bounded linear functional on $L^r(]0,T[;Y)$, which we denote by the same symbol for simplicity. Because of $(L^r(]0,T[,Y))^* = L^{r'}(]0,T[;Y^*)$, see e.g. [11, Theorem IV.1.14], this gives the desired regularity of λ. □

Remark 5.4. *If $\mathcal{D}$ is dense in Y, then (5.2b) clearly also holds for all $v \in Y$. This will be the case in our concrete example in Section 6.*

Remark 5.5. *An inspection of the proof of Theorem 5.3 shows that the arguments cannot be applied, if additional control constraints are given or if* $\mathrm{Rg}(B)$ *is not dense in X. The same observation was made in [25], where strong stationarity for optimal control of the obstacle problem is shown to be necessary for local optimality.*

Remark 5.6. *Let us compare the optimality system (5.2) with the one in (4.27) obtained via regularization. Note first that the multiplier λ coincides with the weak limit of $\{[f'_\varepsilon(y_\varepsilon)]^* p_\varepsilon\}$, which also implies its higher regularity, i.e., $\lambda \in L^{r'}(]0,T[;Y^*)$. We further observe that the sign condition (5.2b) is missing in (4.27). As pointed out in Remark 4.16, it is possible to refine the limit analysis for vanishing regularization, if concrete instances for the nonlinearity f are under consideration. However, the result of Theorem 5.3 is still sharper compared to the optimality systems derived in this way, as we will see in Section 6.*

To see that (5.2) is indeed a comparatively strong optimality condition, we prove the following

Theorem 5.7 (Equivalence between B- and strong stationarity). *Assume that $\bar u \in L^r(]0,T[;U)$ together with its state $\bar y \in \mathbb{W}^r_0(\mathcal{D}, X)$, an adjoint state $p \in \mathbb{W}^{r'}_T(X^*, \mathcal{D}^*)$, and a multiplier $\lambda \in L^{r'}(]0,T[;Y^*)$ satisfy the optimality system (5.2a)–(5.2c). Then it also satisfies the variational inequality in (5.1).*

Proof. Let $h \in L^r(]0,T[;U)$ be arbitrary and set $\eta := S'(\bar u; h)$. Then we test (5.2c) with h, insert (3.3), integrate by parts as in the proof of Theorem 5.3, insert (5.2a) in the arising formula, and use (5.2b) to obtain

$$\begin{aligned}
-\partial_u J(\bar y,\bar u)h &= \int_0^T \langle p(t), B\,h(t)\rangle_X\,dt \\
&= \int_0^T \langle p(t), \dot\eta(t) + A\eta(t) + f'(\bar y(t);\eta(t))\rangle_X\,dt \\
&= \int_0^T \big(\langle \partial_y J(\bar y,\bar u)(t) - \lambda(t), \eta(t)\rangle_{\mathcal D} + \langle p(t), f'(\bar y(t);\eta(t))\rangle_X\big)dt \\
&\le \partial_y J(\bar y,\bar u)\eta,
\end{aligned}$$

which is exactly (5.1). □

Let us finally compare the assertion of Theorem 5.3 with the results known for optimal control of ODEs. To this end set

$$\mathcal D = X = Y = \mathbb R^n, \quad U = \mathbb R^m, \quad A = 0, \quad J(y,u) = \int_0^T F(t, y(t), u(t))\,dt,$$

where $n, m \in \mathbb N$ and $F : \mathbb R \times \mathbb R^n \times \mathbb R^m \mapsto \mathbb R$ is supposed to be smooth and convex w.r.t. y and u. Then the strong stationarity conditions in Theorem 5.3 read

$$-\dot p + \lambda = \partial_y F(t, y(t), u(t)) \quad \text{in }]0,T[, \quad p(T) = 0 \tag{5.8a}$$

$$\langle \lambda(t), v\rangle \le g_y'(t, y(t); v) \quad \forall\, v \in \mathbb R^n \tag{5.8b}$$

$$B^\top p(t) + \partial_u F(t, y(t), u(t)) = 0 \quad \text{a.e. in }]0,T[, \tag{5.8c}$$

where $g : \mathbb R \times \mathbb R^n \to \mathbb R$ is defined by

$$g(t,y) := \langle p(t), f(y)\rangle_{\mathbb R^n} \tag{5.9}$$

and g_y' denotes the directional derivative of g w.r.t. y. It is easily verified that $g_y'(t,y;v) = \langle p(t), f'(y;v)\rangle_{\mathbb R^n}$.

Next we apply [7, Theorem 5.2.1] to the above setting. For this purpose, define the Hamiltonian by

$$\begin{aligned}
&H : \mathbb R \times \mathbb R^n \times \mathbb R^n \times \mathbb R^m \times \mathbb R \to \mathbb R, \\
&H(t, y, p, u, \mu) := \langle -p, Bu - f(y)\rangle_{\mathbb R^n} - \mu\, F(t, y, u)
\end{aligned}$$

Then [7, Theorem 5.2.1] yields the existence of a multiplier $\mu \in \{0,1\}$ such that

$$\dot p(t) \in \partial_y H(t, y(t), p(t), u(t), \mu), \quad \text{a.e. in }]0,T[, \quad p(T) = 0, \tag{5.10}$$

$$H(t, y(t), p(t), u(t), \mu) = \max_{w\in\mathbb R^n} \big\{H(t, y(t), p(t), w, \mu)\big\} \quad \text{a.e. in }]0,T[\tag{5.11}$$

and μ and p must not vanish at the same time. Herein, $\partial_y H$ denotes the partial generalized gradient, cf. [7, Chapter 2]. If we define $\lambda(t) := \dot p(t) + \mu\,\partial_y F(t, y(t), u(t))$, then, thanks to the sum rule for generalized gradients, (5.10) can be rephrased by

$$\lambda(t) \in \partial_y\big(\langle -p(t), Bu(t) - f(y(t))\rangle\big) \quad \Longleftrightarrow \quad \langle \lambda(t), v\rangle \le g_y^\circ(t, y(t); v) \quad \forall\, v \in \mathbb R^n$$

with g as defined in (5.9) and g_y° denoted the generalized directional derivative w.r.t. y. Since F is convex, we obtain

$$\begin{aligned}(5.11) \quad &\iff \quad \partial_u H(t, y(t), p(t), u(t), \mu) = 0 \text{ a.e. in }]0,T[\\ &\iff \quad B^\top p(t) + \mu\, \partial_u F(t, y(t), u(t)) = 0 \text{ a.e. in }]0,T[.\end{aligned}$$

Altogether, we arrive at the following Fritz-John-type conditions:

$$-\dot{p} + \lambda = \mu\, \partial_y F(t, y(t), u(t)) \text{ in }]0,T[, \quad p(T) = 0 \tag{5.12a}$$

$$\langle \lambda(t), v\rangle \leq g_y^\circ(t, y(t); v) \quad \forall\, v \in \mathbb{R}^n \tag{5.12b}$$

$$B^\top p(t) + \mu\, \partial_u F(t, y(t), u(t)) = 0 \text{ a.e. in }]0,T[. \tag{5.12c}$$

Because of

$$g_y^\circ(t, y; v) = \limsup_{\substack{z\to y\\ \tau\searrow 0}} \frac{g(t, z + \tau v) - g(z)}{\tau} \geq g_y'(t, y; v),$$

we see that the conditions in (5.8) are indeed sharper compared to (5.12), even in the qualified case $\mu = 1$. We point out however that Clarke's result covers a much broader class of ODE-problems compared to the specific example considered here.

6. Application to a concrete setting. In this section we apply the results of the previous sections to the following specific setting:

Let Ω be a bounded domain in $\mathbb{R}^n$, $n \in \mathbb{N}$. Then we set

$$\begin{aligned} &X = H^{-1}(\Omega), \quad \mathcal{D} = H_0^1(\Omega), \quad U = Y = L^2(\Omega),\\ &A = -\triangle, \qquad f = -\max, \qquad B = E : L^2(\Omega) \hookrightarrow H^{-1}(\Omega),\end{aligned} \tag{6.1}$$

where E denotes the embedding operator, i.e., $\langle Eu, v\rangle = \int_\Omega u\, v\, dx$. Moreover, with a little abuse of notation, we denote by $\max : L^2(\Omega) \to L^2(\Omega) \hookrightarrow H^{-1}(\Omega)$ the Nemytskii-operator associated with $\mathbb{R} \ni x \mapsto \max\{0, x\} \in \mathbb{R}$. Note that $\max\{0, .\}$ satisfies the Carathéodory-condition and is globally Lipschitz with constant 1 so that the associated Nemytskii-operator maps all of $L^2(\Omega)$ into $L^2(\Omega)$. In summary, the optimal control problem, considered in this concrete example, reads as follows

$$\left.\begin{aligned} \min \quad & J(y, u)\\ \text{s.t.} \quad & \dot{y}(t) - \triangle y(t) - \max(y(t)) = u(t) \quad \text{in }]0,T[\\ & y(0) = 0.\end{aligned}\right\} \tag{P$_{\text{ex}}$}$$

We impose the following hypothesis on the objective in (P$_{\text{ex}}$):

Assumption 6.1. *The objective functional*

$$J : L^r(]0,T[; H_0^1(\Omega)) \times L^r(]0,T[; L^2(\Omega)) \to \mathbb{R} \quad \textit{with } r > 2 \tag{6.2}$$

is assumed to be convex and continously differentiable.

A typical example, which also fulfills the additional assumptions of Proposition 2.11, is

$$J(y, u) = \frac{1}{2}\int_0^T \int_\Omega |\nabla y - z|^2\, dx\, dt + \frac{\alpha}{r}\int_0^T \Big(\int_\Omega |u(x,t)|^2\, dx\Big)^{r/2} dt$$

with a desired state gradient $z \in L^2(]0,T[;H_0^1(\Omega))$ and a Tikhonov parameter $\alpha > 0$.

LEMMA 6.2. *Under Assumption 6.1, in particular $r > 2$, the setting in (6.1) fulfills Assumption 2.1 for all $\theta \in]1/2,(r-1)/r[$. Moreover, the Lipschitz and boundedness conditions in (2.1), (3.4), and (3.5) are fulfilled with constants L, Q, and K independent of M so that $f = -\max$ and its directional derivative are globally Lipschitz continuous.*
Furthermore, $A = -\triangle$ satisfies maximal parabolic $L^r(]0,T[;H^{-1}(\Omega))$-regularity for every $r \in]1,\infty[$.

Proof. As most of the assumptions are obvious to see, we concentrate on the non-trivial conditions. We start with the assumptions on $Y = L^2(\Omega)$. In view of [2, Sec. 1], [23, Lemma 3.7 i)], and [12, Lemma 3.4], it holds for all $\theta > 1/2$ that

$$(H^{-1}(\Omega), H_0^1(\Omega))_{\theta,\infty} \hookrightarrow [H^{-1}(\Omega), H_0^1(\Omega)]_{\frac{1}{2}} = L^2(\Omega) = Y, \tag{6.3}$$

which is the condition in Assumption 2.1.4. Herein, $[X,\mathcal{D}]_{1/2}$ denotes the complex interpolation space, [31]. Note that, due to $r > 2$, the interval $]1/2,(r-1)/r[$ is non-empty.

Moreover, $\mathbb{R} \ni x \mapsto \max\{x,0\} \in \mathbb{R}$ is clearly globally Lipschitz continuous with constant 1 and satisfies $|\max\{x,0\}| \le |x|$ for all $x \in \mathbb{R}$. These properties readily transfer to the associated Nemytskii-operator as a mapping from $L^2(\Omega)$ to $L^2(\Omega)$, giving in turn that (2.1) and (2.2) are fulfilled. Due to global Lipschitz continuity, the constant in (2.1) is independent of M in this example. As $\mathbb{R} \ni x \mapsto \max\{x,0\} \in \mathbb{R}$ is directionally differentiable with

$$\max{}'(y;v) = \begin{cases} v, & \text{if } y > 0 \\ \max\{v,0\}, & \text{if } y = 0 \\ 0, & \text{if } y < 0 \end{cases} \tag{6.4}$$

and there holds $|\max{}'(y;v)| \le |v|$ for all $v \in \mathbb{R}$, Lebesgue's dominated convergence theorem implies that the Nemytskii-operator $\max : L^2(\Omega) \to L^2(\Omega)$ is directionally differentiable, too. For convenience of the reader, we included the proof in Appendix D. The directional derivative is given by (6.4) evaluated pointwise a.e. in Ω. Therefore, it satisfies (3.4) and (3.5) globally, i.e., with constants Q and K independent of M. Clearly, all these properties also hold, if max is considered as an operator with range in $X = H^{-1}(\Omega)$.

It is well known that $A = -\triangle$ satisfies maximal parabolic $L^2(]0,T[;H^{-1}(\Omega))$-regularity, see for instance [14]. Moreover, according to [8], if an operator satisfies maximal parabolic $L^2(]0,T[;X)$-regularity, then it also satisfies maximal parabolic $L^r(]0,T[;X)$-regularity for every $r \in]1,\infty[$. Since $-\triangle$ satisfies maximal parabolic regularity, it is automatically a generator of an analytic semigroup, see e.g. [2, Rem. 3.1(b)]. As Poincaré's inequality yields $0 \notin \sigma(-\triangle)$, this gives that Assumption 2.1.3 is fulfilled, too.

Finally, $L^2(\Omega)$, equipped with its natural norm, i.e., $\|u\|_{L^2(\Omega)} = (\int_\Omega u^2\,dx)^{1/2}$, is (locally) uniformly convex, see Remark 2.2 and [29, Example 4.7.7], so that Assumption 2.1.2 is also satisfied. □

REMARK 6.3. *One could also discuss ($\mathrm{P_{ex}}$) in another setting, where $X = Y = L^2(\Omega)$. According to [13, Thm. 7.2] the Laplacian satisfies maximal parabolic $L^r(]0,T[;L^2(\Omega))$-regularity under mild assumptions on the boundary $\partial\Omega$, even if mixed boundary conditions are imposed. Then Assumption 2.1.4 is fulfilled by every $\theta > 0$, which allows* *us to set $r = 2$, cf. (2.6), such that we obtain a Hilbert-space for the control. However, the domain of A restricted to $X = L^2(\Omega)$ is hard to characterize in general, especially when the $\partial\Omega$ is non-smooth and a general divergence-type operator is considered instead of the Laplacian. Therefore, we chose $X = H^{-1}(\Omega)$ as the functional analytic framework.*

6.1. Regularization. First we apply Theorem 4.15 to our specific setting and sharpen the result by employing the special structure of the example under consideration. For this purpose, we need the following

ASSUMPTION 6.4 (Regularization of max). *There exists a family of functions $\{\max_\varepsilon\}_{\varepsilon>0}$, $\max_\varepsilon : \mathbb{R} \to \mathbb{R}$, with the following properties:*

1. *For all $\varepsilon > 0$ it holds $\max_\varepsilon \in C^1(\mathbb{R})$ with $\max_\varepsilon{}'(x) \geq 0$ for all $x \in \mathbb{R}$,*
2. *For all $x \in \mathbb{R}$ there holds $\max_\varepsilon(x) \to \max\{0, x\}$ for $\varepsilon \searrow 0$,*
3. *There exist constants $\gamma, \beta \geq 0$, independent of ε, such that the following growth condition is fulfilled*

$$|\max_\varepsilon(x)| \leq \gamma + \beta|x| \quad \forall\, x \in \mathbb{R},\ \varepsilon > 0. \tag{6.5}$$

4. *There is a constant $\kappa > 0$ such that $|\max_\varepsilon{}'(x)| \leq \kappa$ for all $x \in \mathbb{R}$.*

There are numerous possibilities to construct a family of functions satisfying Assumption 6.4. We only refer to the regularized max-functions used in [25] and [28], respectively, which are

$$\max{}_\varepsilon^{(1)}(x) := \begin{cases} 0, & x \leq 0, \\ \frac{1}{2\varepsilon}\, x^2, & x \in]0, \varepsilon[\,, \\ x - \frac{\varepsilon}{2}, & x \geq \varepsilon, \end{cases} \qquad \max{}_\varepsilon^{(2)}(x) := \begin{cases} \max\{x, 0\}, & |x| \geq \varepsilon, \\ \frac{1}{16\varepsilon^3}(x+\varepsilon)^3(3\varepsilon - x), & |x| < \varepsilon. \end{cases}$$

It is easiliy seen that these functions satisfy the conditions in Assumption 6.4.

In the sequel, we again denote the Nemytskii-operator associated with $\max_\varepsilon$ by the same symbol. Similarly to the corresponding part of the proof of 6.2, one shows the following

LEMMA 6.5. *Given Assumption 6.4, the Nemytskii-operator $f_\varepsilon := -\max_\varepsilon : L^2(\Omega) \to L^2(\Omega)$ fulfills Assumption 4.2. The conditions (4.1) and (4.3) are satisfied with constants independent of M.*

Proof. By assumption $\max_\varepsilon$ fulfills the Carathéodory condition and is globally Lipschitz continuous with constant κ so that the associated Nemytskii-operator is well defined and globally Lipschitz continuous from $L^2(\Omega)$ to $L^2(\Omega)$. The boundedness condition in (4.2) follows immediately from (6.5). The Gâteaux-differentiability can be deduced completely analogously to the proof of Lemma D.1. The boundedness of the Gâteaux-derivative in (4.3) is ensured by Assumption 6.4.4. Finally, Assumption 4.2.1 follows from Assumption 6.4.2 in combination with $|\max\{0, x\}| \leq |x|$, (6.5), and Lebesgue's dominated convergence theorem. ∎

THEOREM 6.6 (Optimality system after passing to the limit). *Assume in addition to Assumption 6.1 that $r \in]2, 4[$, and let $\bar{u}$ be a local solution of (P_{ex}). Then there exist unique $\bar{y} \in \mathbb{W}_0^r(H_0^1(\Omega), H^{-1}(\Omega))$, $p \in \mathbb{W}_T^{r'}(H_0^1(\Omega), H^{-1}(\Omega))$, and $\lambda \in$*

$L^{r'}(]0,T[;L^2(\Omega))$ *such that*

$$\dot{\bar{y}} - \triangle \bar{y} - \max(\bar{y}) = \bar{u}, \quad \bar{y}(0) = 0 \tag{6.6a}$$

$$-\dot{p} - \triangle p + \lambda = \partial_y J(\bar{y}, \bar{u}), \quad p(T) = 0 \tag{6.6b}$$

$$p(t,x) + \partial_u J(\bar{y}, \bar{u})(t,x) = 0 \quad \text{a.e. in }]0,T[\times\Omega \tag{6.6c}$$

$$\int_0^T \int_\Omega \lambda\, p\, dx\, dt \leq 0. \tag{6.6d}$$

Proof. From Theorem 4.15 we know that there exist unique $\bar{y} \in \mathbb{W}_0^r(H_0^1(\Omega), H^{-1}(\Omega))$, $p \in \mathbb{W}_T^{r'}(H_0^1(\Omega), H^{-1}(\Omega))$, and $\lambda \in L^{r'}(]0,T[;L^2(\Omega))$ such that (6.6a)–(6.6c) holds. We have also seen that there exist sequences $\{\lambda_\varepsilon\} \subset L^{r'}(]0,T[;L^2(\Omega))$ and $\{p_\varepsilon\} \subset \mathbb{W}_T^{r'}(H_0^1(\Omega), H^{-1}(\Omega))$ such that

$$\lambda_\varepsilon = [f_\varepsilon'(y_\varepsilon)]^* p_\varepsilon \rightharpoonup \lambda \quad \text{in } L^{r'}(]0,T[;L^2(\Omega)), \tag{6.7}$$

$$p_\varepsilon \rightharpoonup p \quad \text{in } \mathbb{W}_T^{r'}(H_0^1(\Omega), H^{-1}(\Omega)), \tag{6.8}$$

cf. Lemma 4.14 and (4.28). By [2, Eq. (1.2)], there holds

$$\mathbb{W}_T^{r'}(H_0^1(\Omega), H^{-1}(\Omega)) \hookrightarrow L^q(]0,T[;(H^{-1}(\Omega), H_0^1(\Omega))_{\theta,1}) \qquad \forall\, q \in [1,\infty] \text{ with } 1/q > \theta - 1/r > 0$$

and, due to the compactness of $H_0^1(\Omega) \hookrightarrow\hookrightarrow H^{-1}(\Omega)$, this embedding is compact as well. Since $r > 2$ and $\theta > 1/2$, see Lemma 6.2, the above sign condition reduces to $\theta < 1/q + 1/r$, and the additional assumption $r < 4$ guarantees that this inequality is satisfied for $q = r$, if we choose θ sufficiently close to $1/2$. Since $(H^{-1}(\Omega), H_0^1(\Omega))_{\theta,1} \hookrightarrow (H^{-1}(\Omega), H_0^1(\Omega))_{\theta,\infty} \hookrightarrow L^2(\Omega)$ for $\theta > 1/2$, see [2, Eq. (1.1)] and (6.3), $\mathbb{W}_T^{r'}(H_0^1(\Omega), H^{-1}(\Omega))$ is therefore compactly embedded in $L^r(]0,T[;L^2(\Omega)$ so that (6.8) yields

$$p_\varepsilon \to p \text{ in } L^r(]0,T[;L^2(\Omega)). \tag{6.9}$$

Since in our concrete setting $\lambda_\varepsilon = [-{\max}_\varepsilon'(y_\varepsilon)]^* p_\varepsilon$, the monotonicity of $\max_\varepsilon$ by Assumption 6.4.1 together with (6.7) and (6.9) implies

$$\begin{aligned} 0 &\geq -\int_0^T \int_\Omega {\max}_\varepsilon'(y_\varepsilon(x,t))\, p_\varepsilon(x,t)^2\, dx\, dt \\ &= \langle \lambda_\varepsilon, p_\varepsilon \rangle_{L^r(]0,T[;L^2(\Omega))} \overset{\varepsilon\searrow 0}{\longrightarrow} \langle \lambda, p \rangle_{L^r(]0,T[;L^2(\Omega))}, \end{aligned} \tag{6.10}$$

which is the desired sign condition in (6.6d). □

Remark 6.7. *As already indicated in Remark 4.16, the result of Theorem 6.6 is indeed sharper than the general result of Theorem 4.15, as it additionally contains the sign condition in* (6.6d).

6.2. Strong Stationarity. In this section we apply Theorem 5.3 to our concrete setting. Making use of the special structure of max the result of Theorem 5.3 can significantly be sharpened, which is demonstrated in the following

Theorem 6.8 (Strong stationarity). *Let $\bar{u} \in L^r(]0,T[;L^2(\Omega))$ be locally optimal with associated state $\bar{y} = S(\bar{u})$. Then there exists a unique adjoint state $p \in \mathbb{W}_T^{r'}(H_0^1(\Omega), H^{-1}(\Omega))$*

and a unique multiplier $\lambda \in L^{r'}(]0,T[;L^2(\Omega))$ *such that*

$$-\dot p - \triangle p + \lambda = \partial_y J(\bar y, \bar u), \quad p(T) = 0 \tag{6.11a}$$

$$\left.\begin{aligned} \lambda(t,x) &= -p(t,x), && \text{if } \bar y(t,x) > 0 \\ -p(t,x) \le \lambda(t,x) &\le 0, && \text{if } \bar y(t,x) = 0 \\ \lambda(t,x) &= 0, && \text{if } \bar y(t,x) < 0 \end{aligned}\right\} \text{ f.a.a. } (t,x) \in]0,T[\times\Omega \tag{6.11b}$$

$$p(t,x) + \partial_u J(\bar y, \bar u)(t,x) = 0 \quad \text{a.e. in }]0,T[\times\Omega. \tag{6.11c}$$

Proof. From Theorem 5.3 we know that there exists a unique adjoint state $p \in \mathbb{W}_T^{r'}(H_0^1(\Omega), H^{-1}(\Omega))$ and a unique multiplier $\lambda \in L^{r'}(]0,T[;L^2(\Omega))$ such that

$$-\dot p - \triangle p + \lambda = \partial_y J(\bar y, \bar u), \quad p(T) = 0 \tag{6.12}$$

$$\langle \lambda(t), v\rangle_{H_0^1(\Omega)} \ge \langle -\max{}'(\bar y(t); v), p(t)\rangle_{H_0^1(\Omega)} \quad \forall\, v \in H_0^1(\Omega), \text{ f.a.a. } t \in]0,T[\tag{6.13}$$

$$p(t,x) + \partial_u J(\bar y, \bar u)(t,x) = 0 \quad \text{a.e. in }]0,T[\times\Omega. \tag{6.14}$$

It remains to show that (6.13) implies (6.11b). In view of $\lambda(t), \max'(\bar y(t); v) \in L^2(\Omega)$ f.a.a. $t \in]0,T[$, the density of $H_0^1(\Omega)$ in $L^2(\Omega)$ yields

$$(\lambda(t), v)_{L^2(\Omega)} \ge (-\max{}'(\bar y(t); v), p(t))_{L^2(\Omega)} \quad \forall\, v \in L^2(\Omega), \text{ f.a.a. } t \in]0,T[, \tag{6.15}$$

where we used that $L^2(\Omega) \ni v \mapsto \max'(\bar y(t); v) \in L^2(\Omega)$ is continuous, see also Assumption 2.1.5. Next define for almost all $t \in]0,T[$ the following sets (up to sets of zero measure):

$$\begin{aligned} \Omega_t^- &:= \{x \in \Omega : \bar y(t,x) < 0\}, \\ \Omega_t^0 &:= \{x \in \Omega : \bar y(t,x) = 0\}, \\ \Omega_t^+ &:= \{x \in \Omega : \bar y(t,x) > 0\}. \end{aligned} \tag{6.16}$$

Then (6.15) can be continued as

$$(\lambda(t), v)_{L^2(\Omega)} \ge (-\chi_{\Omega_t^+} v - \chi_{\Omega_t^0} \max\{v; 0\}, p(t))_{L^2(\Omega)} \quad \forall\, v \in L^2(\Omega), \text{ f.a.a. } t \in]0,T[.$$

Now let $v \in L^2(\Omega)$ with $v(x) \ge 0$ a.e. in Ω be arbitrary. Then testing with v and $-v$ leads to

$$(-\chi_{\Omega_t^+} v, p(t))_{L^2(\Omega)} \ge (\lambda(t), v)_{L^2(\Omega)} \ge (-\chi_{\Omega_t^+ \cup \Omega_t^0} v, p(t))_{L^2(\Omega)} \quad \text{f.a.a. } t \in]0,T[$$

and, since $v \ge 0$ was arbitrary, the fundamental lemma of the calculus of variations implies

$$-\chi_{\Omega_t^+ \cup \Omega_t^0}(x)\, p(t,x) \le \lambda(t,x) \le -\chi_{\Omega_t^+}(x)\, p(t,x) \quad \text{f.a.a. } (t,x) \in]0,T[\times\Omega,$$

whence (6.11b). □

Although we sharpened the limit analysis for vanishing regularization in Theorem 6.6, the result of the previous theorem is still more rigorous as the following corollary shows.

Corollary 6.9. *Let* $r \in]2,4[$ *and assume that* $\bar u \in L^r(]0,T[;L^2(\Omega))$ *together with its state* $\bar y \in \mathbb{W}_0^r(H_0^1(\Omega), H^{-1}(\Omega))$*, an adjoint state* $p \in \mathbb{W}_T^{r'}(H_0^1(\Omega), H^{-1}(\Omega))$*, and a* *multiplier* $\lambda \in L^{r'}(]0,T[;L^2(\Omega))$ *satisfy the optimality system* (6.11a)–(6.11c)*. Then it also satisfies* (6.6)*.*

Proof. One just needs to show that (6.11b) implies (6.6d). This can be seen by testing (6.11b) with $p(t) \in L^2(\Omega)$ and integrating over Ω and $]0,T[$. Note that $p \in L^r(]0,T[;L^2(\Omega))$ thanks to the embedding $\mathbb{W}_T^{r'}(H_0^1(\Omega),H^{-1}(\Omega)) \hookrightarrow L^r(]0,T[;L^2(\Omega))$ for $r \in]2,4[$, see the proof of Theorem 6.6, $\lambda \in L^{r'}(]0,T[;L^2(\Omega))$ implies that the product λp is integrable over $]0,T[\times\Omega$. Using the sign conditions in (6.11b) then leads to (6.6d). □

As in the general case the strong stationarity conditions in Theorem 6.8 are again equivalent to B-stationarity, as the following result shows.

Theorem 6.10. *Assume that* $\bar u \in L^r(]0,T[;L^2(\Omega))$ *together with its state* $\bar y \in \mathbb{W}_0^r(H_0^1(\Omega),H^{-1}(\Omega))$*, an adjoint state* $p \in \mathbb{W}_T^{r'}(H_0^1(\Omega),H^{-1}(\Omega))$*, and a multiplier* $\lambda \in L^{r'}(]0,T[;L^2(\Omega))$ *satisfy the optimality system* (6.11a)–(6.11c)*. Then it also satisfies the variational inequality* (6.17)*, i.e.,*

$$\partial_y J(\bar y,\bar u)S'(\bar u;h) + \partial_u J(\bar y,\bar u)h \geq 0 \quad \forall\, h \in L^r(]0,T[;L^2(\Omega)). \tag{6.17}$$

Proof. In view of Theorem 5.7 we only need to show that (6.11b) implies

$$(\lambda(t),v)_{L^2(\Omega)} \geq (-\max{}'(\bar y(t);v),p(t))_{L^2(\Omega)} \quad \forall\, v \in L^2(\Omega), \text{ f.a.a. } t \in]0,T[. \tag{6.18}$$

For this purpose rewrite (6.11b) by means of the characteristic functions associated with the sets in (6.16) to obtain

$$\left.\begin{aligned} \chi_{\Omega_t^+}(x)\,\lambda(t,x) &= -\chi_{\Omega_t^+}(x)\,p(t,x)\\ -\chi_{\Omega_t^0}(x)\,p(t,x) \leq \chi_{\Omega_t^0}(x)\,\lambda(t,x) &\leq 0\\ \chi_{\Omega_t^-}(x)\,\lambda(t,x) &= 0 \end{aligned}\right\} \quad \text{f.a.a. } (t,x) \in]0,T[\times\Omega.$$

Multiplying with $v \in L^2(\Omega)$, $v \geq 0$ a.e. in Ω, and integrating over Ω leads to

$$\left.\begin{aligned} \int_\Omega \chi_{\Omega_t^+}\,\lambda(t)\,v\,dx &= \int_\Omega -\chi_{\Omega_t^+}\,p(t)\,v\,dx\\ -\int_\Omega \chi_{\Omega_t^0}\,p(t)\,v\,dx \leq \int_\Omega \chi_{\Omega_t^0}\,\lambda(t)\,v\,dx &\leq 0\\ \int_\Omega \chi_{\Omega_t^-}\,\lambda(t)\,v\,dx &= 0 \end{aligned}\right\} \quad \text{f.a.a. } t \in]0,T[.$$

Adding these inequalities gives

$$(-\chi_{\Omega_t^+}v,p(t))_{L^2(\Omega)} \geq (\lambda(t),v)_{L^2(\Omega)} \geq (-\chi_{\Omega_t^+\cup\Omega_t^0}v,p(t))_{L^2(\Omega)} \quad \text{f.a.a. } t \in]0,T[. \tag{6.19}$$

Now let $w \in L^2(\Omega)$ be arbitrary, but fixed. We test the first inequality in (6.19) with $v = -\min\{0,w\}$ and the second one with $\max\{0,w\}$, respectively, to obtain

$$(-\chi_{\Omega_t^+}\min\{0,w\},p(t))_{L^2(\Omega)} \leq (\lambda(t),\min\{0,w\})_{L^2(\Omega)}, \tag{6.20}$$

$$(\lambda(t),\max\{0,w\})_{L^2(\Omega)} \geq (-\chi_{\Omega_t^+\cup\Omega_t^0}\max\{0,w\},p(t))_{L^2(\Omega)}. \tag{6.21}$$

In view of $w = \min\{0, w\} + \max\{0, w\}$, adding (6.20) and (6.21) yields

$$(\lambda(t), w)_{L^2(\Omega)} \geq (-\chi_{\Omega_t^+} w - \chi_{\Omega_t^0} \max\{w; 0\}, p(t))_{L^2(\Omega)}, \tag{6.22}$$

which, on account of (6.4) and (6.16), gives in turn (6.18). □

Appendix A. Global Existence for the State Equation.

Proof of Proposition 2.5. (i) Existence for Lipschitz continuous right hand sides:
We first assume that $u \in C^{0,1}([0,T];U)$ and apply [26, Ch. 6, Thm. 3.3]. To this end we rewrite (2.8) as

$$\dot{y}(t) + Ay(t) = f_u(t, y(t)), \quad y(0) = 0$$

with $f_u(t,y) := Bu(t) - f(y)$. Then Assumption 2.1.4 gives for every $M > 0$ that

$$\begin{aligned} \|f_u(t_1,y_1) - f_u(t_2,y_2)\|_X &\leq L(M)\|y_1 - y_2\|_Y + \|B\|_{\mathcal{L}(U,X)}\, L_u\, |t_1 - t_2| \\ &\leq c\,\big(|t_1 - t_2| + \|y_1 - y_2\|_{\mathcal{D}(A^\theta)}\big) \end{aligned}$$

for all $t_1, t_2 \in \mathbb{R}$ and all $y_1, y_2 \in \overline{B_Y(0,M)} \cap \mathcal{D}(A^\theta)$. Herein $L_u > 0$ denotes the Lipschitz constant of u. Moreover, Assumption 2.1.4 yields

$$\begin{aligned} \|f_u(t,y)\|_X &\leq K(1 + \|y\|_Y) + \|B\|_{L(U,X)}\|u\|_{C([0,T];U)} \\ &\leq c(1 + \|y\|_{\mathcal{D}(A^\theta)}) \quad \forall\, y \in \mathcal{D}(A^\theta). \end{aligned}$$

Therefore, all assumptions of [26, Ch. 6, Thm. 3.3] are satisfied giving in turn the existence of a unique (classical) solution $y \in C([0,T];X) \cap C^1(]0,T];X)$ with $y(t) \in \mathcal{D}$, $t \in [0,T]$, of (2.8). It is well known that this solution also satisfies the integral equation (2.9).

(ii) Boundedness and continuity of solutions:
Next we prove that y is uniformly bounded in Y with a bound depending on u. We start with integral equation, which, for an arbitrary $t \in [0,T]$, results in

$$\begin{aligned} \|y(t)\|_Y &\leq \int_0^t \big\|e^{-(t-s)A}\big(Bu(s) - f(y(s))\big)\big\|_Y\, ds \\ &\leq \int_0^T \|e^{-sA}\|_{\mathcal{L}(X,Y)}^{r'}\, ds^{1/r'}\, \|B\|_{\mathcal{L}(U,X)}\, \|u\|_{L^r(]0,T[;U)} \\ &\qquad + K \int_0^T \|e^{-sA}\|_{\mathcal{L}(X,Y)}\, ds + K \int_0^t \|e^{-(t-s)A}\|_{\mathcal{L}(X,Y))}\|y(s)\|_Y\, ds, \end{aligned}$$

where we used Assumption 2.1.4 for the last estimate. Note that $r' = r/(r-1) < \theta^{-1}$ by (2.6). Thus Lemma 2.4 is applicable, which, together with Gronwall's Lemma, implies the existence of a constant $C > 0$ such that

$$\begin{aligned} \|y(t)\|_Y &\leq C\, \exp\Big(K \int_0^t \|e^{-(t-s)A}\|_{\mathcal{L}(X,Y)}\, ds\Big)\big(1 + \|u\|_{L^r(]0,T[;U)}\big) \\ &\leq C\big(1 + \|u\|_{L^r(]0,T[;U)}\big) \end{aligned} \tag{A.1}$$

for all $t \in [0,T]$.

Now consider two arbitrary time points $t_1, t_2 \in [0,T]$ with $t_2 > t_1$. This time the integral equation yields

$$\begin{aligned}\|y(t_2) - y(t_1)\|_Y \leq \int_0^{t_1} \big\|\big(e^{-(t_2-t_1)A} - I\big) e^{-(t_1-s)A} x(s)\big\|_Y ds \\ + \int_{t_1}^{t_2} \big\|e^{-(t_2-s)A} x(s)\big\|_Y ds,\end{aligned} \tag{A.2}$$

where we abbreviated $x(s) := Bu(s) - f(y(s)) \in X$. In view of

$$\|f(y(t))\|_X \leq K(1 + \|y(t)\|_Y) \leq K\big(1 + C(1 + \|u\|_{L^r(]0,T[;U)})\big) \quad \forall\, t \in [0,T] \tag{A.3}$$

and Lemma 2.4, the second integrand in (A.2) an element of $L^1(]0,T[)$ and therefore vanishes if $t_1 \to t_2$. For the first integral we argue as follows: Since $\theta < 1$ and $r > \frac{1}{1-\theta}$ there is an $\varepsilon > 0$ such that $\theta + \varepsilon < 1$ and $r'(\theta + \varepsilon) < 1$. Then an estimate analogous to (2.7) and [26, Thm. 6.8(d) and Thm. 6.13(d)] yield

$$\begin{aligned}&\int_0^{t_1} \big\|\big(e^{-(t_2-t_1)A} - I\big) e^{-(t_1-s)A} x(s)\big\|_Y ds \\ &\quad\leq c \int_0^{t_1} \big\|A^\theta \big(e^{-(t_2-t_1)A} - I\big) e^{-(t_1-s)A} x(s)\big\|_X ds \\ &\quad\leq c \int_0^{t_1} \big\|\big(e^{-(t_2-t_1)A} - I\big) A^\theta\, e^{-(t_1-s)A} x(s)\big\|_X ds \\ &\quad\leq c(t_2 - t_1)^\varepsilon \int_0^{t_1} \big\|\underbrace{A^\varepsilon A^\theta}_{=A^{\theta+\varepsilon}}\, e^{-(t_1-s)A} x(s)\big\|_X ds \\ &\quad\leq c(t_2 - t_1)^\varepsilon \int_0^{t_1} (t_1 - s)^{-\theta-\varepsilon} \|x(s)\|_X\, ds \\ &\quad\leq c(t_2 - t_1)^\varepsilon \Big(\int_0^{t_1} (t_1 - s)^{-r'(\theta+\varepsilon)}\, ds\Big)^{\frac{1}{r'}} \|x\|_{L^r(]0,T[;X)} \to 0 \quad \text{for } t_1 \to t_2.\end{aligned}$$

In view of (A.2) this gives the desired continuity.

(iii) Local Lipschitz continuity in $C([0,T];Y)$:
Let $t \in [0,T]$ and $u_1, u_2 \in C^{0,1}([0,T];U)$ be arbitrary and denote the associated solutions of (2.8) by y_1, y_2. Moreover, set $R := \max_{j=1,2} \|u_j\|_{L^r(]0,T[;U)}$. In view of (A.1) we then have

$$\|y_i(t)\|_Y \leq C(1 + R) =: M_R, \quad i = 1, 2.$$

Completely analogously to (ii), it follows from the integral equation that

$$\begin{aligned}\|y_1(t) - y_2(t)\|_Y \leq \int_0^T \|e^{-sA}\|_{\mathcal{L}(X,Y)}^{r'}\, ds^{1/r'}\, \|B\|_{\mathcal{L}(U,X)}\, \|u_1 - u_2\|_{L^r(]0,T[;U)} \\ + L(M_R) \int_0^t \|e^{-(t-s)A}\|_{\mathcal{L}(X,Y))} \|y_1(s) - y_2(s)\|_Y\, ds,\end{aligned} \tag{A.4}$$

where we used the local Lipschitz continuity of f according to Assumption 2.1.4. Using again Lemma 2.4 and Gronwall's lemma yields

$$\|y_1(t) - y_2(t)\|_Y \leq \mathcal{L}(R)\, \|u_1 - u_2\|_{L^r(]0,T[;U)} \tag{A.5}$$

with $\mathcal{L}(R) := c \exp(L(M_R) \int_0^T \|e^{-tA}\|_{\mathcal{L}(X,Y))} dt)$.

(iv) Existence for $u \in L^r(]0,T[;U)$:

To finish the proof we now turn to non-smooth right hand sides. So let $u \in L^r(]0,T[;U)$ be arbitrary. Then, from Lemma B.1 we know that, there is a sequence $\{u_n\} \subset C^{0,1}([0,T];U)$ such that $u_n \to u$ in $L^r(]0,T[;U)$. Thus the sequence is bounded in $L^r(]0,T[;U)$ and hence the local Lipschitz continuity from (iii) yields that the associated sequence of states $\{y_n\}$ is Cauchy in $C([0,T];Y)$. As this space is complete, it therefore converges to an element $y \in C([0,T];Y)$. Due to Lemma 2.4 and the local Lipschitz continuity of f, this allows to pass to the limit in the integral equation (2.9) such that the limit $y \in C([0,T];Y)$ is indeed the mild solution associated with $u \in L^r(]0,T[;U)$. Moreover, it is easily seen that the arguments leading to (A.5) also apply to right hand sides in $L^r(]0,T[;U)$. This finally ensures that the mild solution associated with $u \in L^r(]0,T[;U)$ is unique. □

Appendix B. Density in Bochner spaces.

Lemma B.1. *Let $r \in [1,\infty)$. Let U, X be normed spaces such that $U \overset{d}{\hookrightarrow} X$. Then $C_c^\infty([0,T];U) \overset{d}{\hookrightarrow} L^r(]0,T[;X)$.*

Proof. We show this by means of [21, Thm. 9.5]. For this purpose let $f \in L^{r'}(]0,T[;X^*)$ with

$$\langle f, v\rangle_{L^r(]0,T[;X)} = 0 \quad \forall\, v \in C_c^\infty([0,T];U) \tag{B.1}$$

be arbitrary. Now, test the above equation with $v = \varphi u$ with $\varphi \in C_c^\infty([0,T])$ and $u \in U$. Then we arrive at

$$\int_0^T \varphi(t)\langle f(t), u\rangle_X \, dt = 0 \quad \forall\, \varphi \in C_c^\infty([0,T]),\ u \in U,$$

and by means of Fundamental Lemma we have that for every $u \in U$ it holds

$$\langle f(t), u\rangle_X = 0 \qquad \text{f.a.a. } t \in]0,T[. \tag{B.2}$$

Since $U \overset{d}{\hookrightarrow} X$, (B.2) yields that for any $x \in X$ it holds

$$\langle f(t), x\rangle_X = 0 \qquad \text{f.a.a. } t \in]0,T[,$$

and therefore $f = 0$. Since this was followed from (B.1), [21, Thm. 9.5] gives the assertion. □

Appendix C. Differentiability of Powers of Norms.

Lemma C.1. *Assume that U is reflexive, separable, and equipped with a norm such that U and U^* are locally uniformly convex. Let $r \in (1,\infty)$ and $\bar{u} \in L^r(]0,T[;U)$. Then the mapping*

$$F : L^r(]0,T[;U) \ni u \mapsto \frac{1}{r}\,\|u - \bar{u}\|^r_{L^r(]0,T[;U)} \in \mathbb{R}$$

is continuously Fréchet-differentiable. Moreover $F'(\bar{u}) = 0$.

Proof. Let $u \in L^r(]0,T[;U)$ be arbitrary. We distinguish between two cases:

Case (i): $u = \bar{u}$
Straight-forward computation yields

$$\frac{|F(\bar{u}+h) - F(\bar{u})|}{\|h\|_{L^r(]0,T[;U)}} = \frac{1}{r}\,\|h\|^{r-1}_{L^r(]0,T[;U)} \to 0 \quad \text{as } \|h\|_{L^r(]0,T[;U)} \to 0,$$

hence the Fréchet-differentiability with

$$F'(\bar{u}) = 0. \tag{C.1}$$

Case (ii): $u \neq \bar{u}$
Note that $F = g \circ w$, with $g : \mathbb{R} \to \mathbb{R}$, $g(x) := \frac{1}{r}\,x^r$, and $w : L^r(]0,T[;U) \to \mathbb{R}$, $w(u) := \|u - \bar{u}\|_{L^r(]0,T[;U)}$. Clearly, g is continuously Fréchet differentiable with

$$g'(x) = x^{r-1}. \tag{C.2}$$

Since U is reflexive and separable and $r \in]1,\infty[$, one has that $L^r(]0,T[;U)$ is reflexive with $(L^r(]0,T[;U))^* = L^{r'}(]0,T[;U^*)$, cf. [11, Theorem IV.1.14]. By [30, Theorem 2], the locally uniform convexity of U^* ensures that $L^{r'}(]0,T[;U^*)$ is locally uniformly convex. Therefore, one can apply [29, Prop. 4.7.10, 3.4.2] to see that w is continuously Fréchet-differentiable at u. Moreover, [29, Prop. 4.7.1] gives

$$\|w'(u)\|_{L^{r'}(]0,T[;U^*)} = 1 \quad \forall\, u \neq \bar{u}. \tag{C.3}$$

Therefore, the chain rule implies that $F = g \circ w$ is continuously Fréchet-differentiable at $u \neq \bar{u}$, and in view of (C.2) it holds

$$F'(u)h = \|u - \bar{u}\|^{r-1}_{L^r(]0,T[;U)} w'(u)h \quad \forall\, h \in L^r(]0,T[;U). \tag{C.4}$$

It now remains to show that F is continuously Fréchet-differentiable at $\bar{u}$. Let $u_n \to \bar{u}$ in $L^r(]0,T[;U)$. Consider just those u_n with $u_n \neq \bar{u}$. Then in view of (C.1) and (C.4)

$$\begin{aligned}\|F'(u_n) - F'(\bar{u})\|_{L^{r'}(]0,T[;U^*)} &\leq \|u_n - \bar{u}\|^{r-1}_{L^r(]0,T[;U)} \|w'(u_n)\|_{L^{r'}(]0,T[;U^*)} \\ &= \|u_n - \bar{u}\|^{r-1}_{L^r(]0,T[;U)} \to 0.\end{aligned}$$

For the last equality we used (C.3). ☐

Appendix D. Direction Differentiability of the max-Operator.

Lemma D.1. *The* max*-operator, defined at the beginning of Section 6, is directionally differentiable from $L^2(\Omega)$ to $L^2(\Omega)$.*

Proof. Let $y, h \in L^2(\Omega)$ be arbitrary. For simplicity, we denote the Nemytskii-operator associated with (6.4) by the same symbol. Note that $\max'(y;h)(x) = \max'(y(x);h(x))$ a.e. in Ω. Since $\mathbb{R} \ni x \mapsto \max\{x,0\} \in \mathbb{R}$ is directionally differentiable, see (6.4), we obtain the following pointwise convergence

$$\Big|\frac{\max(y(x)+\tau h(x)) - \max(y(x))}{\tau} - \max{}'(y(x);h(x))\Big| \to 0 \quad \text{f.a.a. } x \in \Omega. \tag{D.1}$$

Since $|\max'(y;h)| \leq |h|$, cf. (6.4), the global Lipschitz continuity of $\mathbb{R} \ni x \mapsto \max\{x,0\} \in \mathbb{R}$ with constant 1 implies that

$$\Big|\frac{\max(y(x)+\tau h(x)) - \max(y(x))}{\tau} - \max{}'(y(x);h(x))\Big| \leq 2\,|h(x)| \quad \text{f.a.a. } x \in \Omega. \tag{D.2}$$

Now, in view of (D.1) and (D.2), Lebesgue's dominated convergence theorem yields the result. ☐

REFERENCES

[1] H. Amann. *Linear and quasilinear parabolic problems.* Birkhäuser, Basel, 1995.

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

[3] M. Bergounioux and H. Zidani. A fully discrete approximation for control problems governed by parabolic variational inequalities. *SIAM J. Numer. Anal.*, 39(6):2014–2033, 2002.

[4] J.F. Bonnans and A. Shapiro. *Pertubation Analysis of Optimization Problems.* Springer, New York, 2000.

[5] E. Casas and F. Tröltzsch. Error estimates for the finite-element approximation of a semilinear elliptic control problem. *Control and Cybernetics*, 31:695–712, 2002.

[6] F.H. Clarke. The maximum principle under minimal hypotheses. *SIAM J. Control. Optim.*, 14(6):1078–1091, 1976.

[7] F.H. Clarke. *Optimization and Nonsmooth Analysis*, volume 5 of *Classics in Applied Mathematics.* SIAM, Philadelphia, 1990.

[8] G. Dore. L^p-regularity for abstract differential equations. In *Functional Analysis and Related Topics*, volume 1540 of *Lect. Notes Math.*, pages 25–38. Springer, Berlin, 1993. Proc. of the international conference in memory of K. Yosida, Kyoto University, Japan, 1991.

[9] M.H. Farshbaf-Shaker. A penalty approach to optimal control of Allen-Cahn variational inequalities: MPEC-view. *Numerical Functional Analysis and Optimization*, 33(11):1321–1349, 2012.

[10] A. Friedman. Optimal control for parabolic variational inequalities. *SIAM Journal on Control and Optimization*, 25(2):482–497, 1987.

[11] H. Gajewski, K. Gröger, and K. Zacharias. *Nichtlineare Operatorgleichungen und Operatordifferentialgleichungen.* Akademie-Verlag, Berlin, 1974.

[12] J. A. Griepentrog, K. Gröger, H.-C. Kaiser, and J. Rehberg. Interpolation for function spaces related to mixed boundary value problems. *Mathematische Nachrichten*, 241:110–120, 2002.

[13] J.A. Griepentrog, H.-C. Kaiser, and J. Rehberg. Resolvent and heat kernel properties for second order elliptic differential operators with general boundary conditions on L^p. *Advances in Mathematical Sciences and Applications*, 11(1):87–112, 2001.

[14] K. Gröger. $W^{1,p}$–estimates of solutions to evolution equations corresponding to nonsmooth second order elliptic differential operators. *Nonlinear Analysis*, 18:569—577, 1992.

[15] Z.X. He. State constrained control problems governed by variational inequalities. *SIAM Journal on Control and Optimization*, 25:1119–1144, 1987.

[16] R. Herzog, C. Meyer, and A. Schäfer. Existence of solutions of a (nonsmooth) thermoviscoplastic model and associated optimal control problems. Ergebnisberichte des Instituts für Angewandte Mathematik 521, TU Dortmund, Faculty for Mathematics, 2015.

[17] R. Herzog, C. Meyer, and G. Wachsmuth. B- and strong stationarity for optimal control of static plasticity with hardening. *SIAM J. Optim.*, 23(1):321–352, 2013.

[18] M. Hintermüller and I. Kopacka. Mathematical programs with complementarity constraints in function space: C- and strong stationarity and a path-following algorithm. *SIAM Journal on Optimization*, 20(2):868–902, 2009.

[19] M. Hintermüller and D. Wegner. Optimal control of a semi-discrete cahn-hilliard-navier-stokes system. *SIAM J. Control and Optimization*, 52(1):747–772, 2014.

[20] K. Ito and K. Kunisch. Optimal control of parabolic variational inequalities. *Journal de Mathématiques Pures et Appliqués*, 93(4):329–360, 2010.

[21] W. Kaballo. *Grundkurs Funktionalanalysis.* Spektrum Akademischer Verlag, Heidelberg, 2011.

[22] P. D. Loewen and R. T. Rockafellar. The adjoint arc in nonsmooth optimization. *Transactions of the American Mathematical Society*, 325(1):39–72, 1991.

[23] P. D. Loewen and R. T. Rockafellar. The adjoint arc in nonsmooth optimization. *Transactions of the American Mathematical Society*, 325(1):39–72, 1991.

[24] F. Mignot and J.-P. Puel. Contrôle optimal d'un système gouverné par une inéquation variationelle parabolique. *Comptes Rendus de l'Académie des Sciences Paris*, 298(12):277–280, 1984.

[25] F. Mignot and J.-P. Puel. Optimal control in some variational inequalities. *SIAM Journal on Control and Optimization*, 22(3):466–476, 1984.

[26] A. Pazy. *Semigroups of Linear Operators and Application to Partial Differential Equations.* Springer, New York, 1983.

[27] H. Scheel and S. Scholtes. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, 25(1):1–22, 2000.

[28] A. Schiela and D. Wachsmuth. Convergence analysis of smoothing methods for optimal control

of stationary variational inequalities with control constraints. *ESAIM: M2AN*, 47:771–787, 2013.

[29] W. Schirotzek. *Nonsmooth Analysis*. Springer, Berlin, 2007.

[30] A. Smith, M. and B. Turett. Rotundity in Lebesgue-Bochner function spaces. *Transactions of the American Mathematical Society*, 237(1):105–118, 1980.

[31] H. Triebel. *Interpolation Theory, Function Spaces, Differential Operators*. North-Holland, Amsterdam, 1978.

[32] F. Tröltzsch. *Optimal Control of Partial Differential Equations*, volume 112 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, 2010. Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels.