Zeitschrift für Analysis und ihre Anwendungen
Journal of Analysis and its Applications
Volume 34 (2015), 391–418
DOI: 10.4171/ZAA/1546

Optimal Control of Quasistatic Plasticity with Linear Kinematic Hardening II: Regularization and Differentiability

Gerd Wachsmuth

Abstract. We consider an optimal control problem governed by an evolution variational inequality arising in quasistatic plasticity with linear kinematic hardening. A regularization of the time-discrete problem is derived. The regularized forward problem can be interpreted as system of coupled quasilinear PDEs whose principal parts depend on the gradient of the state. We show the Fréchet differentiability of the solution map of this quasilinear system. As a consequence, we obtain a first order necessary optimality system. Moreover, we address certain convergence properties of the regularization.

Keywords. Complementarity condition, quasistatic plasticity, time-dependent variational inequality, mathematical program with complementarity constraints, evolution variational inequality, rate-independent

Mathematics Subject Classification (2010). Primary 49J40, secondary 74P10, 74C05

1. Introduction

In this paper we consider an optimal control problem for the problem of small-strain, quasistatic elastoplasticity. Since the forward problem is governed by a time-dependent variational inequality (VI), see [8, Chapter 8], the control-to-state map is not, in general, differentiable. Moreover, it is known from finite-dimensional problems, that the associated Karush-Kuhn-Tucker system is not a necessary optimality system for optimization problems constrained by a VI.

Due to these reasons, one considers regularizations of VIs, see [2]. Any reasonable regularization has to possess two important properties: its control-to-state map should be differentiable and the solutions of the regularized problems should converge towards the solution of the unregularized problem. If both

G. Wachsmuth: Technische Universität Chemnitz, Faculty of Mathematics, Professorship Numerical Methods (Partial Differential Equations), 09107 Chemnitz, Germany; gerd.wachsmuth@mathematik.tu-chemnitz.de

properties are satisfied, one may be able to pass to the limit in the first order necessary optimality system of the regularized problems in order to obtain a optimality system for the unregularized problem, see, e.g., [2,12,14] for optimal control of the obstacle problem, and [11] for optimal control of static plasticity.

The aim of this paper is to provide a regularization scheme for the time-discrete system of quasistatic plasticity. We show the Fréchet differentiability of the regularized solution map (Subsection 3.2) as well as the approximability of unregularized optimal controls by solutions of the regularized control problems (Subsection 4.2). The passage to the limit with the regularization parameter in the full optimality system goes beyond the scope of this paper and is addressed in a subsequent publication, see [19].

We will see in Subsection 3.1, see in particular (28), that the regularized forward equation is a system of *quasilinear* PDEs whose principal parts depend on the gradient of the state. This renders the analysis of the regularized optimal control problem challenging. Note that in the simpler case of the obstacle problem, regularization leads to a *semilinear* PDE, whose analysis is simpler than it is here.

The paper is organized as follows. In the remainder of the introduction we present the notation (Subsection 1.1), the time-discrete forward problem (Subsection 1.2), and the optimal control problem (Subsection 1.3). Details on the derivation of the time-discrete forward problem of quasistatic plasticity can be found in [8, Chapter 8] and [17, Section 1]. The regularization is derived in Section 2. By applying the abstract differentiability result of Appendix A, we show the Fréchet differentiability of the solution map of the regularized forward problem in Section 3. Finally, we address the convergence of the regularization in Section 4.

1.1. Notation and standing assumptions. Our notation follows [8].

Function spaces. Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain with boundary $\Gamma = \partial\Omega$ in dimension $d = 3$. The boundary consists of two disjoint parts Γ_N and Γ_D. Concerning the assumptions on the regularity of Ω, we refer to Assumption 1.1. We point out that the presented analysis is not restricted to the case $d = 3$, but for reasons of physical interpretation we focus on the three dimensional case. In dimension $d = 2$, the interpretation of the forward equation has to be slightly modified, depending on whether one considers the plane strain or plane stress formulation.

By $\mathbb{S} := \mathbb{R}^{d\times d}_{\text{sym}}$ we denote the space of symmetric d-by-d matrices, endowed with the inner product $\boldsymbol{\sigma} : \boldsymbol{\tau} = \sum_{i,j=1}^{d} \sigma_{ij}\tau_{ij}$, and we define

$$V = H^1_D(\Omega;\mathbb{R}^d) = \{\boldsymbol{u} \in H^1(\Omega;\mathbb{R}^d) : \boldsymbol{u} = \boldsymbol{0} \text{ on } \Gamma_D\}, \quad S = L^2(\Omega;\mathbb{S})$$

as the spaces for the displacement $\boldsymbol{u}$; and for the stress $\boldsymbol{\sigma}$, and back stress $\boldsymbol{\chi}$, respectively. The generalized stress $\boldsymbol{\Sigma} = (\boldsymbol{\sigma}, \boldsymbol{\chi})$ belongs to S^2. The control $\boldsymbol{g}$

acts as a boundary force and belongs to the space

$$U = L^2(\Gamma_N; \mathbb{R}^d).$$

The control operator $E : U \to V'$, $\boldsymbol{g} \mapsto \ell$, which maps boundary forces (i.e., controls) $\boldsymbol{g} \in U$ to functionals (i.e., right-hand sides of the forward equation) $\ell \in V'$ is given by

$$\langle \boldsymbol{v},\, E\boldsymbol{g}\rangle_{V,V'} := -\int_{\Gamma_N} \boldsymbol{v} \cdot \boldsymbol{g}\, \mathrm{d}s \quad \text{for all } \boldsymbol{v} \in V. \tag{1}$$

Hence, E is the negative adjoint of the trace operator from V to $U = L^2(\Gamma_N; \mathbb{R}^d)$. Clearly, $E : U \to V'$ is compact.

Yield function and admissible stresses. We restrict our discussion to the von Mises yield function. In the context of linear kinematic hardening, it reads

$$\phi(\boldsymbol{\Sigma}) = \frac{1}{2}\big(|\boldsymbol{\sigma}^D + \boldsymbol{\chi}^D|^2 - \tilde{\sigma}_0^2\big) \tag{2}$$

for $\boldsymbol{\Sigma} = (\boldsymbol{\sigma}, \boldsymbol{\chi}) \in S^2$, where $|\cdot|$ denotes the pointwise Frobenius norm of matrices and

$$\boldsymbol{\sigma}^D = \boldsymbol{\sigma} - \frac{1}{d}\,(\operatorname{trace} \boldsymbol{\sigma})\, \boldsymbol{I} \tag{3}$$

is the deviatoric part of $\boldsymbol{\sigma}$. Here, $\boldsymbol{I} \in \mathbb{S}$ is the identity matrix. The yield function gives rise to the set of admissible generalized stresses

$$\mathcal{K} = \{\boldsymbol{\Sigma} \in S^2 : \phi(\boldsymbol{\Sigma}) \le 0 \text{ almost everywhere (a.e.) in } \Omega\}.$$

Let us mention that the structure of the yield function ϕ given in (2) implies the *shift invariance*

$$\boldsymbol{\Sigma} \in \mathcal{K} \quad \Leftrightarrow \quad \boldsymbol{\Sigma} + (\boldsymbol{\tau}, -\boldsymbol{\tau}) \in \mathcal{K} \quad \text{for all } \boldsymbol{\tau} \in S. \tag{4}$$

This property is exploited quite often in the analysis.

Due to the structure of the yield function ϕ, $\boldsymbol{\sigma}^D + \boldsymbol{\chi}^D$ appears frequently and we abbreviate it and its adjoint by

$$\mathcal{D}\boldsymbol{\Sigma} = \boldsymbol{\sigma}^D + \boldsymbol{\chi}^D \quad \text{and} \quad \mathcal{D}^\star\boldsymbol{\sigma} = \begin{pmatrix} \boldsymbol{\sigma}^D \\ \boldsymbol{\sigma}^D \end{pmatrix} \tag{5}$$

for matrices $\boldsymbol{\Sigma} \in \mathbb{S}^2$ as well as for functions $\boldsymbol{\Sigma} \in S^2$ and tuples of functions $\boldsymbol{\Sigma} \in (S^2)^N$. When considered as an operator in function space, $\mathcal{D}$ maps S^2 and $(S^2)^N$ continuously into S and S^N, respectively. For later reference, we also remark that

$$\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\sigma}^D + \boldsymbol{\chi}^D \\ \boldsymbol{\sigma}^D + \boldsymbol{\chi}^D \end{pmatrix} \quad \text{and} \quad (\mathcal{D}^\star\mathcal{D})^2 = 2\,\mathcal{D}^\star\mathcal{D}$$

holds. Due to the definition of the operator $\mathcal{D}$, the constraint $\phi(\mathbf{\Sigma}) \le 0$ can be formulated as $\|\mathcal{D}\mathbf{\Sigma}\|_{L^\infty(\Omega;\mathbb{S})} \le \tilde\sigma_0$. Hence, we obtain

$$\mathbf{\Sigma} \in \mathcal{K} \quad \Rightarrow \quad \mathcal{D}\mathbf{\Sigma} \in L^\infty(\Omega;\mathbb{S}).$$

Here and in the sequel we denote linear operators, e.g., $\mathcal{D} : \mathbb{S}^2 \to \mathbb{S}$, and the induced Nemytskii operators, e.g., $\mathcal{D} : S^2 \to S$ and $\mathcal{D} : (S^2)^N \to S^N$, with the same symbol. This will cause no confusion, since the meaning will be clear from the context.

Operators. The linear operators $A : S^2 \to S^2$ and $B : S^2 \to V'$ are defined as follows. For $\mathbf{\Sigma} = (\boldsymbol{\sigma}, \boldsymbol{\chi}) \in S^2$ and $\boldsymbol{T} = (\boldsymbol{\tau}, \boldsymbol{\mu}) \in S^2$, let $A\mathbf{\Sigma}$ be defined through

$$\langle \boldsymbol{T},\, A\mathbf{\Sigma}\rangle_{S^2} = \int_\Omega \boldsymbol{\tau} : \mathbb{C}^{-1}\boldsymbol{\sigma}\,\mathrm{d}x + \int_\Omega \boldsymbol{\mu} : \mathbb{H}^{-1}\boldsymbol{\chi}\,\mathrm{d}x.$$

The term $\frac{1}{2}\langle A\mathbf{\Sigma},\, \mathbf{\Sigma}\rangle_{S^2}$ corresponds to the energy associated with the stress state $\mathbf{\Sigma}$. Here $\mathbb{C}^{-1}(x)$ and $\mathbb{H}^{-1}(x)$ are linear maps from $\mathbb{S}$ to $\mathbb{S}$ (i.e., they are fourth order tensors) which may depend on the spatial variable x. For $\mathbf{\Sigma} = (\boldsymbol{\sigma}, \boldsymbol{\chi}) \in S^2$ and $\boldsymbol{v} \in V$, let

$$\langle B\mathbf{\Sigma},\, \boldsymbol{v}\rangle_{V',V} = -\int_\Omega \boldsymbol{\sigma} : \boldsymbol{\varepsilon}(\boldsymbol{v})\,\mathrm{d}x.$$

We recall that $\boldsymbol{\varepsilon}(\boldsymbol{v}) = \frac{1}{2}\big(\nabla\boldsymbol{v} + (\nabla\boldsymbol{v})^\top\big) \in S$ denotes the (linearized) strain tensor.

For the construction of the regularization we will also need the operator $B_1 : S \to V'$, defined by

$$\langle B_1\boldsymbol{\sigma},\, \boldsymbol{v}\rangle_{V',V} = -\int_\Omega \boldsymbol{\sigma} : \boldsymbol{\varepsilon}(\boldsymbol{v})\,\mathrm{d}x,$$

where $\boldsymbol{\sigma} \in S$ and $\boldsymbol{v} \in V$. Note that $B^\star\boldsymbol{v} = (B_1^\star\boldsymbol{v}, \mathbf{0}) = -(\boldsymbol{\varepsilon}(\boldsymbol{v}), \mathbf{0})$ for all $\boldsymbol{v} \in V$.

Standing assumptions. Throughout the paper, we require

Assumption 1.1.

(1) The domain $\Omega \subset \mathbb{R}^d$, $d = 3$ is a bounded Lipschitz domain in the sense of [5, Chapter 1.2]. The boundary of Ω, denoted by Γ, consists of two disjoint measurable parts Γ_N and Γ_D such that $\Gamma = \Gamma_N \cup \Gamma_D$. While Γ_N is a relatively open subset, Γ_D is a relatively closed subset of Γ. Furthermore Γ_D is assumed to have positive measure. In addition, the set $\Omega \cup \Gamma_N$ is regular in the sense of Gröger, cf. [6]. A characterization of regular domains for the case $d \in \{2, 3\}$ can be found in [7, Section 5]. This class of domains covers a wide range of geometries.

(2) The yield stress $\tilde{\sigma}_0$ is assumed to be a positive constant. It equals $\sqrt{\frac{2}{3}}\,\sigma_0$, where σ_0 is the uni-axial yield stress.
(3) $\mathbb{C}^{-1}$ is a uniformly coercive element of $L^\infty(\Omega; \mathcal{L}(\mathbb{S}, \mathbb{S}))$, where $\mathcal{L}(\mathbb{S}, \mathbb{S})$ denotes the space of linear operators $\mathbb{S} \to \mathbb{S}$. Moreover, we assume that $\mathbb{C}^{-1}(x)$ is symmetric, i.e., $\boldsymbol{\tau} : \mathbb{C}^{-1}(x)\,\boldsymbol{\sigma} = \boldsymbol{\sigma} : \mathbb{C}^{-1}(x)\,\boldsymbol{\tau}$.
(4) The hardening modulus satisfies $\mathbb{H}^{-1}(x) = k_1^{-1}(x)\,\mathbb{I}$, where the hardening parameter $k_1^{-1} \in L^\infty(\Omega)$ is uniformly positive in Ω and $\mathbb{I}$ is the identity map on $\mathbb{S} = \mathbb{R}^{d\times d}_{\text{sym}}$.

Assumption 1.1(1) enables us to apply the regularity results in [10] pertaining to systems of nonlinear elasticity. The latter appear in the time-discrete forward problem and its regularizations. Additional regularity leads to a norm gap, which is needed to prove the differentiability of the control-to-state map.

Moreover, Assumption 1.1(1) implies that Korn's inequality holds on Ω, i.e.,

$$\|\boldsymbol{u}\|^2_{H^1(\Omega;\mathbb{R}^d)} \le c_K \left(\|\boldsymbol{u}\|^2_{L^2(\Gamma_D;\mathbb{R}^d)} + \|\boldsymbol{\varepsilon}(\boldsymbol{u})\|^2_S\right) \tag{6}$$

for all $\boldsymbol{u} \in H^1(\Omega;\mathbb{R}^d)$, see, e.g., [10, Lemma C.1]. Note that (6) entails in particular that $\|\boldsymbol{\varepsilon}(\boldsymbol{u})\|_S$ is a norm on $H^1_D(\Omega;\mathbb{R}^d)$ equivalent to the standard $H^1(\Omega;\mathbb{R}^d)$ norm. A further consequence is that $B^\star$ satisfies the inf-sup condition

$$\|\boldsymbol{u}\|_V \le \sqrt{c_K}\,\|B^\star \boldsymbol{u}\|_{S^2} \quad \text{for all } \boldsymbol{u} \in V. \tag{7}$$

Assumption 1.1(3) is satisfied, e.g., for isotropic and homogeneous materials, for which

$$\mathbb{C}^{-1}\boldsymbol{\sigma} = \frac{1}{2\,\mu}\boldsymbol{\sigma} - \frac{\lambda}{2\,\mu\,(2\,\mu + d\,\lambda)}\,\text{trace}(\boldsymbol{\sigma})\,\boldsymbol{I}$$

with the identity matrix $\boldsymbol{I} \in \mathbb{S}$ and Lamé constants μ and λ, provided that $\mu > 0$ and $d\,\lambda + 2\,\mu > 0$ hold. These constants appear only here and there is no risk of confusion with the plastic multiplier λ.

Clearly, Assumption 1.1(3),(4) show that $\langle A\boldsymbol{\Sigma},\, \boldsymbol{\Sigma}\rangle_{S^2} \ge \underline{\alpha}\,\|\boldsymbol{\Sigma}\|^2_{S^2}$ for some $\underline{\alpha} > 0$ and all $\boldsymbol{\Sigma} \in \mathbb{S}^2$. Hence, the operator A is S^2-elliptic.

1.2. The forward problem. The time-discrete problem of quasistatic plasticity is defined as follows, see [8, p. 196] or [17, Section 3.1]: given $\ell^\tau \in (V')^N$, find $(\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau) \in (S^2 \times V)^N$ such that $\boldsymbol{\Sigma}^\tau_i \in \mathcal{K}$ and

$$\langle A(\boldsymbol{\Sigma}^\tau_i - \boldsymbol{\Sigma}^\tau_{i-1}) + B^\star(\boldsymbol{u}^\tau_i - \boldsymbol{u}^\tau_{i-1}),\, \boldsymbol{T} - \boldsymbol{\Sigma}^\tau_i\rangle_{S^2} \ge 0 \quad \text{for all } \boldsymbol{T} \in \mathcal{K}, \tag{8a}$$

$$B\boldsymbol{\Sigma}^\tau_i = \ell^\tau_i \quad \text{in } V', \tag{8b}$$

holds for all $i \in \{1, \ldots, N\}$, where $(\boldsymbol{\Sigma}^\tau_0, \boldsymbol{u}^\tau_0) = \boldsymbol{0}$. The time step size $\tau > 0$ and the number of time steps N are fixed throughout the paper. Since the time-continuous problem is rate-independent, the time step size does not appear explicitly in (8). The VI (8a) is the time discretization of the plastic flow law, whereas (8b) is the weak formulation of the balance of forces. Due to the homogeneous initial conditions, (8b) could also be written in incremental form $B(\mathbf{\Sigma}_i^\tau - \mathbf{\Sigma}_{i-1}^\tau) = \ell_i^\tau - \ell_{i-1}^\tau$, where $\ell_0^\tau = \mathbf{0}$.

The unique solvability of (8) is shown in [8, Proof of Theorem 8.12, p. 196]. We denote the solution operator of (8) by $\mathcal{G}^\tau : (V')^N \to (S^2 \times V)^N$, $\ell^\tau \mapsto (\mathbf{\Sigma}^\tau, \boldsymbol{u}^\tau)$.

By introducing the plastic multiplier $\lambda^\tau \in L^2(\Omega)^N$, the system (8) can be formulated equivalently as a complementarity system

$$A(\mathbf{\Sigma}_i^\tau - \mathbf{\Sigma}_{i-1}^\tau) + B^\star(\boldsymbol{u}_i^\tau - \boldsymbol{u}_{i-1}^\tau) + \tau\, \lambda_i^\tau\, \mathcal{D}^\star \mathcal{D} \mathbf{\Sigma}_i^\tau = \mathbf{0} \quad \text{in } S^2, \tag{9a}$$

$$B\mathbf{\Sigma}_i^\tau = \ell_i^\tau \quad \text{in } V', \tag{9b}$$

$$0 \le \lambda_i^\tau \quad \perp \quad \phi(\mathbf{\Sigma}_i^\tau) \le 0 \quad \text{a.e. in } \Omega, \tag{9c}$$

see [17, Section 3.1] and [9, Section 2]. As usual, $0 \le \lambda_i^\tau \perp \phi(\mathbf{\Sigma}_i^\tau) \le 0$ is short for $\lambda_i^\tau \ge 0$, $\phi(\mathbf{\Sigma}_i^\tau) \le 0$, and $\lambda_i^\tau\, \phi(\mathbf{\Sigma}_i^\tau) = 0$ a.e. in Ω.

1.3. The optimal control problem. It remains to specify the optimal control problem under consideration. The control acts as boundary force on the Neumann boundary Γ_N and belongs to $U^N = L^2(\Gamma_N; \mathbb{R}^d)^N$. The control operator $E : U \to V'$ maps controls $\boldsymbol{g}^\tau$ to right hand sides ℓ^τ of the state equation (8), see (1). The optimal control problem is given by

$$\left.\begin{aligned} &\text{Minimize} && F(\boldsymbol{u}^\tau, \boldsymbol{g}^\tau) = \psi^\tau(\boldsymbol{u}^\tau) + \frac{\nu}{2}\|\boldsymbol{g}^\tau\|_{U^N}^2 \\ &\text{such that} && (\mathbf{\Sigma}^\tau, \boldsymbol{u}^\tau) = \mathcal{G}^\tau(E\boldsymbol{g}^\tau) \\ &\text{and} && \boldsymbol{g}^\tau \in U_{\text{ad}}^\tau, \end{aligned}\right\} \tag{$\mathbf{P}^\tau$}$$

see also [17, Section 3.4]. Here, $\|\cdot\|_{U^N}$ can be any norm such that U^N is a Hilbert space. For simplicity of the presentation, the objective includes only the displacements $\boldsymbol{u}^\tau$ and not the stresses $\mathbf{\Sigma}^\tau$.

Throughout this paper we assume

Assumption 1.2.

(1) The function $\psi^\tau : V^N \to \mathbb{R}$ is weakly lower semicontinuous, bounded from below, and continuously Fréchet differentiable. We denote the partial derivatives w.r.t. $\boldsymbol{u}_i^\tau$ by $\psi_i^\tau(\boldsymbol{u}^\tau) \in V'$.
(2) The cost parameter ν is a positive, real number.
(3) The admissible set U_{ad}^τ is nonempty, convex and closed in U^N.

Based on these assumptions, the existence of a solution of $(\mathbf{P}^\tau)$ can be shown by standard arguments, see also [17, Lemma 3.7].

The most important examples for ψ^τ and U_{ad}^τ are

$$\psi^\tau(\boldsymbol{u}^\tau) = \frac{1}{2}\,\|\boldsymbol{u}_N^\tau - \boldsymbol{u}_d\|_{L^2(\Omega;\mathbb{R}^d)}^2 \quad \text{and} \quad U_{\text{ad}}^\tau = \{\boldsymbol{g}^\tau \in U^N : \boldsymbol{g}_N^\tau = \mathbf{0}\},$$

where $\boldsymbol{u}_d \in L^2(\Omega;\mathbb{R}^d)$ is a desired displacement field. Here, $\boldsymbol{u}_N^\tau$ and $\boldsymbol{g}_N^\tau$ refer to the last value of the time-discrete variables $\boldsymbol{u}^\tau$ and $\boldsymbol{g}^\tau$, respectively. The constraint $\boldsymbol{g}_N^\tau = \boldsymbol{0}$ implies that the body Ω is unloaded in the final time step, i.e., $B\boldsymbol{\Sigma}_N^\tau = \boldsymbol{0}$. Due to the observation of the final displacement $\boldsymbol{u}_N^\tau$ in the objective ψ^τ, this combination of objective and control constraints corresponds to controlling the springback of the solid body. This is of great interest in applications, e.g., deep-drawing of metal sheets.

2. Regularization of the time-discrete forward problem

Since the forward problem (8) consists of a system of VIs, the control-to-state map $\mathcal{G}^\tau \circ E$ is not differentiable. Therefore, one considers regularizations of the forward problem. A regularization of (8) which is adapted to the special structure of this problem is presented in Subsection 2.1. By introducing the regularized counterpart of the plastic multiplier, we will find that the regularization relaxes the complementarity $\lambda_i^\tau\,\phi(\boldsymbol{\Sigma}_i^\tau) = 0$ whereas the sign constraints $\lambda_i^\tau \ge 0$ and $\phi(\boldsymbol{\Sigma}_i^\tau) \le 0$ are maintained, see Subsection 2.2.

2.1. Derivation of the regularization. In this section we derive a regularized counterpart of the time-discrete system (8). We will employ Assumption 1.1(4) to derive a regularization which is tailored to the problem (8). If Assumption 1.1(4) does not hold, one may use a penalization strategy similar to [11, Section 2.2].

Let $\bar{\mathcal{K}} = \{\boldsymbol{\tau} \in S : (\boldsymbol{\tau}, \boldsymbol{0}) \in \mathcal{K}\}$ be the restriction of the admissible set $\mathcal{K}$ to the first variable. Due to the shift-invariance (4), $\boldsymbol{T} = (\boldsymbol{\tau}, \boldsymbol{\mu}) \in \mathcal{K}$ is equivalent to $\boldsymbol{\tau}+\boldsymbol{\mu} \in \bar{\mathcal{K}}$. As usual, the components of the generalized stress $\boldsymbol{\Sigma}_i^\tau$ are denoted by $\boldsymbol{\sigma}_i^\tau$ and $\boldsymbol{\chi}_i^\tau$, i.e., $\boldsymbol{\Sigma}_i^\tau = (\boldsymbol{\sigma}_i^\tau, \boldsymbol{\chi}_i^\tau)$. Given an arbitrary $\boldsymbol{\mu} \in \bar{\mathcal{K}} - \boldsymbol{\sigma}_i^\tau - \boldsymbol{\chi}_{i-1}^\tau$, the reasoning above shows that $\boldsymbol{T} = (\boldsymbol{\sigma}_i^\tau, \boldsymbol{\mu} + \boldsymbol{\chi}_{i-1}^\tau) \in \mathcal{K}$. Testing the time-discrete forward problem (8a) with $\boldsymbol{T}$ yields

$$\langle \mathbb{H}^{-1}(\boldsymbol{\chi}_i^\tau - \boldsymbol{\chi}_{i-1}^\tau),\, \boldsymbol{\mu} - (\boldsymbol{\chi}_i^\tau - \boldsymbol{\chi}_{i-1}^\tau)\rangle_S \ge 0 \quad \text{for all } \boldsymbol{\mu} \in \bar{\mathcal{K}} - \boldsymbol{\sigma}_i^\tau - \boldsymbol{\chi}_{i-1}^\tau.$$

This is equivalent to

$$\boldsymbol{\chi}_i^\tau - \boldsymbol{\chi}_{i-1}^\tau = \operatorname{Proj}^{\mathbb{H}^{-1}}_{\bar{\mathcal{K}} - \boldsymbol{\sigma}_i^\tau - \boldsymbol{\chi}_{i-1}^\tau}(\boldsymbol{0}) = \operatorname{Proj}^{\mathbb{H}^{-1}}_{\bar{\mathcal{K}}}(\boldsymbol{\sigma}_i^\tau + \boldsymbol{\chi}_{i-1}^\tau) - (\boldsymbol{\sigma}_i^\tau + \boldsymbol{\chi}_{i-1}^\tau),$$

where $\operatorname{Proj}^{\mathbb{H}^{-1}}_{\bar{\mathcal{K}}}$ is the orthogonal projection in S onto the set $\bar{\mathcal{K}} \subset S$ with respect to the norm induced by $\mathbb{H}^{-1}$. This observation gives rise to the definition of the function $\Delta\boldsymbol{\chi} : S \to S$ by

$$\Delta\boldsymbol{\chi}(\boldsymbol{\tau}) := \operatorname{Proj}^{\mathbb{H}^{-1}}_{\bar{\mathcal{K}}}(\boldsymbol{\tau}) - \boldsymbol{\tau}. \tag{10}$$

Using this definition, we find $\boldsymbol{\chi}_i^\tau - \boldsymbol{\chi}_{i-1}^\tau = \Delta\boldsymbol{\chi}(\boldsymbol{\sigma}_i^\tau + \boldsymbol{\chi}_{i-1}^\tau)$ for all $i \in \{1, \ldots, N\}$.

By employing Assumption 1.1 we can derive an explicit formula of the projection in the definition of the function $\Delta\boldsymbol{\chi}$. Indeed, using $\mathbb{H}^{-1} = k_1^{-1}\,\mathbb{I}$, where $\mathbb{I}$ is the identity map on $\mathbb{S} = \mathbb{R}^{d\times d}_{\text{sym}}$, we obtain that the norm induced by $\mathbb{H}^{-1}$ is a pointwise scaled variant of the usual norm in S. Since the restriction in the admissible set $\mathcal{K}$ and in turn in $\bar{\mathcal{K}}$ is a pointwise restriction, the projection in (10) can be evaluated pointwise. Using that the admissible set $\bar{\mathcal{K}}$ is a cylinder in S, a straightforward computation shows

$$\Delta\boldsymbol{\chi}(\boldsymbol{\tau}) = -\max\left\{0, 1 - \frac{\tilde{\sigma}_0}{|\boldsymbol{\tau}^D|}\right\}\boldsymbol{\tau}^D. \tag{11}$$

Obviously and expectedly, this function is not differentiable. A smoothed version of this relation is given by

$$\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\tau}) := -\max{}^\varepsilon\left(1 - \frac{\tilde{\sigma}_0}{|\boldsymbol{\tau}^D|}\right)\boldsymbol{\tau}^D, \tag{12}$$

where $\varepsilon > 0$ and $\max^\varepsilon$ is a smooth regularization of $\max\{0,\cdot\}$. We will not fix a particular choice of $\max^\varepsilon$ here, but use the following abstract assumption.

Assumption 2.1. For all $\varepsilon > 0$, the function $\max^\varepsilon : \mathbb{R} \to \mathbb{R}$ is of class $C^{1,1}$ and satisfies

(1) $\max^\varepsilon(x) \geq \max\{0, x\}$ for all $x \in \mathbb{R}$,
(2) $\max^\varepsilon$ is monotone increasing and convex,
(3) $\max^\varepsilon(x) = \max\{0, x\}$ for $|x| \geq \varepsilon$.

Clearly, for all $\varepsilon > 0$ there are functions $\max^\varepsilon$ satisfying this assumption, e.g., the convolution of $\max\{0,\cdot\}$ with some differentiable function. Since the term appearing inside $\max^\varepsilon$ is smaller than 1, we will assume $\varepsilon \in (0,1)$.

The function $\Delta\boldsymbol{\chi}^\varepsilon$ will be used to define the increment of $\boldsymbol{\chi}_i^\varepsilon$. In order to find a formula for $\boldsymbol{\sigma}_i^\varepsilon$, we test (8) by $\boldsymbol{T} = \boldsymbol{\Sigma}_i^\tau + (\boldsymbol{\tau}, -\boldsymbol{\tau})$, which is feasible for all $\boldsymbol{\tau} \in S$ due to the shift-invariance (4), and obtain

$$\mathbb{C}^{-1}(\boldsymbol{\sigma}_i^\tau - \boldsymbol{\sigma}_{i-1}^\tau) - \mathbb{H}^{-1}(\boldsymbol{\chi}_i^\tau - \boldsymbol{\chi}_{i-1}^\tau) + B_1^\star(\boldsymbol{u}_i^\tau - \boldsymbol{u}_{i-1}^\tau) = \boldsymbol{0}. \tag{13}$$

The arguments above give rise to the following regularized version of (8): given the loads $\ell^\varepsilon \in (V')^N$, find $(\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon) \in (S^2 \times V)^N$ satisfying

$$\mathbb{C}^{-1}(\boldsymbol{\sigma}_i^\varepsilon - \boldsymbol{\sigma}_{i-1}^\varepsilon) - \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}_i^\varepsilon + \boldsymbol{\chi}_{i-1}^\varepsilon) + B_1^\star(\boldsymbol{u}_i^\varepsilon - \boldsymbol{u}_{i-1}^\varepsilon) = \boldsymbol{0} \quad \text{in } S, \tag{14a}$$

$$B_1\boldsymbol{\sigma}_i^\varepsilon = \ell_i^\varepsilon \quad \text{in } V', \tag{14b}$$

$$\boldsymbol{\chi}_i^\varepsilon = \boldsymbol{\chi}_{i-1}^\varepsilon + \Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}_i^\varepsilon + \boldsymbol{\chi}_{i-1}^\varepsilon) \quad \text{in } S \tag{14c}$$

for all $i \in \{1,\ldots,N\}$, and with the initial condition $(\boldsymbol{\Sigma}_0^\varepsilon, \boldsymbol{u}_0^\varepsilon) = (\boldsymbol{0}, \boldsymbol{0})$. In a time-stepping scheme, we can first solve (14a) and (14b) for $(\boldsymbol{\sigma}_i^\varepsilon, \boldsymbol{u}_i^\varepsilon)$, and $\boldsymbol{\chi}_i^\tau$ can be evaluated afterwards.

Using the Browder-Minty theorem, we infer the unique solvability of (14a) and (14b) w.r.t. $(\boldsymbol{\sigma}_i^\varepsilon, \boldsymbol{u}_i^\varepsilon)$, see, e.g., [11, Lemma A.1] for a proof. This implies the unique solvability of the system (14). Let us denote the solution operator of (14) mapping $\ell^\varepsilon \in (V')^N$ to $(\boldsymbol{\sigma}^\varepsilon, \boldsymbol{\chi}^\varepsilon, \boldsymbol{u}^\varepsilon)$ by $\mathcal{G}^\varepsilon$. Moreover, the Browder-Minty theorem implies the global Lipschitz continuity of $\mathcal{G}^\varepsilon : (V')^N \to (S^2 \times V)^N$.

By replacing the time-discrete solution map $\mathcal{G}^\tau$ of (8) with its regularization $\mathcal{G}^\varepsilon$, see (14), we obtain the regularized control problem

$$\left.\begin{aligned} \text{Minimize}\quad & F(\boldsymbol{u}^\varepsilon, \boldsymbol{g}^\varepsilon) = \psi^\tau(\boldsymbol{u}^\varepsilon) + \frac{\nu}{2}\|\boldsymbol{g}^\varepsilon\|_{U^N}^2 \\ \text{such that}\quad & (\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon) = \mathcal{G}^\varepsilon(E\boldsymbol{g}^\varepsilon), \\ \text{and}\quad & \boldsymbol{g}^\varepsilon \in U_{\text{ad}}^\tau. \end{aligned}\right\} \tag{$\mathbf{P}^\varepsilon$}$$

Due to the Lipschitz continuity of $\mathcal{G}^\varepsilon$, we obtain

Lemma 2.2. *There exists a global minimizer of* $(\mathbf{P}^\varepsilon)$.

2.2. Reformulation of the forward system involving the plastic multiplier. The aim of this section is the introduction of the regularized counterparts of the plastic multiplier λ^τ in the forward problem.

Let us recall that the deviatoric part of a matrix $\boldsymbol{\sigma} \in \mathbb{S}$ is given by $\boldsymbol{\sigma}^D = \boldsymbol{\sigma} - \frac{1}{d}\,\text{trace}(\boldsymbol{\sigma})\,\boldsymbol{I}$, see (3). By (12), (14c), and the initial datum $\boldsymbol{\chi}_0^\varepsilon = \boldsymbol{0}$, we infer $\boldsymbol{\chi}_i^\varepsilon = (\boldsymbol{\chi}_i^\varepsilon)^D$ for all $i = 0, \ldots, N$. Hence, we can omit the superscript D on the variable $\boldsymbol{\chi}^\varepsilon$. For convenience, we define

$$\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon := (\boldsymbol{\sigma}_i^\varepsilon)^D + \boldsymbol{\chi}_{i-1}^\varepsilon.$$

In contrast to $\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon = (\boldsymbol{\sigma}_i^\varepsilon)^D + \boldsymbol{\chi}_i^\varepsilon$, see (5), $\boldsymbol{\chi}^\varepsilon$ is taken from the previous time step $i-1$.

Moreover, we introduce an abbreviation for the scalar factor appearing in $\Delta\boldsymbol{\chi}^\varepsilon(\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon)$, cf. (12),

$$\alpha_i^\varepsilon := \max{}^\varepsilon\left(1 - \frac{\tilde{\sigma}_0}{|\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon|}\right). \tag{15}$$

By Assumption 2.1 we infer $\alpha_i^\varepsilon \in [0, 1)$. Adding $(\boldsymbol{\sigma}_i^\varepsilon)^D$ on both sides of (14c) yields

$$\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon = \tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon - \alpha_i^\varepsilon\,\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon = (1 - \alpha_i^\varepsilon)\,\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon. \tag{16}$$

Hence, the definition of $\Delta\boldsymbol{\chi}^\varepsilon$, see (12), implies

$$-\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}_i^\varepsilon + \boldsymbol{\chi}_{i-1}^\varepsilon) = k_1^{-1}\alpha_i^\varepsilon\,((\boldsymbol{\sigma}_i^\varepsilon)^D + \boldsymbol{\chi}_{i-1}^\varepsilon) = k_1^{-1}\alpha_i^\varepsilon\,\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon = k_1^{-1}\frac{\alpha_i^\varepsilon}{1 - \alpha_i^\varepsilon}\,\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon.$$

Comparing (14a) and (14c) with (9a) gives rise to the definition of λ_i^ε by

$$\lambda_i^\varepsilon := \tau^{-1}\,k_1^{-1}\frac{\alpha_i^\varepsilon}{1 - \alpha_i^\varepsilon}. \tag{17}$$

The $L^2(\Omega)$-regularity of λ_i^ε is shown in Corollary 4.4. Using the definition of λ_i^ε, the forward system (14) becomes

$$A(\boldsymbol{\Sigma}_i^\varepsilon - \boldsymbol{\Sigma}_{i-1}^\varepsilon) + B^\star(\boldsymbol{u}_i^\varepsilon - \boldsymbol{u}_{i-1}^\varepsilon) + \tau\,\lambda_i^\varepsilon\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon = \mathbf{0}, \tag{18a}$$

$$B\boldsymbol{\Sigma}_i^\varepsilon = \ell_i^\varepsilon. \tag{18b}$$

In Assumption 2.1 we require $\max^\varepsilon(x) = \max\{0, x\}$ if $x \notin (-\varepsilon, \varepsilon)$. Hence it is natural to split Ω into three disjoint sets in dependence whether the argument of $\max^\varepsilon$ in (12) is smaller than $-\varepsilon$, larger than ε or in $(-\varepsilon, \varepsilon)$.

$$A_i^{\varepsilon,-} := \Big\{x \in \Omega : |\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon| \le \frac{\tilde\sigma_0}{1+\varepsilon}\Big\} = \Big\{x \in \Omega : 1 - \frac{\tilde\sigma_0}{|\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon|} \le -\varepsilon\Big\}, \tag{19a}$$

$$A_i^{\varepsilon,+} := \Big\{x \in \Omega : |\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon| \ge \frac{\tilde\sigma_0}{1-\varepsilon}\Big\} = \Big\{x \in \Omega : 1 - \frac{\tilde\sigma_0}{|\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon|} \ge \varepsilon\Big\}, \tag{19b}$$

$$A_i^{\varepsilon,0} := \Omega \setminus (A_i^{\varepsilon,-} \cup A_i^{\varepsilon,+}). \tag{19c}$$

Using these definitions, the equations (15), (16) and (17) imply

$$\begin{aligned} 0 &= \lambda_i^\varepsilon, & \phi(\boldsymbol{\Sigma}_i^\varepsilon) &< 0 & &\text{on } A_i^{\varepsilon,-}, \\ \Big(0, \frac{\varepsilon}{1-\varepsilon}\,\tau^{-1}\,k_1^{-1}\Big) &\ni \lambda_i^\varepsilon, & \phi(\boldsymbol{\Sigma}_i^\varepsilon) &\in \Big(\frac{-2\,\varepsilon}{(1+\varepsilon)^2}, 0\Big)\,\tilde\sigma_0^2 & &\text{on } A_i^{\varepsilon,0}, \\ 0 &< \lambda_i^\varepsilon, & \phi(\boldsymbol{\Sigma}_i^\varepsilon) &= 0 & &\text{on } A_i^{\varepsilon,+}. \end{aligned}$$

Hence, the plastic multiplier λ_i^ε and the generalized stress $\boldsymbol{\Sigma}_i^\varepsilon$ still satisfy the sign conditions in (9c), but they do not satisfy the complementarity condition in (9c). Thus our regularization approach can be viewed as a problem-tailored version of the relaxation strategy given in [15].

3. Differentiability of the forward problem and optimality conditions

This section is devoted to the proof of the Fréchet differentiability of the control-to-state map $\mathcal{G}^\varepsilon \circ E$. Since the forward system is equivalent to a system of quasilinear PDEs, this is a novel result. In particular, it relies on the regularity of solutions to quasilinear systems of infinitesimal elasticity proven in [10].

First, we address the differentiability of the solution of one time step in Subsection 3.1 and conclude the differentiability of $\mathcal{G}^\varepsilon \circ E$ in Subsection 3.2. Finally, we state the first order necessary conditions for $(\mathbf{P}^\varepsilon)$ in Subsection 3.3.

Throughout this section, we consider a fixed regularization parameter $\varepsilon \in (0, 1)$.

3.1. Differentiability of one time step. In order to simplify the notation, we use an abstract notation. For all time steps $i \in \{1, \ldots, N\}$, the system (14a), (14b) has a common structure: given $(\mathscr{L}, \boldsymbol{\chi}, \ell) \in S \times S \times V'$, find $(\boldsymbol{\sigma}, \Delta \boldsymbol{u}) \in S \times V$ satisfying

$$\mathbb{C}^{-1}\boldsymbol{\sigma} - \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^{\varepsilon}(\boldsymbol{\sigma} + \boldsymbol{\chi}) + B^{\star}\Delta\boldsymbol{u} = \mathscr{L} \quad \text{in } S, \tag{20a}$$

$$B\boldsymbol{\sigma} = \ell \quad \text{in } V'. \tag{20b}$$

We denote the solution operator of this system by $G^{\varepsilon} : (\mathscr{L}, \boldsymbol{\chi}, \ell) \mapsto (\boldsymbol{\sigma}, \Delta\boldsymbol{u})$. Then one time step of (14) is equivalent to

$$(\boldsymbol{\sigma}_i^{\varepsilon}, \Delta\boldsymbol{u}_i^{\varepsilon}) := (\boldsymbol{\sigma}_i^{\varepsilon}, \boldsymbol{u}_i^{\varepsilon} - \boldsymbol{u}_{i-1}^{\varepsilon}) = G^{\varepsilon}(\mathbb{C}^{-1}\boldsymbol{\sigma}_{i-1}^{\varepsilon}, \boldsymbol{\chi}_{i-1}^{\varepsilon}, \ell_i^{\varepsilon}), \tag{21a}$$

$$\boldsymbol{\chi}_i^{\varepsilon} = \boldsymbol{\chi}_{i-1}^{\varepsilon} + \Delta\boldsymbol{\chi}^{\varepsilon}(\boldsymbol{\sigma}_i^{\varepsilon} + \boldsymbol{\chi}_{i-1}^{\varepsilon}). \tag{21b}$$

By $H^{\varepsilon} : (\boldsymbol{\sigma}_{i-1}^{\varepsilon}, \boldsymbol{\chi}_{i-1}^{\varepsilon}, \ell_i^{\varepsilon}) \mapsto (\boldsymbol{\sigma}_i^{\varepsilon}, \boldsymbol{\chi}_i^{\varepsilon}, \Delta\boldsymbol{u}_i^{\varepsilon})$ we denote the solution of a single time step. The aim of this section is to show that

$$H^{\varepsilon} : L^{p_2}(\Omega; \mathbb{S})^2 \times W^{-1,p_2}(\Omega; \mathbb{R}^d) \to L^{p_1}(\Omega; \mathbb{S})^2 \times V$$

is Fréchet differentiable for any pair of exponents satisfying $p_2 > p_1 \geq 2$. First, we prove the differentiability of G^{ε}, giving in turn the differentiability of H^{ε}, see Theorem 3.7. The differentiability of G^{ε} will be proven by applying the abstract result Theorem A.6 concerning nonlinear saddle-point problems with the setting

$$\left.\begin{aligned}
&X = S = L^2(\Omega; \mathbb{S}), && V = H^1_D(\Omega; \mathbb{R}^d), && A = \mathbb{C}^{-1}, \\
&Y = L^{p_1}(\Omega; \mathbb{S}), && W' = W^{-1,p_2}(\Omega; \mathbb{R}^d), && J = -\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^{\varepsilon}, \\
&Z = L^{p_2}(\Omega; \mathbb{S}). && &&
\end{aligned}\right\} \tag{22}$$

Here,

$$\begin{aligned}
W_D^{1,p_2'}(\Omega; \mathbb{R}^d) &:= \{\boldsymbol{u} \in W^{1,p_2'}(\Omega; \mathbb{R}^d) : \boldsymbol{u} = \boldsymbol{0} \text{ on } \Gamma_D\} \quad \text{and} \\
W^{-1,p_2}(\Omega; \mathbb{R}^d) &:= \left(W_D^{1,p_2'}(\Omega; \mathbb{R}^d)\right)',
\end{aligned}$$

where $p_2' < 2$ is the exponent conjugate to $p_2 > 2$. Due to $V \hookrightarrow W = W_D^{1,p_2'}(\Omega; \mathbb{R}^d)$ we obtain the embedding $W' \hookrightarrow V'$.

Let us comment on the prerequisites of Theorem A.6. Surely, Assumption A.1 is satisfied, in particular the monotonicity of $-\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^{\varepsilon}$ follows easily from Assumption 2.1. Assumption A.3, which concerns the differentiability of the nonlinear term, follows from standard results on the differentiability of Nemytskii operators, see Lemma 3.1. In order to satisfy Assumption A.5, we have to prove the local Lipschitz continuity of the solution operator of (20) and its linearization w.r.t. stronger norms. This is the main work of this section and it is done in Proposition 3.5 and Proposition 3.6. Finally, the application of Theorem A.6 yields the Fréchet differentiability of H^ε, see Theorem 3.7.

As announced, we start by addressing the Fréchet differentiability of the Nemytskii operator $\Delta\boldsymbol{\chi}^\varepsilon$.

Lemma 3.1. *The operator $\Delta\boldsymbol{\chi}^\varepsilon$ is Fréchet differentiable from $L^{p_2}(\Omega;\mathbb{S})$ to $L^{p_1}(\Omega;\mathbb{S})$ for all $p_2 > p_1 \geq 2$. The derivative at $\boldsymbol{\tau}$ in direction $\delta\boldsymbol{\tau}$ is given by*

$$(\Delta\boldsymbol{\chi}^\varepsilon)'(\boldsymbol{\tau})\delta\boldsymbol{\tau} = -(\max{}^\varepsilon)'\left(1-\frac{\tilde{\sigma}_0}{|\boldsymbol{\tau}^D|}\right)\frac{\tilde{\sigma}_0}{|\boldsymbol{\tau}^D|^3}(\boldsymbol{\tau}^D{:}\delta\boldsymbol{\tau}^D)\boldsymbol{\tau}^D - \max{}^\varepsilon\left(1-\frac{\tilde{\sigma}_0}{|\boldsymbol{\tau}^D|}\right)\delta\boldsymbol{\tau}^D. \quad (23)$$

The operator $(\Delta\boldsymbol{\chi}^\varepsilon)'(\boldsymbol{\tau}) : L^{p_2}(\Omega;\mathbb{S}) \to L^{p_1}(\Omega;\mathbb{S})$ can be extended to a bounded linear and positive semi-definite operator $S \to S$, i.e., $\langle\delta\boldsymbol{\tau},\, (\Delta\boldsymbol{\chi}^\varepsilon)'(\boldsymbol{\tau})\,\delta\boldsymbol{\tau}\rangle_S \geq 0$ holds for all $\delta\boldsymbol{\tau} \in S$.

Proof. The result follows from general differentiability results for nonlinear Nemytskii operators, e.g., [4, Theorem 7], [16, Section 4.3.3], see also [11, Proposition 2.11]. □

The linearized version of (20) reads, cf. (45),

$$\mathbb{C}^{-1}\boldsymbol{\sigma} - \mathbb{H}^{-1}(\Delta\boldsymbol{\chi}^\varepsilon)'(\hat{\boldsymbol{\sigma}})\,\boldsymbol{\sigma} + B_1^\star\Delta\boldsymbol{u} = \mathscr{L} \quad \text{in } S \quad (24a)$$

$$B_1\boldsymbol{\sigma} = \ell \quad \text{in } V' \quad (24b)$$

We denote its solution operator by $\tilde{G}^\varepsilon : (\mathscr{L}, \ell) \mapsto (\boldsymbol{\sigma}, \boldsymbol{u})$.

Now we are going to prove the Lipschitz continuity of the solution maps of (20) (w.r.t. $\mathscr{L}$, $\boldsymbol{\chi}$ and ℓ) and of (24) (w.r.t. $\mathscr{L}$ and ℓ) in the norms of $Z = L^{p_2}(\Omega;\mathbb{S})$ and $Y = L^{p_1}(\Omega;\mathbb{S})$, see Assumption A.5. Standard regularity theory using the Browder-Minty theorem yields only the Lipschitz continuity with respect to the norm in $X = S = L^2(\Omega;\mathbb{S})$. We are going to apply the regularity result [10, Proposition 1.2].

To allow for a uniform treatment of both systems, we define a set $\mathcal{Q}$ of those mappings acting on $\boldsymbol{\sigma}$ which appear on the left hand sides of (20a) and (24a). To be precise, we define $\mathcal{Q}$ to contain all operators mapping $\Omega \times \mathbb{S} \to \mathbb{S}$,

$$(x, \boldsymbol{\sigma}) \mapsto \mathbb{C}^{-1}(x)\,\boldsymbol{\sigma} - \mathbb{H}^{-1}(x)\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma} + \boldsymbol{\chi}(x)) \quad \text{for all } \boldsymbol{\chi} \in S, \quad (25a)$$

$$\text{and} \quad (x, \boldsymbol{\sigma}) \mapsto \mathbb{C}^{-1}(x)\,\boldsymbol{\sigma} - \mathbb{H}^{-1}(x)(\Delta\boldsymbol{\chi}^\varepsilon)'(\hat{\boldsymbol{\sigma}}(x))\,\boldsymbol{\sigma} \quad \text{for all } \hat{\boldsymbol{\sigma}} \in S. \quad (25b)$$

For every $Q \in \mathcal{Q}$ we will denote the induced Nemytskii operator with the same symbol. Due to the definition of $\mathcal{Q}$, both systems (20) and (24) can be written as

$$Q(\boldsymbol{\sigma}) + B_1^\star\boldsymbol{u} = \mathscr{L}, \quad (26a)$$

$$B_1\boldsymbol{\sigma} = \ell, \quad (26b)$$

with the corresponding $Q \in \mathcal{Q}$. We show, that the solution mapping of (26) is Lipschitz continuous with respect to $\mathscr{L}$ and ℓ. For the system (20), the Lipschitz dependence on $\boldsymbol{\chi}$ is proven afterwards.

First, we state the Lipschitz continuity and strong monotonicity of $Q \in \mathcal{Q}$.

Lemma 3.2. *For all $Q \in \mathcal{Q}$, the inequalities*

$$(Q(x, \boldsymbol{\varepsilon}) - Q(x, \hat{\boldsymbol{\varepsilon}})) : (\boldsymbol{\varepsilon} - \hat{\boldsymbol{\varepsilon}}) \geq \underline{\alpha}\, |\boldsymbol{\varepsilon} - \hat{\boldsymbol{\varepsilon}}|^2,$$
$$|Q(x, \boldsymbol{\varepsilon}) - Q(x, \hat{\boldsymbol{\varepsilon}})| \leq (\overline{\alpha} + 2\,\overline{h})\, |\boldsymbol{\varepsilon} - \hat{\boldsymbol{\varepsilon}}|$$

are satisfied for almost all $x \in \Omega$ and all $\boldsymbol{\varepsilon}, \hat{\boldsymbol{\varepsilon}} \in \mathbb{S}$. Here $\underline{\alpha}$ is the uniform coercivity constant of $\mathbb{C}^{-1}$ and $\overline{\alpha}$, $\overline{h}$ are the uniform boundedness constants of $\mathbb{C}^{-1}$ and $\mathbb{H}^{-1}$, respectively.

Proof. Due to the uniform coercivity of $\mathbb{C}^{-1}$, the first assertion follows easily by the monotonicity of $\Delta\boldsymbol{\chi}^{\varepsilon}$ and $(\Delta\boldsymbol{\chi}^{\varepsilon})'$. The Lipschitz continuity of $\Delta\boldsymbol{\chi}^{\varepsilon}$ and $(\Delta\boldsymbol{\chi}^{\varepsilon})'$ can be proved easily, starting from (12) and (23), respectively. □

By the Browder-Minty theorem, we infer that for all $Q \in \mathcal{Q}$ and almost all $x \in \Omega$, the operator $Q(x, \cdot) : \mathbb{S} \to \mathbb{S}$ is invertible. We define the pointwise inverse $Q^{-1} : \Omega \times \mathbb{S} \to \mathbb{S}$, $Q^{-1}(x, \cdot) = Q(x, \cdot)^{-1}$. We obtain for all $Q \in \mathcal{Q}$ and almost all $x \in \Omega$

$$(Q^{-1}(x, \boldsymbol{\varepsilon}) - Q^{-1}(x, \hat{\boldsymbol{\varepsilon}})) : (\boldsymbol{\varepsilon} - \hat{\boldsymbol{\varepsilon}}) \geq m\, |\boldsymbol{\varepsilon} - \hat{\boldsymbol{\varepsilon}}|^2, \tag{27a}$$
$$|Q^{-1}(x, \boldsymbol{\varepsilon}) - Q^{-1}(x, \hat{\boldsymbol{\varepsilon}})| \leq M\, |\boldsymbol{\varepsilon} - \hat{\boldsymbol{\varepsilon}}|, \tag{27b}$$

where m, $M > 0$ are constants. This implies that the Nemytskii operator Q^{-1} maps $L^p(\Omega; \mathbb{S}) \to L^p(\Omega; \mathbb{S})$ for all $p \geq 2$, provided that $Q^{-1}(\cdot, \mathbf{0}) \in L^p(\Omega; \mathbb{S})$. If Q is of the form (25a), this requires $\boldsymbol{\chi} \in L^p(\Omega; \mathbb{S})$. For Q defined by (25b) this imposes no restriction, since $Q(\cdot, \mathbf{0}) = \mathbf{0}$ holds independently of $\hat{\boldsymbol{\sigma}}$.

By applying Q^{-1} to (26), we obtain that $\boldsymbol{u}$ solves

$$B_1\, Q^{-1}\, (-B_1^\star \boldsymbol{u} + \mathscr{L}) = \ell. \tag{28}$$

This is a quasilinear system of infinitesimal elasticity. Higher integrability of solutions to such systems is the topic of [10].

Lemma 3.3. *There exists an exponent $\bar{p} > 2$ such that for all $p \in [2, \bar{p}]$ and all $Q \in \mathcal{Q}$ satisfying $Q^{-1}(\cdot, \mathbf{0}) \in L^p(\Omega; \mathbb{S})$, the solution map of (28) is Lipschitz continuous w.r.t. $\ell \in W^{-1,p}(\Omega; \mathbb{R}^d) = W_D^{1,p'}(\Omega; \mathbb{R}^d)'$ for fixed $\mathscr{L} \in L^p(\Omega; \mathbb{S})$. The Lipschitz constant does not depend on the particular choice of Q.*

Proof. In order to apply [10, Proposition 1.2], we define the family of nonlinearities

$$\mathcal{F}_p := \{\boldsymbol{b} : \Omega \times \mathbb{S} \to \mathbb{S} : \exists \mathscr{L} \in L^p(\Omega;\mathbb{S}) \text{ and } Q \in \mathcal{Q} \text{ with } Q^{-1}(\cdot,\mathbf{0}) \in L^p(\Omega;\mathbb{S}), \\ \text{s.t. } \boldsymbol{b}(x,\boldsymbol{\varepsilon}) = Q^{-1}(x, \boldsymbol{\varepsilon} + \mathscr{L}(x)) \text{ f.a.a. } x \in \Omega \text{ and all } \boldsymbol{\varepsilon} \in \mathbb{S}\}.$$

By (27), we obtain that [10, Assumption 1.5(2)] is fulfilled uniformly for all $Q \in \mathcal{Q}$. Note that by [10, Remark 1.6 (2)] it is sufficient to ensure $Q^{-1}(\cdot,\mathbf{0}) \in L^p(\Omega;\mathbb{S})$.

According to the regularity result [10, Proposition 1.2], there exists an exponent $\bar{p}$ such that for all $p \in [2,\bar{p}]$, the solution map of (28) is Lipschitz continuous from $W^{-1,p}(\Omega;\mathbb{R}^d) = W_D^{1,p'}(\Omega;\mathbb{R}^d)'$ to $W_D^{1,p}(\Omega;\mathbb{R}^d)$, for all $\mathscr{L} \in L^p(\Omega;\mathbb{S})$ and all $Q \in \mathcal{Q}$ with $Q^{-1}(\cdot,\mathbf{0}) \in L^p(\Omega;\mathbb{S})$. All these solution maps share a common Lipschitz constant, i.e., we obtain

$$\|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_{W_D^{1,p}(\Omega;\mathbb{R}^d)} \le L \, \|\ell_1 - \ell_2\|_{W^{-1,p}(\Omega;\mathbb{R}^d)},$$

where $\boldsymbol{u}_i$ solves $\ell_i = B_1 \, Q^{-1}(-B_1^\star \boldsymbol{u}_i + \mathscr{L})$. The constant L does not depend on ℓ_i, $\mathscr{L}$, Q and p, but only on Ω, Γ_D, $\mathbb{C}^{-1}$, $\mathbb{H}^{-1}$, and $\bar{p}$. □

In the sequel, $\bar{p}$ always refers to the regularity exponent given in Lemma 3.3. In the next proposition, we show that $\boldsymbol{u}$ in (28) depends also Lipschitz continuously on the data $\mathscr{L}$. We also show the Lipschitz dependence of $\boldsymbol{\sigma}$ in (26) on ℓ and $\mathscr{L}$.

Proposition 3.4. *For all $p \in [2,\bar{p}]$ and $Q \in \mathcal{Q}$ satisfying $Q^{-1}(\cdot,\mathbf{0}) \in L^p(\Omega;\mathbb{S})$, the solution mapping of* (26) *is Lipschitz continuous w.r.t. $\mathscr{L} \in L^p(\Omega;\mathbb{S})$ and $\ell \in W^{-1,p}(\Omega;\mathbb{R}^d)$. The Lipschitz constant does not depend on Q.*

Proof. Let $p \in [2,\bar{p}]$ be given. The Lipschitz dependence of $\boldsymbol{u}$ on ℓ has been shown in Lemma 3.3.

Step (1): We consider the Lipschitz dependence of $\boldsymbol{u}$ on $\mathscr{L}$. For $i \in \{1,2\}$, let $\ell_i \in W^{-1,p}(\Omega;\mathbb{R}^d)$ and $\mathscr{L}_i \in L^p(\Omega;\mathbb{S})$ be given. Define $\boldsymbol{u}_i$ as the solution of $\ell_i = B_1 \, Q^{-1}(-B_1^\star \boldsymbol{u}_i + \mathscr{L}_i)$.

Using $\boldsymbol{\tau} = Q^{-1}(-B_1^\star \boldsymbol{u}_2 + \mathscr{L}_2) - Q^{-1}(-B_1^\star \boldsymbol{u}_2 + \mathscr{L}_1)$ we obtain $\ell_2 - B_1 \boldsymbol{\tau} = B_1 \, Q^{-1}(-B_1^\star \boldsymbol{u}_2 + \mathscr{L}_1)$. This shows that $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ solve the systems

$$B_1 \, Q^{-1}(-B_1^\star \boldsymbol{u}_1 + \mathscr{L}_1) = \ell_1 \quad \text{and} \quad B_1 \, Q^{-1}(-B_1^\star \boldsymbol{u}_2 + \mathscr{L}_1) = \ell_2 - B_1 \boldsymbol{\tau}.$$

An application of Lemma 3.3 yields

$$\|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_{W_D^{1,p}(\Omega;\mathbb{R}^d)} \le L \, \|\ell_1 - (\ell_2 - B_1 \boldsymbol{\tau})\|_{W^{-1,p}(\Omega;\mathbb{R}^d)}.$$

By definition of $\boldsymbol{\tau}$ and (27b) we obtain $\|\boldsymbol{\tau}\|_{L^p(\Omega;\mathbb{S})} \le M \, \|\mathscr{L}_1 - \mathscr{L}_2\|_{L^p(\Omega;\mathbb{S})}$. This implies

$$\|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_{W_D^{1,p}(\Omega;\mathbb{R}^d)} \le C \left(\|\mathscr{L}_1 - \mathscr{L}_2\|_{L^p(\Omega;\mathbb{S})} + \|\ell_1 - \ell_2\|_{W^{-1,p}(\Omega;\mathbb{R}^d)} \right).$$

Step (2): We consider the Lipschitz dependence of $\boldsymbol{\sigma}$. By (26) we infer

$$\boldsymbol{\sigma} = Q^{-1}(-B_1^\star \boldsymbol{u} + \mathscr{L}).$$

Since Q^{-1} maps $L^p(\Omega;\mathbb{S})$ Lipschitz continuously into itself by (27b), and since $\boldsymbol{u}$ depends Lipschitz continuously on ℓ and $\mathscr{L}$, we obtain

$$\|\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2\|_{L^p(\Omega;\mathbb{S})} \leq C\left(\|\mathscr{L}_1 - \mathscr{L}_2\|_{L^p(\Omega;\mathbb{S})} + \|\ell_1 - \ell_2\|_{W^{-1,p}(\Omega;\mathbb{R}^d)}\right). \quad \square$$

The next proposition translates the abstract result of Proposition 3.4 to the systems (20) and (24). It is an immediate consequence using the definition of $\mathcal{Q}$ in (25).

Proposition 3.5. *Let $p \in [2, \bar{p}]$ be given. The solution maps of*
(i) (20)*, i.e., G^ε, for fixed $\boldsymbol{\chi} \in L^p(\Omega;\mathbb{S})$ and*
(ii) (24)*, i.e., $\tilde{G}^\varepsilon$, for fixed $\hat{\boldsymbol{\sigma}} \in S$*
are Lipschitz continuous w.r.t. $\mathscr{L} \in L^p(\Omega;\mathbb{S})$ and $\ell \in W^{-1,p}(\Omega;\mathbb{R}^d)$. They share a common Lipschitz constant, i.e.,

$$\begin{aligned}&\|\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2\|_{L^p(\Omega;\mathbb{S})} + \|\Delta\boldsymbol{u}_1 - \Delta\boldsymbol{u}_2\|_{W_D^{1,p}(\Omega;\mathbb{R}^d)}\\ &\leq C\left(\|\mathscr{L}_1 - \mathscr{L}_2\|_{L^p(\Omega;\mathbb{S})} + \|\ell_1 - \ell_2\|_{W^{-1,p}(\Omega;\mathbb{R}^d)}\right),\end{aligned}$$

where $(\boldsymbol{\sigma}_i, \Delta\boldsymbol{u}_i)$ are the solutions of (20) *or* (24) *respectively with the data $(\ell_i, \mathscr{L}_i)$, $i = 1, 2$.*

It remains to prove the Lipschitz dependence of the solution of (20) on $\boldsymbol{\chi}$.

Proposition 3.6. *Let $p \in [2, \bar{p}]$ be given. Then, in addition to Proposition* 3.5*, G^ε is also Lipschitz continuous w.r.t. $\boldsymbol{\chi} \in L^p(\Omega;\mathbb{S})$, i.e.,*

$$\begin{aligned}&\|\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2\|_{L^p(\Omega;\mathbb{S})} + \|\Delta\boldsymbol{u}_1 - \Delta\boldsymbol{u}_2\|_{W_D^{1,p}(\Omega;\mathbb{R}^d)}\\ &\leq C\left(\|\mathscr{L}_1 - \mathscr{L}_2\|_{L^p(\Omega;\mathbb{S})} + \|\boldsymbol{\chi}_1 - \boldsymbol{\chi}_2\|_{L^p(\Omega;\mathbb{S})} + \|\ell_1 - \ell_2\|_{W^{-1,p}(\Omega;\mathbb{R}^d)}\right),\end{aligned}$$

where $(\boldsymbol{\sigma}_i, \Delta\boldsymbol{u}_i)$ is the solution of (20) *with the data $(\mathscr{L}_i, \boldsymbol{\chi}_i, \ell_i)$, $i = 1, 2$.*

Proof. For $i \in \{1, 2\}$, let $\ell_i \in W^{-1,p}(\Omega;\mathbb{R}^d)$ and $\boldsymbol{\chi}_i, \mathscr{L}_i \in L^p(\Omega;\mathbb{S})$ be given. Define $(\boldsymbol{\sigma}_i, \Delta\boldsymbol{u}_i) = G^\varepsilon(\mathscr{L}_i, \boldsymbol{\chi}_i, \ell_i)$ as the solutions of (20). We define $\delta\boldsymbol{\chi} = \boldsymbol{\chi}_2 - \boldsymbol{\chi}_1$. We infer from (20)

$$\begin{aligned}\mathbb{C}^{-1}(\boldsymbol{\sigma}_2 + \delta\boldsymbol{\chi}) - \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon((\boldsymbol{\sigma}_2 + \delta\boldsymbol{\chi}) + \boldsymbol{\chi}_1) + B_1^\star\Delta\boldsymbol{u}_2 &= \mathscr{L}_2 + \mathbb{C}^{-1}\delta\boldsymbol{\chi},\\ B_1(\boldsymbol{\sigma}_2 + \delta\boldsymbol{\chi}) &= \ell_2 + B_1\delta\boldsymbol{\chi}.\end{aligned}$$

Therefore, $(\boldsymbol{\sigma}_2 + \delta\boldsymbol{\chi}, \Delta\boldsymbol{u}_2)$ solves the same system (i.e., with the same $\boldsymbol{\chi} = \boldsymbol{\chi}_1$) as $(\boldsymbol{\sigma}_1, \Delta\boldsymbol{u}_1)$, but with a different right hand side. Thus, the Lipschitz estimate from Proposition 3.5 yields

$$\begin{aligned}&\|\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2 - \delta\boldsymbol{\chi}\|_{L^p(\Omega;\mathbb{S})} + \|\Delta\boldsymbol{u}_1 - \Delta\boldsymbol{u}_2\|_{W^{1,p}_D(\Omega;\mathbb{R}^d)}\\ &\le C\left(\|\mathscr{L}_1 - \mathscr{L}_2 - \mathbb{C}^{-1}\delta\boldsymbol{\chi}\|_{L^p(\Omega;\mathbb{S})} + \|\ell_1 - \ell_2 - B_1\delta\boldsymbol{\chi}\|_{W^{-1,p}(\Omega;\mathbb{R}^d)}\right).\end{aligned}$$

Using the triangle inequality and the boundedness of the operators B_1 and $\mathbb{C}^{-1}$ we infer

$$\begin{aligned}&\|\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2\|_{L^p(\Omega;\mathbb{S})} + \|\Delta\boldsymbol{u}_1 - \Delta\boldsymbol{u}_2\|_{W^{1,p}_D(\Omega;\mathbb{R}^d)}\\ &\le C\left(\|\mathscr{L}_1 - \mathscr{L}_2\|_{L^p(\Omega;\mathbb{S})} + \|\boldsymbol{\chi}_1 - \boldsymbol{\chi}_2\|_{L^p(\Omega;\mathbb{S})} + \|\ell_1 - \ell_2\|_{W^{-1,p}(\Omega;\mathbb{R}^d)}\right). \quad \square\end{aligned}$$

As announced in the beginning of this section, we use Theorem A.6 with the setting (22) to infer the Fréchet differentiability of H^ε.

Theorem 3.7. *Let p_1, p_2 be given such that $2 \le p_1 < p_2 \le \bar{p}$. Then the operator H^ε defined in (21) is Fréchet differentiable from $L^{p_2}(\Omega;\mathbb{S})^2 \times W^{-1,p_2}(\Omega;\mathbb{R}^d)$ to $L^{p_1}(\Omega;\mathbb{S})^2 \times V$.*

At the point $(\boldsymbol{\sigma}_i^\varepsilon, \boldsymbol{\chi}_i^\varepsilon, \Delta\boldsymbol{u}_i^\varepsilon) = H^\varepsilon(\boldsymbol{\sigma}_{i-1}^\varepsilon, \boldsymbol{\chi}_{i-1}^\varepsilon, \ell_i^\varepsilon)$ the directional derivative of H^ε in direction $(\delta\boldsymbol{\sigma}_{i-1}^\varepsilon, \delta\boldsymbol{\chi}_{i-1}^\varepsilon, \delta\ell_i^\varepsilon)$ is denoted by $(\delta\boldsymbol{\sigma}_i^\varepsilon, \delta\boldsymbol{\chi}_i^\varepsilon, \delta\Delta\boldsymbol{u}_i^\varepsilon)$. The directional derivative $(\delta\boldsymbol{\sigma}_i^\varepsilon, \delta\Delta\boldsymbol{u}_i^\varepsilon)$ solves the system

$$(\mathbb{C}^{-1} - \mathbb{H}^{-1}(\Delta\boldsymbol{\chi}^\varepsilon)'(\boldsymbol{\sigma}_i^\varepsilon + \boldsymbol{\chi}_{i-1}^\varepsilon))\delta\boldsymbol{\sigma}_i^\varepsilon + B_1^\star\delta\Delta\boldsymbol{u}_i^\varepsilon = \delta\mathscr{L}_i, \tag{29a}$$

$$B_1\delta\boldsymbol{\sigma}_i^\varepsilon = \delta\ell_i^\varepsilon, \tag{29b}$$

where

$$\delta\mathscr{L}_i = \mathbb{C}^{-1}\delta\boldsymbol{\sigma}_{i-1}^\varepsilon + \mathbb{H}^{-1}(\Delta\boldsymbol{\chi}^\varepsilon)'(\boldsymbol{\sigma}_i^\varepsilon + \boldsymbol{\chi}_{i-1}^\varepsilon)\delta\boldsymbol{\chi}_{i-1}^\varepsilon. \tag{29c}$$

The directional derivative $\delta\boldsymbol{\chi}_i^\varepsilon$ is given by

$$\delta\boldsymbol{\chi}_i^\varepsilon = \delta\boldsymbol{\chi}_{i-1}^\varepsilon + \mathbb{H}\,(\mathbb{C}^{-1}(\delta\boldsymbol{\sigma}_i^\varepsilon - \delta\boldsymbol{\sigma}_{i-1}^\varepsilon) + B_1^\star\delta\Delta\boldsymbol{u}_i^\varepsilon). \tag{29d}$$

Proof. Invoking Theorem A.6 with the setting (22) yields that G^ε is differentiable from $L^{p_2}(\Omega;\mathbb{S})^2 \times W^{-1,p_2}(\Omega;\mathbb{R}^d)$ to $L^{p_1}(\Omega;\mathbb{S}) \times V$, i.e., $(\boldsymbol{\sigma}_i^\varepsilon, \Delta\boldsymbol{u}_i^\varepsilon)$ depends differentiably on $(\boldsymbol{\sigma}_{i-1}^\varepsilon, \boldsymbol{\chi}_{i-1}^\varepsilon, \ell_i^\varepsilon)$ and the derivative solves (29a)–(29b).

Rewriting (21b) we obtain

$$\begin{aligned}\boldsymbol{\chi}_i^\varepsilon &= \boldsymbol{\chi}_{i-1}^\varepsilon + \mathbb{H}\,(\mathbb{C}^{-1}(\boldsymbol{\sigma}_i^\varepsilon - \boldsymbol{\sigma}_{i-1}^\varepsilon) + B_1^\star\Delta\boldsymbol{u}_i^\varepsilon)\\ &= \boldsymbol{\chi}_{i-1}^\varepsilon + \mathbb{H}\,(\mathbb{C}^{-1}(G^{\varepsilon,\boldsymbol{\sigma}}(\boldsymbol{\sigma}_{i-1}^\varepsilon, \boldsymbol{\chi}_{i-1}^\varepsilon, \ell_i^\varepsilon) - \boldsymbol{\sigma}_{i-1}^\varepsilon) + B_1^\star G^{\varepsilon,\Delta\boldsymbol{u}}(\boldsymbol{\sigma}_{i-1}^\varepsilon, \boldsymbol{\chi}_{i-1}^\varepsilon, \ell_i^\varepsilon)).\end{aligned}$$

This implies that also $\boldsymbol{\chi}_i^\varepsilon$ depends differentiably on $(\boldsymbol{\sigma}_{i-1}^\varepsilon, \boldsymbol{\chi}_{i-1}^\varepsilon, \ell_i^\varepsilon)$. Thus, $H^\varepsilon : L^{p_2}(\Omega;\mathbb{S})^2 \times W^{-1,p_2}(\Omega;\mathbb{R}^d) \to L^{p_1}(\Omega;\mathbb{S})^2 \times V$ is differentiable and the derivative is given by the solution of (29). $\square$

3.2. Differentiability of the regularized problem. In this section we address the differentiability of the solution map of the *entire* system (14). To be precise, we show that $\mathcal{G}^\varepsilon \circ E : U^N \to (S^2 \times V)^N$ is Fréchet differentiable, where $E : U = L^2(\Gamma_N;\mathbb{R}^d) \to V'$ is the control operator given by (1). Due to the trace theorem, the operator E maps $U = L^2(\Gamma_N;\mathbb{R}^d)$ continuously into $W^{-1,p}(\Omega;\mathbb{R}^d) = \big(W_D^{1,p'}(\Omega;\mathbb{R}^d)\big)'$ for all $p \in \big[2, 2\frac{d}{d-1}\big]$.

Every solution of a time step of the system (14) is equivalent to an application of H^ε, see (21). In Theorem 3.7 the differentiability of

$$H^\varepsilon : L^{p_{i-1}}(\Omega;\mathbb{S})^2 \times W^{-1,p_{i-1}}(\Omega;\mathbb{R}^d) \ni (\boldsymbol{\sigma}_{i-1}^\varepsilon, \boldsymbol{\chi}_{i-1}^\varepsilon, E\boldsymbol{g}_i^\varepsilon) \\ \mapsto (\boldsymbol{\sigma}_i^\varepsilon, \boldsymbol{\chi}_i^\varepsilon, \Delta\boldsymbol{u}_i^\varepsilon) \in L^{p_i}(\Omega;\mathbb{S})^2 \times V$$

is proven, provided that $2 \le p_i < p_{i-1} \le \bar{p}$. The requisite $E\boldsymbol{g}_i^\varepsilon \in W^{-1,p_{i-1}}(\Omega;\mathbb{R}^d)$ would be implied by $p_{i-1} \le 2\frac{d}{d-1}$. Therefore, we choose a strictly decreasing sequence $\{p_i\}_{i=1}^N$ such that

$$2 \le p_N < p_{N-1} < \cdots < p_2 < p_1 \le \min\left\{\bar{p}, 2\frac{d}{d-1}\right\} \tag{30}$$

and we define

$$\tilde{S}^N := L^{p_1}(\Omega;\mathbb{S}) \times L^{p_2}(\Omega;\mathbb{S}) \times \cdots \times L^{p_{N-1}}(\Omega;\mathbb{S}) \times L^{p_N}(\Omega;\mathbb{S}).$$

This yields the Fréchet differentiability of $\mathcal{G}^\varepsilon \circ E : U^N \to \tilde{S}^N \times \tilde{S}^N \times V^N$. Since $\tilde{S}^N \hookrightarrow S^N$ holds by (30), $\mathcal{G}^\varepsilon \circ E$ is Fréchet differentiable from U^N to $(S^2 \times V)^N$.

The directional derivative $(\delta\boldsymbol{\sigma}^\varepsilon, \delta\boldsymbol{\chi}^\varepsilon, \delta\boldsymbol{u}^\varepsilon) = (\mathcal{G}^\varepsilon)'(E\boldsymbol{g}^\varepsilon)E\delta\boldsymbol{g}^\varepsilon$ is given by the solution of the system

$$\mathbb{C}^{-1}(\delta\boldsymbol{\sigma}_i^\varepsilon - \delta\boldsymbol{\sigma}_{i-1}^\varepsilon) + B_1^\star(\delta\boldsymbol{u}_i^\varepsilon - \delta\boldsymbol{u}_{i-1}^\varepsilon) + J_i^\varepsilon(\delta\boldsymbol{\sigma}_i^\varepsilon + \delta\boldsymbol{\chi}_{i-1}^\varepsilon) = \mathbf{0}, \tag{31a}$$

$$\mathbb{H}^{-1}(\delta\boldsymbol{\chi}_i^\varepsilon - \delta\boldsymbol{\chi}_{i-1}^\varepsilon) + J_i^\varepsilon(\delta\boldsymbol{\sigma}_i^\varepsilon + \delta\boldsymbol{\chi}_{i-1}^\varepsilon) = \mathbf{0}, \tag{31b}$$

$$B_1(\delta\boldsymbol{\sigma}_i^\varepsilon - \delta\boldsymbol{\sigma}_{i-1}^\varepsilon) = E(\delta\boldsymbol{g}_i^\varepsilon - \delta\boldsymbol{g}_{i-1}^\varepsilon), \tag{31c}$$

where $(\delta\boldsymbol{\sigma}_0^\varepsilon, \delta\boldsymbol{\chi}_0^\varepsilon, \delta\boldsymbol{u}_0^\varepsilon) = \mathbf{0}$ and

$$J_i^\varepsilon := -\mathbb{H}^{-1}(\Delta\boldsymbol{\chi}^\varepsilon)'(\boldsymbol{\sigma}_i^\varepsilon(\boldsymbol{g}^\varepsilon) + \boldsymbol{\chi}_{i-1}^\varepsilon(\boldsymbol{g}^\varepsilon)). \tag{32}$$

3.3. Regularized upper level problem: Optimality conditions. Using the control-to-state map $\mathcal{G}^\varepsilon$, we define the reduced objective

$$f^\varepsilon(\boldsymbol{g}^\varepsilon) = F(\mathcal{G}^{\varepsilon,\boldsymbol{u}}(E\boldsymbol{g}^\varepsilon), \boldsymbol{g}^\varepsilon).$$

Here, $\boldsymbol{u}^\varepsilon = \mathcal{G}^{\varepsilon,\boldsymbol{u}}(E\boldsymbol{g}^\varepsilon)$ refers to the second component of $(\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon) = \mathcal{G}^\varepsilon(E\boldsymbol{g}^\varepsilon)$. Due to the differentiability of the objective F and the control-to-state map $\mathcal{G}^\varepsilon \circ E$, the reduced objective f^ε is differentiable. Since the admissible set

U_{ad}^τ is convex, we obtain for a local optimum $\boldsymbol{g}^\varepsilon$ with associated displacements $\boldsymbol{u}^\varepsilon = \mathcal{G}^{\varepsilon,\boldsymbol{u}}(E\boldsymbol{g}^\varepsilon)$ the necessary optimality condition

$$\sum_{i=1}^N \langle \psi_i^\tau(\boldsymbol{u}^\varepsilon),\, \delta\boldsymbol{u}_i^\varepsilon\rangle_{V',V} + \nu\,\langle \boldsymbol{g}^\varepsilon,\, \tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\varepsilon\rangle_{U^N} \ge 0 \quad \text{for all } \tilde{\boldsymbol{g}}^\tau \in U_{\mathrm{ad}}^\tau. \tag{33}$$

where $(\delta\boldsymbol{u}_1^\varepsilon, \ldots, \delta\boldsymbol{u}_N^\varepsilon)$ is the solution of the linearized state equation (31) with right hand side $\delta\boldsymbol{g}^\varepsilon = \tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\varepsilon$, and ψ_i^τ are the partial derivatives of ψ, see Assumption 1.1(2). We compute the adjoint of (31) and define the adjoint state $(\boldsymbol{\Upsilon}^\varepsilon, \boldsymbol{w}^\varepsilon) = (\boldsymbol{v}^\varepsilon, \boldsymbol{\zeta}^\varepsilon, \boldsymbol{w}^\varepsilon) \in S^N \times S^N \times V^N$ as the solution of the system

$$\mathbb{C}^{-1}(\boldsymbol{v}_i^\varepsilon - \boldsymbol{v}_{i+1}^\varepsilon) + B_1^\star(\boldsymbol{w}_i^\varepsilon - \boldsymbol{w}_{i+1}^\varepsilon) + J_i^\varepsilon(\boldsymbol{v}_i^\varepsilon + \boldsymbol{\zeta}_{i+1}^\varepsilon) = \boldsymbol{0}, \tag{34a}$$

$$\mathbb{H}^{-1}(\boldsymbol{\zeta}_i^\varepsilon - \boldsymbol{\zeta}_{i+1}^\varepsilon) + J_i^\varepsilon(\boldsymbol{v}_i^\varepsilon + \boldsymbol{\zeta}_{i+1}^\varepsilon) = \boldsymbol{0}, \tag{34b}$$

$$B_1(\boldsymbol{v}_i^\varepsilon - \boldsymbol{v}_{i+1}^\varepsilon) = \psi_i^\tau(\boldsymbol{u}^\varepsilon), \tag{34c}$$

for $i = N, \ldots, 1$, where $(\boldsymbol{v}_{N+1}^\varepsilon, \boldsymbol{\zeta}_{N+1}^\varepsilon, \boldsymbol{w}_{N+1}^\varepsilon) = \boldsymbol{0}$ and J_i^ε is given by (32). Testing

(31a) with $\boldsymbol{v}_i^\varepsilon$, (31b) with $\boldsymbol{\zeta}_{i+1}^\varepsilon$, (31c) with $\boldsymbol{w}_i^\varepsilon$,
(34a) with $\delta\boldsymbol{\sigma}_i^\varepsilon$, (34b) with $\delta\boldsymbol{\chi}_{i-1}^\varepsilon$, (34c) with $\delta\boldsymbol{u}_i^\varepsilon$,

and summing everything over $i = 1, \ldots, N$ yields

$$\sum_{i=1}^N \langle \psi_i^\tau(\boldsymbol{u}^\varepsilon),\, \delta\boldsymbol{u}_i^\varepsilon\rangle_{V',V} = \sum_{i=1}^N \big\langle E^\star\boldsymbol{w}_i^\varepsilon,\, \tilde{\boldsymbol{g}}_i^\tau - \tilde{\boldsymbol{g}}_{i-1}^\tau - (\bar{\boldsymbol{g}}_i^\varepsilon - \bar{\boldsymbol{g}}_{i-1}^\varepsilon)\big\rangle_{L^2(\Gamma_N;\mathbb{R}^d)}.$$

Hence, the variational inequality (33) becomes

$$\sum_{i=1}^N \big\langle E^\star\boldsymbol{w}_i^\varepsilon,\, \tilde{\boldsymbol{g}}_i^\tau - \tilde{\boldsymbol{g}}_{i-1}^\tau - (\boldsymbol{g}_i^\varepsilon - \boldsymbol{g}_{i-1}^\varepsilon)\big\rangle_{L^2(\Gamma_N;\mathbb{R}^d)} + \langle \nu\,\boldsymbol{g}^\varepsilon,\, \tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\varepsilon\rangle_{U^N} \ge 0 \tag{35}$$

for all $\tilde{\boldsymbol{g}}^\tau \in U_{\mathrm{ad}}^\tau$. This proves the following necessary optimality condition of first order.

Theorem 3.8. *Let* $(\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon, \boldsymbol{g}^\varepsilon) \in (S^2)^N \times V^N \times U_{\mathrm{ad}}^\tau$ *be a local solution of* $(\mathbf{P}^\varepsilon)$. *Then* (35) *is satisfied, where the adjoint state* $\boldsymbol{w}^\varepsilon$ *is defined as the unique solution of* (34).

This theorem will be the starting point in [19] to prove necessary conditions for the time-discrete, unregularized problem $(\mathbf{P}^\tau)$.

4. Convergence of the regularization

In this section we show that the regularization given in Section 2 approximates the original problem. First, we address the convergence of the regularized states in Subsection 4.1. Afterwards, we show that every local solution of the time-discrete optimal control problem $(\mathbf{P}^\tau)$ can be approximated by local solutions of regularized problems, see Subsection 4.2. This will be a main ingredient for showing a necessary optimality system for $(\mathbf{P}^\tau)$ in the subsequent work [19].

4.1. Convergence of the forward problem. This section is devoted to the proof of convergence of the regularized solutions $(\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon, \lambda^\varepsilon)$ of (14) to the unregularized solutions $(\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau, \lambda^\tau)$ as $\varepsilon \searrow 0$ for any fixed number of time steps. We consider a sequence of regularization parameters $\{\varepsilon_k\}_{k=1}^\infty$ and a sequence of loads $\{\ell^{\varepsilon_k}\}_{k=1}^\infty$. For simplicity of the presentation, we drop the index k and refer to the convergence $\ell^{\varepsilon_k} \to \ell^\tau$ by "$\ell^\varepsilon \to \ell^\tau$ as $\varepsilon \searrow 0$".

We start by considering the system of one time step

$$\mathbb{C}^{-1}\boldsymbol{\sigma}^\tau - \mathbb{H}^{-1}\Delta\boldsymbol{\chi}(\boldsymbol{\sigma}^\tau + \boldsymbol{\chi}^\tau) + B_1^\star \Delta \boldsymbol{u}^\tau = \mathscr{L}^\tau \quad \text{in } S, \tag{36a}$$

$$B_1 \boldsymbol{\sigma}^\tau = \ell^\tau \quad \text{in } V', \tag{36b}$$

cf. (13), and its regularization

$$\mathbb{C}^{-1}\boldsymbol{\sigma}^\varepsilon - \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon) + B_1^\star \Delta \boldsymbol{u}^\varepsilon = \mathscr{L}^\varepsilon \quad \text{in } S, \tag{37a}$$

$$B_1 \boldsymbol{\sigma}^\varepsilon = \ell^\varepsilon \quad \text{in } V', \tag{37b}$$

cf. (14a), (14b).

First, we prove a convergence estimate of the nonlinear term $\Delta\boldsymbol{\chi}^\varepsilon(\cdot)$.

Corollary 4.1. *For matrices $\boldsymbol{\sigma}, \boldsymbol{\tau} \in \mathbb{S}$ we have*

$$|\Delta\boldsymbol{\chi}(\boldsymbol{\sigma}) - \Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\tau})|_{\mathbb{S}} \le L\,|\boldsymbol{\sigma} - \boldsymbol{\tau}|_{\mathbb{S}} + \varepsilon\,|\boldsymbol{\tau}|_{\mathbb{S}}, \tag{38}$$

where L is the Lipschitz constant of $\Delta\boldsymbol{\chi}$, see (10). *Similarly, we obtain the estimate*

$$\|\Delta\boldsymbol{\chi}(\boldsymbol{\sigma}) - \Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\tau})\|_S \le L\,\|\boldsymbol{\sigma} - \boldsymbol{\tau}\|_S + \varepsilon\,\|\boldsymbol{\tau}\|_S \tag{39}$$

for matrix functions $\boldsymbol{\sigma}, \boldsymbol{\tau} \in S$.

Proof. The triangle inequality implies

$$|\Delta\boldsymbol{\chi}(\boldsymbol{\sigma}) - \Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\tau})|_{\mathbb{S}} \le |\Delta\boldsymbol{\chi}(\boldsymbol{\sigma}) - \Delta\boldsymbol{\chi}(\boldsymbol{\tau})|_{\mathbb{S}} + |\Delta\boldsymbol{\chi}(\boldsymbol{\tau}) - \Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\tau})|_{\mathbb{S}}.$$

Using the Lipschitz continuity of $\Delta\boldsymbol{\chi}$ we obtain

$$|\Delta\boldsymbol{\chi}(\boldsymbol{\sigma}) - \Delta\boldsymbol{\chi}(\boldsymbol{\tau})|_{\mathbb{S}} \le L\,|\boldsymbol{\sigma} - \boldsymbol{\tau}|_{\mathbb{S}}.$$

Assumption 2.1 implies $|\max(0,x) - \max^\varepsilon(x)| \le \varepsilon$ for all $x \in \mathbb{R}$. By definition (12) of $\Delta\boldsymbol{\chi}^\varepsilon$ this yields

$$|\Delta\boldsymbol{\chi}(\boldsymbol{\tau}) - \Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\tau})|_{\mathbb{S}} \le \varepsilon\,|\boldsymbol{\tau}|_{\mathbb{S}}.$$

Altogether we obtain (38). For $\boldsymbol{\sigma}, \boldsymbol{\tau} \in S$, we achieve (38) pointwise. Taking the $L^2(\Omega)$ norm yields (39). □

Using this result, we obtain the convergence of the solution operator of (37) towards the solution operator of (36) as $\varepsilon \searrow 0$.

Lemma 4.2. *Let a sequence* $\{(\boldsymbol{\chi}^\varepsilon, \mathscr{L}^\varepsilon, \ell^\varepsilon)\}_{\varepsilon>0} \subset S \times S \times V'$ *be given. We denote the solutions of* (37) *by* $\{(\boldsymbol{\sigma}^\varepsilon, \Delta\boldsymbol{u}^\varepsilon)\}_{\varepsilon>0} \subset S \times V$. *Moreover, let* $(\boldsymbol{\sigma}^\tau, \Delta\boldsymbol{u}^\tau) \in S \times V$ *be the solution of* (36) *with given data* $(\boldsymbol{\chi}^\tau, \mathscr{L}^\tau, \ell^\tau) \in S \times S \times V'$. *For all* $\varepsilon > 0$ *we obtain*

$$\begin{aligned}\|\boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon\|_S^2 &\le c\big(\|\mathscr{L}^\tau - \mathscr{L}^\varepsilon\|_S^2 + \|\boldsymbol{\chi}^\tau - \boldsymbol{\chi}^\varepsilon\|_S^2 + \varepsilon^2\,\|\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon\|_S^2 \\ &\qquad + \|\Delta\boldsymbol{u}^\tau - \Delta\boldsymbol{u}^\varepsilon\|_V\,\|\ell^\tau - \ell^\varepsilon\|_{V'}\big), \\ \|\Delta\boldsymbol{u}^\tau - \Delta\boldsymbol{u}^\varepsilon\|_V &\le c\Big(\|\mathscr{L}^\tau - \mathscr{L}^\varepsilon\|_S + \|\boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon\|_S \\ &\qquad + \|\boldsymbol{\chi}^\tau - \boldsymbol{\chi}^\varepsilon\|_S + \varepsilon\,\|\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon\|_S\Big).\end{aligned}$$

The constant $c > 0$ *depends only on the operators* A *and* B_1.

In particular, the convergence of the data $(\boldsymbol{\chi}^\varepsilon, \mathscr{L}^\varepsilon, \ell^\varepsilon) \to (\boldsymbol{\chi}^\tau, \mathscr{L}^\tau, \ell^\tau)$ *as* $\varepsilon \searrow 0$ *implies the convergence of the solutions* $(\boldsymbol{\sigma}^\varepsilon, \Delta\boldsymbol{u}^\varepsilon) \to (\boldsymbol{\sigma}^\tau, \Delta\boldsymbol{u}^\tau)$ *as* $\varepsilon \searrow 0$.

Proof. Testing (36a) and (37a) with $\boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon$ and taking differences yields

$$\begin{aligned}\underline{\alpha}\,\|\boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon\|_S^2 \le{}& \|\boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon\|_S\,\|\mathscr{L}^\tau - \mathscr{L}^\varepsilon\|_S - \langle \boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon,\, B_1^\star\Delta\boldsymbol{u}^\tau - B_1^\star\Delta\boldsymbol{u}^\varepsilon\rangle_S \\ &- \langle \boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon,\, -\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\tau(\boldsymbol{\sigma}^\tau + \boldsymbol{\chi}^\tau) + \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon)\rangle_S,\end{aligned}$$

where $\underline{\alpha} > 0$ is the coercivity constant of $\mathbb{C}^{-1}$. Using (36b) and (37b) we obtain

$$\begin{aligned}\underline{\alpha}\,\|\boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon\|_S^2 \le{}& \|\boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon\|_S\,\|\mathscr{L}^\tau - \mathscr{L}^\varepsilon\|_S + \|\Delta\boldsymbol{u}^\tau - \Delta\boldsymbol{u}^\varepsilon\|_V\,\|\ell^\tau - \ell^\varepsilon\|_{V'} \\ &- \langle \boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon,\, -\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\tau(\boldsymbol{\sigma}^\tau + \boldsymbol{\chi}^\tau) + \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon)\rangle_S.\end{aligned}$$

The monotonicity of $-\Delta\boldsymbol{\chi}$ implies

$$\begin{aligned}&- \langle \boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon,\, -\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\tau(\boldsymbol{\sigma}^\tau + \boldsymbol{\chi}^\tau) + \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon)\rangle_S \\ &= -\langle \boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon,\, -\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\tau(\boldsymbol{\sigma}^\tau + \boldsymbol{\chi}^\tau) + \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\tau(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\tau)\rangle_S \\ &\quad - \langle \boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon,\, -\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\tau(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\tau) + \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon)\rangle_S \\ &\le -\langle \boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon,\, -\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\tau(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\tau) + \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon)\rangle_S.\end{aligned}$$

Using Corollary 4.1 yields

$$\big\|-\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\tau(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\tau) + \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon)\big\|_S \le L\,\|\boldsymbol{\chi}^\tau - \boldsymbol{\chi}^\varepsilon\|_S + \varepsilon\,\|\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon\|_S.$$

This implies

$$\begin{aligned}\underline{\alpha}\,\|\boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon\|_S^2 \le{}& \|\boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon\|_S\,\big(\|\mathscr{L}^\tau - \mathscr{L}^\varepsilon\|_S + L\,\|\boldsymbol{\chi}^\tau - \boldsymbol{\chi}^\varepsilon\|_S + \varepsilon\,\|\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon\|_S\big) \\ &+ \|\Delta\boldsymbol{u}^\tau - \Delta\boldsymbol{u}^\varepsilon\|_V\,\|\ell^\tau - \ell^\varepsilon\|_{V'}.\end{aligned}$$

Young's inequality completes the estimate of $\boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon$.

It remains to verify the estimate of $\Delta\boldsymbol{u}^\tau - \Delta\boldsymbol{u}^\varepsilon$. Testing (36a) and (37a) with $\boldsymbol{\tau} \in S$, $\|\boldsymbol{\tau}\|_S \le 1$, and taking differences yields

$$\begin{aligned}&\langle B_1\boldsymbol{\tau},\, \Delta\boldsymbol{u}^\tau - \Delta\boldsymbol{u}^\varepsilon\rangle_{V',V}\\ &\le c\Big(\|\mathscr{L}^\tau - \mathscr{L}^\varepsilon\|_S + \|\boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon\|_S + \|-\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\tau(\boldsymbol{\sigma}^\tau + \boldsymbol{\chi}^\tau) + \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon)\|_S\Big).\end{aligned}$$

Invoking Corollary 4.1 and the inf-sup condition (7) implies

$$\|\Delta\boldsymbol{u}^\tau - \Delta\boldsymbol{u}^\varepsilon\|_V \le c\Big(\|\mathscr{L}^\tau - \mathscr{L}^\varepsilon\|_S + \|\boldsymbol{\sigma}^\tau - \boldsymbol{\sigma}^\varepsilon\|_S + \|\boldsymbol{\chi}^\tau - \boldsymbol{\chi}^\varepsilon\|_S + \varepsilon\|\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon\|_S\Big). \quad \square$$

Easily this result carries over to the entire problem (14) for any fixed number N of time steps.

Theorem 4.3. *Let a sequence $\{\ell^\varepsilon\}_{\varepsilon>0} \subset (V')^N$ be given, such that $\ell^\varepsilon \to \ell^\tau \in (V')^N$ as $\varepsilon \searrow 0$. Denote by $(\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau)$ and $(\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon)$ the solutions of* (8) *and* (14), *i.e.,*

$$(\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau) = \mathcal{G}^\tau(\ell^\tau) \quad \text{and} \quad (\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon) = \mathcal{G}^\varepsilon(\ell^\varepsilon).$$

Then $(\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon) \to (\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau) \in (S^2 \times V)^N$ as $\varepsilon \searrow 0$.

Finally, we show the convergence of the plastic multiplier.

Corollary 4.4. *Let a sequence $\{\ell^\varepsilon\}_{\varepsilon>0} \subset (V')^N$ be given, such that $\ell^\varepsilon \to \ell^\tau \in (V')^N$ as $\varepsilon \to 0$. We denote by λ^ε the associated plastic multipliers according to* (17) *and by λ^τ the multiplier given by* (9). *Then $\lambda^\varepsilon \in L^2(\Omega)^N$ and $\lambda^\varepsilon \to \lambda^\tau$ in $L^2(\Omega)^N$.*

Proof. Step (1): From (9a) and (18a) we infer

$$\begin{aligned}\tau\,\lambda_i^\tau\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_i^\tau &= -A(\boldsymbol{\Sigma}_i^\tau - \boldsymbol{\Sigma}_{i-1}^\tau) - B^\star(\boldsymbol{u}_i^\tau - \boldsymbol{u}_{i-1}^\tau),\\ \text{and}\quad \tau\,\lambda_i^\varepsilon\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon &= -A(\boldsymbol{\Sigma}_i^\varepsilon - \boldsymbol{\Sigma}_{i-1}^\varepsilon) - B^\star(\boldsymbol{u}_i^\varepsilon - \boldsymbol{u}_{i-1}^\varepsilon),\end{aligned}$$

for all $i = 1, \ldots, N$. By Theorem 4.3 we obtain

$$\tau\,\lambda_i^\varepsilon\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon \to \tau\,\lambda_i^\tau\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_i^\tau \quad \text{in } S^2 \text{ for all } i = 1, \ldots, N. \tag{40}$$

Step (2): By using Assumption 2.1 and the definition of $A_i^{\varepsilon,-}$ in (19a) we obtain $\lambda_i^\varepsilon = 0$ on $A_i^{\varepsilon,-}$. Moreover, the definition of α_i^ε in (15) implies $\alpha_i^\varepsilon \in [0,\varepsilon]$ on $A_i^{\varepsilon,0}$ and $\alpha_i^\varepsilon = 1 - \frac{\tilde\sigma_0}{|\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon|}$ on $A_i^{\varepsilon,+}$. Altogether we obtain $\|\lambda_i^\varepsilon\,\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon\|_S^2 = \int_{\Omega\setminus A_i^{\varepsilon,-}} |\lambda_i^\varepsilon|^2 |\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon|^2\,\mathrm{d}x = \int_{\Omega\setminus A_i^{\varepsilon,-}} |\lambda_i^\varepsilon|^2 (1-\alpha_i^\varepsilon)^2 |\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon|^2\,\mathrm{d}x \ge \frac{(1-\varepsilon)^2\tilde\sigma_0^2}{(1+\varepsilon)^2}\int_{A_i^{\varepsilon,0}} |\lambda_i^\varepsilon|^2\,\mathrm{d}x + \int_{A_i^{\varepsilon,+}} |\lambda_i^\varepsilon|^2 \tilde\sigma_0^2\,\mathrm{d}x \ge \big(\frac{1-\varepsilon}{1+\varepsilon}\big)^2 \tilde\sigma_0^2 \|\lambda_i^\varepsilon\|_{L^2(\Omega)}^2$. Using (40) we obtain the boundedness of λ_i^ε in $L^2(\Omega)$ as $\varepsilon \searrow 0$. Therefore, a subsequence (denoted by the same symbol) converges weakly towards some $\tilde{\lambda}_i \in L^2(\Omega)$. By $\boldsymbol{\Sigma}_i^\varepsilon \to \boldsymbol{\Sigma}_i^\tau$ in S^2 we infer $\lambda_i^\varepsilon \, \mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon \rightharpoonup \tilde{\lambda}_i \, \mathcal{D}\boldsymbol{\Sigma}_i^\tau$ in $L^1(\Omega; \mathbb{S})$. Hence, (40) implies

$$\tilde{\lambda}_i = \lambda_i^\tau \quad \text{on } \{x \in \Omega : |\mathcal{D}\boldsymbol{\Sigma}_i^\tau| \neq 0\}. \tag{41}$$

Step (3): We have

$$\begin{aligned} \|\tilde{\lambda}_i\|_{L^2(\Omega)} &\leq \liminf \|\lambda_i^\varepsilon\|_{L^2(\Omega)} \\ &\leq \liminf \Big(\frac{1+\varepsilon}{1-\varepsilon}\Big) \frac{1}{\tilde{\sigma}_0} \|\lambda_i^\varepsilon \, \mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon\|_S = \frac{1}{\tilde{\sigma}_0} \|\lambda_i^\tau \, \mathcal{D}\boldsymbol{\Sigma}_i^\tau\|_S = \|\lambda_i^\tau\|_{L^2(\Omega)}, \end{aligned} \tag{42}$$

in particular $\|\tilde{\lambda}_i\|_{L^2(\Omega)} \leq \|\lambda_i^\tau\|_{L^2(\Omega)}$. By (41) and $\lambda_i^\tau = 0$ on $\{x \in \Omega : |\mathcal{D}\boldsymbol{\Sigma}_i^\tau| = 0\}$, we infer $\lambda_i^\tau = \tilde{\lambda}_i$. Now, (40) and (42) imply the convergence of norms

$$\|\lambda_i^\varepsilon\|_{L^2(\Omega)} \to \|\lambda_i^\tau\|_{L^2(\Omega)}$$

and hence $\lambda_i^\varepsilon \to \lambda_i^\tau$ in $L^2(\Omega)$. Since the limit of λ_i^ε is independent of the subsequence chosen in the second step, the whole sequence λ_i^ε converges towards λ_i^τ strongly in $L^2(\Omega)$. □

4.2. Approximability of solutions. In this section we will study which optima of the time-discrete problem $(\mathbf{P}^\tau)$ can be approximated by optima of its regularized counterparts $(\mathbf{P}^\varepsilon)$. Let us recall that by Assumption 1.2

(1) the admissible set U_{ad}^τ is nonempty, convex and closed in U^N, and
(2) the objective $F : (V \times U)^N$ is weakly lower semicontinuous.

Lemma 4.5. *Let* $\{\boldsymbol{g}^\varepsilon\}_{\varepsilon>0}$ *be a sequence of global solutions to* $(\mathbf{P}^\varepsilon)$.

(i) *There exists an accumulation point* $\boldsymbol{g}^\tau$ *of* $\{\boldsymbol{g}^\varepsilon\}_{\varepsilon>0}$.
(ii) *Every weak accumulation point of* $\{\boldsymbol{g}^\varepsilon\}_{\varepsilon>0}$ *is a strong accumulation point and a global solution of* $(\mathbf{P}^\tau)$.

We are going to use standard arguments. For convenience, we included the proof.

Proof. By assumption, U_{ad}^τ is nonempty. Hence, there exists some $\tilde{\boldsymbol{g}}^\tau \in U_{\text{ad}}^\tau$. Hence, by Proposition 3.5 the corresponding displacements $\tilde{\boldsymbol{u}}^\varepsilon = \mathcal{G}^{\varepsilon,\boldsymbol{u}}(\tilde{\boldsymbol{g}}^\tau)$ converges in V^N. Since $\boldsymbol{g}^\varepsilon$ is a global optimum of $(\mathbf{P}^\varepsilon)$, we have $F(\boldsymbol{u}^\varepsilon, \boldsymbol{g}^\varepsilon) \leq F(\tilde{\boldsymbol{u}}^\varepsilon, \tilde{\boldsymbol{g}}^\tau)$. This implies the boundedness of $\{\boldsymbol{g}^\varepsilon\}_{\varepsilon>0}$ in U^N. Hence, there exists a weakly convergent subsequence. Therefore, assertion (1) follows by assertion (2).

To prove assertion (2), suppose that $\{\boldsymbol{g}^\varepsilon\}_{\varepsilon>0}$ converges weakly towards $\boldsymbol{g}$. We denote by $(\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon) = \mathcal{G}^\varepsilon(E\boldsymbol{g}^\varepsilon)$ the (regularized and time-discrete) solution to (14) and by $(\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau) = \mathcal{G}^\tau(E\boldsymbol{g})$ the solution to (8). Since E embeds U^N compactly into $(V')^N$, Theorem 4.3 implies $\boldsymbol{u}^\varepsilon \to \boldsymbol{u}^\tau$ in V^N.

Let $\tilde{\boldsymbol{g}}^\tau \in U^\tau_{\text{ad}}$ with corresponding unregularized displacement $\tilde{\boldsymbol{u}}^\tau = \mathcal{G}^{\tau,\boldsymbol{u}}(\tilde{\boldsymbol{g}}^\tau)$ be arbitrary. We denote the corresponding regularized displacements by $\tilde{\boldsymbol{u}}^\varepsilon = \mathcal{G}^{\varepsilon,\boldsymbol{u}}(\tilde{\boldsymbol{g}}^\tau)$. By Theorem 4.3 we infer $\tilde{\boldsymbol{u}}^\varepsilon \to \tilde{\boldsymbol{u}}^\tau$. We have

$$\begin{aligned} F(\boldsymbol{u}^\tau, \boldsymbol{g}^\tau) &\le \liminf F(\boldsymbol{u}^\varepsilon, \boldsymbol{g}^\varepsilon) && \text{(by lower semicontinuity of } F) \\ &\le \liminf F(\tilde{\boldsymbol{u}}^\varepsilon, \tilde{\boldsymbol{g}}^\tau) && \text{(by global optimality of } (\boldsymbol{u}^\varepsilon, \boldsymbol{g}^\varepsilon)) \\ &= F(\tilde{\boldsymbol{u}}^\tau, \tilde{\boldsymbol{g}}^\tau). && \text{(by convergence of } \tilde{\boldsymbol{u}}^\varepsilon) \end{aligned}$$

This shows that $\boldsymbol{g}^\tau$ is a global optimal solution of $(\mathbf{P}^\tau)$. Inserting $\tilde{\boldsymbol{g}}^\tau = \boldsymbol{g}^\tau$ yields the convergence of the norms of $\boldsymbol{g}^\varepsilon$ and hence the strong convergence of $\boldsymbol{g}^\varepsilon$ in U^N. □

Lemma 4.6. *Let $\boldsymbol{g}^\tau$ be a strict local optimum of $(\mathbf{P}^\tau)$ w.r.t. the topology of U^N. Then, for every sequence $\{\varepsilon_k\}_{k\in\mathbb{N}}$ tending to 0, there is a sequence $\{\boldsymbol{g}^{\varepsilon_k}\}$ of local solutions to $(\mathbf{P}^{\varepsilon_k})$ such that $\boldsymbol{g}^{\varepsilon_k} \to \boldsymbol{g}^\tau$ strongly in U^N.*

Proof. Since $\boldsymbol{g}^\tau$ is the only global minimum w.r.t. the feasible set $U_{\text{ad}} \cap B_\delta(\boldsymbol{g}^\tau)$ for some $\delta > 0$, this result can be proven similarly to Lemma 4.5. □

Finally, we address the approximability of a (not necessarily strict) local minimum. Let $\boldsymbol{g}^\tau$ be a local optimum of $(\mathbf{P}^\tau)$ w.r.t. the topology of U^N. We define the modified problem, see also [1,3],

$$\left.\begin{aligned} &\text{Minimize} && F_{\boldsymbol{g}^\tau}(\boldsymbol{u}^\tau, \tilde{\boldsymbol{g}}^\tau) = F(\boldsymbol{u}^\tau, \tilde{\boldsymbol{g}}^\tau) + \frac{1}{2}\|\tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\tau\|^2_{U^N} \\ &\text{such that} && (\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau) = \mathcal{G}^\tau(E\tilde{\boldsymbol{g}}^\tau), \\ &\text{and} && \tilde{\boldsymbol{g}}^\tau \in U^\tau_{\text{ad}}. \end{aligned}\right\} \qquad (\mathbf{P}^\tau_{\boldsymbol{g}^\tau})$$

Clearly, $\boldsymbol{g}^\tau$ becomes a *strict* local optimum of $(\mathbf{P}^\tau_{\boldsymbol{g}^\tau})$. This enable us to approximate $\boldsymbol{g}^\tau$ by solutions to regularized problems. Analogously to $(\mathbf{P}^\varepsilon)$ we define the regularized approximation $(\mathbf{P}^\varepsilon_{\boldsymbol{g}^\tau})$.

Corollary 4.7. *Let $\boldsymbol{g}^\tau$ be a local optimum of $(\mathbf{P}^\tau)$ w.r.t. the topology of U^N. Then, for every sequence $\{\varepsilon_k\}_{k\in\mathbb{N}}$ tending to 0, there is a sequence $\{\boldsymbol{g}^{\varepsilon_k}\}$ of local solutions to the modified problems $(\mathbf{P}^{\varepsilon_k}_{\boldsymbol{g}^\tau})$, such that $\boldsymbol{g}^{\varepsilon_k} \to \boldsymbol{g}^\tau$ strongly in U^N.*

Proof. Since the additional term $\|\tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\tau\|^2_{U^N}$ is weakly lower semicontinuous, this result follows analogously to Lemma 4.6. □

Similar to the necessary optimality condition of $(\mathbf{P}^\varepsilon)$ we find for a local optimum $\boldsymbol{g}^\varepsilon \in U^\tau_{\text{ad}}$ of the modified problem $(\mathbf{P}^\varepsilon_{\boldsymbol{g}^\tau})$

$$\sum_{i=1}^N \big\langle E^\star \boldsymbol{w}^\varepsilon_i,\, \tilde{\boldsymbol{g}}^\tau_i - \tilde{\boldsymbol{g}}^\tau_{i-1} - (\boldsymbol{g}^\varepsilon_i - \boldsymbol{g}^\varepsilon_{i-1})\big\rangle_{L^2(\Gamma_N;\mathbb{R}^d)} + \langle \nu\, \boldsymbol{g}^\varepsilon + \boldsymbol{g}^\varepsilon - \boldsymbol{g}^\tau,\, \tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\varepsilon\rangle_{U^N} \ge 0 \quad (43)$$

holds for all $\tilde{\boldsymbol{g}}^\tau \in U^\tau_{\text{ad}}$, where the adjoint displacement $\boldsymbol{w}^\varepsilon$ is given by the solution of (34).

Corollary 4.7 will be the starting point for showing a necessary optimality system for $(\mathbf{P}^\tau)$. This is addressed in a subsequent publication [19].

A. Differentiability of an abstract saddle point problem

In this section we derive a differentiability result for an abstract, nonlinear saddle point problem. We generalize the results given in [11, Appendix A]. We mention that the results of this appendix could also be obtained by invoking [18, Theorem 2.1]. However, it is simpler to prove the result of Theorem A.6 directly than to rephrase [18, Theorem 2.1] in our saddle-point notation.

Throughout this section we consider the abstract system

$$A\,\boldsymbol{\sigma} + J(\boldsymbol{\sigma} + \mathscr{L}_2) + B^\star \boldsymbol{u} = \mathscr{L}_1, \tag{44a}$$
$$B\,\boldsymbol{\sigma} = \ell, \tag{44b}$$

where $(\mathscr{L}_1, \mathscr{L}_2, \ell)$ is the data and $(\boldsymbol{\sigma}, \boldsymbol{u})$ is the solution. We will show in Theorem A.6 that the solution map of (44) is Fréchet differentiable under certain assumptions. The functional analytic setting is made precise in the following assumption.

Assumption A.1 (Basic functional analytic setting). The space X is a Hilbert space and V is a reflexive Banach space. The linear operators $A : X \to X$ and $B : X \to V'$ are bounded. Furthermore, A is coercive and $B^\star$ satisfies the inf-sup condition, i.e.,

$$\|B^\star \boldsymbol{u}\|_X \ge \beta\, \|\boldsymbol{u}\|_V \quad \text{for all } \boldsymbol{u} \in V.$$

The (possibly nonlinear) operator $J : X \to X$ is monotone and continuous.

It is a standard result that the system (44) admits a unique solution $(\boldsymbol{\sigma}, \boldsymbol{u}) \in X \times V$ for all $\mathscr{L}_1, \mathscr{L}_2 \in X$ and $\ell \in V'$, if Assumption A.1 is satisfied. A proof can be found in [11, Lemma A.1].

Lemma A.2 (Nonlinear saddle point problem). *Let Assumption* A.1 *be satisfied. Then, for all* $(\mathscr{L}_1, \mathscr{L}_2, \ell) \in X \times X \times V'$, *the system* (44) *has a unique solution* $(\boldsymbol{\sigma}, \boldsymbol{u}) \in X \times V$. *Moreover, the solution map* $G : X \times X \times V' \to X \times V$ *is Lipschitz continuous.*

Next, we assume J to be differentiable and state a linearization of (44). We consider the case that J is only differentiable with a norm gap, i.e., we have to choose a weaker norm in the image space or a stronger norm in the domain of J. This is the typical case if J is a nonlinear Nemytskii operator, see [16, Section 4.3.2] or [13].

Assumption A.3 (Fréchet differentiability of J). In addition to Assumption A.1, Y and Z are normed linear spaces with continuous embeddings $Z \hookrightarrow Y \hookrightarrow X$. Moreover, J is Fréchet differentiable as a mapping $Z \to Y$. At any $\boldsymbol{\sigma} \in Z$, the derivative $J'(\boldsymbol{\sigma})$ possesses a positive semi-definite extension which maps $X \to X$, i.e., $\langle J'(\boldsymbol{\sigma})\,\delta\boldsymbol{\sigma},\, \delta\boldsymbol{\sigma}\rangle_X \ge 0$ holds for all $\delta\boldsymbol{\sigma} \in X$.

Using Assumption A.3, we state the linearization of (44),

$$(A + J'(\boldsymbol{\sigma} + \mathscr{L}_2))\,\delta\boldsymbol{\sigma} + B^\star\delta\boldsymbol{u} = \delta\mathscr{L}, \tag{45a}$$

$$B\,\delta\boldsymbol{\sigma} = \delta\ell. \tag{45b}$$

Here, $(\delta\mathscr{L}, \delta\ell)$ is the data and $(\delta\boldsymbol{\sigma}, \delta\boldsymbol{u})$ is the solution.

Lemma A.4 (Solvability of the linearized problem)**.** *Let Assumption* A.3 *be satisfied. Then, for all* $(\boldsymbol{\sigma}, \mathscr{L}_2) \in Z \times Z$ *and* $(\delta\mathscr{L}, \delta\ell) \in X \times V'$ *the system* (45) *has a unique solution* $(\delta\boldsymbol{\sigma}, \delta\boldsymbol{u}) \in X \times V$*. Moreover, the solution map* $\tilde{G}(\boldsymbol{\sigma}, \mathscr{L}_2) : X \times V' \to X \times V$ *is Lipschitz continuous.*

Proof. Follows by standard arguments for linear saddle point problems. □

The last ingredient for the proof of Theorem A.6 is the assumption that G and $\tilde{G}$ are also Lipschitz continuous w.r.t. stronger norms, both in the domain and in the image space.

Assumption A.5 (Lipschitz continuity in stronger norms). In addition to Assumption A.3, W is a normed linear space with continuous embedding $W' \hookrightarrow V'$. The partial solution map $G^{\boldsymbol{\sigma}} : (\mathscr{L}_1, \mathscr{L}_2, \ell) \to \boldsymbol{\sigma}$ of (44) is locally Lipschitz as a function $Z \times Z \times W' \to Z$. Moreover, the solution map $\tilde{G}(\boldsymbol{\sigma}, \mathscr{L}_2)$ of (45) which maps $(\delta\mathscr{L}, \delta\ell)$ to $(\delta\boldsymbol{\sigma}, \delta\boldsymbol{u})$ maps $Y \times W' \to Y \times V$.

Under this assumption, we show that the solution mapping of (44) is Fréchet differentiable and the derivative $(\delta\boldsymbol{\sigma}, \delta\boldsymbol{u})$ in direction $(\delta\mathscr{L}_1, \delta\mathscr{L}_2, \delta\ell)$ is given by the solution of the linearization (45) with $\delta\mathscr{L} = \delta\mathscr{L}_1 - J'(\boldsymbol{\sigma} + \mathscr{L}_2)\,\delta\mathscr{L}_2$.

Theorem A.6 (Differentiability)**.** *Let Assumption* A.5 *be satisfied. Then* G *is Fréchet differentiable as a function* $Z \times Z \times W' \to Y \times V$*. The derivative* $(\delta\boldsymbol{\sigma}, \delta\boldsymbol{u})$ *at* $(\mathscr{L}_1, \mathscr{L}_2, \ell)$ *in the direction* $(\delta\mathscr{L}_1, \delta\mathscr{L}_2, \delta\ell)$ *is given by the unique solution of* (45)*, with* $\delta\mathscr{L} = \delta\mathscr{L}_1 - J'(\boldsymbol{\sigma} + \mathscr{L}_2)\delta\mathscr{L}_2$*.*

Proof. Let $\mathscr{L}_i, \delta\mathscr{L}_i \in Z$ $(i = 1, 2)$ and $\ell, \delta\ell \in W'$ be given and set $\mathscr{L}_i' = \mathscr{L}_i + \delta\mathscr{L}_i$ $(i = 1, 2)$, $\ell' = \ell + \delta\ell$ as well as

$$\begin{aligned}
(\boldsymbol{\sigma}, \boldsymbol{u}) &= G(\mathscr{L}_1, \mathscr{L}_2, \ell),\\
(\boldsymbol{\sigma}', \boldsymbol{u}') &= G(\mathscr{L}_1', \mathscr{L}_2', \ell'),\\
(\delta\boldsymbol{\sigma}, \delta\boldsymbol{u}) &= \tilde{G}(\boldsymbol{\sigma}, \mathscr{L}_2)(\delta\mathscr{L}, \delta\ell),
\end{aligned}$$

with $\delta\mathscr{L} = \delta\mathscr{L}_1 - J'(\boldsymbol{\sigma} + \mathscr{L}_2)\delta\mathscr{L}_2$. The remainder is given by

$$\begin{pmatrix}\boldsymbol{\sigma}_r\\ \boldsymbol{u}_r\end{pmatrix} = G(\mathscr{L}_1', \mathscr{L}_2', \ell') - G(\mathscr{L}_1, \mathscr{L}_2, \ell) - \tilde{G}(\boldsymbol{\sigma}, \mathscr{L}_2)(\delta\mathscr{L}, \delta\ell) = \begin{pmatrix}\boldsymbol{\sigma}'\\ \boldsymbol{u}'\end{pmatrix} - \begin{pmatrix}\boldsymbol{\sigma}\\ \boldsymbol{u}\end{pmatrix} - \begin{pmatrix}\delta\boldsymbol{\sigma}\\ \delta\boldsymbol{u}\end{pmatrix}.$$

We have to verify the estimate of the remainder

$$\|\boldsymbol{\sigma}_r\|_Y + \|\boldsymbol{u}_r\|_V = o\,(\|\delta\mathscr{L}_1\|_Z + \|\delta\mathscr{L}_2\|_Z + \|\delta\ell\|_{W'}).$$

A simple calculation shows that the remainder $(\boldsymbol{\sigma}_r, \boldsymbol{u}_r)$ satisfies

$$\begin{aligned}(A + J'(\boldsymbol{\sigma} + \mathscr{L}_2))\,\boldsymbol{\sigma}_r + B^\star \boldsymbol{u}_r &= -\big(J(\boldsymbol{\sigma}' + \mathscr{L}_2') - J(\boldsymbol{\sigma} + \mathscr{L}_2) \\ &\qquad - J'(\boldsymbol{\sigma} + \mathscr{L}_2)(\boldsymbol{\sigma}' - \boldsymbol{\sigma} + \delta\mathscr{L}_2)\big), \\ B\boldsymbol{\sigma}_r &= \mathbf{0}.\end{aligned}$$

By definition of $\tilde{G}(\boldsymbol{\sigma}, \mathscr{L}_2)$, see Lemma A.4, this can be expressed as

$$(\boldsymbol{\sigma}_r, \boldsymbol{u}_r) = -\tilde{G}(\boldsymbol{\sigma}, \mathscr{L}_2)\big(J(\boldsymbol{\sigma}' + \mathscr{L}_2') - J(\boldsymbol{\sigma} + \mathscr{L}_2) - J'(\boldsymbol{\sigma} + \mathscr{L}_2)(\boldsymbol{\sigma}' - \boldsymbol{\sigma} + \delta\mathscr{L}_2), \mathbf{0}\big).$$

The assumption on $\tilde{G}$ yields

$$\|\boldsymbol{\sigma}_r\|_Y + \|\boldsymbol{u}_r\|_V \le C\,\|J(\boldsymbol{\sigma}' + \mathscr{L}_2') - J(\boldsymbol{\sigma} + \mathscr{L}_2) - J'(\boldsymbol{\sigma} + \mathscr{L}_2)(\boldsymbol{\sigma}' - \boldsymbol{\sigma} + \delta\mathscr{L}_2)\|_Y.$$

Since $J : Z \to Y$ is assumed to be Fréchet differentiable, we obtain

$$\|J(\boldsymbol{\sigma}' + \mathscr{L}_2') - J(\boldsymbol{\sigma} + \mathscr{L}_2) - J'(\boldsymbol{\sigma} + \mathscr{L}_2)(\boldsymbol{\sigma}' - \boldsymbol{\sigma} + \delta\mathscr{L}_2)\|_Y = o\,(\|\boldsymbol{\sigma}' - \boldsymbol{\sigma}\|_Z).$$

Due to the local Lipschitz continuity of $G^{\boldsymbol{\sigma}} : Z \times Z \times W' \to Z$, the term on the right hand side is of order $o\,(\|\delta\mathscr{L}_1\|_Z + \|\delta\mathscr{L}_2\|_Z + \|\delta\ell\|_{W'})$, and the combination of all estimates leads to

$$\|\boldsymbol{\sigma}_r\|_Y + \|\boldsymbol{u}_r\|_V = o\,(\|\delta\mathscr{L}_1\|_Z + \|\delta\mathscr{L}_2\|_Z + \|\delta\ell\|_{W'}),$$

which concludes the proof. □

We remark that the result of Theorem A.6 does not simply follow from the implicit function theorem. In order to apply the implicit function theorem to (44), we would need the assumption that $\tilde{G}$ maps $Y \times V'$ to $Z \times V$. This is, however, not satisfied for the situation in which we apply Theorem A.6 in Subsection 3.1.

Acknowledgment. The author would like to express his gratitude to Roland Herzog and Christian Meyer for helpful discussions on the topic of this paper.

This work was supported by a DFG grant within the Priority Program SPP 1253 (*Optimization with Partial Differential Equations*), which is gratefully acknowledged.

References

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

[2] Barbu, V., *Optimal Control of Variational Inequalities.* Res. Notes Math. 100. Boston: Pitman 1984.

[3] Casas, E. and Tröltzsch, F., Second order necessary and sufficient conditions for optimization problems and applications to control theory. *SIAM J. Optim.* 13 (2002), 406 – 431.

[4] Goldberg, H., Kampowsky, W. and Tröltzsch, F., On Nemytskij operators in L_p-spaces of abstract functions. *Math. Nachr.* 155 (1992), 127 – 140.

[5] Grisvard, P., *Elliptic Problems in Nonsmooth Domains.* Boston: Pitman 1985.

[6] Gröger, K., A $W^{1,p}$-estimate for solutions to mixed boundary value problems for second order elliptic differential equations. *Math. Ann.* 283 (1989), 679 – 687.

[7] Haller-Dintelmann, R., Meyer, C., Rehberg, J. and Schiela, A., Hölder continuity and optimal control for nonsmooth elliptic problems. *Appl. Math. Optim.* 60 (2009)(3), 397 – 428.

[8] Han, W. and Reddy, B. D., *Plasticity.* New York: Springer 1999.

[9] Herzog, R., Meyer, C. and Wachsmuth, G., Existence and regularity of the plastic multiplier in static and quasistatic plasticity. *GAMM Reports* 34 (2011)(1), 39 – 44.

[10] Herzog, R., Meyer, C. and Wachsmuth, G., Integrability of displacement and stresses in linear and nonlinear elasticity with mixed boundary conditions. *J. Math. Anal. Appl.* 382 (2011)(2), 802 – 813.

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

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

[13] Krasnoselskii, M. A., Zabreiko, P. P., Pustylnik, E. I. and Sobolevskii, P. E., *Integral Operators in Spaces of Summable Functions.* Leyden: Noordhoff 1976.

[14] Schiela, A. and Wachsmuth, D., Convergence analysis of smoothing methods for optimal control of stationary variational inequalities with control constraints. *ESAIM Math. Model. Numer. Anal.* 47 (2013)(3), 771 – 787.

[15] Steffensen, S. and Ulbrich, M., A new relaxation scheme for mathematical programs with equilibrium constraints. *SIAM J. Optim.* 20 (2010)(5), 2504 – 2539.

[16] Tröltzsch, F., *Optimal Control of Partial Differential Equations. Theory, Methods and Applications* (translated from the 2005 German original by J. Sprekels). Grad. Stud. Math. 112. Providence (RI): Amer. Math. Soc. 2010.

[17] Wachsmuth, G., Optimal control of quasistatic plasticity with linear kinematic hardening, part I: Existence and discretization in time. *SIAM J. Control Optim.* 50 (2012)(5); 2836 – 2861.

[18] Wachsmuth; G., Differentiability of implicit functions. *J. Math. Anal. Appl.* 414 (2014)(1), 259 – 272.

[19] Wachsmuth, G., Optimal control of quasistatic plasticity with linear kinematic hardening III: Optimality conditions. *Z. Anal. Anwend.* 35 (2016) (to appear).

Received April 7, 2014; revised July 21, 2015