SIAM J. OPTIM.
Vol. 29, No. 4, pp. 3069–3099

STRONG STATIONARITY FOR OPTIMAL CONTROL OF A NONSMOOTH COUPLED SYSTEM: APPLICATION TO A VISCOUS EVOLUTIONARY VARIATIONAL INEQUALITY COUPLED WITH AN ELLIPTIC PDE*

LIVIA M. BETZ†

Abstract. This paper is mainly concerned with an optimal control problem governed by a nonsmooth coupled system of equations. The nonsmooth 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. The abstract result is then applied to prove strong stationarity for optimal control of a coupled system consisting of a viscous evolutionary variational inequality (EVI) and an elliptic PDE. To this end, we show that EVIs with viscosity can be formulated as nonsmooth ODEs in Hilbert space in a general setting. The nonsmooth nonlinearity appearing in the ODE turns out to be the solution operator of an elliptic variational inequality, for which we can give an explicit formula.

Key words. optimal control of coupled systems, nonsmooth optimization, strong stationarity, evolutionary variational inequalities with viscosity, optimal control of PDEs, viscous damage evolution

AMS subject classifications. 34G25, 34K35, 49J20, 49J27

DOI. 10.1137/18M1216778

1. Introduction.

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

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

where J is a smooth objective, f is a nonsmooth mapping, while Φ and Ψ are smooth nonlinearities. The precise assumptions on the data are stated in Assumption 2.1 below. The essential feature of the problem under consideration is that the nonlinearity 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 nonsmooth structure in $(*)$. This gives rise to additional challenges. We deal with two state variables (y,u) which depend nonlinearly on each other. Moreover, the nonsmooth mapping does not act directly on either one of the states, but on a nonlinear coupling involving them. In this context, an interesting application is the optimal control of viscous evolutionary variational inequalities (EVIs) coupled with elliptic PDEs, since, as we will see, the evolution of viscous processes is described by

*Received by the editors September 26, 2018; accepted for publication (in revised form) August 26, 2019; published electronically December 12, 2019.
https://doi.org/10.1137/18M1216778
Funding: This work was supported by the DFG under grant YO 159/2-1 within the priority programme SPP 1962, "Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization."
†Faculty of Mathematics, University of Duisburg-Essen, Essen, Germany 45127 (livia.betz@uni-due.de).

nonsmooth ODEs such as the one in ($*$). 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 special attention is given to mathematical programs with complementarity constraints. In [39] a detailed overview of various optimality conditions of different strengths was introduced; see also [25] for the infinite-dimensional 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., [47] and the references therein), fewer papers deal with nonsmooth problems. Most of these papers resort to regularization or relaxation techniques to smooth the problem; see, e.g., [1, 2, 13, 14, 20, 22, 27, 29] 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 nonsmooth problems requires direct approaches, which employ the limited differentiability properties of the control-to-state map. In this context, there are even fewer contributions. Based on the pioneering work [34] (strong stationarity for optimal control of elliptic variational inequalities (VIs) of obstacle type), most of them focus on elliptic VIs [10, 23, 26, 35, 48]. Recently, strong stationarity for optimal control of parabolic VIs of the first kind was proven in [5]. Regarding strong stationarity for optimal control of nonsmooth PDEs, the literature is very scarce and the only papers known to the author addressing this issue so far are [31] (parabolic PDE), [6] (elliptic PDE), and the more recent [9] (elliptic quasi-linear PDE). We point out that, in contrast to our problem, all the above-mentioned contributions which investigate strong stationarity deal with only one state, except [23].

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 nonsmooth nonlinearity. 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 [31], which is to employ a "surjectivity" trick. As it turns out, it all comes down to the following aspect: the set of directions into which the nonsmooth mapping f is differentiated—in the "linearized" state equation associated with the local optimum—must be dense in a suitable space (see Lemma 2.8 and Remark 2.12).

In section 3, we show that viscous EVIs are nonsmooth ODEs in Hilbert space. It is indeed known (in particular, in the context of Lebesgue spaces [24, 45]) 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 solution operator is nothing else than the nonsmooth nonlinearity 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 nonsmooth 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 a viscous two-field gradient damage model. 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 nonsmooth coupled PDE system of the type $(*)$; see (4.4) below. It consists of a nonsmooth ODE in Hilbert space (viscous damage evolution) and an elliptic PDE. 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 damage model. For convenience of the reader, we included an essential proof regarding the improved regularity of the multiplier appearing in the adjoint system in Appendix B.

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 densely embedded 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 closed ball in X around $x \in X$ with radius $\alpha > 0$ is denoted by $B_X(x, \alpha)$. If X is a Hilbert space, we write $(\cdot, \cdot)_X$ for the associated scalar product. 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 with 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., [37]. With a little abuse of notation, the Nemystkii operators associated with the mappings considered in this paper will be described by the same symbol, even when considered with different domains and ranges. By $\max(\cdot, 0)$ we denote the positive part function, while $\max'(x; h)$ indicates its directional derivative in the point x in direction h. Similarly, $\min(\cdot, 0)$ stands for the negative part function.

2. Strong stationarity for optimal control of nonsmooth 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 [31] 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 into which the nonsmooth mapping f is differentiated in the linearized state equation (associated with a local minimum); 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 constraint qualification is satisfied by the optimal control of the viscous two-field damage model considered in section 4 below.

Our optimal control problem reads as follows:

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

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 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
(2.1)
$$\|\Phi'(y,u)\|_{\mathcal{L}(Y\times U;Y^*)} \le L, \quad \|\Psi'(y,u)\|_{\mathcal{L}(Y\times U;U^*)} \le L \quad \forall\,(y,u) \in Y \times U.$$
3. The nonsmooth function $f : Y^* \to Y$ is assumed to be Lipschitz continuous and directionally differentiable, i.e.,
$$\Big\|\frac{f(x+\tau\, h) - f(x)}{\tau} - f'(x;h)\Big\|_Y \overset{\tau \searrow 0}{\longrightarrow} 0 \quad \forall\, x, h \in Y^*. \tag{2.2}$$
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 with 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 2.1.3 combined with Lebesgue's dominated convergence theorem. In view of the latter and Assumption 2.1.2, we also deduce that the Nemytskii operators

$$\Phi : L^2(0,T;Y) \times L^2(0,T;U) \to L^2(0,T;Y^*),$$
$$\Psi : L^2(0,T;Y) \times L^2(0,T;U) \to L^2(0,T;U^*)$$

are Gâteaux differentiable, with

$$\partial_y\Phi(y,u)(\delta y) = \partial_y\Phi(y(\cdot),u(\cdot))(\delta y(\cdot)), \quad \partial_u\Phi(y,u)(\delta u) = \partial_u\Phi(y(\cdot),u(\cdot))(\delta u(\cdot)),$$
$$\partial_y\Psi(y,u)(\delta y) = \partial_y\Psi(y(\cdot),u(\cdot))(\delta y(\cdot)), \quad \partial_u\Psi(y,u)(\delta u) = \partial_u\Psi(y(\cdot),u(\cdot))(\delta u(\cdot))$$

for all $(y,u), (\delta y, \delta u) \in L^2(0,T;Y \times U)$. In the proof of Theorem 2.11 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*

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

for 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 demonstration concise, we do not discuss the unique solvability of the state equation nor 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 2.11) 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. The mapping $S : L^2(0,T;V) \to L^2(0,T;Y) \times L^2(0,T;U)$ is directionally differentiable, i.e.,

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

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

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

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

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

3. For any $\ell \in L^2(0,T;V)$, there exists a constant $K > 0$ so that

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

   If $(\hat{\delta y}, \hat{\delta u}) \in H_0^1(0,T;Y) \times L^2(0,T;U)$ solves (2.5) with right-hand side (r.h.s.) $\hat{\delta\ell} \in L^2(0,T;U^*)$ and if there exists a sequence $\{\delta\ell_n\}_n \subset L^2(0,T;V)$ with $\delta\ell_n \to \hat{\delta\ell}$ in $L^2(0,T;U^*)$, then $S'(\ell;\delta\ell_n) \to (\hat{\delta y}, \hat{\delta u})$ in $L^2(0,T;Y\times U)$.

Remark 2.4. Assumption 2.3.3 is, for instance, guaranteed if for any $\ell \in L^2(0,T;V)$ there exist constants $\alpha > 0$ and $K > 0$ so that for all $\ell_1, \ell_2 \in B_{L^2(0,T;V)}(\ell,\alpha)$ 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

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

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

In particular, this means that Assumption 2.3.3 is satisfied when $V = U^*$ and $S : L^2(0,T;V) \to L^2(0,T;Y\times U)$ is locally Lipschitz continuous.

Note that the estimate (2.6) is needed only in the first part of the proof of Theorem 2.11 below and can thus be dropped if, e.g., $\Psi'(y(\cdot),u(\cdot)) \in \mathcal{L}(Y\times U;V)$ a.e. in $(0,T)$ or if J is partially Frechet differentiable w.r.t. ℓ on $L^2(0,T;U^*)$.

Finally, let us point out that in the upcoming analysis we will employ Assumptions 2.3.2–2.3.3 just for local minimizers of (P).

In the following, 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 [23, 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 (constraint qualification). For any local optimum $\bar\ell$ of (P), we assume that $\mathrm{Rg}(\partial_u \Phi(\bar y,\bar u)) \overset{d}{\hookrightarrow} L^2(0,T;Y^*)$, 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 nonsmooth mapping f. Roughly speaking, it is the price to pay for having two states on which the nonsmoothness acts. We emphasize that the claim concerning the local minimizer in Assumption 2.6 is essential for deriving the strong stationary optimality system (2.19) below and it thus plays the role of a constraint qualification; cf., e.g., [47, sect. 6]. This terminology has its roots in finite-dimensional nonlinear optimization, where it describes a condition for the (unknown) local optimizer which guarantees the existence of Lagrange multipliers such that a KKT-system is satisfied; see, e.g., [15, sect. 2]. In the nonsmooth case, the KKT conditions correspond to the strong stationary optimality conditions; see Remark 2.14 below. For these reasons, we sometimes refer to the statement in Assumption 2.6 as constraint qualification in the following.

(ii) Furthermore, let us point out that Assumption 2.6 is satisfied at any ℓ by the example considered in section 4; see (4.21) below.

The next result is crucial for proving the strong stationarity result in Theorem 2.11.

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). We observe that $\hat y$ fulfills

$$\frac{d}{dt}\hat y(t) = f'\big(\Phi(\bar y(t),\bar u(t)); \partial_y \Phi(\bar y(t),\bar u(t))\hat y(t) + \rho(t) - \partial_y \Phi(\bar y(t),\bar u(t))\hat y(t)\big) \quad \text{a.e. in } (0,T). \tag{2.11}$$

In view of Assumption 2.6, there exists a sequence $\{\hat u_m\}_m \subset L^2(0,T;U)$ such that

$$\partial_u \Phi(\bar y,\bar u)\hat u_m \to \rho - \partial_y \Phi(\bar y,\bar u)\hat y \quad \text{in } L^2(0,T;Y^*) \;\text{ as } m\to\infty. \tag{2.12}$$

For any $m \in \mathbb{N}$, consider the equation

$$\frac{d}{dt}\hat{y}_m(t) = f'\big(\Phi(\bar{y}(t),\bar{u}(t)); \Phi'(\bar{y}(t),\bar{u}(t))(\hat{y}_m(t),\hat{u}_m(t))\big) \quad \text{a.e. in } (0,T),\ \hat{y}_m(0) = 0. \tag{2.13}$$

Due to Assumption 2.1.3, the mapping $f'(x;\cdot) : L^2(0,T;Y^*) \to L^2(0,T;Y)$ is Lipschitz continuous for any $x \in L^2(0,T;Y^*)$ with the same Lipschitz constant as f; this follows from the definition of the directional derivative and the Lipschitz continuity of f. In view of the assumptions on Φ (see Assumption 2.1.2), we can employ a classical contraction argument (cf., e.g., [12, Thm. 7.2.3]), to show that (2.13) admits a unique solution $\hat{y}_m \in H^1(0,T;Y)$. Now, we define

$$\hat{\ell}_m := \Psi'(\bar{y}(\cdot),\bar{u}(\cdot))(\hat{y}_m(\cdot),\hat{u}_m(\cdot)) \in L^2(0,T;U^*) \tag{2.14}$$

such that the pair $(\hat{y}_m,\hat{u}_m) \in H_0^1(0,T;Y) \times L^2(0,T;U)$ solves the system (2.5) associated with $\bar{\ell}$ with r.h.s. $\hat{\ell}_m \in L^2(0,T;U^*)$, i.e.,

$$\begin{aligned} \frac{d}{dt}\hat{y}_m(t) &= f'\big(\Phi(\bar{y}(t),\bar{u}(t)); \Phi'(\bar{y}(t),\bar{u}(t))(\hat{y}_m(t),\hat{u}_m(t))\big) \quad \text{a.e. in } (0,T),\ \hat{y}_m(0) = 0, \\ &\Psi'(\bar{y}(t),\bar{u}(t))(\hat{y}_m(t),\hat{u}_m(t)) = \hat{\ell}_m(t) \quad \text{a.e. in } (0,T). \end{aligned} \tag{2.15}$$

Owing to the Lipschitz continuity of the directional derivative of f (w.r.t. direction) and (2.1) combined with Gronwall's inequality, we further obtain from (2.11) and (2.13)

$$\|\hat{y}_m - \hat{y}\|_{H^1(0,T;Y)} \le c\,\|\partial_u\Phi(\bar{y},\bar{u})\hat{u}_m - (\rho - \partial_y\Phi(\bar{y},\bar{u})\hat{y})\|_{L^2(0,T;Y^*)} \to 0 \quad \text{as } m \to \infty, \tag{2.16}$$

where $c > 0$ is a constant dependent only on the given data. Note that the convergence in (2.16) is due to (2.12). By relying on the Gâteaux differentiability of Φ, we have

$$\partial_y\Phi(\bar{y},\bar{u})\hat{y}_m \to \partial_y\Phi(\bar{y},\bar{u})\hat{y} \quad \text{in } L^2(0,T;Y^*) \ \text{ as } m \to \infty \tag{2.17}$$

as a result of (2.16). Combining (2.12) and (2.17) gives in turn the convergence

$$\Phi'(\bar{y},\bar{u})(\hat{y}_m,\hat{u}_m) \to \rho \quad \text{in } L^2(0,T;Y^*) \ \text{ as } m \to \infty. \tag{2.18}$$

Moreover, thanks to $V \overset{d}{\hookrightarrow} U^*$ (cf. Assumption 2.1.1), it holds that $L^2(0,T;V) \overset{d}{\hookrightarrow} L^2(0,T;U^*)$; see [31, Lem. A.1]. Thus, for any $m \in \mathbb{N}$ there exists a sequence $\{\delta\ell_n^m\}_n \subset L^2(0,T;V)$ so that $\delta\ell_n^m \to \hat{\ell}_m$ in $L^2(0,T;U^*)$ as $n \to \infty$. Since we established above that $(\hat{y}_m,\hat{u}_m) \in H_0^1(0,T;Y) \times L^2(0,T;U)$ solves (2.5) with r.h.s. $\hat{\ell}_m \in L^2(0,T;U^*)$, we can apply Assumption 2.3.3, which tells us that $S'(\bar{\ell};\delta\ell_n^m) \to (\hat{y}_m,\hat{u}_m)$ in $L^2(0,T;Y \times U)$ as $n \to \infty$ for any $m \in \mathbb{N}$. Hence, by the continuity of $\Phi'(\bar{y},\bar{u}) : L^2(0,T;Y \times U) \to L^2(0,T;Y^*)$, we have the convergence

$$\Phi'(\bar{y},\bar{u})S'(\bar{\ell};\delta\ell_n^m) \to \Phi'(\bar{y},\bar{u})(\hat{y}_m,\hat{u}_m) \quad \text{in } L^2(0,T;Y^*) \ \text{ as } n \to \infty$$

for any $m \in \mathbb{N}$. In view of (2.18), we can finally construct a sequence $\{\delta\ell_{n(m)}^m\}_m \subset L^2(0,T;V)$ such that

$$\Phi'(\bar{y},\bar{u})S'(\bar{\ell};\delta\ell_{n(m)}^m) \to \rho \quad \text{in } L^2(0,T;Y^*) \ \text{ as } m \to \infty.$$

Since $\rho \in L^2(0,T;Y^*)$ was arbitrary, the proof is now complete. $\square$

Assumption 2.9. For any local optimum $\bar{\ell}$ of (P), we assume that there exists $\lambda \in L^2(0,T;Y)$ so that $-\partial_u \Phi(\bar{y},\bar{u})^*\lambda = \partial_u J(\bar{y},\bar{u},\bar{\ell}) + \partial_u \Psi(\bar{y},\bar{u})^* \partial_\ell J(\bar{y},\bar{u},\bar{\ell})$, where $(\bar{y},\bar{u}) := S(\bar{\ell})$.

Remark 2.10. Assumption 2.9 is needed only for the solvability of the adjoint system (2.19a)–(2.19b)–(2.19d) below. Let us point out that, given a concrete setting, this can be checked by regularizing the nonsmooth problem [1] (unless the assertion in Assumption 2.9 follows immediately, e.g., if $\partial_u \Phi(\bar{y},\bar{u})^* : L^2(0,T;Y) \to L^2(0,T;U^*)$ is surjective). The solution (ξ,λ,w) then arises as the limit of the sequence of solutions $(\xi_\varepsilon,\lambda_\varepsilon,w_\varepsilon)$ of the regularized adjoint system for regularization parameter $\varepsilon \searrow 0$. By choosing a suitable smooth approximation of f, the regularity of $(\xi_\varepsilon,\lambda_\varepsilon,w_\varepsilon)$ may be preserved in the limit so that $(\xi,\lambda,w) \in H^1_T(0,T;Y^*) \times L^2(0,T;Y) \times L^2(0,T;U)$ follows. This will be the case in section 4 below. There, the solvability of the adjoint system can be directly infered from (2.19a)–(2.19b)–(2.19d), except the desired regularity of λ, for which we resort to a regularization approach as explained above; see also Remark 4.4 below. Note that in [31] too, the desired regularity of the adjoint state (which was crucial for deriving strong stationarity) cannot be simply deduced from the unregularized adjoint system [31, (5.2)] and it is a consequence of the limit analysis concerning the regularized adjoint system [31, (4.16)]; see [31, Thm. 4.16].

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

THEOREM 2.11 (strong stationarity). *Suppose that Assumptions* 2.6 *and* 2.9 *are 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*

$$-\dot{\xi} - \partial_y \Phi(\bar{y},\bar{u})^*\lambda + \partial_y \Psi(\bar{y},\bar{u})^* w = \partial_y J(\bar{y},\bar{u},\bar{\ell}) \quad \text{in } L^2(0,T;Y^*), \quad \xi(T) = 0, \tag{2.19a}$$

$$-\partial_u \Phi(\bar{y},\bar{u})^*\lambda + \partial_u \Psi(\bar{y},\bar{u})^* w = \partial_u J(\bar{y},\bar{u},\bar{\ell}) \quad \text{in } L^2(0,T;U^*), \tag{2.19b}$$

$$\langle \xi(t), f'(\Phi(\bar{y}(t),\bar{u}(t));v)\rangle_Y \geq \langle \lambda(t), v\rangle_{Y^*} \quad \forall\, v \in Y^* \text{ a.e. in } (0,T), \tag{2.19c}$$

$$w + \partial_\ell J(\bar{y},\bar{u},\bar{\ell}) = 0 \quad \text{in } L^2(0,T;U). \tag{2.19d}$$

Proof. (1) We first prove that there exists a unique tuple (ξ,w,λ) which satisfies (2.19a), (2.19b), and (2.19d) 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 the 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 2.9, there exists $\lambda \in L^2(0,T;Y)$ such that (2.19b) is fulfilled. Moreover, in view of Assumption 2.6, $\partial_u \Phi(\bar{y},\bar{u})^* : L^2(0,T;Y) \to L^2(0,T;U^*)$ is injective, which yields the uniqueness of λ. Further, Lemma 2.2 and Assumption 2.1.4 imply that the expression

$$\nu := \partial_y J(\bar{y},\bar{u},\bar{\ell}) + \partial_y \Phi(\bar{y},\bar{u})^*\lambda - \partial_y \Psi(\bar{y},\bar{u})^* w \tag{2.20}$$

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

$$[0,T] \ni t \mapsto \xi(t) \in Y^*, \quad \xi(t) := \int_t^T \nu(s)\, ds, \tag{2.21}$$

such that $\xi(T) = 0$, $\xi \in H^1(0,T;Y^*)$, by the regularity of ν, and $-\dot{\xi}(t) = \nu(t)$ for a.a. $t \in (0,T)$. We now observe that the tuple (ξ, w, λ) satisfies (2.19a), (2.19b), and (2.19d). Moreover, it has the desired regularity and is unique.

(2) It remains to show that the VI (2.19c) is true. To this end, we extend the basic idea from [31, 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

$$\underbrace{\Phi'(\bar{y},\bar{u})(\delta y_n, \delta u_n)}_{:=\rho_n} \to \rho \quad \text{in } L^2(0,T;Y^*) \quad \text{as } n \to \infty, \tag{2.22}$$

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 (2.19a), (2.19b), and (2.19d) with δy_n, δu_n, and $\delta\ell_n$, respectively, we obtain

$$\begin{aligned}
0 &\le \partial_y J(\bar{y},\bar{u},\bar{\ell})\delta y_n + \partial_u J(\bar{y},\bar{u},\bar{\ell})\delta u_n + \partial_\ell J(\bar{y},\bar{u},\bar{\ell})\delta\ell_n \\
&= -\int_0^T \langle \dot{\xi}(t), \delta y_n(t)\rangle_Y \, dt - \langle \partial_y\Phi(\bar{y},\bar{u})^*\lambda - \partial_y\Psi(\bar{y},\bar{u})^* w, \delta y_n\rangle_{L^2(0,T;Y)} \\
&\quad + \langle \partial_u\Psi(\bar{y},\bar{u})^* w - \partial_u\Phi(\bar{y},\bar{u})^*\lambda, \delta u_n\rangle_{L^2(0,T;U)} - \langle \delta\ell_n, w\rangle_{L^2(0,T;U)} \\
&= \int_0^T \langle \xi(t), \dot{\delta y}_n(t)\rangle_Y \, dt + \langle \partial_y\Psi(\bar{y},\bar{u})\delta y_n + \partial_u\Psi(\bar{y},\bar{u})\delta u_n - \delta\ell_n, w\rangle_{L^2(0,T;U)} \\
&\quad - \langle \partial_y\Phi(\bar{y},\bar{u})\delta y_n + \partial_u\Phi(\bar{y},\bar{u})\delta u_n, \lambda\rangle_{L^2(0,T;Y)} \\
&= \int_0^T \langle \xi(t), f'(\Phi(\bar{y}(t),\bar{u}(t));\rho_n(t))\rangle_Y \, dt - \int_0^T \langle \lambda(t), \rho_n(t)\rangle_{Y^*} \, dt \quad \forall\, n \in \mathbb{N},
\end{aligned} \tag{2.23}$$

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 (2.23) leads to

$$0 \le \int_0^T \langle \xi(t), f'(\Phi(\bar{y}(t),\bar{u}(t));\rho(t))\rangle_Y \, dt - \int_0^T \langle \lambda(t), \rho(t)\rangle_{Y^*} \, dt \quad \forall\, \rho \in L^2(0,T;Y^*) \tag{2.24}$$

in view of (2.22). 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 2.1.3, see also (2.3).

Now, consider $v \in Y^*$ and let $\varphi \in C_0^\infty(0,T)$ with $\varphi \ge 0$ be arbitrary. By setting $\rho := \varphi v \in L^2(0,T;Y^*)$ in (2.24) and by employing the positive homogeneity of the directional derivative, we get

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

The fundamental lemma of the calculus of variations then gives (2.19c). The proof is now complete. ∎

Remark 2.12 (density of the set of arguments of $f'(\Phi(\bar y,\bar u);\cdot)$).

(i) The proof of Theorem 2.11 (see (2.23)), shows that it is essential that the set of directions into which the nonsmooth mapping f is differentiated—in the linearized state equation associated with $\bar\ell$—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 nonsmooth 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 read completely differently, but it should imply that the set of directions into which f is differentiated—in the linearized state equation—is dense in an entire space.

(iii) Finally, let us remark that the observations made here are consistent with the results in [31]. Therein, the direction into which one differentiates 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., [31, Lem. 5.2]. In [31], 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 [31, Assump. 2.1.6] can be regarded as such, in view of Remark 2.16 below.

To see that (2.19) is indeed of strong stationary type (cf. section 1), we prove the following.

THEOREM 2.13 (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* (2.19a)–(2.19d). *Then, it also satisfies the VI* (2.9). *If Assumptions* 2.6 *and* 2.9 *are satisfied,* (2.9) *is equivalent to* (2.19a)–(2.19d).

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 (2.23) to obtain

$$\begin{aligned}&\partial_y J(\bar y,\bar u,\bar\ell)\delta y+\partial_u J(\bar y,\bar u,\bar\ell)\delta u+\partial_\ell J(\bar y,\bar u,\bar\ell)\delta\ell\\&\quad=\int_0^T\langle\xi(t),f'\big(\Phi(\bar y(t),\bar u(t));\Phi'(\bar y(t),\bar u(t))(\delta y(t),\delta u(t))\big)\rangle_Y\,dt\\&\qquad-\int_0^T\langle\lambda(t),\Phi'(\bar y(t),\bar u(t))(\delta y(t),\delta u(t))\rangle_{Y^*}\,dt.\end{aligned}\tag{2.25}$$

Note that one does not need local optimality for $\bar\ell$ or Assumptions 2.6 and 2.9 to prove (2.25). The VI (2.9) follows by testing (2.19c) with $v:=\Phi'(\bar y,\bar u)(\delta y,\delta u)(t)\in Y^*$ for a.a. $t\in(0,T)$ and by using the resulting inequalities on the r.h.s. of (2.25). Moreover, if Assumptions 2.6 and 2.9 are satisfied, then (2.9) implies (2.19a)–(2.19d); see the proof of Theorem 2.11. This shows the second assertion. ∎

Remark 2.14. If f is Gâteaux differentiable at $\Phi(\bar y(t),\bar u(t))$ a.e. in $(0,T)$, then (2.19c) is equivalent to

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

by the linearity of $f'(\Phi(\bar y(t),\bar u(t)))$. Thus, the optimality system in Theorem 2.11 reduces to the very same optimality conditions which one obtains when directly applying the KKT theory to (P) (cf. [47]), i.e., (2.19) 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 2.15. If $Y = L^q(\Omega)$ with $q \in (1, \infty)$ and if $f : Y^* \to Y$ is the Nemytskii operator associated with some mapping $f : \mathbb{R} \to \mathbb{R}$, then (2.19c) is equivalent to a similar sign condition which holds a.e. *in* $(0, T) \times \Omega$. This will be the case in section 4, where we apply the result of Theorem 2.11 on a concrete setting; see (4.19c) below.

Remark 2.16. The entire analysis in this section carries on if we consider an additional (nonlinear) 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, 31, 35]. 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.

Remark 2.17 (absence of control constraints). An inspection of the proof of Theorem 2.11 shows that the arguments cannot be applied if the controls are restricted by additional constraints. The same observation was made in [35], where strong stationarity for optimal control of the obstacle problem is shown to be necessary for local optimality. Let us, however, mention [48] 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 EVIs as nonsmooth ODEs. This section focuses on proving that the following viscous evolution

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

is equivalent to a nonsmooth ODE in the Hilbert space Y (see Theorem 3.7 below). In the case of coupled systems consisting of an EVI and an elliptic PDE, such as (4.1) (see also (4.9)) below, the mapping g contains the solution operator of the PDE, so that the nonsmooth ODE (3.8) below is obtained by a reduction of one of the states; cf. the upcoming section 4 (in particular, the proof of Lemma 4.1.1). Note that (EVI) is a generalization of the classical evolutionary VI with viscosity; see, e.g., [43, Chap. 4]. In addition, we are concerned with the differentiability properties of the solution map associated with (EVI). The nonsmooth nonlinearity appearing in the ODE (3.8) is the solution operator of an elliptic VI of the second kind, for which we give an explicit formula. Moreover, we state a condition which is necessary and sufficient for the directional differentiability thereof (cf. Corollary 3.4 below).

The results established in this section will be combined later on with the findings from section 2, in order to establish strong stationarity for the optimal control of a concrete viscous EVI (coupled with an elliptic PDE); 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 the following:

1. The nonsmooth functional $R : Y \to (-\infty, \infty]$ is proper, convex, lower semicontinuous, and positively 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 self-adjoint, 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 $g : Y \times \mathcal{H} \to Y^*$ is directionally differentiable and Lipschitz continuous.

In the following, 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 nonsmooth nonlinearity). *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 (metric) projection onto the set* $\partial R(0)$ *w.r.t. the inner product* $\langle \mathcal{V}^{-1}\cdot, \cdot\rangle_{Y^*}$*, i.e.,* $P_{\partial R(0)}\omega$ *is the unique solution of*

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

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 alternative proofs. The first one is based on convex analysis tools and it involves the dual formulation of (3.3). Moreover, it highlights the background of Definition 3.2 (see (3.5) below), by showing step by step how one arrives at the formula (3.1) for the solution operator of (3.3). The second proof is just a consequence of [8, Prop. 2.1].

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

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

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

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

By (3.4) and the sum rule for conjugate functionals (see [28, Thm. 3.3.4.1]), it holds

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

where for the second identity we used $R^* = I_{\partial R(0)}$, which is due to the positive homogeneity of R. Note that if the operator $\mathcal{V}$ is a positive scalar (and $Y = Y^*$), then (3.5) coincides with the Yosida approximation of $I_{\partial R(0)}$; see, e.g., [38, (5.46)]. Further, in view of [21, Lem. 4.1] and (3.5), in combination with (3.1), we have

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

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

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

in view of (3.4) and the sum rule for subdifferentials. Thus, $\mathcal{F}(\omega) \in Y$ solves (3.3). As (3.3) is uniquely solvable (see, e.g., [17]), 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 self-adjoint, 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.

COROLLARY 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 criterion 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 differentiable.

Remark 3.6. With the result in Corollary 3.4 at hand, the analysis in some papers addressing the differentiability of VIs of the second kind via a polyhedricity condition [3, 26, 44] could have been reduced. Once the polyhedricity of $\partial R(0)$ was shown or assumed, there is no need to provide uniform bounds and different relations for the difference quotients [26] as well as dual formulations of the original VI [3, 44], since the desired differentiability follows then from [19, Thm. 2] and Corollary 3.4. Note that this is in accordance with [8, Thm. 2.3].

However, there are other contributions such as [7], where the authors make assumptions which guarantee the directional differentiability of the solution operator of the considered VI of the second kind, without resorting to polyhedricity; see also [10]. As a consequence of Corollary 3.4 and [7, Thm. 4.14], [7, Assump. 4.3] is, for instance, sufficient for the directional differentiability of the metric projection $P_{\partial R(0)} : L^p(\Omega) \to H_0^1(\Omega)^*$ with $p > n/2$, where $\Omega \subset \mathbb{R}^n$, $n \in \{2, 3\}$, is a bounded Lipschitz domain, $R = \|\cdot\|_{L^1(\Omega)}$ and $\mathcal{V} := -\triangle$. A similar assertion concerning the weak directional differentiability of the above metric projection follows from [10, Assumps. 3.1 and 3.2] combined with Corollary 3.4.

Theorem 3.7 (viscous EVIs are nonsmooth ODEs in Hilbert space).

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

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

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 r.h.s. $\ell \in L^2(0,T;\mathcal{H})$.*

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 $\mathcal{F} : Y^* \to Y$ is directionally differentiable at $g(\bar{y}(t), \bar{\ell}(t))$ for a.a. $t \in (0,T)$ or, equivalently, if $P_{\partial R(0)} : Y^* \to Y^*$ does so, 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*

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

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

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

where $G : Y \times (0,T) \ni (\eta, t) \mapsto \mathcal{F}(g(\eta, \ell(t)) \in Y$. By the global Lipschitz continuity of $\mathcal{F}$ and g, we have that G maps $L^2(0,T;Y)$ to $L^2(0,T;Y)$. Moreover, $G(\cdot, t)$ is Lipschitz continuous for 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., [12, 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 $\delta y \in H_0^1(0,T;Y)$.

Further, we observe that $\mathcal{F}$ is Hadamard directionally differentiable at $g(\bar{y}(t), \bar{\ell}(t))$ for a.a. $t \in (0,T)$ [40, Def. 3.1.1], as a result of [40, Lem. 3.1.2(b)]. Note that here we employed again the Lipschitz continuity of $\mathcal{F}$. Since g is directionally differentiable, by Assumption 3.1.3, chain rule [41, Prop. 3.6(i)] implies that

$$\widehat{G} := \mathcal{F} \circ g$$

is (Hadamard) directionally differentiable at $(\bar{y}(t), \bar{\ell}(t))$ with

$$\widehat{G}'((\bar{y}(t), \bar{\ell}(t)); h) = \mathcal{F}'\big(g(\bar{y}(t), \bar{\ell}(t)); g'((\bar{y}(t), \bar{\ell}(t)); h)\big) \quad \forall\, h \in Y \times \mathcal{H}, \quad \text{for a.a. } t \in (0,T). \tag{3.10}$$

For simplicity, in the following we abbreviate $\bar{y}^\tau := \mathcal{S}(\bar{\ell} + \tau\,\delta\ell)$, where $\tau > 0$ and $\delta\ell \in L^2(0,T;\mathcal{H})$ are arbitrary, but fixed. Due to the above and by combining the equations for $\bar{y}^\tau$, $\bar{y}$, and (3.9), we obtain

$$\begin{aligned} \frac{d}{dt}\Big(\frac{\bar{y}^\tau - \bar{y}}{\tau} - \delta y\Big)(t) &= \frac{\widehat{G}(\bar{y}^\tau(t), \bar{\ell}(t) + \tau\,\delta\ell(t)) - \widehat{G}(\bar{y}(t), \bar{\ell}(t))}{\tau} \\ &\qquad - \widehat{G}'\big((\bar{y}(t), \bar{\ell}(t)); (\delta y(t), \delta\ell(t))\big) \quad \text{a.e. in } (0,T), \\ \Big(\frac{\bar{y}^\tau - \bar{y}}{\tau} - \delta y\Big)(0) &= 0. \end{aligned} \tag{3.11}$$

This implies

(3.12)

$$
\begin{aligned}
&\Big\|\Big(\frac{\bar y^\tau - \bar y}{\tau} - \delta y\Big)(t)\Big\|_Y \\
&\quad\le \int_0^t \Big\|\frac{\widehat G(\bar y^\tau(s), \bar\ell(s) + \tau\,\delta\ell(s)) - \widehat G\big((\bar y(s), \bar\ell(s)) + \tau(\delta y(s), \delta\ell(s))\big)}{\tau}\Big\|_Y \\
&+ \underbrace{\Big\|\frac{\widehat G\big((\bar y(s), \bar\ell(s)) + \tau(\delta y(s), \delta\ell(s))\big) - \widehat G(\bar y(s), \bar\ell(s))}{\tau} - \widehat G'\big((\bar y(s), \bar\ell(s)); (\delta y(s), \delta\ell(s))\big)\Big\|_Y}_{=:A_\tau(s)} ds \\
&\quad\le \int_0^t L_{\widehat G}\Big\|\frac{\bar y^\tau(s) - \bar y(s)}{\tau} - \delta y(s)\Big\|_Y + A_\tau(s)\,ds \qquad \forall\, t \in [0,T],
\end{aligned}
$$

where $L_{\widehat G} > 0$ is the Lipschitz constant of $\widehat G : Y \times \mathcal H \to Y$; recall that $\mathcal F$ and g are globally Lipschitz continuous. Applying Gronwall's inequality in (3.12) then yields

$$
\Big\|\Big(\frac{\bar y^\tau - \bar y}{\tau} - \delta y\Big)(t)\Big\|_Y \le c\int_0^t A_\tau(s)\,ds \qquad \forall\, t \in [0,T], \tag{3.13}
$$

where $c > 0$ is a constant dependent only on the given data. Now, (3.11) and estimating as in (3.12), in combination with (3.13), leads to

$$
\Big\|\frac{\bar y^\tau - \bar y}{\tau} - \delta y\Big\|_{H^1(0,T;Y)} \le c\,\|A_\tau\|_{L^2(0,T)}. \tag{3.14}
$$

On the other hand, we recall the definition of A_τ in (3.12) and the fact that $\widehat G$ is directionally differentiable at $(\bar y(t), \bar\ell(t))$ for a.a. $t \in (0,T)$, from which we deduce the convergence $A_\tau(t) \to 0$ for a.a. $t \in (0,T)$, as $\tau \searrow 0$. Moreover, by relying again on the global Lipschitz continuity of $\widehat G$, we have

$$
A_\tau(t) \le 2L_{\widehat G}\,\|(\delta y(t), \delta\ell(t))\|_{Y\times\mathcal H} \quad \text{for a.a. } t \in (0,T),
$$

so that Lebesgue's dominated convergence theorem implies

$$
\|A_\tau\|_{L^2(0,T)} \to 0 \quad \text{as } \tau \searrow 0.
$$

Finally, the desired assertion follows from (3.14). This completes the proof. ∎

4. Application to a damage model. Based on the results from the previous sections, we next derive strong stationary optimality conditions for the optimal control of a two-field gradient damage model. This type of model approximates the classical single-field damage model (cf. [33]), and is frequently employed in computational mechanics (see, e.g., [11] and the references therein). It involves two damage variables, a "local" and a "nonlocal" one, which are connected through a penalty term in the stored energy. The problem considered in this section describes the evolution of damage under the influence of a time-dependent load $\ell : [0,T] \to H^1(\Omega)^*$ (control) acting on a body occupying the bounded Lipschitz domain $\Omega \subset \mathbb R^N$, $N \in \{2,3\}$. The induced local and nonlocal damage are expressed in terms of $\varphi : [0,T] \to H^1(\Omega)$ and $d : [0,T] \to L^2(\Omega)$, respectively (states). For more details, see [32, sect. 2.1–2.2]. Note that, for simplicity reasons, we do not take a displacement variable into account.

In the following, we investigate the viscous two-field gradient damage model

$$\left.\begin{aligned} \varphi(t) \in \underset{\varphi\in H^1(\Omega)}{\arg\min}\, \mathcal{E}(t,\varphi,d(t)),\\ -\partial_d\mathcal{E}(t,\varphi(t),d(t)) \in \partial\mathcal{R}_\epsilon(\dot d(t)), \quad d(0)=0 \end{aligned}\right\} \tag{4.1}$$

a.e. in $(0,T)$, where the stored energy $\mathcal{E} : [0,T]\times H^1(\Omega)\times L^2(\Omega)\to\mathbb{R}$ is given by

$$\mathcal{E}(t,\varphi,d) := \frac{\alpha}{2}\|\nabla\varphi\|^2_{L^2(\Omega)} + \frac{\beta}{2}\|\varphi-d\|^2_{L^2(\Omega)} - \langle\ell(t),\varphi\rangle_{H^1(\Omega)} \tag{4.2}$$

with $\alpha,\beta>0$ fixed. The viscous dissipation $\mathcal{R}_\epsilon : L^2(\Omega)\to[0,\infty]$ is defined as

$$\mathcal{R}_\epsilon(\eta) := \begin{cases} r\int_\Omega \eta\,dx + \frac{\epsilon}{2}\|\eta\|^2_{L^2(\Omega)} & \text{if } \eta\geq 0 \text{ a.e. in } \Omega,\\ \infty & \text{otherwise,}\end{cases} \tag{4.3}$$

where $r>0$ is the fracture toughness and $\epsilon>0$ stands for the viscosity parameter.

As we will next see, the problem (4.1) consists of a viscous EVI (viscous damage evolution) and an elliptic PDE (equation for the nonlocal damage). 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 of strong stationary type [6, 31].

With a little abuse of notation, we use in the following the Laplace symbol for the operator $\Delta : H^1(\Omega)\to H^1(\Omega)^*$ defined by

$$\langle\Delta\eta,\psi\rangle_{H^1(\Omega)} := -\int_\Omega \nabla\eta\nabla\psi\,dx \quad \forall\,\psi\in H^1(\Omega).$$

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

1. *For every $\ell\in L^2(0,T;H^1(\Omega)^*)$, the viscous damage problem* (4.1) *admits a unique solution $(d,\varphi)\in H^1_0(0,T;L^2(\Omega))\times L^2(0,T;H^1(\Omega))$, which is characterized by the following PDE system*

$$\dot d(t) = \frac{1}{\epsilon}\max(-\beta(d(t)-\varphi(t))-r,0) \quad \textit{in } L^2(\Omega), \quad d(0)=0, \tag{4.4a}$$

$$-\alpha\Delta\varphi(t)+\beta\,\varphi(t) = \beta d(t)+\ell(t) \quad \textit{in } H^1(\Omega)^* \tag{4.4b}$$

a.e. in $(0,T)$.

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

$$S : L^2(0,T;H^1(\Omega)^*) \ni \ell \mapsto (d,\varphi)\in H^1_0(0,T;L^2(\Omega))\times L^2(0,T;H^1(\Omega))$$

is directionally differentiable. Its directional derivative $(\delta d,\delta\varphi) := S'(\ell;\delta\ell)$ at $\ell\in L^2(0,T;H^1(\Omega)^*)$ in direction $\delta\ell\in L^2(0,T;H^1(\Omega)^*)$ is the unique solution of

$$\begin{aligned} \dot{\delta d}(t) &= \frac{1}{\epsilon}\max{}'(-\beta(d(t)-\varphi(t))-r;-\beta(\delta d(t)-\delta\varphi(t))) \ \textit{in } L^2(\Omega),\ \delta d(0)=0,\\ &-\alpha\Delta\delta\varphi(t)+\beta\,\delta\varphi(t) = \beta\delta d(t)+\delta\ell(t) \quad \textit{in } H^1(\Omega)^* \end{aligned} \tag{4.5}$$

a.e. in $(0,T)$, where we abbreviate $(d,\varphi) := S(\ell)$.

Proof. 1. Let $t \in [0,T]$ and $\hat{d} : [0,T] \to L^2(\Omega)$ be arbitrary, but fixed. Since $\mathcal{E}(t, \cdot, \hat{d}(t))$ is strictly convex, continuous, and radially unbounded (see (4.2)), the minimization problem $\min_{\varphi \in H^1(\Omega)} \mathcal{E}(t, \cdot, \hat{d}(t))$ admits a unique solution $\hat{\varphi}(t)$ characterized by $\partial_\varphi \mathcal{E}(t, \hat{\varphi}(t), \hat{d}(t)) = 0$ in $H^1(\Omega)^*$. In view of (4.2), this means that

$$\hat{\varphi}(t) \in \operatorname*{arg\,min}_{\varphi \in H^1(\Omega)} \mathcal{E}(t, \varphi, \hat{d}(t)) \Longleftrightarrow \hat{\varphi}(t) = \phi(\hat{d}(t), \ell(t)), \tag{4.6}$$

where $\phi : L^2(\Omega) \times H^1(\Omega)^* \ni (\tilde{d}, \tilde{\ell}) \mapsto \tilde{\varphi} \in H^1(\Omega)$ is the solution operator of

$$-\alpha \Delta \tilde{\varphi} + \beta\, \tilde{\varphi} = \beta \tilde{d} + \tilde{\ell} \quad \text{in } H^1(\Omega)^*. \tag{4.7}$$

With the map ϕ at hand, the evolution in (4.1) reads

$$-\partial_d \mathcal{E}(t, \phi(d(t), \ell(t)), d(t)) \in \partial \mathcal{R}_\epsilon(\dot{d}(t)) \quad \text{a.e. in } (0,T). \tag{4.8}$$

In the light of (4.2), (4.3), and the sum rule for convex subdifferentials, (4.8) is further equivalent to

$$\mathcal{R}(v) - \mathcal{R}(\dot{d}(t)) + \epsilon\, (\dot{d}(t), v - \dot{d}(t))_{L^2(\Omega)} \geq \beta(\phi(d(t), \ell(t)) - d(t), v - \dot{d}(t))_{L^2(\Omega)} \quad \forall v \in L^2(\Omega), \tag{4.9}$$

a.e. in $(0,T)$, where

$$\mathcal{R} : L^2(\Omega) \to [0, \infty], \quad \mathcal{R}(\eta) := \begin{cases} r \int_\Omega \eta \, dx & \text{if } \eta \geq 0 \text{ a.e. in } \Omega, \\ \infty & \text{otherwise.} \end{cases} \tag{4.10}$$

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

$$Y := L^2(\Omega), \quad \mathcal{H} := H^1(\Omega)^*, \tag{4.11a}$$

$$R := \mathcal{R}, \quad \mathcal{V} := \epsilon\, \mathbb{I}, \quad g(\tilde{d}, \tilde{\ell}) := \beta(\phi(\tilde{d}, \tilde{\ell}) - \tilde{d}), \tag{4.11b}$$

then (EVI) coincides with (4.9). Note that, due to $\phi \in \mathcal{L}(L^2(\Omega) \times H^1(\Omega)^*; H^1(\Omega))$, we have $g \in \mathcal{L}(Y \times \mathcal{H}; Y^*)$ with

$$g'(\tilde{d}, \tilde{\ell})(\delta d, \delta \ell) = \beta(\phi(\delta d, \delta \ell) - \delta d) \quad \forall\, (\tilde{d}, \tilde{\ell}), (\delta d, \delta \ell) \in Y \times \mathcal{H}. \tag{4.12}$$

Since the quantities in (4.11) satisfy Assumption 3.1, we can apply Theorem 3.7.1, which yields that (4.9) is equivalent to

$$\dot{d}(t) = \mathcal{F}\big(g(d(t), \ell(t))\big) \quad \text{a.e. in } (0,T), \tag{4.13}$$

where $\mathcal{F} = \frac{1}{\epsilon}(\mathbb{I} - P_{\partial \mathcal{R}(0)})$. As a result of

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

it holds $P_{\partial \mathcal{R}(0)}(\omega) = \min(\omega, r)$, so that

$$\mathcal{F}(\omega) = \frac{1}{\epsilon} \max(\omega - r, 0) \quad \forall\, \omega \in L^2(\Omega). \tag{4.14}$$

Moreover, Theorem 3.7.1 tells us that (4.9) (with the initial condition $d(0) = 0$) admits a unique solution $d \in H^1_0(0,T; L^2(\Omega))$ for every $\ell \in L^2(0,T; H^1(\Omega)^*)$. Further, since $\varphi(\cdot) = \phi(d(\cdot), \ell(\cdot))$ (cf. (4.6)), we deduce from (4.7) that $\varphi \in L^2(0,T; H^1(\Omega))$. To summarize, we obtained that (4.1) admits a unique solution $(d,\varphi) \in H_0^1(0,T;L^2(\Omega)) \times L^2(0,T;H^1(\Omega))$, which, owing to (4.6), (4.13), and (4.14), is characterized by (4.4). The proof of this step is now complete.

2. We begin by noticing that the mapping $\mathcal{F} : Y^* \to Y$ from (4.14) is directionally differentiable. This follows by Lebesgue's dominated convergence theorem or, alternatively, by [4, Prop. 6.33] and [19, Thm. 2], which yield that $P_{\partial\mathcal{R}(0)} : Y^* \to Y$ is directionally differentiable (recall here Corollary 3.4). We denote the first component of the operator S by S_1, i.e., $S_1 : L^2(0,T;H^1(\Omega)^*) \ni \ell \mapsto d \in H_0^1(0,T;L^2(\Omega))$ is the solution map associated with (4.9) or, equivalently, with (4.13). From Theorem 3.7.2 we infer that S_1 is directionally differentiable. Its directional derivative $\delta d := S_1'(\ell;\delta\ell)$ at ℓ in direction $\delta\ell$ is the unique solution of

$$\dot{\delta d}(t) = \mathcal{F}'\big(g(d(t),\ell(t)); g'(d(t),\ell(t))(\delta d(t),\delta\ell(t))\big) \quad \text{a.e. in } (0,T),\ \delta d(0) = 0, \tag{4.15}$$

where $d := S_1(\ell)$. In view of (4.14), the definition of g (see (4.11)) and (4.12), (4.15) reads

$$\dot{\delta d}(t) = \frac{1}{\epsilon} \max{}'\big(\beta(\phi(d(t),\ell(t)) - d(t)) - r; \beta(\phi(\delta d(t),\delta\ell(t)) - \delta d(t))\big) \qquad \text{a.e. in } (0,T),\ \delta d(0) = 0. \tag{4.16}$$

Further, from (4.6) we have $S_2(\ell) = \phi(S_1(\ell),\ell)$ for all $\ell \in L^2(0,T;H^1(\Omega)^*)$, where S_2 is the second component of the operator S, i.e., $S_2 : L^2(0,T;H^1(\Omega)^*) \ni \ell \mapsto \varphi \in L^2(0,T;H^1(\Omega))$. Thus, S_2 is directionally differentiable as well, since $\phi \in \mathcal{L}(L^2(\Omega) \times H^1(\Omega)^*; H^1(\Omega))$ and S_1 is directionally differentiable. Its directional derivative $\delta\varphi := S_2'(\ell;\delta\ell)$ at ℓ in direction $\delta\ell$ is given by $\delta\varphi = \phi(S_1'(\ell;\delta\ell),\delta\ell)$. On account of (4.16), the proof is now complete. ◻

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

$$\text{(Q)} \qquad \left.\begin{aligned} \min_{\ell \in L^2(0,T;L^2(\Omega))} \quad & \mathcal{J}(d,\varphi,\ell) \\ \text{s.t.} \quad & (d,\varphi) \text{ solves (4.1) with r.h.s. } \ell. \end{aligned}\right\}$$

In what follows, the objective $\mathcal{J}$ is supposed to fulfill the following.

Assumption 4.2. The functional

$$\mathcal{J} : L^2(0,T;L^2(\Omega)) \times L^2(0,T;H^1(\Omega)) \times L^2(0,T;L^2(\Omega)) \to \mathbb{R}$$

is continuously Fréchet differentiable and Lipschitz continuous on bounded sets, i.e., for all $M > 0$ there exists $L_M > 0$ so that

$$|\mathcal{J}(v_1) - \mathcal{J}(v_2)| \le L_M \,\|v_1 - v_2\|_X \quad \forall\, v_1, v_2 \in B_X(0,M), \tag{4.17}$$

where we abbreviate $X := L^2(0,T;L^2(\Omega)) \times L^2(0,T;H^1(\Omega)) \times L^2(0,T;L^2(\Omega))$.

Note that Assumption 4.2 is satisfied by classical objectives of tracking type such as

$$\mathcal{J}_{ex}(d,\varphi,\ell) := \frac{1}{2}\,\|\varphi - \varphi_d\|^2_{L^2(0,T;H^1(\Omega))} + \frac{\alpha_1}{2}\|d\|^2_{L^2(0,T;L^2(\Omega))} + \frac{\alpha_2}{2}\|\ell - \ell_d\|^2_{L^2(0,T;L^2(\Omega))},$$

where $\varphi_d \in L^2(0,T;H^1(\Omega))$, $\ell_d \in L^2(0,T;L^2(\Omega))$ and $\alpha_1, \alpha_2 > 0$.

Before stating the strong stationary optimality conditions, we check that Assumption 2.9 is satisfied in our setting. As it will turn out in the proof of Theorem 4.5 below, this is indeed the case, as a result of the following.

LEMMA 4.3. *For any local optimum $\bar{\ell}$ of* (Q), *there exists a pair*

$$(\lambda, w) \in L^2(0,T;L^2(\Omega)) \times L^2(0,T;H^1(\Omega))$$

so that

$$-\beta\lambda(t) - \alpha\Delta w(t) + \beta w(t) = \partial_\varphi \mathcal{J}(S(\bar{\ell}),\bar{\ell})(t) \quad \text{in } H^1(\Omega)^*, \tag{4.18a}$$

$$w(t) + \partial_\ell \mathcal{J}(S(\bar{\ell}),\bar{\ell})(t) = 0 \quad \text{in } H^1(\Omega), \quad \text{a.e. in } (0,T), \tag{4.18b}$$

where S is the solution operator associated with (4.1)*; see Lemma* 4.1.2.

Proof. For convenience of the reader, the proof is given in Appendix B. ∎

Remark 4.4. Note that the existence of

$$(\xi,\lambda,w) \in H^1_T(0,T;H^1(\Omega)^*) \times L^2(0,T;H^1(\Omega)^*) \times L^2(0,T;H^1(\Omega))$$

satisfying the adjoint system (4.19a)–(4.19b)–(4.19d) below follows directly, without employing Lemma 4.3. It is the $L^2(0,T;L^2(\Omega))$-regularity of the multiplier λ which cannot be immediately deduced from (4.19b). This additional information stated in Lemma 4.3 is proven by a regularization approach; see Appendix B. We refer here to [31] for a similar situation; see also Remark 2.10. Let us underline that the improved space regularity of λ is essential for deriving the sign condition in (4.19c) a.e. in $(0,T)\times\Omega$; see the proof of Theorem 4.5 below.

The main result of this section reads as follows.

THEOREM 4.5 (strong stationarity for the optimal control of the viscous two-field gradient damage model). *Let $\bar{\ell} \in L^2(0,T;L^2(\Omega))$ be locally optimal for* (Q) *with associated states*

$$\bar{d} \in H^1_0(0,T;L^2(\Omega)) \quad \text{and} \quad \bar{\varphi} \in L^2(0,T;H^1(\Omega)).$$

Then, there exist unique adjoint states

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

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

$$-\dot{\xi} + \beta\lambda - \beta w = \partial_d \mathcal{J}(\bar{d},\bar{\varphi},\bar{\ell}) \quad \text{in } L^2(0,T;L^2(\Omega)), \quad \xi(T) = 0, \tag{4.19a}$$

$$-\beta\lambda - \alpha\Delta w + \beta w = \partial_\varphi \mathcal{J}(\bar{d},\bar{\varphi},\bar{\ell}) \quad \text{in } L^2(0,T;H^1(\Omega)^*), \tag{4.19b}$$

$$\left.\begin{aligned} &\lambda(t,x) = \frac{1}{\epsilon}\chi_{\{\bar{z}>r\}}(t,x)\xi(t,x) \quad \text{a.e. where } \bar{z}(t,x) \neq r, \\ &\lambda(t,x) \in \big[0, \frac{1}{\epsilon}\xi(t,x)\big] \quad \text{a.e. where } \bar{z}(t,x) = r, \end{aligned}\right\} \tag{4.19c}$$

$$w + \partial_\ell \mathcal{J}(\bar{d},\bar{\varphi},\bar{\ell}) = 0 \quad \text{in } L^2(0,T;H^1(\Omega)), \tag{4.19d}$$

where we abbreviate $\bar{z} := -\beta(\bar{d}-\bar{\varphi})$.

Proof. We aim to apply the strong stationarity result given by Theorem 2.11 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, 2.6, and 2.9. Indeed, with the notations from section 2, we see that if we set

$$V := L^2(\Omega),\ Y := L^2(\Omega),\ U := H^1(\Omega),\ J := \mathcal{J}, \tag{4.20a}$$

$$f : Y^* \to Y, \quad f(\omega) = \frac{1}{\epsilon}\max(\omega - r, 0), \tag{4.20b}$$

$$\Phi : Y \times U \ni (d,\varphi) \mapsto -\beta(d-\varphi) \in Y^*, \tag{4.20c}$$

$$\Psi : Y \times U \ni (d,\varphi) \mapsto -\alpha\Delta\varphi + \beta\varphi - \beta d \in U^*, \tag{4.20d}$$

then (P) coincides with (Q), thanks to Lemma 4.1.1. Notice that $V \overset{d}{\hookrightarrow} U^*$ so that Assumption 2.1.1 is satisfied. Since $\Phi \in \mathcal{L}(Y \times U; Y^*)$ and $\Psi \in \mathcal{L}(Y \times U; U^*)$, Assumption 2.1.2 is fulfilled as well. Thus, the entire Assumption 2.1 is satisfied by the quantities in (4.20); cf. also Assumption 4.2.

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., $S : L^2(0,T;H^1(\Omega)^*) \ni \ell \mapsto (d,\varphi) \in H_0^1(0,T;L^2(\Omega)) \times L^2(0,T;H^1(\Omega))$ 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 (4.5) and, thus, of (2.5), whence Assumption 2.3.2 follows. From (4.4), the Lipschitz continuity of f, and Gronwall's inequality, 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 (4.20).

It remains to check that Assumptions 2.6 and 2.9 are guaranteed. To this end, we observe that

$$\partial_\varphi \Phi(\bar{d},\bar{\varphi}) = \beta\,\mathbb{I} : L^2(0,T;H^1(\Omega)) \to L^2(0,T;L^2(\Omega)). \tag{4.21}$$

As a result of $L^2(0,T;H^1(\Omega)) \overset{d}{\hookrightarrow} L^2(0,T;L^2(\Omega))$, the constraint qualification in Assumption 2.6 is fulfilled. In the light of Lemma 2.2, the definition of the adjoint, and (4.20c)–(4.20d), the adjoints of the partial derivatives of Φ and Ψ are given by

$$\begin{aligned}
\partial_d \Phi(\bar{d},\bar{\varphi})^* &= -\beta\,\mathbb{I} : L^2(0,T;L^2(\Omega)) \to L^2(0,T;L^2(\Omega)),\\
\partial_\varphi \Phi(\bar{d},\bar{\varphi})^* &= \beta\,\mathbb{I} : L^2(0,T;L^2(\Omega)) \to L^2(0,T;H^1(\Omega)^*),\\
\partial_d \Psi(\bar{d},\bar{\varphi})^* &= -\beta\,\mathbb{I} : L^2(0,T;H^1(\Omega)) \to L^2(0,T;L^2(\Omega)),\\
\partial_\varphi \Psi(\bar{d},\bar{\varphi})^* &= -\alpha\Delta + \beta\,\mathbb{I} : L^2(0,T;H^1(\Omega)) \to L^2(0,T;H^1(\Omega)^*).
\end{aligned} \tag{4.22}$$

Now Lemma 4.3 gives in turn that Assumption 2.9 is true for the setting (4.20).

Thus, we can apply Theorem 2.11, which in combination with (4.22) tells us that there exist unique adjoint states $\xi \in H_T^1(0,T;L^2(\Omega))$ and $w \in L^2(0,T;H^1(\Omega))$ and a unique multiplier $\lambda \in L^2(0,T;L^2(\Omega))$ such that

$$-\dot{\xi} + \beta\lambda - \beta w = \partial_d \mathcal{J}(\bar{d},\bar{\varphi},\bar{\ell}) \ \text{ in } L^2(0,T;L^2(\Omega)), \quad \xi(T) = 0, \tag{4.23a}$$

$$-\beta\lambda - \alpha\Delta w + \beta w = \partial_\varphi \mathcal{J}(\bar{d},\bar{\varphi},\bar{\ell}) \quad \text{in } L^2(0,T;H^1(\Omega)^*), \tag{4.23b}$$

$$\big(\xi(t), f'\big(\Phi(\bar{d}(t),\bar{\varphi}(t));v\big)\big)_{L^2(\Omega)} \geq (\lambda(t),v)_{L^2(\Omega)} \quad \forall\, v \in L^2(\Omega) \ \text{ a.e. in } (0,T), \tag{4.23c}$$

$$w + \partial_\ell \mathcal{J}(\bar{d},\bar{\varphi},\bar{\ell}) = 0 \quad \text{in } L^2(0,T;H^1(\Omega)). \tag{4.23d}$$

It remains to show that (4.23c) implies (4.19c). Here, we recall the abbreviation $\bar{z} := -\beta(\bar{d} - \bar{\varphi})$ and (4.20c), i.e., $\bar{z} = \Phi(\bar{d}, \bar{\varphi})$. An argument based on the fundamental lemma of calculus of variations and the positive homogeneity of the directional derivative w.r.t. direction yields

$$\frac{1}{\epsilon}\xi(t,x)\max{}'(\bar{z}(t,x) - r; 1) \geq \lambda(t,x) \geq -\frac{1}{\epsilon}\xi(t,x)\max{}'(\bar{z}(t,x) - r; -1) \quad \text{a.e. in } (0,T)\times\Omega, \tag{4.24}$$

in view of (4.20b). The desired assertion now follows by distinguishing between the sets $\{(t,x) : \bar{z}(t,x) > r\}$, $\{(t,x) : \bar{z}(t,x) < r\}$, and $\{(t,x) : \bar{z}(t,x) = r\}$. ☐

Remark 4.6. If $\bar{z}(t,x) \neq r$ a.e. in $(0,T)\times\Omega$, then (4.19) reduces to the standard KKT conditions; see (4.19c) and Remark 2.14.

The optimality system in Theorem 4.5 is indeed of strong stationary type, as the next result shows.

THEOREM 4.7 (equivalence between B- and strong stationarity). *Assume that $\bar{\ell} \in L^2(0,T;L^2(\Omega))$ together with its states $(\bar{d}, \bar{\varphi}) \in H_0^1(0,T;L^2(\Omega)) \times L^2(0,T;H^1(\Omega))$, some adjoint states $(\xi, w) \in H_T^1(0,T;L^2(\Omega)) \times L^2(0,T;H^1(\Omega))$, and a multiplier $\lambda \in L^2(0,T;L^2(\Omega))$ satisfy the optimality system* (4.19a)–(4.19d). *Then, it also satisfies the VI*

$$\partial_{(d,\varphi)}\mathcal{J}(\bar{d}, \bar{\varphi}, \bar{\ell})S'(\bar{\ell}; \delta\ell) + \partial_\ell \mathcal{J}(\bar{d}, \bar{\varphi}, \bar{\ell})\delta\ell \geq 0 \quad \forall\, \delta\ell \in L^2(0,T;L^2(\Omega)), \tag{4.25}$$

where $S : L^2(0,T;H^1(\Omega)^*) \to H_0^1(0,T;L^2(\Omega)) \times L^2(0,T;H^1(\Omega))$ is the solution mapping associated with (4.1)*; see Lemma* 4.1.2.

Proof. We show the result by means of Theorem 2.13. In the proof of Theorem 4.5, 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 (4.20). According to the proof of Theorem 4.5, the system (2.19) coincides with (4.23) in this particular setting; see (4.22). We also note that (2.9) is just (4.25). Thus, in view of Theorem 2.13, we only need to show that (4.19c) implies (4.23c), which, in view of (4.20b) and (4.20c), reads

$$\Big(\xi(t), \frac{1}{\epsilon}\max{}'(\bar{z}(t) - r; v)\Big)_{L^2(\Omega)} \geq (\lambda(t), v)_{L^2(\Omega)} \quad \forall\, v \in L^2(\Omega) \text{ a.e. in } (0,T), \tag{4.26}$$

where $\bar{z} := -\beta(\bar{d} - \bar{\varphi})$.

To this end, let $v \in L^2(\Omega)$ be arbitrary, but fixed. From the first identity in (4.19c), we know that

$$\begin{aligned} \lambda(t,x)v(x) &= \frac{1}{\epsilon}\chi_{\{\bar{z}>r\}}(t,x)v(x)\xi(t,x) \\ &= \frac{1}{\epsilon}\max{}'(\bar{z}(t,x) - r)v(x)\xi(t,x) \quad \text{a.e. where } \bar{z}(t,x) \neq r. \end{aligned} \tag{4.27}$$

Further, we define $M^+ := \{(t,x) \in (0,T)\times\Omega : \bar{z}(t,x) = r \text{ and } v(x) > 0\}$ and $M^- := \{(t,x) \in (0,T)\times\Omega : \bar{z}(t,x) = r \text{ and } v(x) \leq 0\}$ (up to sets of measure zero). Then, the second identity in (4.19c) yields

$$\begin{aligned} \lambda(t,x)v(x) &\leq \begin{cases} \frac{1}{\epsilon}\xi(t,x)v(x) & \text{a.e. in } M^+, \\ 0 & \text{a.e. in } M^- \end{cases} \\ &= \frac{1}{\epsilon}\max{}'(\bar{z}(t,x) - r; v(x))\xi(t,x) \quad \text{a.e. where } \bar{z}(t,x) = r. \end{aligned} \tag{4.28}$$

Now, (4.26) follows from (4.27) and (4.28). Note that, since Assumptions 2.6 and 2.9 are fulfilled (cf. proof of Theorem 4.5), we have the equivalence (4.25) $\Longleftrightarrow$ (4.19). ◻

Remark 4.8. An essential need of information resulting from the strong stationary system (4.19) is the sign condition

$$\xi(t,x) \geq 0 \quad \text{a.e. where } \bar{z}(t,x) = r, \tag{4.29}$$

which is due to $[0, \frac{1}{\epsilon}\xi(t,x)] \neq \emptyset$ a.e. where $\bar{z}(t,x) = r$; see (4.19c). This is crucial for showing the implication (4.19) $\Rightarrow$ (4.25), which ultimately yields that (4.19) is indeed of strong stationary type (see (4.28) in the proof of Theorem 4.7). Let us point out that the condition (4.29) is equivalent to the regularity (cf. [40, Def. 7.4.1]) of a mapping involving the adjoint state and the nonsmooth nonlinearity; see (4.32) below. We refer here to a similar situation in [31, Rem. 6.9].

Remark 4.9. Optimality systems derived by classical regularization techniques often lack a sign condition for the adjoint state; see, e.g., [6, Thm. 4.4] and [46, Thm. 2.4] (nonsmooth PDEs) (eventually along with other information which gets lost in the limit analysis associated with the regularization; cf. [31, sect. 4]). This is also the case when it comes to the optimal control of VIs; see [35] for instance. Generally speaking, a sign condition for the adjoint state in points (t,x), where the argument of the nonsmoothness 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 [31, sect. 7.2]. Note that, in our case, the argument of the nonsmooth nonlinearity $f = \frac{1}{\epsilon}\max(\cdot - r, 0)$ (cf. (4.20b)), appearing in the state equation (4.4) is $\bar{z} = -\beta(\bar{d} - \bar{\varphi})$; see also (4.20c).

Discussion of the strong stationary optimality system (4.19). Comparison to known results. In this subsection, we rewrite (4.19c) in terms of a Clarke subdifferential and we explain how the sign condition (4.29) is related to the notion of regular functions; cf. [40, Def. 7.4.1]. This will help us highlight the similarities between (4.19) and other strong stationary optimality systems, which were derived for optimal control of a single nonsmooth PDE [6, 31].

In the following, $f : \mathbb{R} \to \mathbb{R}$ is the mapping defined as

$$f(z) := \frac{1}{\epsilon}\max(z - r, 0)$$

and $\partial_\circ f$ denotes its Clarke subdifferential in the sense of [40, Def. 7.3.4]. Since f is piecewise continuously differentiable, it holds that

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

by [40, Thm. 7.3.12]; see also [31, eq. (C.3)], where $f'_+(z) := f'(z;1)$ and $f'_-(z) := -f'(z;-1)$ denote the right- and left-sided derivative of f at $z \in \mathbb{R}$, respectively. It is then straightforward to see that

$$\partial_\circ f(z) = \begin{cases} \{\frac{1}{\epsilon}\} & \text{if } z > r, \\ [0, \frac{1}{\epsilon}] & \text{if } z = r, \\ \{0\} & \text{if } z < r \end{cases} \qquad \forall\, z \in \mathbb{R}. \tag{4.30}$$

Thus, (4.19c) can be equivalently written by means of the Clarke subdifferential of f as follows:

$$\lambda(t,x) = \gamma(t,x)\xi(t,x) \quad \text{a.e. in } (0,T)\times\Omega, \tag{4.31a}$$

$$\gamma(t,x) \in \partial_\circ f(\bar z(t,x)) \quad \text{a.e. in } (0,T)\times\Omega, \tag{4.31b}$$

$$\xi(t,x) \geq 0 \quad \text{a.e. where } \bar z(t,x) = r, \tag{4.31c}$$

where $\bar z := -\beta(\bar d - \bar\varphi)$.

Let us shortly compare (4.19a)–(4.19b)–(4.31)–(4.19d) with the strong stationary optimality system [6, (32)]. First, let us point out that in [6], the state equation is a nonsmooth elliptic PDE with nonlinearity $\tilde f = \max(\cdot, 0)$. If we insert the relation (4.31a) in the adjoint equation (4.19a)–(4.19b), 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 (4.19d), and (4.31b)–(4.31c). We remark that the latter 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 nonsmooth mapping and a sign condition on the adjoint state in points where the argument of the nonsmoothness f, i.e., $\bar z$, is such that f is not differentiable there (notice that, in [6], the argument of $\tilde f$ is the state associated with the local optimum and $\tilde f$ is not differentiable at 0).

The optimality system (4.19a)–(4.19b)–(4.31)–(4.19d) is also consistent with [31, (6.8)]. In view of [31, Rem. 6.9, eq. (C.3)], the relation [31, (6.8b)] can be expressed in a way similar to (4.31). Note that (4.31c) is equivalent to the fact that the map

$$\mathbb{R} \ni z \mapsto \xi(t,x) f(z) \in \mathbb{R} \tag{4.32}$$

is regular at $\bar z(t,x)$ for a.a. $(t,x) \in (0,T)\times\Omega$, which is a consequence of [31, Lem. C.1] combined with $f'_+(r) > f'_-(r)$ and $f'_+(z) = f'_-(z)$ for all $z \in \mathbb{R}\setminus\{r\}$. A similar observation was made in [31, Rem. 6.9, sect. 7.2], where it was pointed out that the information concerning the regularity or, equivalently, the sign condition on the adjoint state, is the essential feature which gets lost when providing optimality systems by resorting to a regularization approach; see Remark 4.9 also.

We conclude the paper by giving some comments regarding strong stationarity in the context of elasto-viscoplasticity.

Remark 4.10. In the case of elasto-viscoplastic problems, the constraint qualification in Assumption 2.6 is not satisfied. However, optimality conditions of strong stationary type can be proven for a control problem governed by a modified model. Let us go into a little more detail.

According to, e.g., [42, sect. 2.7] (see also [36, Chap. 22] and [18, sect. 7.1]), the elasto-viscoplasticity problem reads

(4.33)

$$\begin{aligned}
&\mathcal{D}(\boldsymbol{\eta}) - \mathcal{D}(\dot{\boldsymbol{p}}(t)) + \delta(\dot{\boldsymbol{p}}(t), \boldsymbol{\eta} - \dot{\boldsymbol{p}}(t))_{L^2(\Omega;\mathbb{R}^{n\times n}_{sym})} \\
&\qquad \geq \big(\mathbb{C}\big(\varepsilon(\boldsymbol{u}(t)) - \boldsymbol{p}(t)\big) - \kappa \boldsymbol{p}(t), \boldsymbol{\eta} - \dot{\boldsymbol{p}}(t)\big)_{L^2(\Omega;\mathbb{R}^{n\times n}_{sym})} \ \forall \boldsymbol{\eta} \in L^2(\Omega;\mathbb{Q}_0), \quad \boldsymbol{p}(0) = \boldsymbol{0}, \\
&-\operatorname{div}\big(\mathbb{C}\big(\varepsilon(\boldsymbol{u}(t)) - \boldsymbol{p}(t)\big)\big) = \ell(t) \quad \text{a.e. in } (0,T),
\end{aligned}$$

where $\boldsymbol{u} : [0,T]\times\Omega \to \mathbb{R}^n$ and $\boldsymbol{p} : [0,T]\times\Omega \to \mathbb{R}^{n\times n}_{sym}$ are the displacement and plastic strain (states), respectively, while ℓ is the applied force (control). In (4.33), $\mathcal{D}$ is a nonsmooth functional, div stands for the distributional divergence, and $\mathbb{Q}_0 :=$

$\{\boldsymbol{\eta} \in \mathbb{R}^{n\times n}_{sym} \mid \operatorname{trace}(\boldsymbol{\eta}) = 0\}$. Moreover, $\delta > 0$ is a given viscosity parameter and $\kappa > 0$ is fixed. For the precise functional analytical setting, we refer to [18, sect. 7.1]. As we will immediately see, the quantities of interest for the constraint qualification are the linearized strain tensor $\varepsilon(\boldsymbol{v}) = \frac{1}{2}(\nabla \boldsymbol{v} + \nabla \boldsymbol{v}^\top)$ and the uniformly coercive elasticity tensor $\mathbb{C} \in L^\infty(\Omega; \mathcal{L}(\mathbb{R}^{n\times n}_{sym}))$. After applying the theory from section 3, the EVI in (4.33) can be rewritten as

$$\dot{\boldsymbol{p}}(t) = f\big(\mathbb{C}\big(\varepsilon(\boldsymbol{u}(t)) - \boldsymbol{p}(t)\big) - \kappa \boldsymbol{p}(t)\big) \quad \text{a.e. in } (0,T), \tag{4.34}$$

where f is a nonsmooth nonlinearity. Thus, (4.33) appears to be a problem which fits in the setting of section 2. In this context, the constraint qualification in Assumption 2.6 requires that $\operatorname{Rg} \mathbb{C}\varepsilon$ is dense in $L^2(0,T; L^2(\Omega; \mathbb{R}^{n\times n}_{sym}))$, which is not true.

However, let us emphasize that, if the control no longer appears in the elliptic PDE in (4.33), but on the r.h.s. of the EVI, e.g., as a thermal stress (although, an additional control in the elliptic PDE can be considered as well), then strong stationarity conditions can be provided. This situation is similar to the one in [23, sect. 4], where an additional control has to be considered on the r.h.s. of the respective VI in order to be able to prove strong stationarity. In our modified scenario, the problem can be equivalently written as

$$\begin{aligned} &\dot{\boldsymbol{p}}(t) = f\big(\mathbb{C}\big(\varepsilon(\boldsymbol{u}(t)) - \boldsymbol{p}(t)\big) - \kappa \boldsymbol{p}(t) + \ell(t)\big), \quad \boldsymbol{p}(0) = \mathbf{0}, \\ &-\operatorname{div}\big(\mathbb{C}\big(\varepsilon(\boldsymbol{u}(t)) - \boldsymbol{p}(t)\big)\big) = 0 \quad \text{a.e. in } (0,T). \end{aligned} \tag{4.35}$$

Then, by arguing as in the proof of Lemma 2.8, one can show that the set of directions into which f is differentiated—in the linearized state equation—is dense in the space $L^2(0,T; L^2(\Omega; \mathbb{R}^{n\times n}_{sym}))$. Recall that this is the crucial aspect when deriving strong stationarity; cf. Remark 2.12(i). Indeed, (4.35) does not fit in our general setting from section 2, so that the constraint qualification in Assumption 2.6 does not come into play here. Nevertheless, as mentioned above, the basic idea can be transferred; see also Remark 2.12(ii).

Appendix A. Directional differentiability of $\mathcal{F}$ (section 3). In this section, Assumption 3.1.1–3.1.2 is supposed to hold, while $\mathcal{F}$ and the (metric) projection 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 nonempty, 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 nonempty, closed, convex cone. Thanks to Corollary 3.4, the proof is now complete. Alternatively, the assertion of this lemma can be deduced from [19, Thm. 2] and Corollary 3.4. $\square$

Some canonical examples. For details regarding polyhedric sets and their properties, we refer to the contributions [4, 49]. 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 [49, 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 following, $r \geq 0$ 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. [45] and [30, sect. 4]). If $Y = L^2(\Omega)$ or $Y = H^1(\Omega)$ 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 see next.

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

$$\partial R(0) = \{\mu \in L^2(\Omega) | \, \mu \leq r \text{ a.e. in } \Omega\} \tag{A.4}$$

and [4, 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 $Y = H^1(\Omega)$. Then,

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

where $M := \{v \in H^1(\Omega) | \, v \geq 0 \text{ a.e. in } \Omega\}$. It is well known that $M \subset Y$ is polyhedric (see, e.g., [19, 34]), and by [49, 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

$$\langle P_{\partial R(0)}\bar{\omega} - r, \mathcal{F}(\bar{\omega})\rangle_Y = 0. \tag{A.5}$$

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 plasticity [24] and sweeping processes [16]). If $Y = L^2(\Omega)$ and

$$R(\eta) = r \int_\Omega |\eta| \, dx \quad \forall \, \eta \in Y, \tag{A.6}$$

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

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

and [4, 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; cf. [7, Assumption 4.3]. Such conditions ensure the directional differentiability of $\mathcal{F} : Y^* \to Y$ (with $Y = H_0^1(\Omega)$, R as in (A.6), and $\mathcal{V} := -\triangle$), in view of Lemma 3.3.

Appendix B. Improved regularity of the multiplier $\boldsymbol{\lambda}$ (section 4).

Proof of Lemma 4.3. We resort to a classical regularization approach; see [1], for instance. We define a smooth approximation of the function $\max(\cdot, 0)$, with which we associate a state equation where the solution mapping is Gâteaux differentiable (step (I) below). Then, by arguments inspired by, e.g., [1, 35], it follows that $\bar{\ell}$ can be approximated by a sequence of local minimizers of an optimal control problem governed by the regularized state equation (step (II) below). Passing to the limit in the adjoint system associated with the regularized optimal control problem finally yields the desired assertion (step (III) below). Although many of the arguments are well known, we give a detailed proof, for the convenience of the reader.

(I) Let $\varepsilon > 0$ be arbitrary, but fixed. We begin by defining the smooth PDE system

$$\dot{d}(t) = \frac{1}{\epsilon} \max{}_\varepsilon(-\beta(d(t) - \varphi(t)) - r) \;\text{ in } L^2(\Omega), \quad d(0) = 0, \tag{B.1a}$$

$$-\alpha \Delta \varphi(t) + \beta \, \varphi(t) = \beta d(t) + \ell(t) \quad \text{in } H^1(\Omega)^* \quad \text{a.e. in } (0, T), \tag{B.1b}$$

where, for instance,

$$\max{}_\varepsilon : \mathbb{R} \to \mathbb{R}, \quad \max{}_\varepsilon(x) := \begin{cases} 0, & x \le 0, \\ \frac{1}{2\varepsilon} x^2, & x \in (0, \varepsilon), \\ x - \frac{\varepsilon}{2}, & x \ge \varepsilon; \end{cases}$$

cf. [35]. It is straightforward to see that $\max_\varepsilon : L^2(\Omega) \to L^2(\Omega)$ is Lipschitz continuous with constant 1, Gâteaux differentiable, and

$$|\max{}_\varepsilon(x) - \max(x, 0)| \le \varepsilon \quad \forall \, x \in \mathbb{R}. \tag{B.2}$$

By the Lax–Milgram lemma and the Picard–Lindelöf theorem, one infers that (B.1) admits a unique solution $(d, \varphi) \in H_0^1(0, T; L^2(\Omega)) \times L^2(0, T; H^1(\Omega))$ for every $\ell \in L^2(0, T; H^1(\Omega)^*)$, which allows us to define the regularized solution mapping

$$S_\varepsilon : L^2(0, T; H^1(\Omega)^*) \ni \ell \mapsto (d, \varphi) \in H_0^1(0, T; L^2(\Omega)) \times L^2(0, T; H^1(\Omega)).$$

The operator S_ε is Gâteaux differentiable (by standard arguments) and its derivative at $\ell \in L^2(0,T;H^1(\Omega)^*)$ in direction $\delta\ell \in L^2(0,T;H^1(\Omega)^*)$, i.e., $(\delta d, \delta\varphi) := S_\varepsilon'(\ell)(\delta\ell)$, is the unique solution of

(B.3)

$$\begin{aligned}\dot{\delta d}(t) &= \frac{1}{\epsilon} \max{}_\varepsilon{}'(-\beta(d(t)-\varphi(t)) - r)(-\beta(\delta d(t) - \delta\varphi(t))) \quad \text{in } L^2(\Omega), \quad \delta d(0) = 0,\\ &- \alpha\Delta\delta\varphi(t) + \beta\,\delta\varphi(t) = \beta\delta d(t) + \delta\ell(t) \quad \text{in } H^1(\Omega)^* \quad \text{a.e. in } (0,T),\end{aligned}$$

where we abbreviate $(d,\varphi) := S_\varepsilon(\ell)$. As a result of (B.2), Gronwall's inequality, the Lipschitz continuity of $\max_\varepsilon$ (with constant 1) and $\max(\cdot,0)$, we deduce that $S_\varepsilon, S : L^2(0,T;H^1(\Omega)^*) \to H_0^1(0,T;L^2(\Omega)) \times L^2(0,T;H^1(\Omega))$ are Lipschitz continuous (with constant independent of ε), as well as the convergence

(B.4)

$$S_\varepsilon(\ell_\varepsilon) - S(\ell_\varepsilon) \to 0 \quad \text{in } H_0^1(0,T;L^2(\Omega)) \times L^2(0,T;H^1(\Omega)) \;\; \forall\, \ell_\varepsilon \in L^2(0,T;H^1(\Omega)^*).$$

(II) Next, we focus on proving that $\bar\ell$ can be approximated via local minimizers of optimal control problems governed by (B.1). Since the operator $S_\varepsilon : L^2(0,T;L^2(\Omega)) \to L^2(0,T;L^2(\Omega) \times H^1(\Omega))$ is not necessarily weakly continuous (which is crucial for the existence of solutions for the regularized optimal control problem), we will work with the control space $H^1(0,T;H^1(\Omega)) \hookrightarrow\hookrightarrow L^2(0,T;L^2(\Omega))$; see (Q_n) below. To this end, let $B_{L^2(0,T;L^2(\Omega))}(\bar\ell,\rho)$ be the ball of local optimality of $\bar\ell$ and $\{\ell_n\} \subset H^1(0,T;H^1(\Omega))$ such that

$$\ell_n \to \bar\ell \quad \text{in } L^2(0,T;L^2(\Omega)) \;\; \text{as } n \to \infty. \tag{B.5}$$

Note that the existence of such a sequence is due to $H^1(0,T;H^1(\Omega)) \overset{d}{\hookrightarrow} L^2(0,T;L^2(\Omega))$. For $n \in \mathbb{N}$ fixed (large enough) with

$$\|\ell_n - \bar\ell\|_{L^2(0,T;L^2(\Omega))} \le \rho/2, \tag{B.6}$$

we consider the smooth (reduced) optimal control problem

$$(Q_n) \qquad \left.\begin{aligned} \min_{\ell \in H^1(0,T;H^1(\Omega))} \quad & \mathcal{J}(S_{1/n}(\ell),\ell) + \frac{1}{2}\|\ell - \ell_n\|^2_{H^1(0,T;H^1(\Omega))}\\ \text{s.t.} \quad & \ell \in B_{H^1(0,T;H^1(\Omega))}(\ell_n, \rho/2). \end{aligned}\right\}$$

By employing the direct method of the calculus of variations along with the continuity of $S_{1/n} : L^2(0,T;L^2(\Omega)) \to L^2(0,T;L^2(\Omega) \times H^1(\Omega))$ and the compact embedding $H^1(0,T;H^1(\Omega)) \hookrightarrow\hookrightarrow L^2(0,T;L^2(\Omega))$, we see that (Q_n) admits a global solution $g_n \in H^1(0,T;H^1(\Omega))$. Since this is admissible for (Q_n), we have

$$\|g_n - \bar\ell\|_{L^2(0,T;L^2(\Omega))} \le \rho \tag{B.7}$$

as a consequence of (B.6). For simplicity, we abbreviate in the following:

$$j(\ell) := \mathcal{J}(S(\ell),\ell) \quad \forall\, \ell \in L^2(0,T;L^2(\Omega)), \tag{B.8a}$$

$$j_n(\ell) := \mathcal{J}(S_{1/n}(\ell),\ell) + \frac{1}{2}\|\ell - \ell_n\|^2_{H^1(0,T;H^1(\Omega))} \quad \forall\, \ell \in H^1(0,T;H^1(\Omega). \tag{B.8b}$$

Due to (B.4), the continuity of S, and (B.5), it holds that

(B.9)

$$j(\bar\ell) \overset{\text{(B.8a)}}{=} \mathcal{J}(S(\bar\ell),\bar\ell) = \lim_{n\to\infty} \mathcal{J}(S_{1/n}(\ell_n),\ell_n) \overset{\text{(B.8b)}}{=} \lim_{n\to\infty} j_n(\ell_n) \ge \limsup_{n\to\infty} j_n(g_n),$$

where for the last inequality we relied on the fact that g_n is a global minimizer of (Q_n) and that ℓ_n is admissible for (Q_n). By the definition of j_n, (B.9) can be continued as:

(B.10)
$$\begin{aligned} j(\bar\ell) &\geq \limsup_{n\to\infty} \mathcal{J}(S_{1/n}(g_n), g_n) + \underbrace{\frac{1}{2}\|g_n - \ell_n\|^2_{H^1(0,T;H^1(\Omega))}}_{=:\tilde h_n} \\ &= \limsup_{n\to\infty} \underbrace{\mathcal{J}(S(g_n), g_n)}_{\geq j(\bar\ell)} + \tilde h_n \geq j(\bar\ell) + \limsup_{n\to\infty} \tilde h_n \geq j(\bar\ell) + \liminf_{n\to\infty} \tilde h_n \geq j(\bar\ell). \end{aligned}$$

The identity in (B.10) is a result of (4.17) combined with (B.7) and the global Lipschitz continuity of S and $S_{1/n}$ (with constant independent of n), as well as (B.4). The second inequality in (B.10) follows from (B.7) and (B.8a) (recall that $B_{L^2(0,T;L^2(\Omega))}(\bar\ell, \rho)$ is the ball of local optimality of $\bar\ell$). Now, the series of estimates in (B.10) together with (B.5) gives in turn the convergences

$$g_n \to \bar\ell \quad \text{in } L^2(0,T;L^2(\Omega)) \quad \text{as } n \to \infty, \tag{B.11a}$$
$$g_n - \ell_n \to 0 \quad \text{in } H^1(0,T;H^1(\Omega)) \quad \text{as } n \to \infty. \tag{B.11b}$$

It remains to show that g_n is a local solution for $\min_{\ell \in H^1(0,T;H^1(\Omega))} j_n(\ell)$. To this end, let $\ell \in B_{H^1(0,T;H^1(\Omega))}(g_n, \rho/4)$ be arbitrary, but fixed. Since (B.11b) yields $\|g_n - \ell_n\|_{H^1(0,T;H^1(\Omega))} \leq \rho/4$ for n large enough, we infer that ℓ is admissible for (Q_n), whence

$$j_n(g_n) \leq j_n(\ell) \quad \forall\, \ell \in B_{H^1(0,T;H^1(\Omega))}(g_n, \rho/4)$$

follows by the global optimality of g_n. Thus, g_n is indeed a local solution for $\min_{\ell \in H^1(0,T;H^1(\Omega))} j_n(\ell)$ and, in view of (B.11a), this step of the proof is complete.

(III) In the following, $n \in \mathbb{N}$ remains fixed and large enough. Due to the above established local optimality of g_n, we have $j_n'(g_n)(\delta\ell) = 0$ for all $\delta\ell \in H^1(0,T;H^1(\Omega))$, on account of the differentiability properties of $S_{1/n}$; cf. step (I), and $\mathcal{J}$, see Assumption 4.2 (recall the definition of j_n in (B.8b)). This implies

$$\begin{aligned} &\partial_{(d,\varphi)}\mathcal{J}(S_{1/n}(g_n), g_n)(S'_{1/n}(g_n)(\delta\ell)) \\ &\qquad + \partial_\ell \mathcal{J}(S_{1/n}(g_n), g_n)\delta\ell + (g_n - \ell_n, \delta\ell)_{H^1(0,T;H^1(\Omega))} = 0 \end{aligned} \tag{B.12}$$

for all $\delta\ell \in H^1(0,T;H^1(\Omega))$. On the other hand, the system

$$\begin{aligned} -\dot\xi_n(t) - \beta\left(w_n(t) - \frac{1}{\epsilon}\max{}_{1/n}'(-\beta(d_n(t) - \varphi_n(t)) - r)\xi_n(t)\right) \\ = \partial_d \mathcal{J}(S_{1/n}(g_n), g_n)(t), \quad \xi_n(T) = 0, \end{aligned} \tag{B.13a}$$

$$\begin{aligned} -\alpha\Delta w_n(t) + \beta\left(w_n(t) - \frac{1}{\epsilon}\max{}_{1/n}'(-\beta(d_n(t) - \varphi_n(t)) - r)\xi_n(t)\right) \\ = \partial_\varphi \mathcal{J}(S_{1/n}(g_n), g_n)(t) \end{aligned} \tag{B.13b}$$

a.e. in $(0,T)$, where we abbreviate $(d_n, \varphi_n) := S_{1/n}(g_n)$, admits a unique solution $(\xi_n, w_n) \in H^1_T(0,T;L^2(\Omega)) \times L^2(0,T;H^1(\Omega))$, by standard arguments (see, e.g., the proof of [45, Lem. 5.7]). Note that

$$\partial_d \mathcal{J}(S_{1/n}(g_n), g_n) \to \partial_d \mathcal{J}(S(\bar\ell), \bar\ell) \quad \text{in } L^2(0,T;L^2(\Omega)), \tag{B.14a}$$
$$\partial_\varphi \mathcal{J}(S_{1/n}(g_n), g_n) \to \partial_\varphi \mathcal{J}(S(\bar\ell), \bar\ell) \quad \text{in } L^2(0,T;H^1(\Omega)^*), \tag{B.14b}$$
$$\partial_\ell \mathcal{J}(S_{1/n}(g_n), g_n) \to \partial_\ell \mathcal{J}(S(\bar\ell), \bar\ell) \quad \text{in } L^2(0,T;L^2(\Omega)), \tag{B.14c}$$

in the light of the continuous Fréchet differentiability of $\mathcal{J}$ (Assumption 4.2), the convergences (B.11a), (B.4), and the Lipschitz continuity of S established in step (I). Hence, by the global Lipschitz continuity of $\max_{1/n}$ (with constant 1) and Gronwall's inequality, we can deduce from (B.13) that there exists a constant $c > 0$, independent of n, such that $\|w_n\|_{L^2(0,T;H^1(\Omega))} \leq c$ and $\|\xi_n\|_{C([0,T];L^2(\Omega))} \leq c$. As a consequence, $\lambda_n := \frac{1}{\epsilon} {\max}_{1/n}{}'(-\beta(d_n - \varphi_n) - r)\xi_n$ is uniformly bounded in $L^2(0,T;L^2(\Omega))$. Therefore, we can extract weakly convergent subsequences (denoted by the same symbol) so that

$$\text{(B.15)} \qquad w_n \rightharpoonup w \quad \text{in } L^2(0,T;H^1(\Omega)), \quad \lambda_n \rightharpoonup \lambda \quad \text{in } L^2(0,T;L^2(\Omega)) \quad \text{as } n \to \infty.$$

Testing (B.13) with $S'_{1/n}(g_n)(\delta\ell)$ and (B.3) with (ξ_n, w_n) yields

$$(w_n, \delta\ell)_{L^2(0,T;L^2(\Omega))} = \partial_{(d,\varphi)}\mathcal{J}(S_{1/n}(g_n), g_n)(S'_{1/n}(g_n)(\delta\ell)),$$

which inserted in (B.12) gives

$$\text{(B.16)} \qquad \big(w_n + \partial_\ell \mathcal{J}(S_{1/n}(g_n), g_n), \delta\ell\big)_{L^2(0,T;L^2(\Omega))} + (g_n - \ell_n, \delta\ell)_{H^1(0,T;H^1(\Omega))} = 0$$

for all $\delta\ell \in H^1(0,T;H^1(\Omega))$. Owing to (B.15), (B.14c), and (B.11b), we can pass to the limit $n \to \infty$ in (B.16). This results in

$$\text{(B.17)} \qquad (w + \partial_\ell \mathcal{J}(S(\bar{\ell}), \bar{\ell}), \delta\ell)_{L^2(0,T;L^2(\Omega))} = 0 \quad \forall\, \delta\ell \in H^1(0,T;H^1(\Omega)).$$

By employing a density argument in (B.17) along with passage to the limit in (B.13b) (where we rely on (B.15) and (B.14b)), we arrive at the desired result. ∎

Acknowledgments. We thank the anonymous referees for their helpful comments and for pointing out an inaccuracy in an earlier draft of the manuscript. The author is grateful to Christian Meyer for fruitful discussions on strong stationarity.

REFERENCES

[1] V. Barbu, *Necessary conditions for distributed control problems governed by parabolic variational inequalities*, SIAM J. Control Optim., 19 (1981), pp. 64–86.

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

[3] L. Betz and I. Yousept, *Optimal Control of Elliptic Variational Inequalities with Bounded and Unbounded Operators*, https://duepublico.uni-duisburg-essen.de/servlets/DocumentServlet?id=46819 (2018).

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

[5] C. Christof, *Sensitivity analysis and optimal control of obstacle-type evolution variational inequalities*, SIAM J. Control Optim., 57 (2019), pp. 192–218.

[6] C. Christof, C. Clason, C. Meyer, and S. Walther, *Optimal control of a non-smooth, semilinear elliptic equation*, Math. Control Relat. Fields, 8 (2018), pp. 247–276.

[7] C. Christof and C. Meyer, *Sensitivity analysis for a class of H_0^1-elliptic variational inequalities of the second kind*, Set-Valued Var. Anal., 27 (2018), pp. 469–502. https://doi.org/10.1007/s11228-018-0495-2.

[8] C. Christof and G. Wachsmuth, *On the non-polyhedricity of sets with upper and lower bounds in dual spaces*, GAMM-Mitt., 40 (2018), pp. 339–350.

[9] C. Clason, V. Nhu, and A. Rösch, *Optimal control of a non-smooth quasilinear elliptic equation*, Math. Control Relat. Fields, to appear.

[10] 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*, J. Optim. Theory Appl., 168 (2016), pp. 375–409.

[11] B. Dimitrijevic and K. Hackl, *A method for gradient enhancement of continuum damage models*, Tech. Mech., 28 (2008), pp. 43–52.
[12] E. Emmrich, *Gewöhnliche und Operator Differentialgleichungen*, Vieweg, Wiesbaden, Germany, 2004.
[13] M. Farshbaf-Shaker, *A penalty approach to optimal control of Allen-Cahn variational inequalities: MPEC-view*, Numer. Funct. Anal. Optim., 33 (2012), pp. 1321–1349.
[14] A. Friedman, *Optimal control for parabolic variational inequalities*, SIAM J. Control Optim., 25 (1987), pp. 482–497.
[15] C. Geiger and C. Kanzow, *Theorie und Numerik restringierter Optimierungsaufgaben*, Springer, Berlin, 2002.
[16] T. Geiger and D. Wachsmuth, *Optimal Control of an Evolution Equation with Non-Smooth Dissipation*, preprint, https://arxiv.org/abs/1801.04077 (2018).
[17] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer, Berlin, 2008.
[18] W. Han and B. Reddy, *Plasticity. Mathematical Theory and Numerical Analysis*, Springer, New York, 1999.
[19] A. Haraux, *How to differentiate the projection on a convex set in Hilbert space. Some applications to variational inequalities*, J. Math. Soc. Japan, 29 (1977), pp. 615–631.
[20] Z.-X. He, *State constrained control problems governed by variational inequalities*, SIAM J. Control Optim., 25 (1987), pp. 1119–1144.
[21] R. Herzog and C. Meyer, *Optimal control of static plasticity with linear kinematic hardening*, J. Appl. Math. Mech., 91 (2011), pp. 777–794.
[22] R. Herzog, C. Meyer, and G. Wachsmuth, *C-stationarity for optimal control of static plasticity with linear kinematic hardening*, SIAM J. Optim., 50 (2012), pp. 3052–3082.
[23] R. Herzog, C. Meyer, and G. Wachsmuth, *B- and strong stationarity for optimal control of static plasticity with hardening*, SIAM J. Optim., 23 (2013), pp. 321–352.
[24] R. Herzog and A. Stötzner, *Hadamard differentiability of the solution map in thermoviscoplasticity*, Pure Appl. Funct. Anal., 4 (2019), pp. 271–295.
[25] M. Hintermüller and I. Kopacka, *Mathematical programs with complementarity constraints in function space: C- and strong stationarity and a path-following algorithm*, SIAM J. Optim., 20 (2009), pp. 868–902, https://doi.org/10.1137/080720681.
[26] M. Hintermüller and T. Surowiec, *On the directional differentiability of the solution mapping for a class of variational inequalities of the second kind*, Set-Valued Var. Anal., 26 (2018), pp. 631–642.
[27] M. Hintermüller and D. Wegner, *Optimal control of a semidiscrete Cahn–Hilliard–Navier–Stokes system*, SIAM J. Control Optim., 52 (2014), pp. 747–772.
[28] A. D. Ioffe and V. M. Tihomirov, *Theory of Extremal Problems*, North-Holland, Amsterdam, 1979.
[29] K. Ito and K. Kunisch, *Optimal control of parabolic variational inequalities*, J. Math. Pures Appl. (9), 93 (2010), pp. 329–360, https://doi.org/10.1016/j.matpur.2009.10.005.
[30] D. Knees, R. Rossi, and C. Zanini, *A Vanishing Viscosity Approach to a Rate-Independent Damage Model*, Technical report 1633, WIAS, Berlin, 2011.
[31] C. Meyer and L. M. Susu, *Optimal control of nonsmooth, semilinear parabolic equations*, SIAM J. Control Optim., 55 (2017), pp. 2206–2234.
[32] C. Meyer and L. Susu, *Analysis of a viscous two-field gradient damage model. Part* I*: Existence and uniqueness*, Z. Anal. Anwend., 38 (2019), pp. 249–286.
[33] C. Meyer and L. Susu, *Analysis of a viscous two-field gradient damage model. Part* II*: Penalization limit*, Z. Anal. Anwend., 38 (2019), pp. 439–474.
[34] F. Mignot, *Contrôle dans les inéquations variationelles elliptiques*, J. Funct. Anal., 22 (1976), pp. 130–185.
[35] F. Mignot and J. P. Puel, *Optimal control in some variational inequalities*, SIAM J. Control Optim., 22 (1984), pp. 466–476.
[36] N. Ottosen and M. Ristinmaa, *The Mechanics of Constitutive Modeling*, Elsevier, Amsterdam, 2005.
[37] R. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[38] T. Roubíček, *Nonlinear Partial Differential Equations with Applications*, Internat. Ser. Numer. Math. 153, Birkhäuser, Basel, 2005.
[39] H. Scheel and S. Scholtes, *Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity*, Math. Oper. Res., 25 (2000), pp. 1–22, https://doi.org/10.1287/moor.25.1.1.15213.
[40] W. Schirotzek, *Nonsmooth Analysis*, Springer, Berlin, 2007.
[41] A. Shapiro, *On concepts of directional differentiability*, J. Optim. Theory Appl., 66 (1990), pp. 477–487.

[42] J. SIMO AND T. HUGHES, *Computational Inelasticity*, Springer, New York, 1998.
[43] M. SOFONEA AND M. ANDALUZIA, *Variational Inequalities with Applications*, Springer, New York, 2009.
[44] J. SOKOŁOWSKI, *Sensitivity analysis of contact problems with prescribed friction*, Appl. Math. Optim., 18 (1988), pp. 99–117.
[45] L. SUSU, *Optimal control of a viscous two-field gradient damage model*, GAMM-Mitt., 40 (2018), pp. 287–311.
[46] D. TIBA, *Optimal Control of Nonsmooth Distributed Parameter Systems*, Springer, Berlin, 1990.
[47] F. TRÖLTZSCH, *Optimal Control of Partial Differential Equations*, Grad. Stud. Math. 112, Amer. Math. Soc., Providence, RI, 2010.
[48] G. WACHSMUTH, *Strong stationarity for optimal control of the obstacle problem with control constraints*, SIAM J. Optim., 24 (2014), pp. 1914–1932.
[49] G. WACHSMUTH, *A guided tour of polyhedric sets*, J. Convex Anal., 26 (2019), pp. 153–188.