Deutsche Forschungsgemeinschaft

Priority Programme 1962

Strong Stationarity for Optimal Control of a Non-smooth Coupled System: Application to a Viscous Evolutionary VI Coupled with an Elliptic PDE

Livia Betz

SPP 1962

Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization

Preprint Number SPP1962-083

received on October 11, 2018

Edited by
SPP1962 at Weierstrass Institute for Applied Analysis and Stochastics (WIAS)
Leibniz Institute in the Forschungsverbund Berlin e.V.
Mohrenstraße 39, 10117 Berlin, Germany
E-Mail: spp1962@wias-berlin.de

World Wide Web: http://spp1962.wias-berlin.de/

STRONG STATIONARITY FOR OPTIMAL CONTROL OF A NON-SMOOTH COUPLED SYSTEM: APPLICATION TO A VISCOUS EVOLUTIONARY VI COUPLED WITH AN ELLIPTIC PDE

LIVIA BETZ*

Abstract. This paper is mainly concerned with an optimal control problem governed by a non-smooth coupled system of equations. The non-smooth nonlinearity is Lipschitz-continuous and directionally differentiable, but not Gâteaux-differentiable. We derive a strong stationary optimality system, i.e., an optimality system which is equivalent to the purely primal optimality condition saying that the directional derivative of the reduced objective in feasible directions is nonnegative. On another note, evolutionary VIs (EVIs) with viscosity can be formulated as non-smooth ODEs in Hilbert space. The non-smooth non-linearity appearing in the ODE turns out to be the solution operator of an elliptic VI, for which we can give an explicit formula. As a byproduct, it allows us to characterize the directional differentiability of elliptic VIs of the second kind. All these general results are in the end applied to prove strong stationarity for optimal control of a viscous elastoplastic problem.

Key words. optimal control of coupled systems, nonsmooth optimization, strong stationarity, evolutionary VIs with viscosity, optimal control of PDEs, quasistatic visco-elasto-plasticity

AMS subject classifications. 34K35, 49J21, 49J27

1. Introduction. In this paper, we establish strong stationary optimality conditions for the following optimal control problem:

$$\begin{aligned} \min_{\ell \in L^2(0,T;V)} \quad & J(y,u,\ell) \\ \text{s.t.} \quad & \begin{cases} \dot y(t) = f(\Phi(y(t),u(t))) \quad \text{a.e. in } (0,T), \quad y(0)=0, \\ \Psi(y(t),u(t)) = \ell(t) \quad \text{a.e. in } (0,T), \end{cases} \end{aligned} \tag{$*$}$$

where J is a smooth objective, f is a non-smooth mapping, while Φ and Ψ are smooth non-linearities. The precise assumptions on the data are stated in Assumption 2.1 below. The essential feature of the problem under consideration is that the non-linearity f is not necessarily Gâteaux-differentiable, so that the standard methods for the derivation of qualified optimality conditions are not applicable here. In view of our goal to establish strong stationarity, the main novelty in this paper is the coupled non-smooth structure in ($*$). This gives rise to additional challenges. We deal with two states which depend nonlinearly on each other. Moreover, the non-smooth mapping does not act directly on either one of the states, but on a non-linear coupling involving them. In this context, an interesting application is the optimal control of viscous EVIs coupled with elliptic PDEs, since non-smooth ODEs (in infinite-dimensional spaces) as the one in ($*$) describe the evolution of viscous processes. In the end, we obtain a strong stationary optimality system for this type of problem.

Deriving necessary optimality conditions is a challenging issue even in finite dimensions, where a special attention is given to MPCCs (mathematical programs with complementarity constraints). In [31] a detailed overview of various optimality conditions of different strength was introduced, see also [21] for the infinite dimensional

*Faculty of Mathematics, University of Duisburg-Essen, Germany (former last name: Susu).

case. The most rigorous stationarity concept is strong stationarity. Roughly speaking, the strong stationarity conditions involve an optimality system, which is equivalent to the purely primal conditions saying that the directional derivative of the reduced objective in feasible directions is nonnegative (which is referred to as B stationarity).

While there are plenty of contributions in the field of optimal control of smooth problems, see e.g. [37] and the references therein, less papers are dealing with non-smooth problems. Most of these papers recur to regularization or relaxation techniques to smooth the problem, see e.g. [1, 2, 11, 12, 16, 18, 22, 24] and the references therein. The optimality systems derived in this way are of intermediate strength and are not expected to be of strong stationary type, since one always loses information when passing to the limit in the regularization scheme. Thus, proving strong stationarity for optimal control of non-smooth problems requires direct approaches, which employ the limited differentiability properties of the control-to-state map. In this context, there are even less contributions. Based on the pioneering work [28] (strong stationarity for optimal control of elliptic VIs of obstacle type), most of them focus on elliptic VIs [9, 19, 38]. Recently, strong stationarity for optimal control of parabolic VIs of the first kind was proven in [5]. Regarding strong stationarity for optimal control of non-smooth PDEs, the literature is very scarce and the only papers known to the author addressing this issue so far are [26] (parabolic PDE) and [6] (elliptic PDE). We point out that, in contrast to our problem, all the above mentioned contributions which investigate strong stationarity deal with only one state, excepting [19] (however, therein, one of the two states can be computed independently on the other); in particular, there is no coupling between the non-smooth evolution and an additional equation.

Let us give an overview of the main results in this paper. In Section 2, we show strong stationarity for the optimal control of a *coupled* system of equations involving a non-smooth non-linearity. This is one of the main novelties of this paper. The result can be extended to more complex systems that involve more equations and/or more time derivatives. It is based on the idea from [26], which is to employ a 'surjectivity' trick. As it turns out, it all comes down to the following aspect: the directions at which the non-smooth mapping f is derived - in the 'linearized' state equation associated to the local optimum - must 'cover' an entire space (see Lemma 2.8 and Remark 2.10).

In Section 3, we show that viscous EVIs are non-smooth ODEs in Hilbert space. It is indeed known (in particular, in the context of Lebesgue spaces [20, 35]) that such a formulation exists, but an explicit description which applies to general Hilbert spaces was not found in the literature. As a byproduct, a formula for the solution operator of a general elliptic VI is given (see Lemma 3.3 below). This is nothing else than the non-smooth non-linearity in the ODE. It allows us to establish the following: the solution operator of the classical elliptic VI of the second kind is directionally differentiable if and only if the projection operator onto the convex subdifferential at 0 of the non-smooth functional in the VI is directionally differentiable. Appendix A gives more details and some concrete examples.

Section 4 focuses on strong stationarity for optimal control of quasistatic visco-elasto-plasticity. Here we are concerned with the application of the above mentioned results. We first employ the findings in Section 3 to show that this concrete application is a non-smooth coupled PDE system of the type $(*)$, see (4.5)-(4.6) below. It consists of a non-smooth ODE in Hilbert space (viscoplastic flow rule) and an elliptic PDE (linear elasticity). Then, we make use of the main result in Section 2 to derive strong stationary optimality conditions for an optimal control problem governed by the viscous elastoplastic system. A challenge in this section will be the complexity of the problem arising from the matrix-valued function spaces.

Notation. Throughout the paper, $T > 0$ is a fixed final time. If X and Y are linear normed spaces, then the space of linear and bounded operators from X to Y is denoted by $\mathcal{L}(X, Y)$, and $X \overset{d}{\hookrightarrow} Y$ means that X is dense in Y. The dual space of X will be denoted by X^*. For the dual pairing between X and X^* we write $\langle ., . \rangle_X$. The open ball in X around $x \in X$ with radius $\alpha > 0$ is denoted by $B_X(x, \alpha)$. The following abbreviations will be used throughout the paper:

$$
\begin{aligned}
H_0^1(0, T; X) &:= \{z \in H^1(0, T; X) : z(0) = 0\}, \\
H_T^1(0, T; X) &:= \{z \in H^1(0, T; X) : z(T) = 0\},
\end{aligned}
$$

where X is a Banach space. For the polar cone of a set $M \subset X$ we use the notation $M^\circ := \{x^* \in X^* : \langle x^*, x \rangle_X \leq 0 \quad \forall x \in M\}$. By χ_M we denote the characteristic function associated to the set M. Derivatives w.r.t. time (weak derivatives of vector-valued functions) are frequently denoted by a dot. The symbol ∂ stands for the convex subdifferential, see e.g. [30]. With a little abuse of notation, the Nemystkii-operators associated with the mappings considered in this paper will be denoted by the same symbol, even when considered with different domains and ranges. From the context, it will be always clear which one is meant.

2. Strong stationarity for optimal control of non-smooth coupled systems. This section is devoted to one of the main results of the paper, i.e., the derivation of a strong stationary optimality system for the optimal control of $(*)$. After introducing the functional analytical setting, we start our analysis with a purely primal optimality condition (B-stationarity). Then, we exploit and extend the trick from [26] to derive our strong stationary optimality conditions. As it turns out, generally speaking, the key role is played by the density (in certain spaces) of the set of directions at which the non-smooth mapping f is derived in the 'linearized' state equation, see (2.5) and Lemma 2.8 below. To ensure that this is satisfied, we formulate an assumption on the optimizer ('constraint qualification') which is based on the particular structure of the state equation under consideration. Let us point out that this assumption is always satisfied by the optimal control of quasistatic visco-elastoplasticity, see Section 4 below.

Our optimal control problem reads as follows:

$$
\left.
\begin{aligned}
\min_{\ell \in L^2(0,T;V)} \quad & J(y, u, \ell) \\
\text{s.t.} \quad & \dot{y}(t) = f(\Phi(y(t), u(t))) \quad \text{a.e. in } (0, T), \quad y(0) = 0, \\
& \Psi(y(t), u(t)) = \ell(t) \quad \text{a.e. in } (0, T), \\
& y \in H^1(0, T; Y), \quad u \in L^2(0, T; U).
\end{aligned}
\right\} \tag{P}
$$

Assumption 2.1. *For the quantities in* (P) *we require the following:*

1. *V, Y, and U are real reflexive Banach spaces, such that $V \overset{d}{\hookrightarrow} U^*$.*
2. *The non-smooth function $f : Y^* \to Y$ is assumed to be Lipschitz-continuous and directionally differentiable, i.e.,*

$$
\Big\| \frac{f(x + \tau h) - f(x)}{\tau} - f'(x; h) \Big\|_Y \xrightarrow{\tau \searrow 0} 0 \quad \forall\, x, h \in Y^*. \tag{2.1}
$$

3. *The mappings $\Phi : Y \times U \to Y^*$ and $\Psi : Y \times U \to U^*$ are Gâteaux-differentiable operators. Moreover, they are Lipschitz-continuous, i.e., there exists $L > 0$ such that*

$$\|\Phi'(y,u)\|_{\mathcal{L}(Y\times U;Y^*)} \le L, \quad \|\Psi'(y,u)\|_{\mathcal{L}(Y\times U;U^*)} \le L \quad \forall\,(y,u) \in Y \times U. \tag{2.2}$$

4. *The objective $J : L^2(0,T;Y) \times L^2(0,T;U) \times L^2(0,T;V) \to \mathbb{R}$ is Fréchet-differentiable.*

Let us observe that the Nemytskii operator associated to the function f is directionally differentiable from $L^2(0,T;Y^*)$ to $L^2(0,T;Y)$ with

$$f'(x;h) = f'(x(\cdot);h(\cdot)) \in L^2(0,T;Y) \tag{2.3}$$

for any $x, h \in L^2(0,T;Y^*)$. This is a result of Assumption 2.1.2 combined with Lebesgue's dominated convergence theorem. In view of the latter and Assumption 2.1.3, we also deduce that the Nemytskii operators

$$\begin{aligned}
&\Phi : L^2(0,T;Y) \times L^2(0,T;U) \to L^2(0,T;Y^*),\\
&\Psi : L^2(0,T;Y) \times L^2(0,T;U) \to L^2(0,T;U^*)
\end{aligned}$$

are Gâteaux-differentiable, with

$$\begin{aligned}
\partial_y\Phi(y,u)(\delta y) &= \partial_y\Phi(y(\cdot),u(\cdot))(\delta y(\cdot)), \quad & \partial_u\Phi(y,u)(\delta u) &= \partial_u\Phi(y(\cdot),u(\cdot))(\delta u(\cdot)),\\
\partial_y\Psi(y,u)(\delta y) &= \partial_y\Psi(y(\cdot),u(\cdot))(\delta y(\cdot)), \quad & \partial_u\Psi(y,u)(\delta u) &= \partial_u\Psi(y(\cdot),u(\cdot))(\delta u(\cdot))
\end{aligned}$$

for all $(y,u), (\delta y, \delta u) \in L^2(0,T;Y \times U)$. In the proof of Theorem 2.9 below, it will be useful to keep the following in mind:

Lemma 2.2. *The adjoint operators of the partial derivatives of Φ and Ψ satisfy*

$$\begin{aligned}
\partial_y\Phi(y,u)^* &: L^2(0,T;Y) \to L^2(0,T;Y^*), \quad & \partial_y\Phi(y,u)^*(\eta)(t) &= \partial_y\Phi(y(t),u(t))^*(\eta(t)),\\
\partial_u\Phi(y,u)^* &: L^2(0,T;Y) \to L^2(0,T;U^*), \quad & \partial_u\Phi(y,u)^*(\eta)(t) &= \partial_u\Phi(y(t),u(t))^*(\eta(t)),\\
\partial_y\Psi(y,u)^* &: L^2(0,T;U) \to L^2(0,T;Y^*), \quad & \partial_y\Psi(y,u)^*(v)(t) &= \partial_y\Psi(y(t),u(t))^*(v(t)),\\
\partial_u\Psi(y,u)^* &: L^2(0,T;U) \to L^2(0,T;U^*), \quad & \partial_u\Psi(y,u)^*(v)(t) &= \partial_u\Psi(y(t),u(t))^*(v(t))
\end{aligned}$$

f.a.a. $t \in (0,T)$ and for all $(y,u) \in L^2(0,T;Y \times U)$.

Our focus in this section is to derive strong stationary optimality conditions for the optimal control problem (P). To keep the depiction concise, we do not discuss the unique solvability of the state equation or the differentiability properties of the resulting solution operator. These issues will be addressed in detail for the application considered later on, see Section 4 below.

Here, the properties we need from the control-to-state map in order to prove the main result (Theorem 2.9) are just assumed to be true. We collect them in the following

Assumption 2.3 (Control-to-state map and directional differentiability).

1. *Throughout this section, we assume that for every $\ell \in L^2(0,T;V)$, the state equation*

$$\left.\begin{aligned}
&\dot y(t) = f(\Phi(y(t),u(t))) \quad \text{a.e. in } (0,T),\ y(0) = 0,\\
&\Psi(y(t),u(t)) = \ell(t) \quad \text{a.e. in } (0,T)
\end{aligned}\right\} \tag{2.4}$$

admits a unique solution $(y,u) \in H_0^1(0,T;Y) \times L^2(0,T;U)$ *and denote the associated solution operator by*

$$S : L^2(0,T;V) \ni \ell \mapsto (y,u) \in H_0^1(0,T;Y) \times L^2(0,T;U).$$

2. *For any local optimum* $\bar{\ell}$ *of* (P), $S : L^2(0,T;V) \to L^2(0,T;Y) \times L^2(0,T;U)$ *has a directional derivative* $S'(\bar{\ell};\delta\ell)$ *at* $\bar{\ell} \in L^2(0,T;V)$ *in any direction* $\delta\ell \in L^2(0,T;V)$*, i.e.,*

$$\Big\| \frac{S(\bar{\ell} + \tau\,\delta\ell) - S(\bar{\ell})}{\tau} - S'(\bar{\ell};\delta\ell) \Big\|_{L^2(0,T;Y)\times L^2(0,T;U)} \xrightarrow{\tau \searrow 0} 0 \quad \forall\, \delta\ell \in L^2(0,T;V).$$

Moreover, we suppose that for any $\delta\ell \in L^2(0,T;V)$*, the pair* $(\delta y, \delta u) := S'(\bar{\ell};\delta\ell) \in H_0^1(0,T;Y) \times L^2(0,T;U)$ *is the unique solution of*

$$\dot{\delta y}(t) = f'\big(\Phi(\bar{y}(t),\bar{u}(t)); \Phi'(\bar{y}(t),\bar{u}(t))(\delta y(t),\delta u(t))\big) \quad \text{a.e. in } (0,T),\ \ \delta y(0) = 0, \tag{2.5a}$$

$$\Psi'(\bar{y}(t),\bar{u}(t))(\delta y(t),\delta u(t)) = \delta\ell(t) \quad \text{a.e. in } (0,T), \tag{2.5b}$$

where we abbreviate $(\bar{y},\bar{u}) := S(\bar{\ell})$.

3. *For any local optimum* $\bar{\ell}$ *of* (P)*, there exists a constant* $K > 0$ *so that*

$$\|S'(\bar{\ell};\delta\ell)\|_{L^2(0,T;Y\times U)} \le K\,\|\delta\ell\|_{L^2(0,T;U^*)} \quad \forall\, \delta\ell \in L^2(0,T;V). \tag{2.6}$$

If the equation (2.5) *admits a solution* $(\hat{\delta y}, \hat{\delta u}) \in H_0^1(0,T;Y) \times L^2(0,T;U)$ *for a right hand-side* $\hat{\delta\ell} \in L^2(0,T;U^*)$*, then* $S'(\bar{\ell};\delta\ell_n) \to (\hat{\delta y}, \hat{\delta u})$ *in* $L^2(0,T;Y \times U)$*, provided that* $\{\delta\ell_n\}_n \subset L^2(0,T;V)$ *is a sequence with* $\delta\ell_n \to \hat{\delta\ell}$ *in* $L^2(0,T;U^*)$.

Remark 2.4. *Assumption 2.1.3 is for instance guaranteed if there exist constants* $\alpha > 0$ *and* $K > 0$ *so that for all* $\ell_1, \ell_2 \in B_{L^2(0,T;V)}(\bar{\ell},\alpha)$ *it holds*

$$\|S(\ell_1) - S(\ell_2)\|_{L^2(0,T;Y\times U)} \le K\,\|\ell_1 - \ell_2\|_{L^2(0,T;U^*)} \tag{2.7}$$

and if the solution $(\hat{\delta y}, \hat{\delta u})$ *is unique. To see this, we observe that* (2.7) *implies*

$$\|S'(\bar{\ell};\delta\ell^1) - S'(\bar{\ell};\delta\ell^2)\|_{L^2(0,T;Y\times U)} \le K\,\|\delta\ell^1 - \delta\ell^2\|_{L^2(0,T;U^*)} \quad \forall\, \delta\ell^1, \delta\ell^2 \in L^2(0,T;V), \tag{2.8}$$

by the definition of the directional derivative. Hence, (2.6) *is true. Moreover, in view of* (2.8), $\{S'(\bar{\ell};\delta\ell_n)\}_n \subset L^2(0,T;Y \times U)$ *is a Cauchy sequence, and thus, it converges. Its limit, say* z*, solves* (2.5) *w.r.h.s.* $\hat{\delta\ell}$*, and due to its unique solvability, we have* $z = (\hat{\delta y}, \hat{\delta u})$.

Note that the estimate (2.6) *can be dropped, if e.g.* $\Psi'(\bar{y}(\cdot),\bar{u}(\cdot)) \in \mathcal{L}(Y \times U; V)$ *a.e. in* $(0,T)$ *or if* J *is partially Frechet-differentiable w.r.t.* ℓ *on* $L^2(0,T;U^*)$*, see the first part of the proof of Theorem 2.9 below.*

Clearly, the entire Assumption 2.1.3 is omited when $V = U^*$.

In the sequel, Assumptions 2.1 and 2.3 are tacitly assumed, without mentioning them every time.

Now, we turn our attention to proving our main result. We begin by stating the first order necessary optimality conditions in primal form.

LEMMA 2.5 (B-stationarity). *If $\bar{\ell} \in L^2(0,T;V)$ is locally optimal for* (P)*, then there holds*

$$\partial_{(y,u)} J(S(\bar{\ell}),\bar{\ell}) S'(\bar{\ell};\delta\ell) + \partial_\ell J(S(\bar{\ell}),\bar{\ell})\delta\ell \geq 0 \quad \forall\, \delta\ell \in L^2(0,T;V). \tag{2.9}$$

Proof. In view of Assumptions 2.1.4 and 2.3.2, we deduce from [19, Lemma 3.9] that the composite mapping $L^2(0,T;V) \ni \ell \mapsto J(S(\ell),\ell) \in \mathbb{R}$ is directionally differentiable at $\bar{\ell}$ in any direction $\delta\ell$ with directional derivative $\partial_{(y,u)} J(S(\bar{\ell}),\bar{\ell}) S'(\bar{\ell};\delta\ell) + \partial_\ell J(S(\bar{\ell}),\bar{\ell})\delta\ell$. The result then follows immediately from the local optimality of $\bar{\ell}$. □

ASSUMPTION 2.6. *For any local optimum $\bar{\ell}$ of* (P)*, we assume that $\partial_u \Phi(\bar{y},\bar{u}) : L^2(0,T;U) \to L^2(0,T;Y^*)$ is invertible, where $(\bar{y},\bar{u}) := S(\bar{\ell})$.*

REMARK 2.7. *(i) We point out that Assumption 2.6 is due to the presence of the additional variable u in the argument of the non-smooth mapping f. Roughly speaking, it is the price to pay for having two states on which the non-smoothness acts.*
(ii) The entire analysis in this section carries on if we consider some (non-linear) operator acting on the control, say $\mathcal{G} : V \to U^*$, so that the associated Nemytskii operator $\mathcal{G} : L^2(0,T;V) \to L^2(0,T;U^*)$ is Gâteaux-differentiable. Then, Assumption 2.3 has to be changed accordingly, and the 'constraint qualification' in Assumption 2.6 also contains the condition $\operatorname{Rg}\mathcal{G}'(\bar{\ell}) \overset{d}{\hookrightarrow} L^2(0,T;U^*)$. We emphasize that such a density assumption is to be expected, see [6, 26, 28]. In all these contributions, the operator acting on the control is linear at best (mostly, $\mathcal{G}$ is just an embedding). For simplicity reasons, we also stick to the case when $\mathcal{G}$ is the embedding operator $V \overset{d}{\hookrightarrow} U^*$, cf. Assumption 2.1.1.
(iii) Furthermore, let us point out that Assumption 2.6 is satisfied at any ℓ by the viscous elasto-plastic problem, see Section 4.

The next result is crucial for proving the strong stationarity result in Theorem 2.9:

LEMMA 2.8 (Density of the set of arguments of $f'(\Phi(\bar{y},\bar{u});\cdot)$). *Let $\bar{\ell} \in L^2(0,T;V)$ be a local optimum of* (P) *and $(\bar{y},\bar{u}) := S(\bar{\ell})$. Under Assumption 2.6, it holds*

$$\{\Phi'(\bar{y},\bar{u})(S'(\bar{\ell};\delta\ell)) : \delta\ell \in L^2(0,T;V)\} \overset{d}{\hookrightarrow} L^2(0,T;Y^*).$$

Proof. Let $\rho \in L^2(0,T;Y^*)$ be arbitrary, but fixed. Then, the mapping

$$[0,T] \ni t \mapsto \hat{y}(t) \in Y, \quad \hat{y}(t) := \int_0^t f'\big(\Phi(\bar{y}(s),\bar{u}(s));\rho(s)\big)\, ds \tag{2.10}$$

satisfies $\hat{y}(0) = 0$ and $\hat{y} \in H^1(0,T;Y)$. Note that the regularity of $\hat{y}$ is due to (2.3). Moreover, Assumption 2.6 allows us to define

$$\hat{u} := \partial_u \Phi(\bar{y},\bar{u})^{-1}(\rho - \partial_y \Phi(\bar{y},\bar{u})\hat{y}) \in L^2(0,T;U). \tag{2.11}$$

Now, we set

$$\hat{\ell} := \Psi'(\bar{y}(\cdot),\bar{u}(\cdot))(\hat{y}(\cdot),\hat{u}(\cdot)) \in L^2(0,T;U^*). \tag{2.12}$$

The regularity of $\hat{\ell}$ is a result of Assumption 2.1.3. By construction, the pair $(\hat{y},\hat{u}) \in H_0^1(0,T;Y) \times L^2(0,T;U)$ solves the system (2.5) with right-hand side $\hat{\ell} \in L^2(0,T;U^*)$. To see this, we observe that from (2.11) we have

$$\rho = \partial_u \Phi(\bar{y},\bar{u})\hat{u} + \partial_y \Phi(\bar{y},\bar{u})\hat{y}, \tag{2.13}$$

which inserted in (2.10) gives

$$\hat{y}(t) = \int_0^t f'\big(\Phi(\bar{y}(s), \bar{u}(s)); \Phi'(\bar{y}(s), \bar{u}(s))(\hat{y}(s), \hat{u}(s))\big)\, ds \quad \forall\, t \in [0, T].$$

This together with $\hat{y}(0) = 0$ and (2.12) yields

$$\frac{d}{dt}\hat{y}(t) = f'\big(\Phi(\bar{y}(t), \bar{u}(t)); \Phi'(\bar{y}(t), \bar{u}(t))(\hat{y}(t), \hat{u}(t))\big) \quad \text{a.e. in } (0, T),\ \hat{y}(0) = 0,$$
$$\Psi'(\bar{y}(t), \bar{u}(t))(\hat{y}(t), \hat{u}(t)) = \hat{\ell}(t) \quad \text{a.e. in } (0, T).$$

Thanks to $V \overset{d}{\hookrightarrow} U^*$, cf. Assumption 2.1.1, there exists a sequence $\{\delta\ell_n\}_n \subset L^2(0, T; V)$ with $\delta\ell_n \to \hat{\ell}$ in $L^2(0, T; U^*)$, see also [26, Lem. A.1]. By Assumption 2.3.3, we deduce that $S'(\bar{\ell}; \delta\ell_n) \to (\hat{y}, \hat{u})$ in $L^2(0, T; Y \times U)$. Since $\rho \in L^2(0, T; Y^*)$ was arbitrary, the desired assertion follows now from the continuity of $\Phi'(\bar{y}, \bar{u}) : L^2(0, T; Y \times U) \to L^2(0, T; Y^*)$ and (2.13). ∎

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

THEOREM 2.9 (Strong stationarity). *Suppose that Assumption 2.6 is satisfied. Let $\bar{\ell} \in L^2(0, T; V)$ be locally optimal for (P) with associated state $(\bar{y}, \bar{u}) := S(\bar{\ell})$. Then, there exist unique adjoint states*

$$\xi \in H^1_T(0, T; Y^*) \quad \text{and} \quad w \in L^2(0, T; U)$$

and a unique multiplier $\lambda \in L^2(0, T; Y)$ such that the following system is satisfied

$$-\dot{\xi} - \partial_y\Phi(\bar{y}, \bar{u})^*\lambda + \partial_y\Psi(\bar{y}, \bar{u})^* w = \partial_y J(\bar{y}, \bar{u}, \bar{\ell}) \ \text{ in } L^2(0, T; Y^*), \quad \xi(T) = 0, \tag{2.15a}$$

$$-\partial_u\Phi(\bar{y}, \bar{u})^*\lambda + \partial_u\Psi(\bar{y}, \bar{u})^* w = \partial_u J(\bar{y}, \bar{u}, \bar{\ell}) \ \text{ in } L^2(0, T; U^*), \tag{2.15b}$$

$$\langle \xi(t), f'(\Phi(\bar{y}(t), \bar{u}(t)); v)\rangle_Y \geq \langle \lambda(t), v\rangle_{Y^*} \quad \forall\, v \in Y^*,\ \text{a.e. in } (0, T), \tag{2.15c}$$

$$w + \partial_\ell J(\bar{y}, \bar{u}, \bar{\ell}) = 0 \ \text{ in } L^2(0, T; U). \tag{2.15d}$$

Proof. (1) We first prove that there exists a unique tuple (ξ, w, λ) which satisfies (2.15a), (2.15b) and (2.15d) and has the desired regularity. As a result of (2.9) and (2.6), we get the estimate

$$\begin{aligned} -\partial_\ell J(\bar{y}, \bar{u}, \bar{\ell})\delta\ell &\leq \underbrace{\|\partial_{(y,u)} J(\cdot)\|_{L^2(0,T;Y^*)\times L^2(0,T;U^*)}}_{=:c} \|S'(\bar{\ell}; \delta\ell)\|_{L^2(0,T;Y\times U)} \\ &\leq cK\, \|\delta\ell\|_{L^2(0,T;U^*)} \quad \forall\, \delta\ell \in L^2(0, T; V), \end{aligned}$$

whence $\partial_\ell J(\bar{y}, \bar{u}, \bar{\ell}) \in L^2(0, T; U)$ follows (by Hahn-Banach theorem or the density assumption $V \overset{d}{\hookrightarrow} U^*$). Then, we set $w := -\partial_\ell J(\bar{y}, \bar{u}, \bar{\ell}) \in L^2(0, T; U)$. Owing to Assumption 2.6, the operator $\partial_u\Phi(\bar{y}, \bar{u})^* : L^2(0, T; Y) \to L^2(0, T; U^*)$ is invertible, which allows us to define

$$\lambda := (\partial_u\Phi(\bar{y}, \bar{u})^*)^{-1}(\partial_u\Psi(\bar{y}, \bar{u})^* w - \partial_u J(\bar{y}, \bar{u}, \bar{\ell})) \in L^2(0, T; Y), \tag{2.16}$$

by Lemma 2.2 and Assumption 2.1.4. These also imply that the expression

$$\nu := \partial_y J(\bar{y}, \bar{u}, \bar{\ell}) + \partial_y\Phi(\bar{y}, \bar{u})^*\lambda - \partial_y\Psi(\bar{y}, \bar{u})^* w \tag{2.17}$$

belongs to $L^2(0,T;Y^*)$. Now, let us define

$$[0,T] \ni t \mapsto \xi(t) \in Y^*, \quad \xi(t) := \int_t^T \nu(s)\, ds, \tag{2.18}$$

such that $\xi(T) = 0$, $\xi \in H^1(0,T;Y^*)$, by the regularity of ν, and $-\dot{\xi}(t) = \nu(t)$ f.a.a. $t \in (0,T)$. We now observe that the tuple (ξ, w, λ) satisfies (2.15a), (2.15b), and (2.15d). Moreover, it has the desired regularity and is unique, by construction.

(2) It remains to show that the variational inequality (2.15c) is true. To this end, we extend the basic idea from [26, Proof of Thm. 5.3]. Let $\rho \in L^2(0,T;Y^*)$ be arbitrary, but fixed. According to Lemma 2.8, there exists $\{\delta\ell_n\} \subset L^2(0,T;V)$ such that

$$\underbrace{\Phi'(\bar{y},\bar{u})(\delta y_n, \delta u_n)}_{:=\rho_n} \to \rho \quad \text{in } L^2(0,T;Y^*) \quad \text{as } n \to \infty, \tag{2.19}$$

where we abbreviate $(\delta y_n, \delta u_n) := S'(\bar{\ell}; \delta\ell_n)$ and $\rho_n := \Phi'(\bar{y}(\cdot), \bar{u}(\cdot))(\delta y_n(\cdot), \delta u_n(\cdot))$ for all $n \in \mathbb{N}$. By relying on the B-stationarity, cf. (2.9), and by testing (2.15a), (2.15b), and (2.15d) with δy_n, δu_n, and $\delta\ell_n$, respectively, we obtain

$$\begin{aligned}
0 &\le \partial_y J(\bar{y},\bar{u},\bar{\ell})\delta y_n + \partial_u J(\bar{y},\bar{u},\bar{\ell})\delta u_n + \partial_\ell J(\bar{y},\bar{u},\bar{\ell})\delta\ell_n \\
&= -\int_0^T \langle \dot{\xi}(t), \delta y_n(t)\rangle_Y \, dt - \langle \partial_y \Phi(\bar{y},\bar{u})^*\lambda - \partial_y \Psi(\bar{y},\bar{u})^* w, \delta y_n\rangle_{L^2(0,T;Y)} \\
&\quad + \langle \partial_u \Psi(\bar{y},\bar{u})^* w - \partial_u \Phi(\bar{y},\bar{u})^*\lambda, \delta u_n\rangle_{L^2(0,T;U)} - \langle \delta\ell_n, w\rangle_{L^2(0,T;U)} \\
&= \int_0^T \langle \xi(t), \dot{\delta y}_n(t)\rangle_Y \, dt + \langle \partial_y \Psi(\bar{y},\bar{u})\delta y_n + \partial_u \Psi(\bar{y},\bar{u})\delta u_n - \delta\ell_n, w\rangle_{L^2(0,T;U)} \\
&\quad - \langle \partial_y \Phi(\bar{y},\bar{u})\delta y_n + \partial_u \Phi(\bar{y},\bar{u})\delta u_n, \lambda\rangle_{L^2(0,T;Y)} \\
&= \int_0^T \langle \xi(t), f'(\Phi(\bar{y}(t),\bar{u}(t)); \rho_n(t))\rangle_Y \, dt - \int_0^T \langle \lambda(t), \rho_n(t)\rangle_{Y^*} \, dt \quad \forall\, n \in \mathbb{N},
\end{aligned} \tag{2.20}$$

where the second identity follows from integration by parts, $\delta y_n(0) = 0$, and $\xi(T) = 0$, while the last identity is a result of (2.5a) tested with ξ, (2.5b) tested with w, and the above definition of ρ_n. Letting $n \to \infty$ in (2.20) leads to

$$0 \le \int_0^T \langle \xi(t), f'(\Phi(\bar{y}(t),\bar{u}(t)); \rho(t))\rangle_Y \, dt - \int_0^T \langle \lambda(t), \rho(t)\rangle_{Y^*} \, dt \quad \forall\, \rho \in L^2(0,T;Y^*), \tag{2.21}$$

in view of (2.19). Here we used the fact that $f'(\Phi(\bar{y},\bar{u}); \cdot) : L^2(0,T;Y^*) \to L^2(0,T;Y)$ is continuous, by the Lipschitz continuity of f, cf. Assumption 2.1.2, see also (2.3).

Now, consider $v \in Y^*$ and $\varphi \in C_0^\infty(0,T)$ with $\varphi \ge 0$ be arbitrary. By setting $\rho := \varphi v \in L^2(0,T;Y^*)$ in (2.21) and by employing the positive homogeneity of the directional derivative, we get

$$\int_0^T \langle \xi(t), f'(\Phi(\bar{y}(t),\bar{u}(t)); v)\rangle_Y \varphi(t) \, dt \ge \int_0^T \langle \lambda(t), v\rangle_{Y^*} \varphi(t) \, dt \quad \forall\, v \in Y^*,\ \varphi \in C_0^\infty(0,T).$$

The fundamental lemma of the calculus of variations then gives (2.15c). The proof is now complete. ∎

Remark 2.10 (Density of the set of arguments of $f'(\Phi(\bar{y},\bar{u}); \cdot)$).

(i) The proof of Theorem 2.9, see (2.20), shows that it is essential that the set of directions at which the non-smooth mapping f is derived - in the 'linearized' state equation associated to $\bar{\ell}$ - is dense in a (suitable) Bochner space (which is basically the assertion in Lemma 2.8). Let us point out that this aspect is also crucial when deriving strong stationarity for optimal control of more complex non-smooth coupled systems (which may involve more than two equations and/or more time derivatives acting on the states).
(ii) Note that Assumption 2.6 is due to the structure of the state equation under consideration. In a different, perhaps more complex setting, this 'constraint qualification' may sound completely different, but it should imply that the set of directions at which f is derived - in the 'linearized' state equation - 'covers' an entire space.
(iii) Finally, let us remark that the observations made here are consistent with the results in [26]. Therein, the direction at which one derives f - in the 'linearized' state equation - is the 'linearized' solution operator at $\bar{\ell}$, such that the counterpart of Lemma 2.8 is the density of the image of $S'(\bar{\ell};\cdot)$ in a suitable Bochner space, i.e., [26, Lem. 5.2]. In [26], there is no 'constraint qualification' in the sense of Assumption 2.6, since the authors deal with one state, see Remark 2.7.(i). However, the density assumption [26, Assump. 2.1.6] can be regarded as such, in view of Remark 2.7.(ii).

To see that (2.15) is indeed of strong stationary type, cf. Section 1, we prove the following:

Theorem 2.11 (Equivalence between B- and strong stationarity). *Assume that $\bar{\ell} \in L^2(0,T;V)$ together with its states $(\bar{y},\bar{u}) \in H_0^1(0,T;Y) \times L^2(0,T;U)$, some adjoint states $(\xi,w) \in H_T^1(0,T;Y^*) \times L^2(0,T;U)$, and a multiplier $\lambda \in L^2(0,T;Y)$ satisfy the optimality system (2.15a)–(2.15d). Then, it also satisfies the variational inequality (2.9). If Assumption 2.6 is satisfied, (2.9) is equivalent to (2.15a)–(2.15d).*

Proof. To show the first assertion, let $\delta\ell \in L^2(0,T;V)$ be arbitrary, but fixed and define $(\delta y,\delta u) := S'(\bar{\ell};\delta\ell)$. We proceed as in the proof of (2.20) to obtain

$$\begin{aligned}\partial_y J(\bar{y},\bar{u},\bar{\ell})\delta y &+ \partial_u J(\bar{y},\bar{u},\bar{\ell})\delta u + \partial_\ell J(\bar{y},\bar{u},\bar{\ell})\delta\ell \\ &= \int_0^T \langle \xi(t), f'\big(\Phi(\bar{y}(t),\bar{u}(t)); \Phi'(\bar{y}(t),\bar{u}(t))(\delta y(t),\delta u(t))\big)\rangle_Y \, dt \\ &\quad - \int_0^T \langle \lambda(t), \Phi'(\bar{y}(t),\bar{u}(t))(\delta y(t),\delta u(t))\rangle_{Y^*} \, dt.\end{aligned} \tag{2.22}$$

Note that one does not need local optimality for $\bar{\ell}$ or Assumption 2.6 to prove (2.22). The variational inequality (2.9) follows by testing (2.15c) with $v := \Phi'(\bar{y},\bar{u})(\delta y,\delta u)(t) \in Y^*$ f.a.a. $t \in (0,T)$ and by using the resulting inequalities on the right-hand side of (2.22). Moreover, if Assumption 2.6 is satisfied, then (2.9) implies (2.15a)–(2.15d), see the proof of Theorem 2.9. This shows the second assertion.□

Remark 2.12. *If f is Gâteaux-differentiable at $\Phi(\bar{y}(t),\bar{u}(t))$ a.e. in $(0,T)$, then (2.15c) is equivalent to*

$$\lambda(t) = \Big(f'(\Phi(\bar{y}(t),\bar{u}(t)))\Big)^* \xi(t) \quad a.e. \ \ in \ (0,T),$$

by the linearity of $f'(\Phi(\bar{y}(t),\bar{u}(t)))$. Thus, the optimality system in Theorem 2.9 reduces to the very same optimality conditions which one obtains when directly applying the KKT-theory on (P), cf. [37], i.e., (2.15) is the classical KKT-system, provided *that f is Gâteaux-differentiable at $\Phi(\bar y(t), \bar u(t))$ a.e. in $(0,T)$. Note that this yields the Gâteaux-differentiability of S at $\bar\ell$, in view of (2.5).*

REMARK 2.13. *If $Y = L^q(\Omega)$ with $q \in (1,\infty)$ and if $f : Y^* \to Y$ is the Nemytskii operator associated to some mapping $f : \mathbb{R} \to \mathbb{R}$, then (2.15c) is equivalent to a similar sign condition which holds a.e. in $(0,T)\times\Omega$ (see (4.27) below). This will be the case in Section 4, where we apply the result in Theorem 2.9 on a concrete setting.*

REMARK 2.14 (Distributed controls). *An inspection of the proof of Theorem 2.9 shows that the arguments cannot be applied if the controls are restricted by additional constraints. The same observation was made in [28], where strong stationarity for optimal control of the obstacle problem is shown to be necessary for local optimality. Let us however mention [38] in this context, where pointwise constraints on the control are considered and strong stationarity is proven by requiring that the (unknown) optimizer satisfies certain assumptions.*

3. Formulation of viscous evolutionary VIs as non-smooth ODEs.

This section focuses on proving that the following viscous evolution

$$R(\eta) - R(\dot y(t)) + \langle \mathcal{V}\dot y(t), \eta - \dot y(t)\rangle_Y \geq \langle \varphi(y(t), \ell(t)), \eta - \dot y(t)\rangle_Y \quad \forall\, \eta \in Y, \quad \text{a.e. in } (0,T) \tag{EVI}$$

is equivalent to a non-smooth ODE in the Hilbert space Y (see Theorem 3.6 below). Note that (EVI) is a generalization of the classical evolutionary VI with viscosity, see e.g. [34, Chp. 4]. In addition, we are concerned with the differentiability properties of the solution map associated to (EVI). For the non-smooth non-linearity appearing in the ODE we give an explicit formula. We state a condition on the functional R and on the operator $\mathcal{V}$, which is necessary and sufficient for its directional differentiability. As a byproduct, this allows us to give a characterization of the directional differentiability of the classical elliptic VI of the second kind (cf. Theorem 3.4 below). The results established in this section will be later on combined with the findings from Section 2, in order to establish strong stationarity for the optimal control of a concrete viscous EVI (see Section 4).

In all what follows, $\ell \in L^2(0,T;\mathcal{H})$ is fixed. Here, $\mathcal{H}$ is a real reflexive Banach space, while Y is a real Hilbert space.

ASSUMPTION 3.1. *For the operators in the viscous EVI we require:*

1. *The non-smooth functional $R : Y \to (-\infty,\infty]$ is proper, convex, lower semicontinuous and positive homogeneous, i.e., $R(\alpha\eta) = \alpha R(\eta)$ for all $\alpha > 0$ and all $\eta \in Y$.*
2. *The viscosity operator $\mathcal{V} \in \mathcal{L}(Y, Y^*)$ is coercive, i.e., there exists $\vartheta > 0$ so that $\langle \mathcal{V}\eta, \eta\rangle_Y \geq \vartheta\|\eta\|_Y^2$ for all $\eta \in Y$. Moreover, $\mathcal{V}$ is symmetric, i.e., $\langle \mathcal{V}\eta, y\rangle_Y = \langle \mathcal{V}y, \eta\rangle_Y$ for all $\eta, y \in Y$.*
3. *The mapping $\varphi : Y \times \mathcal{H} \to Y^*$ is continuously Fréchet-differentiable and Lipschitz-continuous.*

In the sequel, Assumption 3.1 is tacitly assumed, without mentioning it every time. Note that, in view of Assumption 3.1.2, the operator $\mathcal{V}$ induces a norm on Y, which will be denoted by $\|\cdot\|_{\mathcal{V}} := \sqrt{\langle \mathcal{V}\cdot,\cdot\rangle_Y}$. Similarly, the operator $\mathcal{V}^{-1}$ induces a norm on Y^*, which we abbreviate $\|\cdot\|_{\mathcal{V}^{-1}} := \sqrt{\langle \mathcal{V}^{-1}\cdot,\cdot\rangle_{Y^*}}$ in the following. We remark that $\|\cdot\|_{\mathcal{V}}$ and $\|\cdot\|_{\mathcal{V}^{-1}}$ are equivalent to $\|\cdot\|_Y$ and $\|\cdot\|_{Y^*}$, respectively.

DEFINITION 3.2 (The non-smooth non-linearity). *Let us define the function $\mathcal{F} : Y^* \to Y$ as*

$$\mathcal{F}(\omega) := \mathcal{V}^{-1}(\omega - P_{\partial R(0)}\omega), \tag{3.1}$$

where $P_{\partial R(0)} : Y^* \to Y^*$ is the projection operator onto the set $\partial R(0)$ w.r.t. the norm $\|\cdot\|_{\mathcal{V}^{-1}}$, i.e., $P_{\partial R(0)}\omega$ is the unique solution of

$$\min_{\mu \in \partial R(0)} \frac{1}{2}\|\omega - \mu\|^2_{\mathcal{V}^{-1}} \tag{3.2}$$

for any $\omega \in Y^*$.

LEMMA 3.3. *The mapping $\mathcal{F} : Y^* \ni \omega \mapsto z \in Y$ is the solution operator of the following elliptic VI*

$$R(\eta) - R(z) + \langle \mathcal{V}z, \eta - z\rangle_Y \geq \langle \omega, \eta - z\rangle_Y \quad \forall\, \eta \in Y. \tag{3.3}$$

Thus, (3.3) is equivalent to $z = \mathcal{V}^{-1}(\omega - P_{\partial R(0)}\omega)$ for any $\omega \in Y^*$.

Proof. We present two different proofs. The first one is based on convex analysis tools, while the second one is due to [8].

(i) Let $\omega \in Y^*$ be arbitrary, but fixed. We define $R_{\mathcal{V}} : Y \to (-\infty, \infty]$ as

$$R_{\mathcal{V}}(\eta) := R(\eta) + \frac{1}{2}\|\eta\|^2_{\mathcal{V}}. \tag{3.4}$$

Straight-forward computation shows that the conjugate of $\frac{1}{2}\|\cdot\|^2_{\mathcal{V}}$ is given by

$$g : Y^* \to \mathbb{R}, \quad g(\mu) = \sup_{\eta \in Y} \langle \mu, \eta\rangle_Y - \frac{1}{2}\|\eta\|^2_{\mathcal{V}} = \frac{1}{2}\|\mu\|^2_{\mathcal{V}^{-1}}.$$

By (3.4) and sum rule for conjugate functionals, see [23, Thm. 3.3.4.1], it holds

$$\begin{aligned} R^*_{\mathcal{V}}(\omega) &= \inf_{\mu \in Y^*} R^*(\mu) + g(\omega - \mu) = \inf_{\mu \in \partial R(0)} \frac{1}{2}\|\omega - \mu\|^2_{\mathcal{V}^{-1}} \\ &= \frac{1}{2}\|\omega - P_{\partial R(0)}\omega\|^2_{\mathcal{V}^{-1}}, \end{aligned} \tag{3.5}$$

where for the second identity we used $R^* = I_{\partial R(0)}$, which is due to the positive homogeneity of R. Further, in view of [17, Lem. 4.1] and (3.5), in combination with (3.1), we have

$$\partial(R^*_{\mathcal{V}})(\omega) = \mathcal{V}^{-1}(\omega - P_{\partial R(0)}\omega) = \mathcal{F}(\omega) \ \text{ in } Y.$$

Since $R_{\mathcal{V}}$ is convex, lower semicontinuous, and proper, we now deduce by a well-known convex analysis result that

$$\omega \in \partial R_{\mathcal{V}}(\mathcal{F}(\omega)) = \partial R(\mathcal{F}(\omega)) + \mathcal{V}\mathcal{F}(\omega) \ \text{ in } Y^*,$$

in view of (3.4) and sum rule for subdifferentials. Thus, $\mathcal{F}(\omega) \in Y$ solves (3.3). As (3.3) is uniquely solvable, see e.g. [14], the proof is now complete.

(ii) The assertion is a direct result of [8, Prop. 2.1] combined with the fact that, for any $\omega \in Y^*$, the projection $P_{\partial R(0)}\omega \in \partial R(0)$ is characterized as the unique solution of

$$\langle \mathcal{V}^{-1}(\omega - P_{\partial R(0)}\omega), \mu - P_{\partial R(0)}\omega\rangle_{Y^*} \leq 0 \quad \forall\, \mu \in \partial R(0). \tag{3.6}$$

Note that here we used the information that $\mathcal{V}^{-1}$ is symmetric, which is a consequence of Assumption 3.1.2. This concludes the proof. □

As an immediate consequence of Lemma 3.3 and Assumption 3.1.2, we have the following

Theorem 3.4. *The solution operator $\mathcal{F} : Y^* \ni \omega \mapsto z \in Y$ of (3.3) is directionally differentiable at $\bar\omega \in Y^*$ if and only if the projection operator $P_{\partial R(0)} : Y^* \to Y^*$ is directionally differentiable at $\bar\omega \in Y^*$. If this is the case, then*

$$\mathcal{F}'(\bar\omega; \delta\omega) = \mathcal{V}^{-1}(\delta\omega - P'_{\partial R(0)}(\bar\omega; \delta\omega)) \quad \forall\, \delta\omega \in Y^*. \tag{3.7}$$

Remark 3.5. *A criteria for the directional differentiability of $\mathcal{F}$ (or equivalently, of $P_{\partial R(0)}$) is given in Lemma A.1. This is formulated in terms of the polyhedricity of the set $\partial R(0)$. In Appendix A we give some concrete examples of functionals R, for which the associated mapping $\mathcal{F}$ is directionally diffferentiable.*

Theorem 3.6 (Viscous EVIs are non-smooth ODEs in Hilbert space).

1. *The viscous problem (EVI) is equivalent to the following ODE*

$$\dot y(t) = \mathcal{F}(\varphi(y(t), \ell(t))) \quad \text{in } Y \quad \text{a.e. in } (0,T), \tag{3.8}$$

where $\mathcal{F}$ is given by (3.1) and $\ell \in L^2(0,T;\mathcal{H})$. If $y(0) = 0$, then (EVI) admits a unique solution $y \in H_0^1(0,T;Y)$ for every right-hand side $\ell \in L^2(0,T;\mathcal{H})$.

2. *The associated solution map $\mathcal{S} : L^2(0,T;\mathcal{H}) \ni \ell \mapsto y \in H_0^1(0,T;Y)$ is directionally differentiable at $\bar\ell \in L^2(0,T;\mathcal{H})$, if $\mathcal{F} : Y^* \to Y$ or $P_{\partial R(0)} : Y^* \to Y^*$ is directionally differentiable at $\varphi(\bar y(t), \bar\ell(t))$ f.a.a. $t \in (0,T)$, where we abbreviate $\bar y := \mathcal{S}(\bar\ell)$. Its directional derivative $\delta y := \mathcal{S}'(\bar\ell; \delta\ell)$ at $\bar\ell$ in direction $\delta\ell \in L^2(0,T;\mathcal{H})$ is the unique solution of*

$$\dot{\delta y}(t) = \mathcal{F}'\big(\varphi(\bar y(t), \bar\ell(t)); \varphi'(\bar y(t), \bar\ell(t))(\delta y(t), \delta\ell(t))\big) \quad \text{a.e. in } (0,T),\ \delta y(0) = 0. \tag{3.9}$$

Proof. 1. The first assertion is due to Lemma 3.3. As a result thereof, (EVI) reduces to

$$\dot y(t) = g(y(t), t) \quad \text{in } Y \quad \text{a.e. in } (0,T),$$

where $g : Y \times (0,T) \ni (\eta, t) \mapsto \mathcal{F}(\varphi(\eta, \ell(t)) \in Y$. By the global Lipschitz-continuity of $\mathcal{F}$ and φ, we have that g maps $L^2(0,T;Y)$ to $L^2(0,T;Y)$. Moreover, $g(\cdot, t)$ is Lipschitz-continuous f.a.a. $t \in (0,T)$, with Lipschitz-constant independent of t. The unique solvability of (EVI) with initial condition $y(0) = 0$, as well as the $H_0^1(0,T;Y)$-regularity, now follows by a contraction argument, see e.g. [10, Thm. 7.2.3].

2. By arguing as in the first part of the proof, we get that for any $\delta\ell \in L^2(0,T;\mathcal{H})$, (3.9) admits a unique solution $\delta y \in H_0^1(0,T;Y)$. The Lipschitz continuity and the directional differentiability of $\mathcal{F}$ in combination with the differentiability and Lipschitz-continuity of φ (see Assumption 3.1.3), Gronwall's inequality and Lebesgue dominated convergence theorem yield

$$\frac{\mathcal{S}(\bar\ell + \tau\,\delta\ell) - \mathcal{S}(\bar\ell)}{\tau} - \delta y \overset{\tau \searrow 0}{\longrightarrow} 0 \quad \text{in } H^1(0,T;Y) \quad \forall\, \delta\ell \in L^2(0,T;\mathcal{H}).$$

The proof is now complete. ◻

4. Application to viscous elasto-plasticity. In this section, we derive strong stationary optimality conditions for an optimal control problem governed by the following:

For $\ell : [0,T] \to H_D^{-1}(\Omega;\mathbb{R}^n)$, find $\boldsymbol{u} : [0,T] \to H_D^1(\Omega;\mathbb{R}^n)$ and $\boldsymbol{p} : [0,T] \to L^2(\Omega;\mathbb{Q}_0)$ with $\boldsymbol{p}(0) = \mathbf{0}$ such that

$$
\begin{aligned}
\mathcal{R}(\boldsymbol{\eta}) - \mathcal{R}(\dot{\boldsymbol{p}}(t)) + \epsilon\,(\dot{\boldsymbol{p}}(t), \boldsymbol{\eta} - \dot{\boldsymbol{p}}(t))_{L^2(\Omega;\mathbb{Q})} &\geq (\mathbb{C}\big(\varepsilon(\boldsymbol{u}(t)) - \boldsymbol{p}(t)\big) - \kappa \boldsymbol{p}(t), \boldsymbol{\eta} - \dot{\boldsymbol{p}}(t))_{L^2(\Omega;\mathbb{Q})} \\
&\qquad \forall\, \boldsymbol{\eta} \in L^2(\Omega;\mathbb{Q}_0), \\
-\operatorname{div}\big(\mathbb{C}\big(\varepsilon(\boldsymbol{u}(t)) - \boldsymbol{p}(t)\big)\big) = \ell(t) \quad \text{in } H_D^{-1}(\Omega;\mathbb{R}^n)
\end{aligned}
\tag{4.1}
$$

a.e. in $(0,T)$. The problem (4.1) models quasistatic visco-elasto-plasticity at small strains with linear kinematic hardening and von Mises yield condition (see for instance [33, Sec. 2.7] and [29, Chp. 22]; note that we do not consider viscous effects in the balance of momentum equation). It consists of a viscous evolutionary VI (viscoplastic flow rule) and an elliptic PDE (balance of momentum). Thus, we can apply our findings from Section 3 to reduce (4.1) to a system of the type (2.4). Then, by employing the main result from Section 2, we derive a strong stationary optimality system for a class of minimization problems governed by (4.1). The section ends with some remarks and a short comparison with other optimality systems [6, 26].

The quantities in (4.1) are as follows:

- $\Omega \subset \mathbb{R}^n$, $n \in \{2,3\}$, is a bounded domain with Lipschitz boundary Γ. This consists of two disjoint measurable parts Γ_N and Γ_D, where Γ_D is a relatively closed set in Γ of positive measure.
- $\mathbb{Q} := \mathbb{R}^{n\times n}_{sym}$ and $\mathbb{Q}_0 := \{\boldsymbol{\eta} \in \mathbb{Q} \,|\, \operatorname{trace}(\boldsymbol{\eta}) = 0\}$. By $\boldsymbol{\eta}^D$ we denote the deviatoric part of $\boldsymbol{\eta} \in \mathbb{Q}$, i.e., $\boldsymbol{\eta}^D := \boldsymbol{\eta} - \frac{\operatorname{trace}(\boldsymbol{\eta})}{n}\mathbb{I}$, where $\mathbb{I}$ is the $n \times n$ identity matrix. The spherical part of $\boldsymbol{\eta}$ is abbreviated $\boldsymbol{\eta}^S := \frac{\operatorname{trace}(\boldsymbol{\eta})}{n}\mathbb{I}$, i.e., $\mathbb{Q}_0 = \{\boldsymbol{\eta} \in \mathbb{Q} \,|\, \boldsymbol{\eta}^S = \mathbf{0}\}$. Straight-forward computation yields the equality
$$\boldsymbol{\eta}_1^D : \boldsymbol{\eta}_2 = \boldsymbol{\eta}_1^D : \boldsymbol{\eta}_2^D \quad \forall\, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \in \mathbb{Q}, \tag{4.2}$$
which will be frequently used in the sequel. The symbol $:$ denotes the scalar product which induces the Frobenius norm on $\mathbb{R}^{n\times n}$.
- The functional $\mathcal{R} : L^2(\Omega;\mathbb{Q}) \to \mathbb{R}$ is given by
$$\mathcal{R}(\boldsymbol{\eta}) = \begin{cases} \sigma_0 \int_\Omega |\boldsymbol{\eta}|\,dx & \text{if } \boldsymbol{\eta} \in L^2(\Omega;\mathbb{Q}_0), \\ \infty & \text{otherwise,} \end{cases} \tag{4.3}$$
where $\sigma_0 > 0$. Here, $|\cdot|$ denotes the Frobenius norm on $\mathbb{R}^{n\times n}$.
- $\epsilon > 0$ is the viscosity parameter, while $\kappa > 0$ is the hardening constant.
- The fourth-order elasticity tensor $\mathbb{C} \in L^\infty(\Omega;\mathcal{L}(\mathbb{Q}))$ is symmetric, i.e., $\mathbb{C}(x)\boldsymbol{\eta} : \boldsymbol{\omega} = \boldsymbol{\eta} : \mathbb{C}(x)\boldsymbol{\omega}$ for all $\boldsymbol{\eta}, \boldsymbol{\omega} \in \mathbb{Q}$, a.e. in Ω. Moreover, $\mathbb{C}$ is uniformly coercive, i.e., there is a constant $\gamma_{\mathbb{C}} > 0$ such that
$$\mathbb{C}(x)\boldsymbol{\eta} : \boldsymbol{\eta} \geq \gamma_{\mathbb{C}} |\boldsymbol{\eta}|^2 \quad \forall\, \boldsymbol{\eta} \in \mathbb{Q} \text{ and f.a.a. } x \in \Omega. \tag{4.4}$$
- $\varepsilon(\boldsymbol{v}) = \frac{1}{2}(\nabla \boldsymbol{v} + \nabla \boldsymbol{v}^\top)$ is the linearized strain tensor and $\operatorname{div} : L^2(\Omega;\mathbb{Q}) \to H_D^{-1}(\Omega;\mathbb{R}^n)$ is the distributional vector-valued divergence, i.e.,
$$\langle \operatorname{div} \boldsymbol{\eta}, \boldsymbol{v}\rangle := -\int_\Omega \boldsymbol{\eta} : \varepsilon(\boldsymbol{v})\,dx \quad \forall\, \boldsymbol{v} \in H_D^1(\Omega;\mathbb{R}^n),$$

where $H^1_D(\Omega;\mathbb{R}^n) := \{\boldsymbol{v} \in H^1(\Omega;\mathbb{R}^n) |\, \boldsymbol{v} = \mathbf{0} \text{ on } \Gamma_D\}$ and $H^{-1}_D(\Omega;\mathbb{R}^n) := H^1_D(\Omega;\mathbb{R}^n)^*$.

Throughout this section, we denote vector valued variables by bold-face case letters. For simplicity, we often make the convention $\frac{0}{0} = 0$.

Lemma 4.1 (Control-to-state map and directional differentiability).

1. *The viscoplastic problem* (4.1) *is equivalent to the following PDE system*

$$\dot{\boldsymbol{p}}(t) = f(\boldsymbol{\sigma}(t) - \kappa \boldsymbol{p}(t)) \quad \text{in } L^2(\Omega;\mathbb{Q}_0), \quad \boldsymbol{p}(0) = \mathbf{0}, \tag{4.5}$$

$$-\operatorname{div} \boldsymbol{\sigma}(t) = \ell(t) \quad \text{in } H^{-1}_D(\Omega;\mathbb{R}^n) \quad \text{a.e. in } (0,T), \tag{4.6}$$

where we abbreviate $\boldsymbol{\sigma} := \mathbb{C}\big(\varepsilon(\boldsymbol{u}) - \boldsymbol{p}\big)$. *Here,* $f : L^2(\Omega;\mathbb{Q}) \to L^2(\Omega;\mathbb{Q}_0)$ *is given by*

$$f(\boldsymbol{\omega}) = \frac{1}{\epsilon}\chi_{\{|\boldsymbol{\omega}^D|>\sigma_0\}}\Big(1 - \frac{\sigma_0}{|\boldsymbol{\omega}^D|}\Big)\boldsymbol{\omega}^D, \tag{4.7}$$

where $\chi_{\{|\boldsymbol{\omega}^D|>\sigma_0\}} : \Omega \to \{0,1\}$ *stands for the characteristic function of the set* $\{x \in \Omega :\, |\boldsymbol{\omega}^D(x)| > \sigma_0\}$. *For every* $\ell \in L^2(0,T;H^{-1}_D(\Omega;\mathbb{R}^n))$, (4.1) *admits a unique solution* $(\boldsymbol{p},\boldsymbol{u}) \in H^1_0(0,T;L^2(\Omega;\mathbb{Q}_0)) \times L^2(0,T;H^1_D(\Omega;\mathbb{R}^n))$.

2. *The solution map associated to* (4.1)

$$S : L^2(0,T;H^{-1}_D(\Omega;\mathbb{R}^n)) \ni \ell \mapsto (\boldsymbol{p},\boldsymbol{u}) \in H^1_0(0,T;L^2(\Omega;\mathbb{Q}_0)) \times L^2(0,T;H^1_D(\Omega;\mathbb{R}^n))$$

is directionally differentiable. Its directional derivative $(\boldsymbol{\delta p},\boldsymbol{\delta u}) := S'(\ell;\delta\ell)$ *at* $\ell \in L^2(0,T;H^{-1}_D(\Omega;\mathbb{R}^n))$ *in direction* $\delta\ell \in L^2(0,T;H^{-1}_D(\Omega;\mathbb{R}^n))$ *is the unique solution of*

$$\begin{aligned}
&\dot{\boldsymbol{\delta p}}(t) = f'(\boldsymbol{\sigma}(t) - \kappa\boldsymbol{p}(t); \boldsymbol{\delta\sigma}(t) - \kappa\boldsymbol{\delta p}(t)) \quad \text{in } L^2(\Omega;\mathbb{Q}_0), \quad \boldsymbol{\delta p}(0) = \mathbf{0},\\
&-\operatorname{div} \boldsymbol{\delta\sigma}(t) = \delta\ell(t) \quad \text{in } H^{-1}_D(\Omega;\mathbb{R}^n) \quad \text{a.e. in } (0,T),
\end{aligned} \tag{4.8}$$

where we abbreviate $(\boldsymbol{p},\boldsymbol{u}) := S(\ell)$, $\boldsymbol{\sigma} := \mathbb{C}(\varepsilon(\boldsymbol{u}) - \boldsymbol{p})$ *and* $\boldsymbol{\delta\sigma} := \mathbb{C}\big(\varepsilon(\boldsymbol{\delta u}) - \boldsymbol{\delta p}\big)$. *Here,*

$$\begin{aligned}
f'(\boldsymbol{\omega};\boldsymbol{\delta\omega}) = {}& \frac{1}{\epsilon}\chi_{\{|\boldsymbol{\omega}^D|>\sigma_0\}}\Big\{\Big(\frac{\sigma_0}{|\boldsymbol{\omega}^D|^3}\boldsymbol{\omega}^D : \boldsymbol{\delta\omega}^D\Big)\boldsymbol{\omega}^D + \Big(1 - \frac{\sigma_0}{|\boldsymbol{\omega}^D|}\Big)\boldsymbol{\delta\omega}^D\Big\}\\
&+ \frac{1}{\epsilon}\chi_{\{|\boldsymbol{\omega}^D|=\sigma_0, \boldsymbol{\omega}^D:\boldsymbol{\delta\omega}^D>0\}}\Big(\frac{1}{\sigma_0^2}\boldsymbol{\omega}^D : \boldsymbol{\delta\omega}^D\Big)\boldsymbol{\omega}^D \quad \forall\, \boldsymbol{\omega},\boldsymbol{\delta\omega} \in L^2(\Omega;\mathbb{Q}),
\end{aligned} \tag{4.9}$$

where $\chi_{\{|\boldsymbol{\omega}^D|=\sigma_0, \boldsymbol{\omega}^D:\boldsymbol{\delta\omega}^D>0\}} : \Omega \to \{0,1\}$ *is the characteristic function of the set* $\{x \in \Omega :\, |\boldsymbol{\omega}^D(x)| = \sigma_0, \boldsymbol{\omega}^D(x) : \boldsymbol{\delta\omega}^D(x) > 0\}$.

Proof. 1. Let us first prove that, for any $\boldsymbol{\omega} \in L^2(\Omega;\mathbb{Q})$, it holds $P_{\partial\mathcal{R}(\mathbf{0})}\boldsymbol{\omega} = g(\boldsymbol{\omega})$, where

$$g(\boldsymbol{\omega}) := \begin{cases} \boldsymbol{\omega}^S + \frac{\sigma_0}{|\boldsymbol{\omega}^D|}\boldsymbol{\omega}^D & \text{a.e. where } |\boldsymbol{\omega}^D| > \sigma_0,\\ \boldsymbol{\omega} & \text{a.e. where } |\boldsymbol{\omega}^D| \le \sigma_0. \end{cases} \tag{4.10}$$

Here, $P_{\partial\mathcal{R}(\mathbf{0})}$ is the projection operator onto the set $\partial\mathcal{R}(\mathbf{0})$ w.r.t. the $L^2(\Omega;\mathbb{Q})-$ norm. We observe that

$$\begin{aligned}
\partial\mathcal{R}(\mathbf{0}) &= \{\boldsymbol{\mu} \in L^2(\Omega;\mathbb{Q}) |\, \int_\Omega \boldsymbol{\mu} : \boldsymbol{\eta}\, dx \le \sigma_0 \int_\Omega |\boldsymbol{\eta}|\, dx \quad \forall\, \boldsymbol{\eta} \in L^2(\Omega;\mathbb{Q}_0)\}\\
&= \{\boldsymbol{\mu} \in L^2(\Omega;\mathbb{Q}) |\, |\boldsymbol{\mu}^D(x)| \le \sigma_0 \text{ f.a.a. } x \in \Omega\},
\end{aligned} \tag{4.11}$$

as a result of (4.3), the identity $\boldsymbol{\mu} : \boldsymbol{\eta} = \boldsymbol{\mu}^D : \boldsymbol{\eta}$ (for $\boldsymbol{\mu} \in \mathbb{Q}$ and $\boldsymbol{\eta} \in \mathbb{Q}_0$, cf. (4.2)) and the fundamental lemma of calculus of variations. It is straight forward to verify that $g(\boldsymbol{\omega}) \in \partial\mathcal{R}(\mathbf{0})$. Moreover, it holds

$$\begin{aligned}&\int_{|\boldsymbol{\omega}^D|>\sigma_0} \Big(\boldsymbol{\omega} - \boldsymbol{\omega}^S - \frac{\sigma_0}{|\boldsymbol{\omega}^D|}\boldsymbol{\omega}^D\Big) : \Big(\boldsymbol{\mu} - \boldsymbol{\omega}^S - \frac{\sigma_0}{|\boldsymbol{\omega}^D|}\boldsymbol{\omega}^D\Big)\, dx + \int_{|\boldsymbol{\omega}^D|\le\sigma_0} (\boldsymbol{\omega} - \boldsymbol{\omega}) : (\boldsymbol{\mu} - \boldsymbol{\omega})\, dx \\ &= \int_{|\boldsymbol{\omega}^D|>\sigma_0} \Big(\boldsymbol{\omega}^D - \frac{\sigma_0}{|\boldsymbol{\omega}^D|}\boldsymbol{\omega}^D\Big) : \Big(\boldsymbol{\mu}^D - \frac{\sigma_0}{|\boldsymbol{\omega}^D|}\boldsymbol{\omega}^D\Big)\, dx \le 0 \quad \forall\, \boldsymbol{\mu} \in \partial\mathcal{R}(\mathbf{0}),\end{aligned} \tag{4.12}$$

where we used (4.2) for the equality. The inequality in (4.12) follows by straight forward computation. Since the projection on $\partial\mathcal{R}(\mathbf{0})$ is characterized as the unique solution of (3.6), we deduce from $g(\boldsymbol{\omega}) \in \partial\mathcal{R}(\mathbf{0})$, (4.10) and (4.12) that

$$P_{\partial\mathcal{R}(\mathbf{0})}\boldsymbol{\omega} = g(\boldsymbol{\omega}) \tag{4.13}$$

is indeed true.

We now show that the problem (4.1) fits in the setting of (EVI). In view of Korn's inequality, the operator $\operatorname{div} \in \mathcal{L}(L^2(\Omega;\mathbb{Q}), H_D^{-1}(\Omega;\mathbb{R}^n))$ is invertible, see e.g. [17, Lem. 2.11]. Thus, we can write (4.1) as

$$\mathcal{R}(\boldsymbol{\eta}) - \mathcal{R}(\dot{\boldsymbol{p}}(t)) + \epsilon\,(\dot{\boldsymbol{p}}(t), \boldsymbol{\eta} - \dot{\boldsymbol{p}}(t))_{L^2(\Omega;\mathbb{Q})} \ge (-\operatorname{div}^{-1}\ell(t) - \kappa\boldsymbol{p}(t), \boldsymbol{\eta} - \dot{\boldsymbol{p}}(t))_{L^2(\Omega;\mathbb{Q})}\ \forall\, \boldsymbol{\eta} \in L^2(\Omega;\mathbb{Q}_0), \tag{4.14}$$

$$-\operatorname{div}\big(\mathbb{C}\big(\varepsilon(\boldsymbol{u}(t)) - \boldsymbol{p}(t)\big)\big) = \ell(t) \quad \text{a.e. in } (0,T). \tag{4.15}$$

Now, with the notations from Section 3, we see that if we set

$$Y := L^2(\Omega;\mathbb{Q}), \quad \mathcal{H} := H_D^{-1}(\Omega;\mathbb{R}^n), \tag{4.16a}$$

$$R := \mathcal{R}, \quad \mathcal{V} := \epsilon\,\mathbb{I}, \quad \varphi(\boldsymbol{p}, \ell) := -\operatorname{div}^{-1}\ell - \kappa\boldsymbol{p}, \tag{4.16b}$$

then (EVI) coincides with (4.14). Note that $\varphi \in \mathcal{L}(Y \times \mathcal{H}; Y^*)$ with

$$\varphi'(\boldsymbol{p}, \ell)(\boldsymbol{\delta p}, \delta\ell) = -\operatorname{div}^{-1}\delta\ell - \kappa\boldsymbol{\delta p} \quad \forall\,(\boldsymbol{p}, \ell), (\boldsymbol{\delta p}, \delta\ell) \in Y \times \mathcal{H}. \tag{4.17}$$

Let now $\ell \in L^2(0,T;H_D^{-1}(\Omega;\mathbb{R}^n))$ be arbitrary, but fixed. Since the quantities in (4.16) satisfy Assumption 3.1, we can apply Theorem 3.6.1, which yields that (4.14) is equivalent to

$$\dot{\boldsymbol{p}}(t) = \mathcal{F}\big(\varphi(\boldsymbol{p}(t), \ell(t))\big) \text{ a.e. in } (0,T),$$

where $\mathcal{F} = \frac{1}{\epsilon}(\mathbb{I} - P_{\partial\mathcal{R}(\mathbf{0})}) = f$, by (4.13), (4.10), and (4.7). Moreover, Theorem 3.6.1 tells us that (4.14) admits a unique solution $\boldsymbol{p} \in H_0^1(0,T;L^2(\Omega;\mathbb{Q}))$ for every $\ell \in L^2(0,T;H_D^{-1}(\Omega;\mathbb{R}^n))$. Note that $\dot{\boldsymbol{p}}(t) \in dom\,\mathcal{R} = L^2(\Omega;\mathbb{Q}_0)$ a.e. in $(0,T)$, in view of (4.3). The linear character of the deviatoric part then yields the Bochner measurabiliy of $\dot{\boldsymbol{p}} : [0,T] \to L^2(\Omega;\mathbb{Q}_0)$, and we can conclude $\boldsymbol{p} \in H_0^1(0,T;L^2(\Omega;\mathbb{Q}_0))$. Further, due to (4.4), we deduce by Lax-Milgram lemma that

$$-\operatorname{div}\mathbb{C}\big(\varepsilon(\boldsymbol{u})\big) = \ell - \operatorname{div}\mathbb{C}\,\boldsymbol{p} \quad \text{in } L^2(0,T;H_D^{-1}(\Omega;\mathbb{R}^n)) \tag{4.18}$$

admits a unique solution $\boldsymbol{u} \in L^2(0,T;H_D^1(\Omega;\mathbb{R}^n))$. As a consequence, $(\boldsymbol{u}, \boldsymbol{p})$ satisfies (4.14)-(4.15). The proof of this step is now complete.

2. Since $f(\boldsymbol{\omega}) = \frac{1}{\epsilon}\max\Big(1 - \frac{\sigma_0}{|\boldsymbol{\omega}^D|}, 0\Big)\boldsymbol{\omega}^D$ for all $\boldsymbol{\omega} \in L^2(\Omega;\mathbb{Q})$ (with the convention $f(\mathbf{0}) = \mathbf{0}$), we can infer from [20, Lem. 2.11] and Lebesgue dominated convergence theorem that $f : L^2(\Omega;\mathbb{Q}) \to L^2(\Omega;\mathbb{Q})$ is directionally differentiable with derivative given by (4.9). Note that here we use the Lipschitz continuity of $f : \mathbb{Q} \to \mathbb{Q}$, which is a result of the inequality $|f'(\boldsymbol{\omega};\boldsymbol{\delta\omega})| \le L|\boldsymbol{\delta\omega}|\ \forall\, \boldsymbol{\omega}, \boldsymbol{\delta\omega} \in \mathbb{Q}$ and [4, Cor. 3.2.2., p. 40].

We denote by S_1 the first component of the operator S, i.e., $S_1 : L^2(0,T;H_D^{-1}(\Omega;\mathbb{R}^n)) \ni \ell \mapsto \boldsymbol{p} \in H_0^1(0,T;L^2(\Omega;\mathbb{Q}_0))$ is the solution map associated to (4.14). From Theorem 3.6.2 we deduce that S_1 is directionally differentiable. Its directional derivative $\boldsymbol{\delta p} := S_1'(\ell;\delta\ell)$ at ℓ in direction $\delta\ell$ is the unique solution of

$$\dot{\boldsymbol{\delta p}}(t) = f'\big(\varphi(\boldsymbol{p}(t),\ell(t)); \varphi'(\boldsymbol{p}(t),\ell(t))(\boldsymbol{\delta p}(t), \delta\ell(t))\big) \quad \text{a.e. in } (0,T),\ \ \boldsymbol{\delta p}(0) = \mathbf{0}, \tag{4.19}$$

where $\boldsymbol{p} := S_1(\ell)$. In view of the definition of φ (see (4.16)) and (4.17), (4.19) reads

$$\dot{\boldsymbol{\delta p}}(t) = f'\big(\underbrace{-\operatorname{div}^{-1}\ell(t)}_{=\boldsymbol{\sigma}(t)} - \kappa\boldsymbol{p}(t); \underbrace{-\operatorname{div}^{-1}\delta\ell(t)}_{=\boldsymbol{\delta\sigma}(t)} - \kappa\boldsymbol{\delta p}(t)\big) \quad \text{a.e. in } (0,T),\ \ \boldsymbol{\delta p}(0) = \mathbf{0}. \tag{4.20}$$

Further, from (4.18) we have

$$S_2(\ell)(t) = \big[-\operatorname{div}\mathbb{C}\big(\varepsilon(\cdot)\big)\big]^{-1}\big(\ell(t) - \operatorname{div}\mathbb{C}\,S_1(\ell)(t)\big) \quad \text{a.e. in } (0,T), \tag{4.21}$$

where S_2 is the second component of the operator S, i.e., $S_2 : L^2(0,T;H_D^{-1}(\Omega;\mathbb{R}^n)) \ni \ell \mapsto \boldsymbol{u} \in L^2(0,T;H_D^1(\Omega;\mathbb{R}^n))$. Thus, S_2 is directionally differentiable as well, since $\big[-\operatorname{div}\mathbb{C}\big(\varepsilon(\cdot)\big)\big]^{-1} \in \mathcal{L}(H_D^{-1}(\Omega;\mathbb{R}^n), H_D^1(\Omega;\mathbb{R}^n))$ and S_1 is directionally differentiable. Its directional derivative $\boldsymbol{\delta u} := S_2'(\ell;\delta\ell)$ at ℓ in direction $\delta\ell$ is given by

$$\boldsymbol{\delta u}(t) = \big(-\operatorname{div}\mathbb{C}\big(\varepsilon(\cdot)\big)\big)^{-1}\big(\delta\ell(t) - \operatorname{div}\mathbb{C}\,S_1'(\ell;\delta\ell)(t)\big) \quad \text{a.e. in } (0,T),$$

as a result of (4.21). This completes the proof. $\square$

Next, we want to apply the strong stationarity result from Section 2 on the following optimal control problem:

$$\left.\begin{aligned} \min_{\ell \in L^2(0,T;L^2(\Omega;\mathbb{R}^n))} \quad & \mathcal{J}(\boldsymbol{p},\boldsymbol{u},\ell) \\ \text{s.t.} \quad & (\boldsymbol{p},\boldsymbol{u}) \text{ solves (4.1) w.r.h.s. } \ell. \end{aligned}\right\} \tag{Q}$$

In the sequel, the objective $\mathcal{J}$ is supposed to fulfill

Assumption 4.2. *The objective $\mathcal{J} : L^2(0,T;L^2(\Omega;\mathbb{Q})) \times L^2(0,T;H_D^1(\Omega;\mathbb{R}^n)) \times L^2(0,T;L^2(\Omega;\mathbb{R}^n)) \to \mathbb{R}$ is Fréchet-differentiable.*

The main result of this section reads as follows.

Theorem 4.3 (Strong stationarity for optimal control of viscous quasistatic elastoplasticity). *Let $\bar\ell \in L^2(0,T;L^2(\Omega;\mathbb{R}^n))$ be locally optimal for (Q) with associated states*

$$\bar{\boldsymbol{p}} \in H_0^1(0,T;L^2(\Omega;\mathbb{Q}_0)) \quad \textit{and} \quad \bar{\boldsymbol{u}} \in L^2(0,T;H_D^1(\Omega;\mathbb{R}^n)).$$

Then, there exist unique adjoint states

$$\boldsymbol{\xi} \in H_T^1(0,T;L^2(\Omega;\mathbb{Q})) \quad \textit{and} \quad \boldsymbol{w} \in L^2(0,T;H_D^1(\Omega;\mathbb{R}^n)),$$

and a unique multiplier $\boldsymbol{\lambda} \in L^2(0,T;L^2(\Omega;\mathbb{Q}_0))$ *such that the following system is satisfied*

$$\left.\begin{aligned} -\dot{\boldsymbol{\xi}} + \mathbb{C}(\boldsymbol{\lambda} - \varepsilon \boldsymbol{w}) + \kappa\boldsymbol{\lambda} &= \partial_{\boldsymbol{p}} \mathcal{J}(\bar{\boldsymbol{p}}, \bar{\boldsymbol{u}}, \bar{\ell}) \quad \text{in } L^2(0,T;L^2(\Omega;\mathbb{Q})), \quad \boldsymbol{\xi}(T) = \mathbf{0}, \\ \operatorname{div}\big(\mathbb{C}(\boldsymbol{\lambda} - \varepsilon\boldsymbol{w})\big) &= \partial_{\boldsymbol{u}} \mathcal{J}(\bar{\boldsymbol{p}}, \bar{\boldsymbol{u}}, \bar{\ell}) \quad \text{in } L^2(0,T;H^{-1}_D(\Omega;\mathbb{R}^n)), \end{aligned}\right\} \tag{4.22a}$$

$$\left.\begin{aligned} &\boldsymbol{\lambda} = \frac{1}{\epsilon}\chi_{\{|\bar{\boldsymbol{z}}^D| > \sigma_0\}} \Big\{ \Big(\frac{\sigma_0}{|\bar{\boldsymbol{z}}^D|^3} \bar{\boldsymbol{z}}^D : \boldsymbol{\xi}^D\Big)\bar{\boldsymbol{z}}^D + \Big(1 - \frac{\sigma_0}{|\bar{\boldsymbol{z}}^D|}\Big)\boldsymbol{\xi}^D \Big\} \quad \text{a.e. where } |\bar{\boldsymbol{z}}(t,x)^D| \neq \sigma_0, \\ &\left.\begin{aligned} &\boldsymbol{\lambda}(t,x) = \theta(t,x)\bar{\boldsymbol{z}}(t,x)^D, \text{ with} \\ &\theta(t,x) \in \Big[0, \frac{\bar{\boldsymbol{z}}(t,x)^D : \boldsymbol{\xi}(t,x)^D}{\epsilon\sigma_0^2}\Big] \end{aligned}\right\} \text{a.e. where } |\bar{\boldsymbol{z}}(t,x)^D| = \sigma_0, \end{aligned}\right\} \tag{4.22b}$$

$$\boldsymbol{w} + \partial_\ell \mathcal{J}(\bar{\boldsymbol{p}}, \bar{\boldsymbol{u}}, \bar{\ell}) = \mathbf{0} \quad \text{in } L^2(0,T;H^1_D(\Omega;\mathbb{R}^n)), \tag{4.22c}$$

where we abbreviate $\bar{\boldsymbol{z}} := \mathbb{C}\big(\varepsilon(\bar{\boldsymbol{u}}) - \bar{\boldsymbol{p}}\big) - \kappa\bar{\boldsymbol{p}}$*.*

Proof. We aim to apply the strong stationarity result given by Theorem 2.9 for the optimal control problem (Q). To this end, we have to check if (Q) fits in the general setting from Section 2. After that, we verify Assumptions 2.3 and 2.6. Indeed, with the notations from Section 2, we see that if we set

$$V := L^2(\Omega;\mathbb{R}^n), \; Y := L^2(\Omega;\mathbb{Q}), \; U := H^1_D(\Omega;\mathbb{R}^n), \; J := \mathcal{J}, \tag{4.23a}$$

$$f : Y^* \ni \boldsymbol{\omega} \mapsto \frac{1}{\epsilon}\chi_{\{|\boldsymbol{\omega}^D| > \sigma_0\}}\Big(1 - \frac{\sigma_0}{|\boldsymbol{\omega}^D|}\Big)\boldsymbol{\omega}^D \in Y, \tag{4.23b}$$

$$\Phi : Y \times U \ni (\boldsymbol{p}, \boldsymbol{u}) \mapsto \mathbb{C}\big(\varepsilon(\boldsymbol{u}) - \boldsymbol{p}\big) - \kappa\boldsymbol{p} \in Y^*, \tag{4.23c}$$

$$\Psi : Y \times U \ni (\boldsymbol{p}, \boldsymbol{u}) \mapsto -\operatorname{div}\big(\mathbb{C}\big(\varepsilon(\boldsymbol{u}) - \boldsymbol{p}\big)\big) \in U^*, \tag{4.23d}$$

then (P) coincides with (Q), thanks to Lemma 4.1.1. Notice that $V \overset{d}{\hookrightarrow} U^*$. The Lipschitz-continuity and the directional differentiability of f was established in the proof of Lemma 4.1.2, so that Assumption 2.1.2 is satisfied. Since $\Phi \in \mathcal{L}(Y \times U; Y^*)$ and $\Psi \in \mathcal{L}(Y \times U; U^*)$, Assumption 2.1.3 is fulfilled as well. Thus, the entire Assumption 2.1 is satisfied by the quantities in (4.23).

Moreover, by employing again Lemma 4.1.1, we see that Assumption 2.3.1 holds. The resulting solution operator of (2.4), i.e., $S : L^2(0,T;H^{-1}_D(\Omega;\mathbb{R}^n)) \ni \ell \mapsto (\boldsymbol{p}, \boldsymbol{u}) \in H^1_0(0,T;L^2(\Omega;\mathbb{Q}_0)) \times L^2(0,T;H^1_D(\Omega;\mathbb{R}^n))$ is directionally differentiable, cf. Lemma 4.1.2. According to the latter, its directional derivative $S'(\ell;\delta\ell)$ at ℓ in direction $\delta\ell$ is the unique solution of (4.8), and thus, of (2.5), whence Assumption 2.3.2 follows. From (4.8), the Lipschitz-continuity of f, and Gronwall's lemma, we further deduce that $S : L^2(0,T;U^*) \to L^2(0,T;Y \times U)$ is Lipschitz-continuous, which implies that Assumption 2.3.3 is verified as well, see Remark 2.4. Hence, the entire Assumption 2.3 is true for the setting (4.23).

It remains to check that Assumption 2.6 is guaranteed, i.e., we have to see if the mapping

$$\partial_{\boldsymbol{u}}\Phi(\bar{\boldsymbol{p}}, \bar{\boldsymbol{u}}) : L^2(0,T;H^1_D(\Omega;\mathbb{R}^n)) \ni \boldsymbol{\delta u} \mapsto \mathbb{C}\varepsilon(\boldsymbol{\delta u}) \in L^2(0,T;L^2(\Omega;\mathbb{Q})) \tag{4.24}$$

is invertible. To this end, we observe that, by the coercivity of $\mathbb{C}$, cf. (4.4), and Korn's inequality, $\mathbb{C}\varepsilon(\cdot) \in \mathcal{L}(H^1_D(\Omega;\mathbb{R}^n), L^2(\Omega;\mathbb{Q}))$ satisfies the condition of Babuška-Brezzi, and is thus, invertible. Hence, given $\boldsymbol{\zeta} \in L^2(0,T;L^2(\Omega;\mathbb{Q}))$, the mapping

$t \mapsto \boldsymbol{\delta u}(t) := (\mathbb{C}\varepsilon(\cdot))^{-1}(\boldsymbol{\zeta}(t)) \in H^1_D(\Omega)$ is the unique solution of $\partial_{\boldsymbol{u}}\Phi(\bar{\boldsymbol{p}}, \bar{\boldsymbol{u}})(\boldsymbol{\delta u})(t) = \boldsymbol{\zeta}(t)$ f.a.a. $t \in (0,T)$. Note that $\boldsymbol{\delta u} \in L^2(0,T;H^1_D(\Omega;\mathbb{R}^n))$. Therefore, Assumption 2.6 is indeed satisfied.

Before applying Theorem 2.9, we compute the adjoints of the partial derivatives of Φ and Ψ. In the light of Lemma 2.2, the definition of the adjoint, the symmetry of $\mathbb{C}$, and (4.23c)-(4.23d), these are given by

$$
\begin{aligned}
\partial_{\boldsymbol{p}}\Phi(\bar{\boldsymbol{p}}, \bar{\boldsymbol{u}})^* &: L^2(0,T;L^2(\Omega;\mathbb{Q})) \ni \boldsymbol{\eta} \mapsto -\mathbb{C}\boldsymbol{\eta} - \kappa\boldsymbol{\eta} \in L^2(0,T;L^2(\Omega;\mathbb{Q})),\\
\partial_{\boldsymbol{u}}\Phi(\bar{\boldsymbol{p}}, \bar{\boldsymbol{u}})^* &: L^2(0,T;L^2(\Omega;\mathbb{Q})) \ni \boldsymbol{\eta} \mapsto -\operatorname{div}\mathbb{C}(\boldsymbol{\eta}) \in L^2(0,T;H^{-1}_D(\Omega;\mathbb{R}^n)),\\
\partial_{\boldsymbol{p}}\Psi(\bar{\boldsymbol{p}}, \bar{\boldsymbol{u}})^* &: L^2(0,T;H^1_D(\Omega;\mathbb{R}^n)) \ni \boldsymbol{\eta} \mapsto -\mathbb{C}(\varepsilon\boldsymbol{\eta}) \in L^2(0,T;L^2(\Omega;\mathbb{Q})),\\
\partial_{\boldsymbol{u}}\Psi(\bar{\boldsymbol{p}}, \bar{\boldsymbol{u}})^* &: L^2(0,T;H^1_D(\Omega;\mathbb{R}^n)) \ni \boldsymbol{\eta} \mapsto -\operatorname{div}\mathbb{C}(\varepsilon\boldsymbol{\eta}) \in L^2(0,T;H^{-1}_D(\Omega;\mathbb{R}^n)).
\end{aligned} \tag{4.25}
$$

Now, Theorem 2.9 in combination with (4.25) tells us that there exist unique adjoint states $\boldsymbol{\xi} \in H^1(0,T;L^2(\Omega;\mathbb{Q}))$ and $\boldsymbol{w} \in L^2(0,T;H^1_D(\Omega;\mathbb{R}^n))$ and a unique multiplier $\boldsymbol{\lambda} \in L^2(0,T;L^2(\Omega;\mathbb{Q}))$ such that

$$-\dot{\boldsymbol{\xi}} + \mathbb{C}(\boldsymbol{\lambda} - \varepsilon\boldsymbol{w}) + \kappa\boldsymbol{\lambda} = \partial_{\boldsymbol{p}}\mathcal{J}(\bar{\boldsymbol{p}}, \bar{\boldsymbol{u}}, \bar{\ell}) \ \text{ in } L^2(0,T;L^2(\Omega;\mathbb{Q})), \quad \boldsymbol{\xi}(T) = \mathbf{0}, \tag{4.26a}$$

$$\operatorname{div}\big(\mathbb{C}(\boldsymbol{\lambda} - \varepsilon\boldsymbol{w})\big) = \partial_{\boldsymbol{u}}\mathcal{J}(\bar{\boldsymbol{p}}, \bar{\boldsymbol{u}}, \bar{\ell}) \ \text{ in } L^2(0,T;H^{-1}_D(\Omega;\mathbb{R}^n)), \tag{4.26b}$$

$$(\boldsymbol{\xi}(t), f'\big(\Phi(\bar{\boldsymbol{p}}(t), \bar{\boldsymbol{u}}(t)); \boldsymbol{v}\big))_{L^2(\Omega;\mathbb{Q})} \geq (\boldsymbol{\lambda}(t), \boldsymbol{v})_{L^2(\Omega;\mathbb{Q})} \quad \forall\, \boldsymbol{v} \in L^2(\Omega;\mathbb{Q}), \text{ a.e. in } (0,T), \tag{4.26c}$$

$$\boldsymbol{w} + \partial_\ell\mathcal{J}(\bar{\boldsymbol{p}}, \bar{\boldsymbol{u}}, \bar{\ell}) = \mathbf{0} \ \text{ in } L^2(0,T;H^1_D(\Omega;\mathbb{R}^n)). \tag{4.26d}$$

It remains to prove that (4.26c) implies (4.22b). An argument based on the fundamental lemma of calculus of variations and the positive homogeneity of the directional derivative w.r.t. direction implies

$$\boldsymbol{\xi}(t,x) : f'(\bar{\boldsymbol{z}}(t,x); \boldsymbol{v}) \geq \boldsymbol{\lambda}(t,x) : \boldsymbol{v} \quad \forall\, \boldsymbol{v} \in \mathbb{Q}, \text{ a.e. in } (0,T) \times \Omega. \tag{4.27}$$

Here, we recall the abbreviation $\bar{\boldsymbol{z}} := \mathbb{C}\big(\varepsilon(\bar{\boldsymbol{u}}) - \bar{\boldsymbol{p}}\big) - \kappa\bar{\boldsymbol{p}}$ and (4.23c), i.e., $\bar{\boldsymbol{z}} = \Phi(\bar{\boldsymbol{p}}, \bar{\boldsymbol{u}})$. From (4.9) we deduce that $f : \mathbb{Q} \to \mathbb{Q}$ is Gâteaux differentiable on the set $\{\boldsymbol{\omega} \in \mathbb{Q} : |\boldsymbol{\omega}^D| \neq \sigma_0\}$ with derivative

$$f'(\boldsymbol{\omega})(\boldsymbol{\delta\omega}) = \frac{1}{\epsilon}\chi_{\{|\boldsymbol{\omega}^D| > \sigma_0\}}\Big\{\Big(\frac{\sigma_0}{|\boldsymbol{\omega}^D|^3}\boldsymbol{\omega}^D : \boldsymbol{\delta\omega}^D\Big)\boldsymbol{\omega}^D + \Big(1 - \frac{\sigma_0}{|\boldsymbol{\omega}^D|}\Big)\boldsymbol{\delta\omega}^D\Big\} \quad \forall\, \boldsymbol{\delta\omega} \in \mathbb{Q}. \tag{4.28}$$

Employing the linearity and the symmetry of the fourth-order tensor $f'(\boldsymbol{\omega}) \in \mathcal{L}(\mathbb{Q},\mathbb{Q})$ (for all $\boldsymbol{\omega} \in \mathbb{Q}$ with $|\boldsymbol{\omega}^D| \neq \sigma_0$) in (4.27), gives in turn

$$
\begin{aligned}
\boldsymbol{\lambda} &= f'(\bar{\boldsymbol{z}})(\boldsymbol{\xi})\\
&= \frac{1}{\epsilon}\chi_{\{|\bar{\boldsymbol{z}}^D| > \sigma_0\}}\Big\{\Big(\frac{\sigma_0}{|\bar{\boldsymbol{z}}^D|^3}\bar{\boldsymbol{z}}^D : \boldsymbol{\xi}^D\Big)\bar{\boldsymbol{z}}^D + \Big(1 - \frac{\sigma_0}{|\bar{\boldsymbol{z}}^D|}\Big)\boldsymbol{\xi}^D\Big\} \quad \text{a.e. where } |\bar{\boldsymbol{z}}(t,x)^D| \neq \sigma_0,
\end{aligned} \tag{4.29}
$$

where the last equality follows from (4.28).

Now, let us define $M := \{(t,x) \in (0,T) \times \Omega : |\bar{\boldsymbol{z}}(t,x)^D| = \sigma_0\}$ (up to sets of measure zero). From (4.9) and (4.27) we infer

$$\frac{1}{\epsilon}\chi_{\{\bar{\boldsymbol{z}}^D : \boldsymbol{v} > 0\}}(t,x)\Big(\frac{1}{\sigma_0^2}\bar{\boldsymbol{z}}(t,x)^D : \boldsymbol{v}\Big)\bar{\boldsymbol{z}}(t,x)^D : \boldsymbol{\xi}(t,x)^D \geq \boldsymbol{\lambda}(t,x) : \boldsymbol{v} \quad \forall\, \boldsymbol{v} \in \mathbb{Q}, \text{ a.e. in } M, \tag{4.30}$$

where we used (4.2). Here, $\chi_{\{\bar{\boldsymbol{z}}^D:\boldsymbol{v}>0\}} : (0,T)\times\Omega \to \{0,1\}$ stands for the characteristic function of the set $\{(t,x) \in (0,T) \times \Omega : \bar{\boldsymbol{z}}^D(t,x) : \boldsymbol{v} > 0\}$ for any $\boldsymbol{v} \in \mathbb{Q}$.

From (4.30) we deduce

$$\Big(\frac{1}{\epsilon\sigma_0^2}\bar{\boldsymbol{z}}(t,x)^D : \boldsymbol{v}\Big)\bar{\boldsymbol{z}}(t,x)^D : \boldsymbol{\xi}(t,x)^D \geq \boldsymbol{\lambda}(t,x) : \boldsymbol{v} \quad \forall\, \boldsymbol{v} \in \{-\bar{\boldsymbol{z}}(t,x)^D\}^\circ, \text{ a.e. in } M,$$

i.e.,

$$\boldsymbol{\lambda}(t,x) - \Big(\frac{\bar{\boldsymbol{z}}(t,x)^D : \boldsymbol{\xi}(t,x)^D}{\epsilon\sigma_0^2}\bar{\boldsymbol{z}}(t,x)^D\Big) \in \{-\bar{\boldsymbol{z}}(t,x)^D\}^{\circ\circ} \text{ a.e. in } M.$$

Therefore, by the bipolar theorem, there exists some $\alpha(t,x) \geq 0$ such that

$$\boldsymbol{\lambda}(t,x) = \Big(\frac{\bar{\boldsymbol{z}}(t,x)^D : \boldsymbol{\xi}(t,x)^D}{\epsilon\sigma_0^2} - \alpha(t,x)\Big)\bar{\boldsymbol{z}}(t,x)^D \;\text{ a.e. in } M. \tag{4.31}$$

In an analogous way, we also obtain from (4.30) the inclusion

$$\boldsymbol{\lambda}(t,x) \in \{\bar{\boldsymbol{z}}(t,x)^D\}^{\circ\circ} \text{ a.e. in } M.$$

By employing again the bipolar theorem, we find some $\theta(t,x) \geq 0$ such that

$$\boldsymbol{\lambda}(t,x) = \theta(t,x)\bar{\boldsymbol{z}}(t,x)^D \;\text{ a.e. in } M. \tag{4.32}$$

Now, relying on (4.31) and (4.32), $|\bar{\boldsymbol{z}}(t,x)^D| \neq 0$, and the fact that $\alpha, \theta \geq 0$ a.e. in M, one has

$$\theta(t,x) \in \Big[0, \frac{\bar{\boldsymbol{z}}(t,x)^D : \boldsymbol{\xi}(t,x)^D}{\epsilon\sigma_0^2}\Big] \;\text{ a.e. in } M. \tag{4.33}$$

Since (4.29), (4.32), and (4.33) are just the relations given in (4.22b), the proof is now complete. Note that $\boldsymbol{\lambda}(t,x) \in \mathbb{Q}_0$ a.e. in $(0,T)\times\Omega$, as a direct result of (4.22b). □

Remark 4.4. *If $|\bar{\boldsymbol{z}}(t,x)^D| \neq \sigma_0$ a.e. in $(0,T)\times\Omega$, then the optimality system in Theorem 4.3 reduces to the standard KKT-conditions, see* (4.29) *and Remark 2.12.*

The optimality system in Theorem 4.3 is indeed of strong stationary type, as the next result shows:

Theorem 4.5 (Equivalence between B- and strong stationarity). *Assume that $\bar{\ell} \in L^2(0,T;L^2(\Omega;\mathbb{R}^n))$ together with its states $(\bar{\boldsymbol{p}}, \bar{\boldsymbol{u}}) \in H_0^1(0,T;L^2(\Omega;\mathbb{Q}_0))\times L^2(0,T;H_D^1(\Omega;\mathbb{R}^n))$, some adjoint states $(\boldsymbol{\xi}, \boldsymbol{w}) \in H_T^1(0,T;L^2(\Omega;\mathbb{Q})) \times L^2(0,T;H_D^1(\Omega;\mathbb{R}^n))$, and a multiplier $\boldsymbol{\lambda} \in L^2(0,T;L^2(\Omega;\mathbb{Q}_0))$ satisfy the optimality system* (4.22a)–(4.22c). *Then, it also satisfies the variational inequality*

$$\partial_{(\boldsymbol{p},\boldsymbol{u})}\mathcal{J}(\bar{\boldsymbol{p}}, \bar{\boldsymbol{u}}, \bar{\ell})S'(\bar{\ell};\delta\ell) + \partial_\ell\mathcal{J}(\bar{\boldsymbol{p}}, \bar{\boldsymbol{u}}, \bar{\ell})\delta\ell \geq 0 \quad \forall\, \delta\ell \in L^2(0,T;L^2(\Omega;\mathbb{R}^n)), \tag{4.34}$$

where $S : L^2(0,T;H_D^{-1}(\Omega;\mathbb{R}^n)) \to H_0^1(0,T;L^2(\Omega;\mathbb{Q}_0)) \times L^2(0,T;H_D^1(\Omega;\mathbb{R}^n))$ is the solution mapping associated to (4.1)*, see Lemma 4.1.*

Proof. We show the result by means of Theorem 2.11. In the proof of Theorem 4.3, we have seen that the problem (Q) fits in the setting from Section 2, i.e., Assumptions 2.1 and 2.3 are satisfied for the quantities in (4.23). According to the proof of Theorem 4.3, the system (2.15) coincides with (4.26) in this particular setting, see (4.25). We also note that (2.9) is just (4.34). Thus, in view of Theorem 2.11, we only need to show that (4.22b) implies (4.26c), i.e.,

$$(\boldsymbol{\xi}(t), f'\big(\bar{\boldsymbol{z}}(t); \boldsymbol{v}\big))_{L^2(\Omega;\mathbb{Q})} \geq (\boldsymbol{\lambda}(t), \boldsymbol{v})_{L^2(\Omega;\mathbb{Q})} \quad \forall\, \boldsymbol{v} \in L^2(\Omega;\mathbb{Q}), \text{ a.e. in } (0,T), \tag{4.35}$$

where f is given by (4.23b) and $\bar{\boldsymbol{z}} := \mathbb{C}\big(\varepsilon(\bar{\boldsymbol{u}}) - \bar{\boldsymbol{p}}\big) - \kappa\bar{\boldsymbol{p}}$.

To this end, let $\boldsymbol{v} \in L^2(\Omega;\mathbb{Q})$ be arbitrary, but fixed. From the first identity in (4.22b), we know that $\boldsymbol{\lambda} = f'(\bar{\boldsymbol{z}})(\boldsymbol{\xi})$ a.e. where $|\bar{\boldsymbol{z}}(t,x)^D| \neq \sigma_0$, in view of (4.29). Thus,

$$\boldsymbol{\lambda}(t,x) : \boldsymbol{v}(x) = f'(\bar{\boldsymbol{z}}(t,x))(\boldsymbol{v}(x)) : \boldsymbol{\xi}(t,x) \quad \text{a.e. where } |\bar{\boldsymbol{z}}(t,x)^D| \neq \sigma_0, \tag{4.36}$$

where we used again the symmetry of the Gâteaux-derivative of f.

Further, we define $M^+ := \{(t,x) \in (0,T)\times\Omega : |\bar{\boldsymbol{z}}(t,x)^D| = \sigma_0 \text{ and } \bar{\boldsymbol{z}}(t,x)^D : \boldsymbol{v}(x)^D > 0\}$ and $M^- := \{(t,x) \in (0,T)\times\Omega : |\bar{\boldsymbol{z}}(t,x)^D| = \sigma_0 \text{ and } \bar{\boldsymbol{z}}(t,x)^D : \boldsymbol{v}(x)^D \leq 0\}$ (up to sets of measure zero). Then, the second identity in (4.22b) yields

$$\begin{aligned}
\boldsymbol{\lambda}(t,x) : \boldsymbol{v}(x) &= \theta(t,x)\bar{\boldsymbol{z}}(t,x)^D : \boldsymbol{v}(x)^D \quad \text{a.e. where } |\bar{\boldsymbol{z}}(t,x)^D| = \sigma_0 \\
&\leq \begin{cases} \dfrac{\bar{\boldsymbol{z}}(t,x)^D : \boldsymbol{\xi}(t,x)^D}{\epsilon\sigma_0^2}\bar{\boldsymbol{z}}(t,x)^D : \boldsymbol{v}(x)^D & \text{a.e. in } M^+ \\ 0 & \text{a.e. in } M^- \end{cases} \\
&= f'(\bar{\boldsymbol{z}}(t,x); \boldsymbol{v}(x)) : \boldsymbol{\xi}(t,x) \quad \text{a.e. where } |\bar{\boldsymbol{z}}(t,x)^D| = \sigma_0,
\end{aligned} \tag{4.37}$$

in view of (4.9) and (4.2). Now, (4.35) follows from (4.36) and (4.37). Note that, since Assumption 2.6 is fulfilled, cf. the proof of Theorem 4.3, we have the equivalence (4.34) $\Longleftrightarrow$ (4.22). ◻

Remark 4.6. *An essential information resulting from the strong stationary system* (4.22) *is the sign condition*

$$\bar{\boldsymbol{z}}(t,x)^D : \boldsymbol{\xi}(t,x)^D \geq 0 \quad \textit{a.e. where } |\bar{\boldsymbol{z}}(t,x)^D| = \sigma_0, \tag{4.38}$$

which is due to $\Big[0, \frac{\bar{\boldsymbol{z}}(t,x)^D : \boldsymbol{\xi}(t,x)^D}{\epsilon\sigma_0^2}\Big] \neq \emptyset$ *a.e. where* $|\bar{\boldsymbol{z}}(t,x)^D| = \sigma_0$*, see* (4.22b)*. This is crucial for showing the implication* (4.22) $\Rightarrow$ (4.34)*, which ultimately yields that* (4.22) *is indeed of strong stationary type (see* (4.37) *in the proof of Theorem 4.5). Let us point out that the condition* (4.38) *is equivalent to the regularity of a mapping involving the adjoint state and the non-smooth nonlinearity, see* (4.41) *below. We refer here to a similar situation [26, Rem. 6.9].*

Remark 4.7. *Optimality systems derived by classical regularization techniques often lack a sign condition for the adjoint state, see e.g. [6, Thm. 4.4] and [36, Thm. 2.4] (non-smooth PDEs) (eventually along with other information which gets lost in the limit analysis associated with the regularization, cf. [26, Sec. 4]). This is also the case when it comes to the optimal control of VIs, see [28]. Generally speaking, a sign condition for the adjoint state in points (t,x) where the argument of the non-smoothness f in the state equation, say $\bar{s}$, is such that f is not differentiable at $\bar{s}(t,x)$ is what ultimately distiguishes a strong stationary optimality system from very 'good' optimality systems obtained via regularization, cf. [6, Thm. 4.4, Thm. 4.12] and [26, Sec. 7.2]. Note that, in our case, the argument of the non-linearity f appearing in the state equation is $\bar{\boldsymbol{z}}$, cf.* (4.5)*.*

In the remaining of the section, we rewrite (4.22b) in terms of a Clarke subdifferential and we explain how the sign condition (4.38) is related to the notion of regular functions, cf. [32, Def. 7.4.1]. Furthermore, in order to get a better understanding of (4.22b), we take a look at the one-dimensional case. This will highlight the similarities between (4.22) and other strong stationary optimality systems, which were derived for optimal control of a single non-smooth PDE [6, 26].

In the sequel, $g : \mathbb{R} \to \mathbb{R}$ is a mapping defined as

$$g(z) := \frac{1}{\epsilon}\chi_{\{|z|>\sigma_0\}}\Big(1 - \frac{\sigma_0}{|z|}\Big)z$$

and $\partial_\circ g$ denotes its Clarke subdifferential in the sense of [32, Def. 7.3.4]. Note that, $g : \mathbb{R} \to \mathbb{R}$ is the one-dimensional version of $f : \mathbb{Q} \to \mathbb{Q}$, cf. (4.7).

Since g is piecewise continuously differentiable, it holds

$$\partial_\circ g(z) = \big[\min\{g'_-(z), g'_+(z)\}, \max\{g'_-(z), g'_+(z)\}\big] \quad \forall z \in \mathbb{R},$$

by [32, Thm. 7.3.12], see also [26, Eq. (C.3)], where $g'_+(z) := g'(z; 1)$ and $g'_-(z) := -g'(z; -1)$ denote the right- and left-sided derivative of g at $z \in \mathbb{R}$, respectively.

It is then straight-forward to see that

$$\partial_\circ g(z) = \begin{cases} \{\frac{1}{\epsilon}\} & \text{if } |z| > \sigma_0 \\ [0, \frac{1}{\epsilon}] & \text{if } |z| = \sigma_0 \\ \{0\} & \text{if } |z| < \sigma_0 \end{cases} \qquad \forall\, z \in \mathbb{R}. \tag{4.39}$$

Thus, (4.22b) can be equivalently written by means of the Clarke subdifferential of g as follows:

$$\left.\begin{aligned} &\boldsymbol{\lambda} = \Big(\frac{\sigma_0\gamma}{|\bar{\boldsymbol{z}}^D|^3}\bar{\boldsymbol{z}}^D : \boldsymbol{\xi}^D\Big)\bar{\boldsymbol{z}}^D + \frac{1}{\epsilon}\chi_{\{|\bar{\boldsymbol{z}}^D|>\sigma_0\}}\Big(1 - \frac{\sigma_0}{|\bar{\boldsymbol{z}}^D|}\Big)\boldsymbol{\xi}^D, \text{ with} \\ &\gamma(t,x) \in \partial_\circ g(|\bar{\boldsymbol{z}}(t,x)^D|) \quad \text{a.e. in } (0,T)\times\Omega, \\ &\bar{\boldsymbol{z}}(t,x)^D : \boldsymbol{\xi}(t,x)^D \geq 0 \quad \text{a.e. where } |\bar{\boldsymbol{z}}(t,x)^D| = \sigma_0, \end{aligned}\right\} \tag{4.40}$$

where we make the convention $\frac{0}{0} = 0$.

The additional information which one obtains via the direct method introduced in [26], and which gets lost when passing to the limit in the classical regularization schemes is the sign condition

$$\bar{\boldsymbol{z}}(t,x)^D : \boldsymbol{\xi}(t,x)^D \geq 0 \quad \text{a.e. where } |\bar{\boldsymbol{z}}(t,x)^D| = \sigma_0,$$

or equivalently, the information that the mapping

$$\mathbb{R} \ni z \mapsto \bar{\boldsymbol{z}}(t,x)^D : \boldsymbol{\xi}(t,x)^D g(z) \in \mathbb{R} \tag{4.41}$$

is regular at $|\bar{\boldsymbol{z}}(t,x)^D|$ f.a.a. $(t,x) \in (0,T)\times\Omega$. This equivalence is due to [26, Lem. C.1] combined with $g'_+(\sigma_0) > g'_-(\sigma_0)$ and $g'_+(z) = g'_-(z)$ for all $z \in \mathbb{R}^+ \setminus \{\sigma_0\}$. The latter follows from the definition of g by straight-forward computation. Thus, similarly to [26], our strong stationary optimality conditions imply the regularity of a mapping involving the adjoint state and the non-smoothness.

To get a better understanding of (4.40), we take a look at its one-dimensional version. Recall that $g = f$, if $n = 1$, cf. (4.7). Then, the system (4.40) (and thus, (4.22b)) reads

$$\left.\begin{aligned} &\lambda(t,x) = \gamma(t,x)\xi(t,x) \quad \text{a.e. in } (0,T)\times\Omega,\\ &\gamma(t,x) \in \partial_\circ f(\bar z(t,x)) \quad \text{a.e. in } (0,T)\times\Omega,\\ &\left.\begin{aligned} &\xi(t,x) \geq 0 \quad \text{a.e. where } \bar z(t,x) = \sigma_0,\\ &\xi(t,x) \leq 0 \quad \text{a.e. where } \bar z(t,x) = -\sigma_0, \end{aligned}\right\} \end{aligned}\right\} \tag{4.42}$$

where we used (4.39) and $\partial_\circ g(z) = \partial_\circ g(|z|)$ for all $z \in \mathbb{R}$.

Note that the sign condition in (4.42) is equivalent to the fact that the map

$$\mathbb{R} \ni z \mapsto \xi(t,x)g(z) \in \mathbb{R}$$

is regular at $\bar z(t,x)$ f.a.a. $(t,x) \in (0,T)\times\Omega$, which is again a consequence of [26, Lem. C.1] combined with $g'_+(\sigma_0) > g'_-(\sigma_0)$, $g'_+(-\sigma_0) < g'_-(-\sigma_0)$ and $g'_+(z) = g'_-(z)$ for all $z \in \mathbb{R}\setminus\{\sigma_0, -\sigma_0\}$.

Let us shortly compare the one-dimensional version (4.22a)-(4.42)-(4.22c) with the strong stationary optimality system [6, (32)]. First, let us point out that in [6], the state equation is a non-smooth elliptic PDE with non-linearity $f = \max\{0,\cdot\}$. If we insert the relation $\lambda = \gamma\xi$ in our adjoint equation (4.22a), then our strong stationarity conditions can be written in terms of the adjoint states and γ, instead of λ. To be more precise, it consists of the same adjoint equation (involving w, ξ and γ), the gradient equation (4.22c), and

$$\left.\begin{aligned} &\gamma(t,x) \in \partial_\circ f(\bar z(t,x)) \quad \text{a.e. in } (0,T)\times\Omega,\\ &\xi(t,x) \geq 0 \text{ a.e. where } \bar z(t,x) = \sigma_0,\\ &\xi(t,x) \leq 0 \text{ a.e. where } \bar z(t,x) = -\sigma_0. \end{aligned}\right\} \tag{4.43}$$

We remark that (4.43) resembles [6, (32b)-(32c)]. Similarly to [6, (32)], our optimality system contains - besides an adjoint equation and a gradient equation - a differential inclusion in terms of the Clarke subdifferential of the non-smooth mapping and a sign condition on the adjoint state in points where the argument of the non-smoothness f, i.e., $\bar z$ (see (4.5)), is such that f is not differentiable there (notice that, in [6], the argument of f is the state associated to the local optimum and f is not differentiable at 0). The optimality system (4.22a)-(4.42)-(4.22c) is also consistent with [26, (6.8)]. In view of [26, Rem. 6.9], the relation [26, (6.8b)] can be expressed in a similar way as (4.42).

Appendix A. Directional differentiability of $\mathcal{F}$ (Section 3). In this section, Assumption 3.1.1-2 is supposed to hold, while $\mathcal{F}$ and the projection operator are given by Definition 3.2.

Lemma A.1. *Let $\bar\omega \in Y^*$ be fixed. Assume that the set $\partial R(0) \subset Y^*$ is polyhedric at $P_{\partial R(0)}\bar\omega$ w.r.t. $\mathcal{F}(\bar\omega)$, i.e.,*

$$\overline{\mathcal{C}(\bar\omega)} \cap [\mathcal{F}(\bar\omega)]^\perp = \overline{\mathcal{C}(\bar\omega) \cap [\mathcal{F}(\bar\omega)]^\perp}, \tag{A.1}$$

where $\mathcal{C}(\bar\omega) := \mathbb{R}^+(\partial R(0) - P_{\partial R(0)}\bar\omega)$ and $[\mathcal{F}(\bar\omega)]^\perp := \{\mu \in Y^* : \langle \mu, \mathcal{F}(\bar\omega)\rangle_Y = 0\}$. Then, $\mathcal{F} : Y^* \to Y$ is directionally differentiable at $\bar\omega$ with

$$\mathcal{F}'(\bar\omega;\delta\omega) = \mathcal{V}^{-1}(\delta\omega - P_{T(\bar\omega)}\delta\omega) \quad \forall\, \delta\omega \in Y^*, \tag{A.2}$$

where $T(\bar\omega) := \overline{\mathcal{C}(\bar\omega)} \cap [\mathcal{F}(\bar\omega)]^\perp$ and $P_{T(\bar\omega)}\delta\omega$ is the unique solution of $\min_{\mu\in T(\bar\omega)} \|\delta\omega - \mu\|_{\mathcal{V}^{-1}}$. Moreover, $P_{\partial R(0)} : Y^* \to Y^*$ is directionally differentiable at $\bar\omega$ as well, with

$$P'_{\partial R(0)}(\bar\omega;\delta\omega) = P_{T(\bar\omega)}\delta\omega \quad \forall\, \delta\omega \in Y^*.$$

Proof. According to Lemma 3.3, the mapping $\mathcal{F}$ is the solution operator of (3.3), and by applying [8, Thm. 2.3], we get that $\mathcal{F} : Y^* \to Y$ is directionally differentiable at $\bar\omega$. Moreover, $\mathcal{F}'(\bar\omega;\cdot) : Y^* \ni \delta\omega \mapsto \delta z \in Y$ is the solution operator of the following elliptic VI

$$I_{T(\bar\omega)^\circ}(\eta) - I_{T(\bar\omega)^\circ}(\delta z) + \langle \mathcal{V}\delta z, \eta - \delta z\rangle_Y \geq \langle \delta\omega, \eta - \delta z\rangle_Y \quad \forall\, \eta \in Y.$$

However, this is again a VI of the type (3.3), since $I_{T(\bar\omega)^\circ}$ satisfies Assumption 3.1.1 (as $T(\bar\omega)^\circ \subset Y$ is a non-empty, closed, convex cone). Thus, by Lemma 3.3 combined with $\partial I_{T(\bar\omega)^\circ}(0) = T(\bar\omega)^{\circ\circ} = T(\bar\omega)$, we have $\mathcal{F}'(\bar\omega;\delta\omega) = \delta z = \mathcal{V}^{-1}(\delta\omega - P_{T(\bar\omega)}\delta\omega)$ for all $\delta\omega \in Y^*$. Notice that here we also used the fact that $T(\bar\omega) \subset Y^*$ is a non-empty, closed, convex cone. Thanks to Theorem 3.4, the proof is now complete. Alternatively, the assertion of this lemma can be deduced from [15, Thm. 2] and Theorem 3.4. ☐

Some canonical examples. For details regarding polyhedric sets and their properties, we refer to the contributions [3, 39]. Let us give some examples of functionals R which are often encountered in applications and for which the polyhedricity of

$$\partial R(0) = \{\mu \in Y^* |\, \langle \mu, v\rangle_Y \leq R(v)\ \forall\, v \in Y\}$$

is guaranteed. We say that $\partial R(0) \subset Y^*$ is polyhedric at $\mu \in \partial R(0)$, if it is polyhedric at $\mu \in \partial R(0)$ w.r.t. any $\eta \in \overline{\mathbb{R}^+(\partial R(0) - \mu)}^\circ$, see [39, Def. 3.1.1]. We also say that $\partial R(0) \subset Y^*$ is polyhedric, if it is polyhedric at any $\mu \in \partial R(0)$.

In the sequel, $r \geq 0$ is fixed and $\Omega \subset \mathbb{R}^n$, $n \in \{2, 3\}$, is a bounded Lipschitz domain.

Example A.1. *[Dissipation functional for damage processes, cf. [25, Sec. 4] and [35]] If $Y = L^2(\Omega)$ or $Y = H^1_0(\Omega)$ and*

$$R(\eta) = \begin{cases} r\int_\Omega \eta\, dx, & \text{if } \eta \geq 0 \text{ a.e. in } \Omega \\ \infty, & \text{otherwise} \end{cases} \tag{A.3}$$

for all $\eta \in Y$, then $\mathcal{F} : Y^* \to Y$ is directionally differentiable, as we will next see.

If $Y = L^2(\Omega)$, then

$$\partial R(0) = \{\mu \in L^2(\Omega) |\, \mu \leq r \text{ a.e. in } \Omega\}$$

and [3, Prop. 6.33] gives the polyhedricity of $\partial R(0) \subset Y^*$. The directional differentiability of $\mathcal{F}$ follows by Lemma A.1.

Let now $Y = H^1_0(\Omega)$. Then,

$$\partial R(0) = \{\mu \in H^1_0(\Omega)^* |\, \langle \mu - r, v\rangle_{H^1_0(\Omega)} \leq 0\ \forall\, v \in H^1_0(\Omega) \text{ with } v \geq 0 \text{ a.e. in } \Omega\} = M^\circ + \{r\},$$

where $M := \{v \in H^1_0(\Omega) |\, v \geq 0 \text{ a.e. in } \Omega\}$. It is well-known that $M \subset Y$ is polyhedric (see e.g. [15, 27]), and by [39, Lem. 3.2], we have that $\partial R(0)$ is polyhedric at any

$\mu \in M^\circ \cap [\eta]^\perp + \{r\}$ w.r.t. any $\eta \in M$. Here we used the fact that $M^\circ + \{r\}$ is polyhedric at $\zeta \in M^\circ + \{r\}$ if and only if M° is polyhedric at $\zeta - r \in M^\circ$. Let now $\bar\omega \in Y^*$ be fixed. From Lemma 3.3 we deduce that $\mathcal{F}(\bar\omega) \in dom(R) = M$. By testing (3.3) with 0 and $\mathcal{F}(\bar\omega)$, respectively, we get

$$\langle P_{\partial R(0)}\bar\omega - r, \mathcal{F}(\bar\omega)\rangle_Y = 0. \tag{A.4}$$

This yields $P_{\partial R(0)}\bar\omega \in M^\circ \cap [\mathcal{F}(\bar\omega)]^\perp + \{r\}$, where we used $\partial R(0) = M^\circ + \{r\}$. Now we can apply Lemma A.1, which tells us that $\mathcal{F} : Y^* \to Y$ is directionally differentiable.

EXAMPLE A.2 (Dissipation functional for sweeping processes [13] and plasticity [20]). *If $Y = L^2(\Omega)$ and*

$$R(\eta) = r \int_\Omega |\eta|\, dx \quad \forall\, \eta \in Y, \tag{A.5}$$

then $\partial R(0) \subset Y^*$ is polyhedric and therefore, $\mathcal{F} : Y^* \to Y$ is directionally differentiable, by Lemma A.1. This is due to

$$\partial R(0) = \{\mu \in L^2(\Omega) |\ -r \le \mu \le r \quad a.e.\ in\ \Omega\}$$

and [3, Prop. 6.33]. The case $Y = H_0^1(\Omega)$ is more delicate, since $\partial R(0) \subset Y^*$ is not polyhedric, cf. [8, Cor. 3.3]. However, the paper [7] provides conditions that guarantee the directional differentiability of the solution operator of the classical elliptic VI of the second kind (i.e., with $Y = H_0^1(\Omega)$, R as in (A.5), and $\mathcal{V} := -\triangle$). Such conditions ensure the directional differentiability of $\mathcal{F} : Y^* \to Y$, in view of Lemma 3.3.

Acknowledgment. This work was supported by the DFG under the grant YO 159/2-1 within the priority programme SPP 1962 'Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization'.

REFERENCES

[1] V. Barbu. Necessary conditions for distributed control problems governed by parabolic variational inequalities. *SIAM Journal on Control and Optimization*, 19(1):64–86, 1981.
[2] 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.
[3] J.F. Bonnans and A. Shapiro. *Pertubation Analysis of Optimization Problems.* Springer, New York, 2000.
[4] H. Cartan. *Differential Calculus.* Kershaw Publishing Company, London, 1971.
[5] C. Christof. Sensitivity analysis and optimal control of obstacle-type evolution variational inequalities. Preprint SPP1962-055, 2018.
[6] C. Christof, C. Clason, C. Meyer, and S. Walther. Optimal control of a non-smooth, semilinear elliptic equation. *Mathematical Control and Related Fields*, 8(1):247–276, 2018.
[7] C. Christof and C. Meyer. Differentiability properties of the solution operator to an elliptic variational inequality of the second kind. Technical Report 527, Ergebnisberichte des Instituts für Angewandte Mathematik, TU Dortmund, 2016.
[8] C. Christof and G Wachsmuth. On the non-polyhedricity of sets with upper and lower bounds in dual spaces. *GAMM Reports*, 40(4):339–350, 2018.
[9] J. C. De los Reyes and C. Meyer. Strong stationarity conditions for a class of optimization problems governed by variational inequalities of the second kind. *Journal of Optimization Theory and Applications*, 168(2):375–409, 2016.
[10] E. Emmrich. *Gewöhnliche und Operator Differentialgleichungen.* Vieweg, Wiesbaden, 2004.
[11] 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.

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

[13] T. Geiger and D. Wachsmuth. Optimal control of an evolution equation with non-smooth dissipation. arXiv:1801.04077, 2018.

[14] R. Glowinski. *Numerical Methods for Nonlinear Variational Problems.* Springer, Berlin Heidelberg, 2008.

[15] A. Haraux. How to differentiate the projection on a convex set in Hilbert space. Some applications to variational inequalities. *Journal of the Mathematical Society of Japan*, 29(4):615–631, 1977.

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

[17] R. Herzog and C. Meyer. Optimal control of static plasticity with linear kinematic hardening. *J. Appl. Math. Mech.*, 91(10):777–794, 2011.

[18] R. Herzog, C. Meyer, and G. Wachsmuth. C-stationarity for optimal control of static plasticity with linear kinematic hardening. *SIAM J. Optim.*, 50(5):3052–3082, 2012.

[19] 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.

[20] R. Herzog and A. Stötzner. Hadamard differentiability of the solution map in thermoviscoplasticity. *Pure and Applied Functional Analysis (PAFA) (to appear)*, 2017.

[21] 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.

[22] 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.

[23] A. D. Ioffe and V. M. Tihomirov. *Theory of Extremal Problems.* North-Holland Publishing Company, Amsterdam, 1979.

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

[25] D. Knees, R. Rossi, and C. Zanini. A vanishing viscosity approach to a rate-independent damage model. Technical Report 1633, WIAS Berlin, 2011, 2011.

[26] C. Meyer and L.M. Susu. Optimal control of nonsmooth, semilinear parabolic equations. *SIAM Journal on Control and Optimization*, 55(4):2206–2234, 2017.

[27] F Mignot. Contrôle dans les inéquations variationelles elliptiques. *Journal of Functional Analysis*, 22(2):130–185, 1976.

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

[29] N.S. Ottosen and M. Ristinmaa. *The Mechanics of Constitutive Modeling.* Elsevier, 2005.

[30] R.T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, 1970.

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

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

[33] J.C. Simo and T.J.R. Hughes. *Computational Inelasticity.* Springer, New York, 1998.

[34] M. Sofonea and M. Andaluzia. *Variational Inequalities with Applications.* Springer, New York, 2009.

[35] L.M. Susu. Optimal control of a viscous two-field gradient damage model. *GAMM-Mitt.*, 40(4):287 – 311, 2018.

[36] D. Tiba. *Optimal control of nonsmooth distributed parameter systems.* Springer, 1990.

[37] 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.

[38] G. Wachsmuth. Strong stationarity for optimal control of the obstacle problem with control constraints. *SIAM Journal on Optimization*, 24(4):1914–1932, 2014.

[39] G. Wachsmuth. A guided tour of polyhedric sets. *Journal of Convex Analysis*, 26(1), 2019.