OPTIMAL CONTROL OF QUASISTATIC PLASTICITY WITH LINEAR KINEMATIC HARDENING
PART II: REGULARIZATION AND OPTIMALITY CONDITIONS

GERD WACHSMUTH

Abstract. In this paper we consider the optimal control problem governed by a time-dependent variational inequality arising in quasistatic plasticity with linear kinematic hardening. In this second part of the paper we address a regularization technique of the time-discrete problems. By showing their Fréchet differentiability we obtain a necessary optimality system. By passing to the limit in the regularization parameter we derive a necessary optimality system of C-stationary type for the time-discrete problems. Finally, by passing to the limit in the time discretization we obtain a necessary optimality system of weakly stationary type.

4 Introduction

In this paper we consider an optimal control problem for the quasistatic problem of small-strain elastoplasticity. The forward system in the stress-based (so-called dual) form is represented by a time-dependent variational inequality (VI) of mixed type: find generalized stresses $\boldsymbol{\Sigma}(t) \in S^2$ and displacements $\boldsymbol{u}(t) \in V$ which satisfy $\boldsymbol{\Sigma}(t) \in \mathcal{K}$ and

$$\begin{aligned} &\langle A\dot{\boldsymbol{\Sigma}}(t) + B^\star \dot{\boldsymbol{u}}(t), \boldsymbol{T} - \boldsymbol{\Sigma}(t)\rangle_{S^2} \geq 0 \quad \text{for all } \boldsymbol{T} \in \mathcal{K}, \\ &B\boldsymbol{\Sigma}(t) = \ell(t) \quad \text{in } V', \end{aligned} \tag{VI}$$

where $A : S^2 \to S^2$ and $B : S^2 \to V'$ are bounded linear operators. Moreover, (VI) is subject to the initial condition $(\boldsymbol{\Sigma}(0), \boldsymbol{u}(0)) = (\boldsymbol{0}, \boldsymbol{0})$. The convex, closed set $\mathcal{K} \subset S^2$ of admissible stresses is determined by the von Mises yield condition. The details are made precise in Section 1 of the first part.

The optimization of elastoplastic systems is of significant importance for industrial deformation processes, e.g., for the control of the springback of deep-drawn metal sheets. A first step in this direction is the optimal control of *static* plasticity which was considered in Herzog et al. [2010, 2011c], see the comments in the introduction of the first part.

Since the derivation of optimality conditions for optimal control of quasistatic plasticity is more involved, this work is split into two parts. Let us give a brief overview over their contents. In the first part, we review some known results concerning the solution map $\mathcal{G} : \ell \mapsto (\boldsymbol{\Sigma}, \boldsymbol{u})$ of the problem (VI) and derive some new continuity properties. Due to the weak continuity of the solution map $\mathcal{G}$, we are able to prove the existence of optimal controls for the optimal control problem (P), whose definition is repeated in Section 7. Moreover, we state a time discretized counterpart of (VI) and of the optimal control problem, and show the convergence of these discretizations.

In this second part, we regularize the time-discrete problem and show the Fréchet differentiability of the associated solution map $\mathcal{G}^\varepsilon$. This result is non-standard and requires some subtle arguments. The regularized time-discrete optimal control problems are differentiable and consequently, optimality conditions can be derived in a straightforward way. The passage to the limit in the regularization parameter ε leads to an optimality system of C-stationary type for the time-discrete problem. Finally, we pass to the limit with respect to the discretization in time. This part also requires new convergence arguments. Due to the weak mode of convergence of the adjoint variables, the sign condition for the multipliers is lost in the limit and we finally obtain a system of weak stationarity for the optimal control of (**VI**).

Let us put our work into perspective. We give some references for optimal control of *time-dependent* VIs. We mention Barbu [1981], Mignot and Puel [1984], Adams and Lenhart [2002], Ito and Kunisch [2010], which deal with optimal control of a *parabolic obstacle problem*. Moreover, Farshbaf-Shaker [2011] and Hintermüller and Wegner [2011] consider optimal control of the *Allen-Cahn* and *Cahn-Hilliard VIs*, respectively. All of these papers use a *penalization* of the VI to obtain a differentiable problem and pass to the limit with the regularization parameter in the optimality system. In contrast, we use a *relaxation* approach in the current paper.

Let us briefly highlight the main contributions of some of these references. In Mignot and Puel [1984] the authors give an idea how to prove an optimality system of *strong stationary* type for the distributed optimal control of a parabolic VI. As for the elliptic obstacle problem, this is limited to the quite restrictive case of ample controls without control constraints, see also the discussion in [Herzog et al., 2011c, Section 4]. To our knowledge, there are no results on optimality systems of C-stationary type for optimal control problems governed by parabolic VIs. In Ito and Kunisch [2010] the authors consider the control in the coefficient of the main part of a parabolic VI. Via a penalization approach they derive a system of *weak stationarity*. All of the other contributions mentioned above derive even weaker optimality systems. Some of them contain sign or complementarity conditions for some of the dual variables, which, however, hold only for approximating sequences, lacking passage to the limit.

Comparing the optimal control of *quasistatic plasticity* to the control of the *parabolic obstacle problem*, we find the regularities in time of the multipliers of both problems to be similar. Indeed, the multiplier (in our notation θ) associated to the constraint in the VI (in our notation $\phi(\boldsymbol{\Sigma}) \le 0$) is not a function, but a measure in time. Moreover, in both problems the adjoint states (in our notation $(\boldsymbol{\Upsilon}, \boldsymbol{w})$) possess no weak derivative w.r.t. time.

Nevertheless, due to the different spatial regularity of the states, adjoints and multipliers we have to employ different techniques as those used for instance in Ito and Kunisch [2010] for control of the parabolic obstacle problem. Moreover, the analysis is rendered more challenging due to the nonlinearity in the set $\mathcal{K}$, see (1.5), and due to the constraint equation (equilibrium of forces) in (**VI**). Another difficulty arises from the fact that there seem to be no existence results for regularized versions of the parabolic variational inequality (**VI**). Therefore, it is more convenient to regularize the discretization in time (3.4) rather than vice versa. The resulting regularized and time-discrete system is a nonlinear saddle-point problem. Showing the Fréchet differentiability of its solution map is a nontrivial task.

In contrast to our analysis, most papers on optimal control of (parabolic) VIs derive conditions which hold only for *accumulation points* of sequences of stationary points for the regularized problems. In order to show that these conditions are satisfied indeed for *all local minima*, one has to prove that all local minima can be approximated by stationary points of regularized problems. To our knowledge, only

Barbu [1981], Mignot and Puel [1984] derive necessary conditions in this sense for time-dependent VIs. We utilize the approximation results of Section 3.4 in order to show that the derived optimality system (7.35)–(7.40) holds for all local minimizers.

Let us sketch the outline of the paper. In Section 5 we derive regularized versions of the time-discrete problems (3.4). Showing the differentiability of the solution maps of the regularizations is the main contribution of this section. Finally, we show the convergence of the regularized solutions in Corollary 5.12. Section 6 is devoted to showing an optimality system for the time-discrete optimal control problem ($\mathbf{P}^\tau$). An implication of the differentiability of the solution maps associated with the regularized problems is the necessary optimality system (6.7) of the regularized optimal control problem ($\mathbf{P}^\varepsilon$). Using the convergence of the regularized solutions we are able to prove a system of C-stationary type, see Theorem 6.15. In Section 7 we pass to the limit with the time discretization parameter τ. To this end, several convergence arguments have to be used. The most difficult task is to prove the weak convergence of the term $\theta\, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}$ in the adjoint system, see Lemmas 7.8 and 7.9. We arrive at the optimality system of weakly stationary type, see Theorem 7.11.

For convenience, the first section in this second part is labeled as Section 4. Similarly, the appendix is denoted by Section B to set it apart from the appendix of part I. The introduction of the notation can be found in Section 1 of the first part. When appropriate, we repeat equations and definitions with their original labels.

5 Regularization of the time-discrete forward problem

This section is devoted to a regularization of the time-discrete forward problem which was introduced in Section 3.1. For convenience, we recall its definition: given $\ell^\tau \in (V')^N$, find $(\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau) \in (S^2 \times V)^N$ such that $\boldsymbol{\Sigma}_i^\tau \in \mathcal{K}$ and

$$\langle A(\boldsymbol{\Sigma}_i^\tau - \boldsymbol{\Sigma}_{i-1}^\tau) + B^\star(\boldsymbol{u}_i^\tau - \boldsymbol{u}_{i-1}^\tau),\, \boldsymbol{T} - \boldsymbol{\Sigma}_i^\tau\rangle \geq 0 \quad \text{for all } \boldsymbol{T} \in \mathcal{K}, \tag{3.4a}$$

$$B\boldsymbol{\Sigma}_i^\tau = \ell_i^\tau \quad \text{in } V', \tag{3.4b}$$

holds for all $i \in \{1, \dots, N\}$, where $(\boldsymbol{\Sigma}_0^\tau, \boldsymbol{u}_0^\tau) = \mathbf{0}$. Here, $\tau = T/N$ is the time step size. The spaces under consideration are

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

where the domain $\Omega \subset \mathbb{R}^d$ satisfies Assumption 1.1 (1) and $\mathbb{S} = \mathbb{R}^{d\times d}_{\text{sym}}$ is the space of symmetric d-by-d matrices. The operators $A : S^2 \to S^2$ and $B : S^2 \to V'$ are bounded and linear. Moreover, A is coercive, see Assumption 1.1 (3), and $B^\star : V \to S^2$ satisfies the inf-sup condition, see (1.13). The set

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

is convex and closed. Here, the function $\phi : S^2 \to L^1(\Omega)$ is given by

$$\phi(\boldsymbol{\Sigma}) := \big(|\mathcal{D}\boldsymbol{\Sigma}|^2 - \tilde{\sigma}_0^2\big)/2 = \big(|\boldsymbol{\sigma}^D + \boldsymbol{\chi}^D|^2 - \tilde{\sigma}_0^2\big)/2, \tag{1.5}$$

for $\boldsymbol{\Sigma} = (\boldsymbol{\sigma}, \boldsymbol{\chi}) \in S^2$, where $|\cdot|$ denotes the pointwise Frobenius norm, $\boldsymbol{\sigma}^D$ is the deviatoric part of $\boldsymbol{\sigma}$, see (1.6), and $\mathcal{D}\boldsymbol{\Sigma} = \boldsymbol{\sigma}^D + \boldsymbol{\chi}^D$, see (1.8). Let us recall

$$\boldsymbol{\Sigma} \in \mathcal{K} \quad \Rightarrow \quad \mathcal{D}\boldsymbol{\Sigma} \in L^\infty(\Omega; \mathbb{S}). \tag{1.9}$$

By introducing the plastic multiplier $\lambda^\tau \in L^2(\Omega)^N$, the system (3.4) can be formulated as a complementarity system

$$A(\boldsymbol{\Sigma}_i^\tau - \boldsymbol{\Sigma}_{i-1}^\tau) + B^\star(\boldsymbol{u}_i^\tau - \boldsymbol{u}_{i-1}^\tau) + \tau\, \lambda_i^\tau\, \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_i^\tau = \mathbf{0} \quad \text{in } S^2, \tag{3.8a}$$

$$B\boldsymbol{\Sigma}_i^\tau = \ell_i^\tau \quad \text{in } V', \tag{3.8b}$$

$$0 \leq \lambda_i^\tau \quad \perp \quad \phi(\boldsymbol{\Sigma}_i^\tau) \leq 0 \quad \text{a.e. in } \Omega, \tag{3.8c}$$

see Section 3.1 and [Herzog et al., 2011b, Section 2].

In Section 5.1 we derive a regularized counterpart of the time-discrete system (3.4), see (5.4). Sections 5.2 and 5.3 are devoted to the differentiability of the associated solution map. Finally, in Section 5.4 we show the convergence of the regularizations.

As explained in Remark 6.8, this regularization is a relaxation of the complementarity condition $\lambda_i^\tau\,\phi(\boldsymbol{\Sigma}_i^\tau) = 0$, which is tailored to the structure of the time-discrete problem.

Throughout the paper we assume

Assumption 5.1 (Isotropic hardening)**.** The hardening modulus satisfies $\mathbb{H}^{-1}(x) = k_1^{-1}(x)\,\mathbb{I}$, where the parameter $k_1^{-1} \in L^\infty(\Omega)$ is uniformly positive in Ω and $\mathbb{I}$ is the identity map on $\mathbb{S} = \mathbb{R}^{d\times d}_{\text{sym}}$.

5.1. Derivation of the regularized formulation. This section is devoted to the derivation of a regularization of the time-discrete system (3.4). We will employ Assumption 5.1 to derive a regularization which is tailored to the problem (3.4). In Remark 6.8 we show that our regularization approach is indeed a relaxation of the complementarity condition $\lambda_i^\tau\,\phi(\boldsymbol{\Sigma}_i^\tau) = 0$, see (3.8c). If Assumption 5.1 does not hold, one can use a penalization strategy similar to [Herzog et al., 2010, Section 2.2].

First, we will derive an equivalent formulation of (3.4). This is done similarly to the arguments used in [Herzog et al., 2011c, Section 2.2]. To this end, let $\bar{\mathcal{K}} = \{\boldsymbol{\tau} \in S : (\boldsymbol{\tau}, \mathbf{0}) \in \mathcal{K}\}$ be the restriction of the admissible set $\mathcal{K}$ to the first variable. Due to the shift-invariance (1.7), $\boldsymbol{T} = (\boldsymbol{\tau}, \boldsymbol{\mu}) \in \mathcal{K}$ 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 (3.4a) with $\boldsymbol{T}$ yields

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

with $\Delta\boldsymbol{\chi}_i^\tau = \boldsymbol{\chi}_i^\tau - \boldsymbol{\chi}_{i-1}^\tau$. This is equivalent to

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

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

$$\Delta\boldsymbol{\chi}(\boldsymbol{\tau}) := \operatorname{Proj}^{\mathbb{H}^{-1}}_{\bar{\mathcal{K}}}(\boldsymbol{\tau}) - \boldsymbol{\tau}. \tag{5.1}$$

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

By employing Assumption 5.1 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 (5.1) can be evaluated pointwise. Using that the admissible set $\bar{\mathcal{K}}$ is a cylinder in S, a straightforward computation shows

$$\Delta\boldsymbol{\chi}(\boldsymbol{\tau}) = -\max\left\{0, 1 - \frac{\tilde{\sigma}_0}{|\boldsymbol{\tau}^D|}\right\}\boldsymbol{\tau}^D. \tag{5.2}$$

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

$$\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\tau}) := -\max{}^\varepsilon\left(1 - \frac{\tilde\sigma_0}{|\boldsymbol{\tau}^D|}\right)\boldsymbol{\tau}^D, \tag{5.3}$$

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

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

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.

We denote by $\Delta\boldsymbol{\chi}^\varepsilon$ both, the operator mapping $\mathbb{S} \to \mathbb{S}$ as well as the Nemytzki operator mapping $S \to S$. For convenience we use the notation $\max^0(\cdot) = \max\{0,\cdot\}$ and $\Delta\boldsymbol{\chi}^0 = \Delta\boldsymbol{\chi}$ when referring to the unregularized case.

The arguments above gives rise to the following regularized version of (3.4): given the loads $\ell^\varepsilon \in (V')^N$, find $(\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon) \in (S^2 \times V)^N$ satisfying

$$\mathbb{C}^{-1}\boldsymbol{\sigma}_i^\varepsilon - \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}_i^\varepsilon + \boldsymbol{\chi}_{i-1}^\varepsilon) + B^\star(\boldsymbol{u}_i^\varepsilon - \boldsymbol{u}_{i-1}^\varepsilon) = \mathbb{C}^{-1}\boldsymbol{\sigma}_{i-1}^\varepsilon \quad \text{in } S \tag{5.4a}$$

$$B\boldsymbol{\sigma}_i^\varepsilon = \ell_i^\varepsilon \quad \text{in } V' \tag{5.4b}$$

$$\boldsymbol{\chi}_i^\varepsilon = \boldsymbol{\chi}_{i-1}^\varepsilon + \Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}_i^\varepsilon + \boldsymbol{\chi}_{i-1}^\varepsilon) \quad \text{in } S \tag{5.4c}$$

for all $i \in \{1, \dots, N\}$, and with the initial condition $(\boldsymbol{\Sigma}_0^\varepsilon, \boldsymbol{u}_0^\varepsilon) = (\mathbf{0}, \mathbf{0})$.

Using the Browder-Minty theorem, we infer the unique solvability of (5.4), see, e.g., [Herzog et al., 2010, Lemma A.1] for a proof. Let us denote the solution operator of (5.4) mapping $\ell^\varepsilon \in (V')^N$ to $(\boldsymbol{\sigma}^\varepsilon, \boldsymbol{\chi}^\varepsilon, \boldsymbol{u}^\varepsilon)$ by $\mathcal{G}^\varepsilon$.

5.2. Differentiability of one time step. This section is devoted to the differentiability of the solution operator of *one* time step of the system (5.4), see Theorem 5.9. In Section 5.3 this is used to show the differentiability of the solution operator $\mathcal{G}^\varepsilon$ of the *entire* system (5.4).

By using an abstract notation we will simplify the presentation. For all time steps $i \in \{1, \dots, N\}$, the system (5.4a)–(5.4b) 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

$$\mathbb{C}^{-1}\boldsymbol{\sigma} - \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma} + \boldsymbol{\chi}) + B^\star\Delta\boldsymbol{u} = \mathscr{L} \quad \text{in } S, \tag{5.5a}$$

$$B\boldsymbol{\sigma} = \ell \quad \text{in } V'. \tag{5.5b}$$

We denote the solution operator of this abstract system by $G^\varepsilon : (\mathscr{L}, \boldsymbol{\chi}, \ell) \mapsto (\boldsymbol{\sigma}, \Delta\boldsymbol{u})$. Then one time step from (5.4) is equivalent to

$$(\boldsymbol{\sigma}_i^\varepsilon, \Delta\boldsymbol{u}_i^\varepsilon) := (\boldsymbol{\sigma}_i^\varepsilon, \boldsymbol{u}_i^\varepsilon - \boldsymbol{u}_{i-1}^\varepsilon) = G^\varepsilon(\mathbb{C}^{-1}\boldsymbol{\sigma}_{i-1}^\varepsilon, \boldsymbol{\chi}_{i-1}^\varepsilon, \ell_i^\varepsilon), \tag{5.6a}$$

$$\boldsymbol{\chi}_i^\varepsilon = \boldsymbol{\chi}_{i-1}^\varepsilon + \Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}_i^\varepsilon + \boldsymbol{\chi}_{i-1}^\varepsilon). \tag{5.6b}$$

We denote the solution of a single time step by $H^\varepsilon : (\boldsymbol{\sigma}_{i-1}^\varepsilon, \boldsymbol{\chi}_{i-1}^\varepsilon, \ell_i^\varepsilon) \mapsto (\boldsymbol{\sigma}_i^\varepsilon, \boldsymbol{\chi}_i^\varepsilon, \Delta\boldsymbol{u}_i^\varepsilon)$. 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 $p_2 > p_1 \geq 2$. Actually, we prove the differentiability of G^ε, see Proposition 5.8, giving in turn the differentiability of H^ε, see Theorem 5.9. The differentiability of G^ε will be proven by applying the abstract result Theorem B.6 concerning nonlinear saddle-point problems with the setting

$$\left.\begin{aligned} &X = S = L^2(\Omega;\mathbb{S}), && V = H^1_D(\Omega;\mathbb{R}^d), && A = \mathbb{C}^{-1}, \\ &Y = L^{p_1}(\Omega;\mathbb{S}), && W' = W^{-1,p_2}(\Omega;\mathbb{R}^d), && J = -\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon, \\ &Z = L^{p_2}(\Omega;\mathbb{S}). \end{aligned}\right\} \tag{5.7}$$

Here,

$$W^{1,p_2'}_D(\Omega;\mathbb{R}^d) := \{\boldsymbol{u} \in W^{1,p_2'}(\Omega;\mathbb{R}^d) : \boldsymbol{u} = \mathbf{0} \text{ on } \Gamma_D\} \quad \text{and}$$
$$W^{-1,p_2}(\Omega;\mathbb{R}^d) := \Big(W^{1,p_2'}_D(\Omega;\mathbb{R}^d)\Big)',$$

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

Let us comment on the prerequisites of Theorem B.6. Surely, Assumption B.1 is satisfied, in particular the monotonicity of $-\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon$ follows easily from Assumption 5.2. Assumption B.3, which concerns the differentiability of the nonlinear term, is proven in Lemma 5.3. In order to satisfy Assumption B.5, we have to prove the local Lipschitz continuity of the solution operator of (5.5) and its linearization w.r.t. stronger norms. This is the main work of this section and it is done in Proposition 5.7 and Proposition 5.8. Finally, the application of Theorem B.6 yields the Fréchet differentiability of H^ε, see Theorem 5.9.

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

Lemma 5.3. For every $\varepsilon > 0$, 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

$$\begin{aligned}(\Delta\boldsymbol{\chi}^\varepsilon)'(\boldsymbol{\tau})\,\delta\boldsymbol{\tau} = -(\max{}^\varepsilon)'\left(1 - \frac{\tilde\sigma_0}{|\boldsymbol{\tau}^D|}\right)\frac{\tilde\sigma_0}{|\boldsymbol{\tau}^D|^3}\,(\boldsymbol{\tau}^D : \delta\boldsymbol{\tau}^D)\,\boldsymbol{\tau}^D \\ - \max{}^\varepsilon\left(1 - \frac{\tilde\sigma_0}{|\boldsymbol{\tau}^D|}\right)\delta\boldsymbol{\tau}^D.\end{aligned} \tag{5.8}$$

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. This can be proven analogously to [Herzog et al., 2010, Proposition 2.11].

The linearized version of (5.5) reads, cf. (B.3),

$$\mathbb{C}^{-1}\boldsymbol{\sigma} - \mathbb{H}^{-1}(\Delta\boldsymbol{\chi}^\varepsilon)'(\hat{\boldsymbol{\sigma}})\,\boldsymbol{\sigma} + B^\star\Delta\boldsymbol{u} = \mathscr{L} \quad \text{in } S \tag{5.9a}$$
$$B\boldsymbol{\sigma} = \ell \quad \text{in } V' \tag{5.9b}$$

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 (5.5) (w.r.t. $\mathscr{L}$, $\boldsymbol{\chi}$ and ℓ) and of (5.9) (w.r.t. $\mathscr{L}$ and ℓ) in the norms of $Z = L^{p_2}(\Omega;\mathbb{S})$ and $Y = L^{p_1}(\Omega;\mathbb{S})$. 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 [Herzog et al., 2011a, 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 (5.5a) and (5.9a). To be precise, we define $\mathcal{Q}$ to contain all operators mapping $\Omega \times \mathbb{S} \to \mathbb{S}$,

$$(x, \boldsymbol{\sigma}) \mapsto \mathbb{C}^{-1}(x)\,\boldsymbol{\sigma} - \mathbb{H}^{-1}(x)\Delta\boldsymbol{\chi}^{\varepsilon}(\boldsymbol{\sigma} + \boldsymbol{\chi}(x)) \quad \text{for all } \varepsilon \geq 0,\ \boldsymbol{\chi} \in S, \tag{5.10a}$$

$$\text{and} \quad (x, \boldsymbol{\sigma}) \mapsto \mathbb{C}^{-1}(x)\,\boldsymbol{\sigma} - \mathbb{H}^{-1}(x)(\Delta\boldsymbol{\chi}^{\varepsilon})'(\hat{\boldsymbol{\sigma}}(x))\,\boldsymbol{\sigma} \quad \text{for all } \varepsilon > 0,\ \hat{\boldsymbol{\sigma}} \in S. \tag{5.10b}$$

We remark, since $\varepsilon = 0$ is admissible in (5.10a), we also included the operator of the unregularized problem (5.4). For every $Q \in \mathcal{Q}$ we will denote the induced Nemytzki operator with the same symbol. Due to the definition of $\mathcal{Q}$, both systems (5.5) and (5.9) can be written as

$$Q(\boldsymbol{\sigma}) + B^{\star}\boldsymbol{u} = \mathscr{L}, \tag{5.11a}$$

$$B\boldsymbol{\sigma} = \ell, \tag{5.11b}$$

with the corresponding $Q \in \mathcal{Q}$. We show, that the solution mapping of (5.11) is Lipschitz continuous with respect to $\mathscr{L}$ and ℓ. For the system (5.5), the Lipschitz dependence on $\boldsymbol{\chi}$ is proven afterwards.

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

Lemma 5.4. For all $Q \in \mathcal{Q}$ we have

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

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

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 $(\Delta\boldsymbol{\chi}^{\varepsilon})'$. By (5.1) we obtain the Lipschitz continuity of $\Delta\boldsymbol{\chi} = \Delta\boldsymbol{\chi}^{0}$. The Lipschitz continuity of $\Delta\boldsymbol{\chi}^{\varepsilon}$ and $(\Delta\boldsymbol{\chi}^{\varepsilon})'$ can be proved easily, starting from (5.3) and (5.8), respectively. Indeed, the constant $\overline{\alpha} + 2\,\overline{h}$ could be improved to $\overline{\alpha} + \overline{h}$, see (6.10) and (6.25).

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 $Q \in \mathcal{Q}$

$$\big(Q^{-1}(x, \boldsymbol{\varepsilon}) - Q^{-1}(x, \hat{\boldsymbol{\varepsilon}})\big) : (\boldsymbol{\varepsilon} - \hat{\boldsymbol{\varepsilon}}) \geq m\,|\boldsymbol{\varepsilon} - \hat{\boldsymbol{\varepsilon}}|^2, \tag{5.12a}$$

$$|Q^{-1}(x, \boldsymbol{\varepsilon}) - Q^{-1}(x, \hat{\boldsymbol{\varepsilon}})| \leq M\,|\boldsymbol{\varepsilon} - \hat{\boldsymbol{\varepsilon}}|, \tag{5.12b}$$

for almost all $x \in \Omega$. Here, we can choose

$$m = \frac{\underline{\alpha}}{(\overline{\alpha} + 2\,\overline{h})^2} \quad \text{and} \quad M = \frac{1}{\underline{\alpha}},$$

cf. [Herzog et al., 2011c, (2.9)].This implies that the Nemytzki 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 (5.10a), this requires $\boldsymbol{\chi} \in L^p(\Omega; \mathbb{S})$. For Q defined by (5.10b) this imposes no restriction, since $Q(\cdot, \mathbf{0}) = \mathbf{0}$ holds independently of $\hat{\boldsymbol{\sigma}}$.

By applying Q^{-1} to (5.11), we obtain that $\boldsymbol{u}$ solves

$$B\,Q^{-1}\,(-B^{\star}\boldsymbol{u} + \mathscr{L}) = \ell. \tag{5.13}$$

This is a quasilinear system of infinitesimal elasticity. Higher integrability of solutions to such systems is the topic of Herzog et al. [2011a].

Theorem 5.5. 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 (5.13) 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 Q.

Proof. In order to apply [Herzog et al., 2011a, 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 (5.12), we obtain that [Herzog et al., 2011a, Assumption 1.5 (2)] is fulfilled uniformly for all $Q \in \mathcal{Q}$. Note that by [Herzog et al., 2011a, Remark 1.6 (2)] it is sufficient to ensure $Q^{-1}(\cdot, \mathbf{0}) \in L^p(\Omega; \mathbb{S})$.

According to [Herzog et al., 2011a, Proposition 1.2], there exists an exponent $\bar{p}$ such that for all $p \in [2, \bar{p}]$, the solution map of (5.13) 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 $\ell_i = B\,Q^{-1}(-B^\star \boldsymbol{u}_i + \mathscr{L})$. 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 following, we show that $\boldsymbol{u}$ in (5.13) depends also Lipschitz continuously on the data $\mathscr{L}$. We also show the Lipschitz dependence of $\boldsymbol{\sigma}$ in (5.11) on ℓ and $\mathscr{L}$.

Proposition 5.6. 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 (5.11) 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 Theorem 5.5.

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 $\ell_i = B\,Q^{-1}(-B^\star \boldsymbol{u}_i + \mathscr{L}_i)$.

Using $\boldsymbol{\tau} = Q^{-1}(-B^\star \boldsymbol{u}_2 + \mathscr{L}_2) - Q^{-1}(-B^\star \boldsymbol{u}_2 + \mathscr{L}_1)$ we obtain

$$\ell_2 - B\boldsymbol{\tau} = B\,Q^{-1}(-B^\star \boldsymbol{u}_2 + \mathscr{L}_1).$$

This shows that $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ solve the systems

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

An application of Theorem 5.5 yields

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

By definition of $\boldsymbol{\tau}$ and (5.12b) 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 (5.11) we infer

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

Since Q^{-1} maps $L^p(\Omega;\mathbb{S})$ Lipschitz continuously into itself by (5.12b), and since $\boldsymbol{u}$ depends Lipschitz continuously on ℓ and $\mathscr{L}$, we obtain

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

The next proposition translates the abstract result of Proposition 5.6 to the systems (5.5) and (5.9). It is an immediate consequence using the definition of $\mathcal{Q}$ in (5.10).

Proposition 5.7. Let $p \in [2,\bar{p}]$. The solution maps of

(1) (5.5), i.e. G^ε, for fixed $\varepsilon \ge 0$ and $\boldsymbol{\chi} \in L^p(\Omega;\mathbb{S})$ and
(2) (5.9), i.e. $\tilde{G}^\varepsilon$, for fixed $\varepsilon > 0$ and $\hat{\boldsymbol{\sigma}} \in S$

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.

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

where $(\boldsymbol{\sigma}_i, \Delta \boldsymbol{u}_i)$ are the solutions of (5.5) or (5.9) respectively with the data $(\ell_i, \mathscr{L}_i)$, $i = 1, 2$.

It remains to prove the Lipschitz dependence of the solution of (5.5) on $\boldsymbol{\chi}$.

Proposition 5.8. Let $p \in [2,\bar{p}]$ and $\varepsilon \ge 0$. Then, in addition to Proposition 5.7, G^ε is also Lipschitz continuous w.r.t. $\boldsymbol{\chi} \in L^p(\Omega;\mathbb{S})$, i.e.

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

where $(\boldsymbol{\sigma}_i, \Delta \boldsymbol{u}_i)$ is the solution of (5.5) 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 (5.5). We define $\delta\boldsymbol{\chi} = \boldsymbol{\chi}_2 - \boldsymbol{\chi}_1$. We infer from (5.5)

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

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 5.7 yields

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

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

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

As announced in the beginning of this section, we use Theorem B.6 with the setting (5.7) to infer the Fréchet differentiability of H^ε.

Theorem 5.9. Let $\bar p$ from Theorem 5.5 be given and $\varepsilon > 0$. Let p_1, p_2 satisfy $2 \le p_1 < p_2 \le \bar p$. Then the operator H^ε defined in (5.6) 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

$$(\mathbb{C}^{-1} - \mathbb{H}^{-1}(\Delta\boldsymbol\chi^\varepsilon)'(\boldsymbol\sigma_i^\varepsilon + \boldsymbol\chi_{i-1}^\varepsilon))\delta\boldsymbol\sigma_i^\varepsilon + B^\star\delta\Delta\boldsymbol u_i^\varepsilon = \delta\mathscr{L}_i, \tag{5.14a}$$

$$B\delta\boldsymbol\sigma_i^\varepsilon = \delta\ell_i^\varepsilon, \tag{5.14b}$$

where

$$\delta\mathscr{L}_i = \mathbb{C}^{-1}\delta\boldsymbol\sigma_{i-1}^\varepsilon + \mathbb{H}^{-1}(\Delta\boldsymbol\chi^\varepsilon)'(\boldsymbol\sigma_i^\varepsilon + \boldsymbol\chi_{i-1}^\varepsilon)\delta\boldsymbol\chi_{i-1}^\varepsilon. \tag{5.14c}$$

The directional derivative $\delta\boldsymbol\chi_i^\varepsilon$ is given by

$$\delta\boldsymbol\chi_i^\varepsilon = \delta\boldsymbol\chi_{i-1}^\varepsilon + \mathbb{H}\,(\mathbb{C}^{-1}(\delta\boldsymbol\sigma_i^\varepsilon - \delta\boldsymbol\sigma_{i-1}^\varepsilon) + B^\star\delta\Delta\boldsymbol u_i^\varepsilon). \tag{5.14d}$$

Proof. Invoking Theorem B.6 with the setting (5.7) 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 (5.14a)–(5.14b).

Rewriting (5.6b) we obtain

$$\begin{aligned}\boldsymbol\chi_i^\varepsilon &= \boldsymbol\chi_{i-1}^\varepsilon + \mathbb{H}\,(\mathbb{C}^{-1}(\boldsymbol\sigma_i^\varepsilon - \boldsymbol\sigma_{i-1}^\varepsilon) + B^\star\Delta\boldsymbol u_i^\varepsilon)\\ &= \boldsymbol\chi_{i-1}^\varepsilon + \mathbb{H}\,(\mathbb{C}^{-1}(G^{\varepsilon,\boldsymbol\sigma}(\boldsymbol\sigma_{i-1}^\varepsilon, \boldsymbol\chi_{i-1}^\varepsilon, \ell_i^\varepsilon) - \boldsymbol\sigma_{i-1}^\varepsilon) + B^\star G^{\varepsilon,\Delta\boldsymbol u}(\boldsymbol\sigma_{i-1}^\varepsilon, \boldsymbol\chi_{i-1}^\varepsilon, \ell_i^\varepsilon)).\end{aligned}$$

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 (5.14).

5.3. Differentiability of the regularized problem. In this section we address the differentiability of the solution map of the *entire* system (5.4). To be precise, we show that $\mathcal{G}^\varepsilon \circ E : U^N \to (S^2 \times V)^N$ is Fréchet differentiable, where the embedding $E : U = L^2(\Gamma_N;\mathbb{R}^d) \to V'$ is given by

$$\langle \boldsymbol v,\, E\boldsymbol g\rangle_{V,V'} := -\int_{\Gamma_N} \boldsymbol v \cdot \boldsymbol g \,\mathrm{d}s \quad \text{for all } \boldsymbol v \in V. \tag{1.2}$$

Due to the trace theorem, the embedding 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 $p \in [2, 2\,d/(d-1)]$.

Every solution of a time step of the system (5.4) is equivalent to an application of H^ε, see (5.6). In Theorem 5.9 the differentiability of

$$\begin{aligned}H^\varepsilon : L^{p_{i-1}}(\Omega;\mathbb{S})^2 \times W^{-1,p_{i-1}}(\Omega;\mathbb{R}^d) \ni (\boldsymbol\sigma_{i-1}^\varepsilon, \boldsymbol\chi_{i-1}^\varepsilon, E\boldsymbol g_i^\varepsilon)&\\ \mapsto (\boldsymbol\sigma_i^\varepsilon, \boldsymbol\chi_i^\varepsilon, \Delta\boldsymbol u_i^\varepsilon) \in L^{p_i}(\Omega;\mathbb{S})^2 \times V&\end{aligned}$$

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 $p_{i-1} \le 2\,d/(d-1)$. Therefore, we choose a strictly decreasing sequence $\{p_i\}_{i=1}^N$ such that

$$2 \le p_N < p_{N-1} < \ldots < p_2 < p_1 \le \min\{\bar p, 2\,d/(d-1)\} \tag{5.15}$$

and we define

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

Due to (5.15), Theorem 5.9 implies 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 (5.15), $\mathcal{G}^\varepsilon \circ E$ 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

$$\mathbb{C}^{-1}(\delta\boldsymbol{\sigma}_i^\varepsilon - \delta\boldsymbol{\sigma}_{i-1}^\varepsilon) + B^\star(\delta\boldsymbol{u}_i^\varepsilon - \delta\boldsymbol{u}_{i-1}^\varepsilon) + J_i^\varepsilon(\delta\boldsymbol{\sigma}_i^\varepsilon + \delta\boldsymbol{\chi}_{i-1}^\varepsilon) = \mathbf{0}, \tag{5.16a}$$

$$\mathbb{H}^{-1}(\delta\boldsymbol{\chi}_i^\varepsilon - \delta\boldsymbol{\chi}_{i-1}^\varepsilon) + J_i^\varepsilon(\delta\boldsymbol{\sigma}_i^\varepsilon + \delta\boldsymbol{\chi}_{i-1}^\varepsilon) = \mathbf{0}, \tag{5.16b}$$

$$B(\delta\boldsymbol{\sigma}_i^\varepsilon - \delta\boldsymbol{\sigma}_{i-1}^\varepsilon) = E(\delta\boldsymbol{g}_i^\varepsilon - \delta\boldsymbol{g}_{i-1}^\varepsilon), \tag{5.16c}$$

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

$$J_i^\varepsilon := -\mathbb{H}^{-1}(\Delta\boldsymbol{\chi}^\varepsilon)'(\boldsymbol{\sigma}_i^\varepsilon(\boldsymbol{g}^\varepsilon) + \boldsymbol{\chi}_{i-1}^\varepsilon(\boldsymbol{g}^\varepsilon)). \tag{5.17}$$

The embedding $\tilde{S}^N \hookrightarrow S^N$, which is ensured by (5.15), implies that $\mathcal{G}^\varepsilon \circ E : U^N \to S^N \times S^N \times V$ is Fréchet differentiable. This shows that the solution operator of the time-discrete and regularized system (5.4) is Fréchet differentiable w.r.t. a stronger norm for the data $\ell_i^\varepsilon = E\boldsymbol{g}_i^\varepsilon$. This will not cause any difficulty later since the control $\boldsymbol{g}^\varepsilon$ belongs to U^N in the regularized control problem ($\mathbf{P}^\varepsilon$).

5.4. Convergence of the regularization. This section is devoted to the proof of convergence of the regularized solutions $(\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon)$ to the unregularized solutions $(\boldsymbol{\Sigma}^0, \boldsymbol{u}^0)$ as $\varepsilon \searrow 0$ in (5.4) for any fixed number of time steps. We consider a sequence of regularization parameters $\{\varepsilon_k\}_{k=1}^\infty$ and a sequence of loads $\{\ell^{\varepsilon_k}\}_{k=1}^\infty$. Similarly to Section 3.2 we drop the index k and refer to the convergence $\ell^{\varepsilon_k} \to \ell$ by "$\ell^\varepsilon \to \ell$ as $\varepsilon \searrow 0$".

We start by considering the system of one time step

$$\mathbb{C}^{-1}\boldsymbol{\sigma}^\varepsilon - \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon) + B^\star\Delta\boldsymbol{u}^\varepsilon = \mathscr{L}^\varepsilon \quad \text{in } S, \tag{5.18a}$$

$$B\boldsymbol{\sigma}^\varepsilon = \ell^\varepsilon \quad \text{in } V'. \tag{5.18b}$$

In case $\varepsilon > 0$ this is the regularized system (5.4a)–(5.4b) and in case $\varepsilon = 0$, we obtain the unregularized system (3.4).

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

Corollary 5.10. For matrices $\boldsymbol{\sigma}, \boldsymbol{\tau} \in \mathbb{S}$ we have

$$|\Delta\boldsymbol{\chi}(\boldsymbol{\sigma}) - \Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\tau})|_{\mathbb{S}} \le L\,|\boldsymbol{\sigma} - \boldsymbol{\tau}|_{\mathbb{S}} + \varepsilon\,|\boldsymbol{\tau}|_{\mathbb{S}}, \tag{5.19}$$

where L is the Lipschitz constant of $\Delta\boldsymbol{\chi}$, see (5.1). Similarly, we obtain for matrix functions $\boldsymbol{\sigma}, \boldsymbol{\tau} \in S$ the estimate

$$\|\Delta\boldsymbol{\chi}(\boldsymbol{\sigma}) - \Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\tau})\|_S \le L\,\|\boldsymbol{\sigma} - \boldsymbol{\tau}\|_S + \varepsilon\,\|\boldsymbol{\tau}\|_S. \tag{5.20}$$

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 5.2 implies $|\max(0,x) - \max^\varepsilon(x)| \le \varepsilon$ for all $x \in \mathbb{R}$. By definition (5.3) 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 (5.19).

For $\boldsymbol{\sigma}, \boldsymbol{\tau} \in S$, we achieve (5.19) pointwise. Taking the $L^2(\Omega)$ norm yields (5.20).

Using this result, we obtain the convergence of the solution operator of (5.18) as $\varepsilon \searrow 0$.

Theorem 5.11. Let a sequence $\{(\boldsymbol{\chi}^\varepsilon, \mathscr{L}^\varepsilon, \ell^\varepsilon)\}_{\varepsilon\geq 0}$ be given. Denote by $\{(\boldsymbol{\sigma}^\varepsilon, \Delta\boldsymbol{u}^\varepsilon)\}_{\varepsilon\geq 0}$ the solutions of the system of one time step (5.18). We obtain for all $\varepsilon > 0$

$$\begin{aligned}\|\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon\|_S^2 &\leq c\big(\|\mathscr{L}^0 - \mathscr{L}^\varepsilon\|_S^2 + \|\boldsymbol{\chi}^0 - \boldsymbol{\chi}^\varepsilon\|_S^2 + \varepsilon^2\,\|\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon\|_S^2 \\ &\qquad + \|\Delta\boldsymbol{u}^0 - \Delta\boldsymbol{u}^\varepsilon\|_V\,\|\ell^0 - \ell^\varepsilon\|_{V'}\big),\\ \|\Delta\boldsymbol{u}^0 - \Delta\boldsymbol{u}^\varepsilon\|_V &\leq c\Big(\|\mathscr{L}^0 - \mathscr{L}^\varepsilon\|_S + \|\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon\|_S \\ &\qquad + \|\boldsymbol{\chi}^0 - \boldsymbol{\chi}^\varepsilon\|_S + \varepsilon\,\|\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon\|_S\Big).\end{aligned}$$

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

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

Proof. Testing (5.18a) for $\varepsilon > 0$ and $\varepsilon = 0$, with $\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon$ and taking differences yields

$$\begin{aligned}\underline{\alpha}\,\|\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon\|_S^2 \leq{}& \|\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon\|_S\,\|\mathscr{L}^0 - \mathscr{L}^\varepsilon\|_S \\ &- \langle\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon,\, B^\star\Delta\boldsymbol{u}^0 - B^\star\Delta\boldsymbol{u}^\varepsilon\rangle \\ &- \langle\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon,\, -\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^0(\boldsymbol{\sigma}^0 + \boldsymbol{\chi}^0) + \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon)\rangle,\end{aligned}$$

where $\underline{\alpha} > 0$ is the coercivity constant of $\mathbb{C}^{-1}$. Using (5.18b) we obtain

$$\begin{aligned}\underline{\alpha}\,\|\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon\|_S^2 \leq{}& \|\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon\|_S\,\|\mathscr{L}^0 - \mathscr{L}^\varepsilon\|_S + \|\Delta\boldsymbol{u}^0 - \Delta\boldsymbol{u}^\varepsilon\|_V\,\|\ell^0 - \ell^\varepsilon\|_{V'} \\ &- \langle\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon,\, -\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^0(\boldsymbol{\sigma}^0 + \boldsymbol{\chi}^0) + \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon)\rangle.\end{aligned}$$

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

$$\begin{aligned}&- \langle\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon,\, -\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^0(\boldsymbol{\sigma}^0 + \boldsymbol{\chi}^0) + \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon)\rangle \\ &\quad = - \langle\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon,\, -\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^0(\boldsymbol{\sigma}^0 + \boldsymbol{\chi}^0) + \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^0(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^0)\rangle \\ &\qquad\;\; - \langle\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon,\, -\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^0(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^0) + \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon)\rangle \\ &\quad \leq - \langle\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon,\, -\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^0(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^0) + \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon)\rangle.\end{aligned}$$

Using Corollary 5.10 yields

$$\big\| - \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^0(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^0) + \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon)\big\|_S \leq L\,\|\boldsymbol{\chi}^0 - \boldsymbol{\chi}^\varepsilon\|_S + \varepsilon\,\|\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon\|_S.$$

This implies

$$\begin{aligned}\underline{\alpha}\,\|\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon\|_S^2 \leq{}& \|\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon\|_S\,\big(\|\mathscr{L}^0 - \mathscr{L}^\varepsilon\|_S + L\,\|\boldsymbol{\chi}^0 - \boldsymbol{\chi}^\varepsilon\|_S + \varepsilon\,\|\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon\|_S\big) \\ &+ \|\Delta\boldsymbol{u}^0 - \Delta\boldsymbol{u}^\varepsilon\|_V\,\|\ell^0 - \ell^\varepsilon\|_{V'}.\end{aligned}$$

Young's inequality completes the estimate of $\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon$.

It remains to verify the estimate of $\Delta\boldsymbol{u}^0 - \Delta\boldsymbol{u}^\varepsilon$. Testing (5.18a) for $\varepsilon > 0$ and $\varepsilon = 0$, with $\boldsymbol{\tau} \in S$, $\|\boldsymbol{\tau}\|_S \leq 1$, and taking differences yields

$$\begin{aligned}\langle B\boldsymbol{\tau},\, \Delta\boldsymbol{u}^0 - \Delta\boldsymbol{u}^\varepsilon\rangle \leq{}& c\Big(\|\mathscr{L}^0 - \mathscr{L}^\varepsilon\|_S + \|\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon\|_S \\ &+ \|-\mathbb{H}^{-1}\Delta\boldsymbol{\chi}^0(\boldsymbol{\sigma}^0 + \boldsymbol{\chi}^0) + \mathbb{H}^{-1}\Delta\boldsymbol{\chi}^\varepsilon(\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon)\|_S\Big).\end{aligned}$$

Invoking Corollary 5.10 and the inf-sup condition (1.13) implies

$$\|\Delta \boldsymbol{u}^0 - \Delta \boldsymbol{u}^\varepsilon\|_V \le c \Big(\|\mathscr{L}^0 - \mathscr{L}^\varepsilon\|_S + \|\boldsymbol{\sigma}^0 - \boldsymbol{\sigma}^\varepsilon\|_S \\ + \|\boldsymbol{\chi}^0 - \boldsymbol{\chi}^\varepsilon\|_S + \varepsilon \, \|\boldsymbol{\sigma}^\varepsilon + \boldsymbol{\chi}^\varepsilon\|_S \Big).$$

Easily this result carries over to the entire problem (5.4) for any fixed number N of time steps.

Corollary 5.12. Let $\ell^\varepsilon \in (V')^N$ be given, such that $\ell^\varepsilon \to \ell^0 \in (V')^N$ as $\varepsilon \searrow 0$. Denote by $(\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon)$, $(\boldsymbol{\Sigma}^0, \boldsymbol{u}^0)$ the solutions of (5.4), i.e.

$$(\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon) = \mathcal{G}^\varepsilon(\ell^\varepsilon) \quad \text{and} \quad (\boldsymbol{\Sigma}^0, \boldsymbol{u}^0) = \mathcal{G}^0(\ell^0).$$

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

We mention that an equivalent reformulation of (5.4) is stated in Section 6.3. This system contains a regularized counterpart λ^ε of the plastic multiplier λ. In Corollary 6.7 the convergence of λ^ε as $\varepsilon \searrow 0$ is shown.

6 C-Stationarity for the time-discrete optimization problem

The aim of this section is to derive an optimality system of C-stationary type for the time-discrete optimal control problem

$$\left.\begin{aligned} \text{Minimize} \quad & F(\boldsymbol{u}^\tau, \boldsymbol{g}^\tau) = \psi(\boldsymbol{u}^\tau) + \frac{\nu}{2}\|\boldsymbol{g}^\tau\|^2_{H^1(0,T;U)} \\ \text{such that} \quad & (\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau) = \mathcal{G}^\tau(E\boldsymbol{g}^\tau) \\ \text{and} \quad & \boldsymbol{g}^\tau \in U^\tau_{\text{ad}}, \end{aligned}\right\} \qquad (\mathbf{P}^\tau)$$

see Section 3.4. Note that we identify $(\boldsymbol{u}^\tau, \boldsymbol{g}^\tau) \in (V \times U)^N$ with its linear interpolant $(\boldsymbol{u}^\tau, \boldsymbol{g}^\tau) \in H^1(0,T; V \times U)$, see (3.1). The admissible set U_{ad} is a convex closed subset of $H^1(0,T;U)$. Here, $\|\cdot\|_{H^1(0,T;U)}$ can be any norm such that $H^1(0,T;U)$ is a Hilbert space, see (1.3). The assumptions on the objective $\psi : H^1(0,T;V) \to \mathbb{R}$ and on the set of admissible controls U_{ad} are fixed in Assumption 3.6. The existence of a solution of $(\mathbf{P}^\tau)$ is shown in Lemma 3.7.

In Section 6.1 we state a regularization $(\mathbf{P}^\varepsilon)$ of $(\mathbf{P}^\tau)$. This regularization is based upon the regularization of the forward problem derived in Section 5. Since the solution operator of this regularized forward problem is Fréchet differentiable, optimality conditions are obtained in a straightforward way. In Section 6.2 we show that not only global solutions but *all* local solutions of $(\mathbf{P}^\tau)$ can be approximated by solutions to a slightly modified version of $(\mathbf{P}^\varepsilon)$, see Corollary 6.6. Section 6.3 is devoted to give alternative formulations of the forward and the adjoint system. Finally, in Section 6.4 we prove the C-stationarity result.

As in Section 5, we consider a fixed number N of time steps of length $\tau = T/N$. For quantities that correspond to the regularized, time-discrete problem, we use a superscript ε (and do not mention explicitly the dependence on τ), whereas for quantities that correspond to the unregularized, time-discrete problem, we use a superscript τ.

At first, we substantiate the assumptions on the objective ψ. Throughout the section we assume

Assumption 6.1 (Differentiability of the objective w.r.t. $\boldsymbol{u}$)**.** For fixed N and $\tau = T/N$ the objective ψ has the form

$$\psi(\boldsymbol{u}^\varepsilon) = \psi^\tau(\boldsymbol{u}_1^\varepsilon, \ldots, \boldsymbol{u}_N^\varepsilon) \quad \text{for all } \boldsymbol{u}^\varepsilon \in V^N.$$

The function $\psi^\tau : V^N \to \mathbb{R}$ is continuously Fréchet differentiable. We denote the partial derivatives w.r.t. $\boldsymbol{u}_i^\varepsilon$ by $\psi_i^\tau(\boldsymbol{u}_1^\varepsilon, \ldots, \boldsymbol{u}_N^\varepsilon) = \psi_i^\tau(\boldsymbol{u}^\varepsilon) \in V'$, for all $i = 1, \ldots, N$.

Proceeding formally, we expect the following system of C-stationarity to hold for local optima of $(\mathbf{P}^\tau)$, cf. Scheel and Scholtes [2000].

$$A(\boldsymbol{\Sigma}_i^\tau - \boldsymbol{\Sigma}_{i-1}^\tau) + B^\star(\boldsymbol{u}_i^\tau - \boldsymbol{u}_{i-1}^\tau) + \tau\, \lambda_i^\tau\, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}_i^\tau = \mathbf{0}, \tag{6.1a}$$

$$B(\boldsymbol{\Sigma}_i^\tau - \boldsymbol{\Sigma}_{i-1}^\tau) = E(\boldsymbol{g}_i^\tau - \boldsymbol{g}_{i-1}^\tau), \tag{6.1b}$$

$$0 \le \lambda_i^\tau \quad \perp \quad \phi(\boldsymbol{\Sigma}_i^\tau) \le 0, \tag{6.1c}$$

$$\begin{aligned} A(\boldsymbol{\Upsilon}_i^\tau - \boldsymbol{\Upsilon}_{i+1}^\tau) &+ B^\star(\boldsymbol{w}_i^\tau - \boldsymbol{w}_{i+1}^\tau) \\ &+ \tau\, \lambda_i^\tau\, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Upsilon}_i^\tau + \tau\, \theta_i^\tau\, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}_i^\tau = \mathbf{0}, \end{aligned} \tag{6.2a}$$

$$B(\boldsymbol{\Upsilon}_i^\tau - \boldsymbol{\Upsilon}_{i+1}^\tau) = \psi_i^\tau(\boldsymbol{u}^\tau), \tag{6.2b}$$

$$\begin{aligned} \sum_{i=1}^N \big\langle E^\star \boldsymbol{w}_i^\tau, \tilde{\boldsymbol{g}}_i^\tau - \tilde{\boldsymbol{g}}_{i-1}^\tau - (\boldsymbol{g}_i^\tau - \boldsymbol{g}_{i-1}^\tau)\big\rangle_{L^2(\Gamma_N;\mathbb{R}^d)} & \\ + \langle \nu\, \boldsymbol{g}^\tau, \tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\tau \rangle_{H^1(0,T;U)} &\ge 0, \end{aligned} \tag{6.3}$$

$$\mathcal{D}\boldsymbol{\Sigma}_i^\tau : \mathcal{D}\boldsymbol{\Upsilon}_i^\tau - \mu_i^\tau = 0, \tag{6.4a}$$

$$\mu_i^\tau\, \lambda_i^\tau = 0, \tag{6.4b}$$

$$\theta_i^\tau\, \phi(\boldsymbol{\Sigma}_i^\tau) = 0, \tag{6.4c}$$

$$\theta_i^\tau\, \mu_i^\tau \ge 0, \tag{6.4d}$$

for $i = 1, \ldots, N$ and for all $\tilde{\boldsymbol{g}}^\tau \in U_{\text{ad}}^\tau$. Here, $(\boldsymbol{\Sigma}_0^\tau, \boldsymbol{u}_0^\tau) = (\mathbf{0}, \mathbf{0})$ and $(\boldsymbol{\Upsilon}_{N+1}^\tau, \boldsymbol{w}_{N+1}^\tau) = (\mathbf{0}, \mathbf{0})$.

Here, (6.1) is the forward system and (6.2) is the adjoint system. The variational inequality (6.3) is the result of passing to the limit in the gradient equation (6.7). The pointwise complementarity conditions on the multipliers (6.4) complete the system of C-stationary type.

To prove the necessity of the system (6.1)–(6.4), we have to perform several steps:

(1) In Section 6.1, we define the regularized time-discrete optimal control problem $(\mathbf{P}^\varepsilon)$ and derive optimality conditions, see Theorem 6.3.
(2) We verify that all local solutions $\boldsymbol{g}^\tau$ of $(\mathbf{P}^\tau)$ can be approximated by local solutions of a modified problem $(\mathbf{P}^\varepsilon_{\boldsymbol{g}^\tau})$, see Section 6.2.
(3) In Section 6.3 we derive alternative formulations of the forward and adjoint systems, which involve the regularized counterparts of the plastic multiplier λ and of the multipliers θ and μ.
(4) Finally, in Section 6.4 we check that the system (6.1)–(6.4) is fulfilled. This is achieved by passing to the limit $\varepsilon \searrow 0$ in the optimality system (6.7) of the regularized time-discrete problem.

6.1. Regularized upper level problem. In this section we consider a regularization of $(\mathbf{P}^\tau)$. We replace the time-discrete solution map $\mathcal{G}^\tau$ of (3.4) by its regularization $\mathcal{G}^\varepsilon$, see (5.4).

$$\left.\begin{aligned} \text{Minimize} \quad & F(\boldsymbol{u}^\varepsilon, \boldsymbol{g}^\varepsilon) = \psi(\boldsymbol{u}^\varepsilon) + \frac{\nu}{2} \|\boldsymbol{g}^\varepsilon\|^2_{H^1(0,T;U)} \\ \text{such that} \quad & (\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon) = \mathcal{G}^\varepsilon(E\boldsymbol{g}^\varepsilon), \\ \text{and} \quad & \boldsymbol{g}^\varepsilon \in U^\tau_{\text{ad}}. \end{aligned}\right\} \tag{$\mathbf{P}^\varepsilon$}$$

Due to the Lipschitz continuity of $\mathcal{G}^\varepsilon$, see Proposition 5.8, we can follow the proof of Lemma 3.7 and obtain

Lemma 6.2. There exists a global minimizer of ($\mathbf{P}^\varepsilon$).

By virtue of 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 $(\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon) = \mathcal{G}^\varepsilon(E\boldsymbol{g}^\varepsilon)$. In Section 5.3 we have proven the Fréchet differentiability of $\mathcal{G}^\varepsilon$. By Assumption 6.1 we conclude the differentiability of ψ. Hence, the reduced objective f^ε is differentiable. Since the admissible set U^τ_{ad} 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

$$\frac{\partial \psi}{\partial \boldsymbol{u}}(\boldsymbol{u}^\varepsilon)\frac{\partial \mathcal{G}^{\varepsilon,\boldsymbol{u}}}{\partial \ell}(E\boldsymbol{g}^\varepsilon)E(\tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\varepsilon) + \nu\, \langle \boldsymbol{g}^\varepsilon, \tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\varepsilon\rangle_{H^1(0,T;U)} \ge 0 \quad \text{for all } \tilde{\boldsymbol{g}}^\tau \in U^\tau_{\text{ad}}, \tag{6.5}$$

where the first addend is the derivative of $\psi(\mathcal{G}^{\varepsilon,\boldsymbol{u}}(E\,\cdot))$ evaluated by the chain rule. The second addend is the derivative of $\nu/2\,\|\cdot\|^2_{H^1(0,T;U)}$. By Assumption 6.1 we find

$$\frac{\partial \psi}{\partial \boldsymbol{u}}(\boldsymbol{u}^\varepsilon)\frac{\partial \mathcal{G}^{\varepsilon,\boldsymbol{u}}}{\partial \ell}(E\boldsymbol{g}^\varepsilon)E(\tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\varepsilon) = \sum_{i=1}^N \langle \psi_i^\tau(\boldsymbol{u}^\varepsilon),\, \delta\boldsymbol{u}_i^\varepsilon\rangle_{V',V},$$

where

$$(\delta\boldsymbol{u}_1^\varepsilon, \ldots, \delta\boldsymbol{u}_N^\varepsilon) = \frac{\partial \mathcal{G}^{\varepsilon,\boldsymbol{u}}}{\partial \ell}(E\boldsymbol{g}^\varepsilon)E(\tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\varepsilon)$$

is the solution of the linearized state equation (5.16) with right hand side $\delta\boldsymbol{g}^\varepsilon = \tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\varepsilon$. We compute the adjoint of (5.16) 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

$$\mathbb{C}^{-1}(\boldsymbol{v}_i^\varepsilon - \boldsymbol{v}_{i+1}^\varepsilon) + B^\star(\boldsymbol{w}_i^\varepsilon - \boldsymbol{w}_{i+1}^\varepsilon) + J_i^\varepsilon(\boldsymbol{v}_i^\varepsilon + \boldsymbol{\zeta}_{i+1}^\varepsilon) = \mathbf{0}, \tag{6.6a}$$
$$\mathbb{H}^{-1}(\boldsymbol{\zeta}_i^\varepsilon - \boldsymbol{\zeta}_{i+1}^\varepsilon) + J_i^\varepsilon(\boldsymbol{v}_i^\varepsilon + \boldsymbol{\zeta}_{i+1}^\varepsilon) = \mathbf{0}, \tag{6.6b}$$
$$B(\boldsymbol{v}_i^\varepsilon - \boldsymbol{v}_{i+1}^\varepsilon) = \psi_i^\tau(\boldsymbol{u}^\varepsilon), \tag{6.6c}$$

for $i = N, \ldots, 1$, where $(\boldsymbol{v}_{N+1}^\varepsilon, \boldsymbol{\zeta}_{N+1}^\varepsilon, \boldsymbol{w}_{N+1}^\varepsilon) = \mathbf{0}$ and J_i^ε is given by (5.17). Testing

(5.16a) with $\boldsymbol{v}_i^\varepsilon$, (5.16b) with $\boldsymbol{\zeta}_{i+1}^\varepsilon$, (5.16c) with $\boldsymbol{w}_i^\varepsilon$,
(6.6a) with $\delta\boldsymbol{\sigma}_i^\varepsilon$, (6.6b) with $\delta\boldsymbol{\chi}_{i-1}^\varepsilon$, (6.6c) 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 (6.5) becomes

$$\sum_{i=1}^N \big\langle E^\star\boldsymbol{w}_i^\varepsilon,\, \tilde{\boldsymbol{g}}_i^\tau - \tilde{\boldsymbol{g}}_{i-1}^\tau - (\boldsymbol{g}_i^\varepsilon - \boldsymbol{g}_{i-1}^\varepsilon)\big\rangle_{L^2(\Gamma_N;\mathbb{R}^d)} + \langle \nu\,\boldsymbol{g}^\varepsilon,\, \tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\varepsilon\rangle_{H^1(0,T;U)} \ge 0 \quad \text{for all } \tilde{\boldsymbol{g}}^\tau \in U^\tau_{\text{ad}}. \tag{6.7}$$

We obtain

Theorem 6.3. Let $(\boldsymbol{u}^\varepsilon, \boldsymbol{g}^\varepsilon) \in V^N \times U^\tau_{\text{ad}}$ be a local solution of ($\mathbf{P}^\varepsilon$). Then (6.7) is satisfied, where the adjoint state $\boldsymbol{w}^\varepsilon$ is defined as the unique solution of (6.6).

6.2. 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 ($\mathbf{P}^\varepsilon$). We will argue as in Section 3.4. Let us recall that by Assumption 3.6

(1) the admissible set U_{ad} is nonempty, convex and closed in $H^1(0, T; U)$,
(2) every admissible control $\boldsymbol{g} \in U_{\text{ad}}$ can be approximated by time-discrete controls $\boldsymbol{g}^\tau \in U^\tau_{\text{ad}}$, and
(3) the objective $F : H^1(0, T; V \times U)$ is weakly lower semicontinuous.

Theorem 6.4. Suppose Assumption 3.6 is fulfilled. Let $\{\varepsilon\}$ be a sequence tending to 0 and let $\boldsymbol{g}^\varepsilon$ denote a global solution to ($\mathbf{P}^\varepsilon$).

(1) There exists an accumulation point $\boldsymbol{g}^\tau$ of $\{\boldsymbol{g}^\varepsilon\}$.
(2) Every weak accumulation point of $\{\boldsymbol{g}^\varepsilon\}$ is a strong accumulation point and a global solution of ($\mathbf{P}^\tau$).

Proof. By Assumption 3.6, U^τ_{ad} is nonempty. Hence, there exists some $\tilde{\boldsymbol{g}}^\tau \in U^\tau_{\text{ad}}$. Hence, by Proposition 5.7 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 $F(\boldsymbol{u}^\varepsilon, \boldsymbol{g}^\varepsilon) \le F(\tilde{\boldsymbol{u}}^\varepsilon, \tilde{\boldsymbol{g}}^\tau)$. This implies the boundedness of $\{\boldsymbol{g}^\varepsilon\}$ in $H^1(0, T; U)$. Hence, there exists a weakly convergent subsequence. Therefore, assertion (1) follows by assertion (2).

To prove assertion (2), suppose that $\{\boldsymbol{g}^\varepsilon\}$ 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 (5.4) and by $(\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau) = \mathcal{G}^\tau(E\boldsymbol{g})$ the solution to (3.4). Since E embeds U^N compactly into $(V')^N$, Corollary 5.12 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 Corollary 5.12 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.

Theorem 6.5. Suppose Assumption 3.6 is fulfilled. 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 $\{\varepsilon\}$ tending to 0, there is a sequence $\{\boldsymbol{g}^\varepsilon\}$ of local solutions to ($\mathbf{P}^\varepsilon$) such that $\boldsymbol{g}^\varepsilon \to \boldsymbol{g}^\tau$ strongly in $H^1(0, T; U)$.

Proof. This result can be proven using the idea of the proof of Lemma 3.9.

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 Casas and Tröltzsch [2002], Barbu [1981],

$$\left.\begin{aligned} &\text{Minimize} && F_{\boldsymbol{g}^\tau}(\boldsymbol{u}^\tau, \tilde{\boldsymbol{g}}^\tau) = F(\boldsymbol{u}^\tau, \tilde{\boldsymbol{g}}^\tau) + \frac{1}{2}\|\tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\tau\|^2_{H^1(0,T;U)} \\ &\text{such that} && (\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau) = \mathcal{G}^\tau(E\tilde{\boldsymbol{g}}^\tau), \\ &\text{and} && \tilde{\boldsymbol{g}}^\tau \in U^\tau_{\text{ad}}. \end{aligned}\right\} \qquad (\mathbf{P}^\tau_{\boldsymbol{g}^\tau})$$

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

Corollary 6.6. Suppose Assumption 3.6 is fulfilled. Let $\boldsymbol{g}^\tau$ be a local optimum of $(\mathbf{P}^\tau)$ w.r.t. the topology of U^N. Then, for every sequence $\{\varepsilon\}$ tending to 0, there is a sequence $\{\boldsymbol{g}^\varepsilon\}$ of local solutions to the modified problems $(\mathbf{P}^\varepsilon_{\boldsymbol{g}^\tau})$, such that $\boldsymbol{g}^\varepsilon \to \boldsymbol{g}^\tau$ strongly in $H^1(0,T;U)$.

Proof. Since the additional term $\|\tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\tau\|^2_{H^1(0,T;U)}$ is weakly lower semicontinuous, this result follows analogously to Theorem 6.5.

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 $(\mathbf{P}^\varepsilon_{\boldsymbol{g}^\tau})$

$$\sum_{i=1}^N \big\langle E^\star \boldsymbol{w}^\varepsilon_i, \tilde{\boldsymbol{g}}^\tau_i - \tilde{\boldsymbol{g}}^\tau_{i-1} - (\boldsymbol{g}^\varepsilon_i - \boldsymbol{g}^\varepsilon_{i-1})\big\rangle_{L^2(\Gamma_N;\mathbb{R}^d)} + \langle \nu\, \boldsymbol{g}^\varepsilon + \boldsymbol{g}^\varepsilon - \boldsymbol{g}^\tau, \tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\varepsilon\rangle_{H^1(0,T;U)} \geq 0 \tag{6.8}$$

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 (6.6).

6.3. Alternative formulation of forward and adjoint systems.

The aim of this section is the introduction of the regularized counterparts of the plastic multiplier λ^τ (in the forward problem) and of the multipliers θ^τ and μ^τ (in the adjoint system). Up to now, these quantities did not appear in the regularized forward problem (5.4) nor in the adjoint system (6.6) nor in the gradient equations (6.5) and (6.7). Therefore, this reformulation is essential in proving that the optimality system (6.1)–(6.4) is satisfied in the limit. The reformulations of the forward and the adjoint system containing the multipliers λ^ε and θ^ε can be found in (6.13) and (6.18), respectively.

Our starting point is the regularized forward system (5.4). Let us recall that the deviatoric part of a matrix $\boldsymbol{\sigma} \in \mathbb{S}$ is given by $\boldsymbol{\sigma}^D = \boldsymbol{\sigma} - \operatorname{trace}(\boldsymbol{\sigma})\, \boldsymbol{I}/d$, see (1.6). By (5.4c) and the initial datum $\boldsymbol{\chi}^\varepsilon_0 = \boldsymbol{0}$, we infer $\boldsymbol{\chi}^\varepsilon_i = (\boldsymbol{\chi}^\varepsilon_i)^D$ for all $i = 0, \ldots, N$. In other words, $\boldsymbol{\chi}^\varepsilon_i$ is trace free. Hence, we can omit the superscript D on the variable $\boldsymbol{\chi}^\varepsilon$. For convenience, we define

$$\tilde{\mathcal{D}}\boldsymbol{\Sigma}^\varepsilon_i := (\boldsymbol{\sigma}^\varepsilon_i)^D + \boldsymbol{\chi}^\varepsilon_{i-1}. \tag{6.9}$$

In contrast to $\mathcal{D}\boldsymbol{\Sigma}^\varepsilon_i = (\boldsymbol{\sigma}^\varepsilon_i)^D + \boldsymbol{\chi}^\varepsilon_i$, see (1.8), $\boldsymbol{\chi}^\varepsilon$ is taken from the previous time step $i-1$.

Moreover, we introduce an abbreviation for two terms appearing in $\Delta\boldsymbol{\chi}^\varepsilon(\tilde{\mathcal{D}}\boldsymbol{\Sigma}^\varepsilon_i)$ and its derivative, cf. (5.3) and (5.8),

$$\alpha^\varepsilon_i := \max{}^\varepsilon\left(1 - \frac{\tilde{\sigma}_0}{|\tilde{\mathcal{D}}\boldsymbol{\Sigma}^\varepsilon_i|}\right), \quad \text{and} \quad \beta^\varepsilon_i := (\max{}^\varepsilon)'\left(1 - \frac{\tilde{\sigma}_0}{|\tilde{\mathcal{D}}\boldsymbol{\Sigma}^\varepsilon_i|}\right)\frac{\tilde{\sigma}_0}{|\tilde{\mathcal{D}}\boldsymbol{\Sigma}^\varepsilon_i|}. \tag{6.10}$$

By Assumption 5.2 we infer $\alpha^\varepsilon_i \in [0,1)$ and $\beta^\varepsilon_i \in [0, 1+\varepsilon)$. By adding $(\boldsymbol{\sigma}^\varepsilon_i)^D$ on both sides of (5.4c) we obtain

$$\mathcal{D}\boldsymbol{\Sigma}^\varepsilon_i = \tilde{\mathcal{D}}\boldsymbol{\Sigma}^\varepsilon_i - \alpha^\varepsilon_i\, \tilde{\mathcal{D}}\boldsymbol{\Sigma}^\varepsilon_i = (1-\alpha^\varepsilon_i)\, \tilde{\mathcal{D}}\boldsymbol{\Sigma}^\varepsilon_i. \tag{6.11}$$

Hence, the definition of $\Delta\boldsymbol{\chi}^\varepsilon$, see (5.3), 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 (5.4a) and (5.4c) with (3.8a) gives rise to the definition of λ_i^ε by

$$\lambda_i^\varepsilon := \tau^{-1}\,k_1^{-1}\frac{\alpha_i^\varepsilon}{1-\alpha_i^\varepsilon}. \tag{6.12}$$

The $L^2(\Omega)$-regularity of λ_i^ε is shown in Corollary 6.7. Using the definition of λ_i^ε, the forward system (5.4) becomes

$$A(\boldsymbol{\Sigma}_i^\varepsilon - \boldsymbol{\Sigma}_{i-1}^\varepsilon) + B^\star(\boldsymbol{u}_i^\varepsilon - \boldsymbol{u}_{i-1}^\varepsilon) + \tau\,\lambda_i^\varepsilon\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon = \boldsymbol{0}, \tag{6.13a}$$

$$B\boldsymbol{\Sigma}_i^\varepsilon = \ell_i^\varepsilon. \tag{6.13b}$$

Let us turn to the adjoint equation (6.6). By (5.8), (5.17) and (6.10), the mapping J_i^ε can be expressed in terms of α_i^ε and β_i^ε. We find

$$J_i^\varepsilon\,\boldsymbol{\tau} = k_1^{-1}\Big(\alpha_i^\varepsilon\,\boldsymbol{\tau}^D + \beta_i^\varepsilon\,\frac{\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon : \boldsymbol{\tau}^D}{|\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon|^2}\,\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon\Big). \tag{6.14}$$

Using (6.6b) and $\boldsymbol{\zeta}_{N+1}^\varepsilon = \boldsymbol{0}$ we infer $\boldsymbol{\zeta}_i^\varepsilon = (\boldsymbol{\zeta}_i^\varepsilon)^D$ for all $i = 1,\ldots,N$. Similar to (6.9), we define

$$\tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon := (\boldsymbol{v}_i^\varepsilon)^D + \boldsymbol{\zeta}_{i+1}^\varepsilon. \tag{6.15}$$

Now we are going to manipulate $J_i^\varepsilon\,\tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon$ in order that the adjoint equation (6.6) resembles the counterpart in the expected C-stationarity system (6.2). We obtain from (6.6b) and (6.14)

$$\mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon - \tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon = \boldsymbol{\zeta}_i^\varepsilon - \boldsymbol{\zeta}_{i+1}^\varepsilon = -\Big(\alpha_i^\varepsilon\,\tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon + \beta_i^\varepsilon\,\frac{\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon : \tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon}{|\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon|^2}\,\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon\Big). \tag{6.16}$$

Dividing by $1-\alpha_i^\varepsilon > 0$ yields

$$\tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon = \frac{1}{1-\alpha_i^\varepsilon}\Big(\mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon + \beta_i^\varepsilon\,\frac{\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon : \tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon}{|\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon|^2}\,\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon\Big).$$

Using (6.14) we proceed by

$$\begin{aligned} J_i^\varepsilon\,\tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon &= k_1^{-1}\,\frac{\alpha_i^\varepsilon}{1-\alpha_i^\varepsilon}\Big(\mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon + \beta_i^\varepsilon\,\frac{\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon : \tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon}{|\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon|^2}\,\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon\Big) + k_1^{-1}\,\beta_i^\varepsilon\,\frac{\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon : \tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon}{|\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon|^2}\,\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon \\ &= \tau\,\lambda_i^\varepsilon\,\mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon + k_1^{-1}\,\frac{\beta_i^\varepsilon}{1-\alpha_i^\varepsilon}\,\frac{\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon : \tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon}{|\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon|^2}\,\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon. \end{aligned}$$

This gives rise to the definition

$$\theta_i^\varepsilon := k_1^{-1}\,\tau^{-1}\,\frac{\beta_i^\varepsilon}{1-\alpha_i^\varepsilon}\,\frac{\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon : \tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon}{|\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon|^2}. \tag{6.17}$$

The $L^2(\Omega)$-regularity of θ_i^ε is shown in Lemma 6.12. The definitions of θ_i^ε implies

$$J_i^\varepsilon\,\tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon = \tau\,\lambda_i^\varepsilon\,\mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon + \tau\,\theta_i^\varepsilon\,\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon.$$

Using the definition of λ_i^ε and θ_i^ε the adjoint system (6.6) becomes

$$A(\boldsymbol{\Upsilon}_i^\varepsilon - \boldsymbol{\Upsilon}_{i+1}^\varepsilon) + \tau\,\lambda_i^\varepsilon\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon + \tau\,\theta_i^\varepsilon\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon + B^\star(\boldsymbol{w}_i^\varepsilon - \boldsymbol{w}_{i+1}^\varepsilon) = \boldsymbol{0}, \tag{6.18a}$$

$$B(\boldsymbol{\Upsilon}_i^\varepsilon - \boldsymbol{\Upsilon}_{i+1}^\varepsilon) = \psi_i^\tau(\boldsymbol{u}^\varepsilon). \tag{6.18b}$$

It remains to define the multiplier μ^ε. According to (6.4a) we define

$$\mu_i^\varepsilon := \mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon : \mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon \in L^2(\Omega). \tag{6.19}$$

Testing (6.16) with $\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon$ implies

$$\mu_i^\varepsilon = \mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon : \mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon = (1-\alpha_i^\varepsilon-\beta_i^\varepsilon)\,\tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon : \mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon. \tag{6.20}$$

This equation is the starting point to estimate the multiplier μ_i^ε, see Lemma 6.9.

6.4. Convergence of the regularization. In this section we verify the time-discrete optimality system (6.1)–(6.4). Let $\boldsymbol{g}^\tau$ be a local solution of $(\mathbf{P}^\tau)$. By Corollary 6.6 there is a sequence $\boldsymbol{g}^\varepsilon$ of local optimal solutions to $(\mathbf{P}^\varepsilon_{\boldsymbol{g}^\tau})$ converging strongly to $\boldsymbol{g}^\tau \in U^N$. Using Corollary 5.12 we find $(\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon) \to (\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau) \in (S^2 \times V)^N$ as $\varepsilon \searrow 0$ for the corresponding stresses and displacements. By $(\boldsymbol{\Upsilon}^\varepsilon, \boldsymbol{w}^\varepsilon)$ we denote the adjoint stresses and displacements satisfying the system (6.6). Moreover, we use the multipliers λ^ε, θ^ε and μ^ε defined in (6.12), (6.17) and (6.20), respectively.

Let us sketch the outline of this section.

(1) We show the convergence of the plastic multipliers λ^ε in Corollary 6.7.
(2) We verify that the sign condition (6.4d) holds for the regularized multipliers, see Lemma 6.10.
(3) In Lemmas 6.11–6.12 we show the boundedness of several variables. This allows for extracting a weakly convergent subsequence.
(4) The regularized counterparts of the complementarity conditions (6.4b) and (6.4c) are shown in Lemmas 6.13 and 6.14.
(5) Finally, we prove that the optimality system (6.1)–(6.4) holds for all local minimizers of the time-discrete optimal control problem $(\mathbf{P}^\tau)$.

In Assumption 5.2 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 (5.3) is smaller than $-\varepsilon$, larger than ε or in $(-\varepsilon, \varepsilon)$.

$$A_i^{\varepsilon,-} := \Big\{x \in \Omega : |\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon| \le \frac{\tilde\sigma_0}{1+\varepsilon}\Big\} = \Big\{x \in \Omega : 1 - \frac{\tilde\sigma_0}{|\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon|} \le -\varepsilon\Big\}, \tag{6.21a}$$

$$A_i^{\varepsilon,+} := \Big\{x \in \Omega : |\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon| \ge \frac{\tilde\sigma_0}{1-\varepsilon}\Big\} = \Big\{x \in \Omega : 1 - \frac{\tilde\sigma_0}{|\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon|} \ge \varepsilon\Big\}, \tag{6.21b}$$

$$A_i^{\varepsilon,0} := \Omega \setminus (A_i^{\varepsilon,-} \cup A_i^{\varepsilon,+}). \tag{6.21c}$$

Here and throughout we assume w.l.o.g. $\varepsilon < 1$.

The first result of this section concerns the convergence of the plastic multiplier λ_i^ε. It is a supplement to the convergence properties in Section 5.4. Since we employ the notation introduced in Section 6.3, this result is not included in Section 5.4.

Corollary 6.7. Let $\ell^\varepsilon \in (V')^N$ 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 (6.12). Then $\lambda^\varepsilon \in L^2(\Omega)^N$ and $\lambda^\varepsilon \to \lambda^\tau$ in $L^2(\Omega)^N$.

Proof. Step (1): We infer from (3.8a) and (6.13a)

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

for all $i = 1, \ldots, N$. By Corollary 5.12 we obtain

$$\tau\,\lambda_i^\varepsilon\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon \to \tau\,\lambda_i^\tau\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_i^\tau \quad \text{in } S^2 \text{ for all } i = 1, \ldots, N. \tag{6.22}$$

Step (2): By using Assumption 5.2 and the definition of $A_i^{\varepsilon,-}$ in (6.21a) we obtain $\lambda_i^\varepsilon = 0$ on $A_i^{\varepsilon,-}$. Moreover, the definition of α_i^ε in (6.10) 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

$$\|\lambda_i^\varepsilon\, \mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon\|_S^2 = \int_{\Omega\setminus A_i^{\varepsilon,-}} |\lambda_i^\varepsilon|^2\, |\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon|^2 \,\mathrm{d}x = \int_{\Omega\setminus A_i^{\varepsilon,-}} |\lambda_i^\varepsilon|^2\, (1-\alpha_i^\varepsilon)^2\, |\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon|^2 \,\mathrm{d}x$$
$$\geq \frac{(1-\varepsilon)^2\, \tilde\sigma_0^2}{(1+\varepsilon)^2} \int_{A_i^{\varepsilon,0}} |\lambda_i^\varepsilon|^2 \,\mathrm{d}x + \int_{A_i^{\varepsilon,+}} |\lambda_i^\varepsilon|^2\, \tilde\sigma_0^2 \,\mathrm{d}x \geq \Big(\frac{1-\varepsilon}{1+\varepsilon}\Big)^2\, \tilde\sigma_0^2\, \|\lambda_i^\varepsilon\|_{L^2(\Omega)}^2.$$

Using (6.22) 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, (6.22) implies

$$\tilde\lambda_i = \lambda_i^\tau \text{ on } \{x\in\Omega : |\mathcal{D}\boldsymbol{\Sigma}_i^\tau| \neq 0\}. \tag{6.23}$$

Step (3): We have

$$\|\tilde\lambda_i\|_{L^2(\Omega)} \leq \liminf \|\lambda_i^\varepsilon\|_{L^2(\Omega)}$$
$$\leq \liminf \Big(\frac{1+\varepsilon}{1-\varepsilon}\Big)\frac{1}{\tilde\sigma_0}\|\lambda_i^\varepsilon\, \mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon\|_S = \frac{1}{\tilde\sigma_0}\|\lambda_i^\tau\, \mathcal{D}\boldsymbol{\Sigma}_i^\tau\|_S = \|\lambda_i^\tau\|_{L^2(\Omega)}, \tag{6.24}$$

in particular $\|\tilde\lambda_i\|_{L^2(\Omega)} \leq \|\lambda_i^\tau\|_{L^2(\Omega)}$. By (6.23) 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, (6.22) and (6.24) 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 $L^2(\Omega)$.

Let us briefly comment on our choice of regularization.

Remark 6.8. Due to the definition of the disjoint partition $\Omega = A_i^{\varepsilon,-} \cup A_i^{\varepsilon,0} \cup A_i^{\varepsilon,+}$, the equations (6.10), (6.11) and (6.12) 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 (3.8c), but they do not satisfy the complementarity condition in (3.8c). Thus our regularization approach can be viewed as a problem-tailored version of the relaxation strategy given in Steffensen and Ulbrich [2010]. Similarly to [Herzog et al., 2010, Section 2.2] we could also use a penalization approach for the time-discrete problem. However, using a relaxation strategy the proofs in this section become more simple compared to those in [Herzog et al., 2010, Section 3], see, e.g., the proofs of Lemmas 6.13 and 6.14 and the corresponding results [Herzog et al., 2010, Lemma 3.11–Proposition 3.13].

As a preparation for the proof of Theorem 6.15 we verify estimates for various quantities introduced in Section 6.3. We start by giving bounds on the term $1 - \alpha_i^\varepsilon - \beta_i^\varepsilon$ which appears in the definition of μ_i^ε, see (6.20).

Lemma 6.9. We have

$$1-\alpha_i^\varepsilon-\beta_i^\varepsilon \in \begin{cases} \{1\}, & \text{on } A_i^{\varepsilon,-}, \\ [0,1], & \text{on } A_i^{\varepsilon,0}, \\ \{0\}, & \text{on } A_i^{\varepsilon,+}, \end{cases} \tag{6.25}$$

for all $i \in \{1,\ldots,N\}$.

Proof. By the definition of α_i^ε and β_i^ε in (6.10), we infer immediately $\alpha_i^\varepsilon = \beta_i^\varepsilon = 0$ on $A_i^{\varepsilon,-}$. On $A_i^{\varepsilon,+}$ we have $\alpha_i^\varepsilon = 1-\frac{\tilde\sigma_0}{|\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon|}$ and $\beta_i^\varepsilon = \frac{\tilde\sigma_0}{|\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon|}$. This implies $1-\alpha_i^\varepsilon-\beta_i^\varepsilon = 0$.

It remains to check the assertion on $A_i^{\varepsilon,0}$. Let us define $\kappa_i^\varepsilon = 1-\tilde\sigma_0/|\tilde{\mathcal{D}}\boldsymbol{\Sigma}_i^\varepsilon|$. On $A_i^{\varepsilon,0}$ we have $\kappa_i^\varepsilon \in (-\varepsilon,\varepsilon)$. By definition of α_i^ε and β_i^ε in (6.10), we have

$$\alpha_i^\varepsilon = \max{}^\varepsilon(\kappa_i^\varepsilon) \quad \text{and} \quad \beta_i^\varepsilon = (\max{}^\varepsilon)'(\kappa_i^\varepsilon)\,(1-\kappa_i^\varepsilon).$$

Let us give a precise upper bound of β_i^ε. The fundamental theorem of calculus yields

$$\int_{\kappa_i^\varepsilon}^{\varepsilon} (\max{}^\varepsilon)'(x)\,\mathrm{d}x = \max{}^\varepsilon(\varepsilon) - \max{}^\varepsilon(\kappa_i^\varepsilon) = \varepsilon - \max{}^\varepsilon(\kappa_i^\varepsilon).$$

Since Assumption 5.2 implies that $(\max^\varepsilon)'$ is monotone increasing, we infer

$$(\varepsilon-\kappa_i^\varepsilon)\,(\max{}^\varepsilon)'(\kappa_i^\varepsilon) \le \varepsilon - \max{}^\varepsilon(\kappa_i^\varepsilon).$$

Hence,

$$\beta_i^\varepsilon = (\max{}^\varepsilon)'(\kappa_i^\varepsilon)\,(1-\kappa_i^\varepsilon) \le \frac{\varepsilon-\max^\varepsilon(\kappa_i^\varepsilon)}{\varepsilon-\kappa_i^\varepsilon}\,(1-\kappa_i^\varepsilon).$$

Now, we obtain

$$\begin{aligned} 1-\alpha_i^\varepsilon-\beta_i^\varepsilon &\ge 1-\max{}^\varepsilon(\kappa_i^\varepsilon) - \frac{\varepsilon-\max^\varepsilon(\kappa_i^\varepsilon)}{\varepsilon-\kappa_i^\varepsilon}\,(1-\kappa_i^\varepsilon) \\ &= \frac{\varepsilon-\kappa_i^\varepsilon-\max^\varepsilon(\kappa_i^\varepsilon)\,(\varepsilon-\kappa_i^\varepsilon)-(\varepsilon-\max^\varepsilon(\kappa_i^\varepsilon))\,(1-\kappa_i^\varepsilon)}{\varepsilon-\kappa_i^\varepsilon} \\ &= \frac{1-\varepsilon}{\varepsilon-\kappa_i^\varepsilon}(\max{}^\varepsilon(\kappa_i^\varepsilon)-\kappa_i^\varepsilon) \ge 0. \end{aligned}$$

By $\alpha_i^\varepsilon \ge 0$ and $\beta_i^\varepsilon \ge 0$ we infer $1-\alpha_i^\varepsilon-\beta_i^\varepsilon \in [0,1]$ on $A_i^{\varepsilon,0}$.

As a simple consequence we obtain the regularized counterpart of the sign condition (6.4d).

Lemma 6.10. For all $i \in \{1,\ldots,N\}$, the condition $\theta_i^\varepsilon\,\mu_i^\varepsilon \ge 0$ is satisfied almost everywhere in Ω.

Proof. By (6.20) we obtain

$$\theta_i^\varepsilon\,\mu_i^\varepsilon = \theta_i^\varepsilon\,(1-\alpha_i^\varepsilon-\beta_i^\varepsilon)\,\mathcal{D}\boldsymbol{\Sigma} : \tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon.$$

Now, Lemma 6.9 and the definition of θ_i^ε in (6.17) imply $\theta_i^\varepsilon\,\mu_i^\varepsilon \ge 0$.

Now we show the boundedness of the adjoint states $(\boldsymbol{\Upsilon}^\varepsilon, \boldsymbol{w}^\varepsilon)$.

Lemma 6.11. The adjoint states $(\boldsymbol{\Upsilon}^\varepsilon, \boldsymbol{w}^\varepsilon)$ satisfy

$$\max_{i=1,\ldots,N}\|\boldsymbol{\Upsilon}_i^\varepsilon\|_{S^2} + \max_{i=1,\ldots,N}\|\boldsymbol{w}_i^\varepsilon\|_V \le C\sum_{i=1}^N \|\psi_i^\tau(\boldsymbol{u}^\varepsilon)\|_{V'}, \tag{6.26}$$

where the constant C depends only on the operators A and B.

Proof. Let us define $\boldsymbol{T} = \boldsymbol{\Upsilon}_i^\varepsilon - \boldsymbol{\Sigma}_{B\boldsymbol{\Upsilon}_i^\varepsilon}$, for the definition of $\boldsymbol{\Sigma}_{(\cdot)}$ we refer to (2.4). This implies

$$\begin{aligned} B\boldsymbol{T} &= \mathbf{0}, && \text{by } B\boldsymbol{\Sigma}_{B\boldsymbol{\Upsilon}_i^\varepsilon} = B\boldsymbol{\Upsilon}_i^\varepsilon, \\ \lambda_i^\varepsilon\, \mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon : \mathcal{D}\boldsymbol{T} = \lambda_i^\varepsilon\, \mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon : \mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon &\geq 0, && \text{by } \lambda_i^\varepsilon \geq 0, \\ \theta_i^\varepsilon\, \mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon : \mathcal{D}\boldsymbol{T} = \theta_i^\varepsilon\, \mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon : \mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon = \theta_i^\varepsilon\, \mu_i^\varepsilon &\geq 0, && \text{by Lemma 6.10.} \end{aligned}$$

Testing (6.18a) with $\boldsymbol{T}$ yields

$$\langle \boldsymbol{\Upsilon}_i^\varepsilon - \boldsymbol{\Upsilon}_{i+1}^\varepsilon, \boldsymbol{T}\rangle_A \leq 0.$$

Here, $\langle\cdot,\cdot\rangle_A$ is the scalar product on S^2 induced by the operator A. Hence

$$\langle \boldsymbol{\Upsilon}_i^\varepsilon - \boldsymbol{\Upsilon}_{i+1}^\varepsilon, \boldsymbol{\Upsilon}_i^\varepsilon\rangle_A \leq \langle \boldsymbol{\Upsilon}_i^\varepsilon - \boldsymbol{\Upsilon}_{i+1}^\varepsilon, \boldsymbol{\Sigma}_{B\boldsymbol{\Upsilon}_i^\varepsilon}\rangle_A.$$

Using

$$\langle \boldsymbol{\Upsilon}_i^\varepsilon - \boldsymbol{\Upsilon}_{i+1}^\varepsilon, \boldsymbol{\Upsilon}_i^\varepsilon\rangle_A = \frac{1}{2}\big(\|\boldsymbol{\Upsilon}_i^\varepsilon\|_A^2 - \|\boldsymbol{\Upsilon}_{i+1}^\varepsilon\|_A^2 + \|\boldsymbol{\Upsilon}_i^\varepsilon - \boldsymbol{\Upsilon}_{i+1}^\varepsilon\|_A^2\big),$$

yields

$$\|\boldsymbol{\Upsilon}_i^\varepsilon\|_A^2 - \|\boldsymbol{\Upsilon}_{i+1}^\varepsilon\|_A^2 + \|\boldsymbol{\Upsilon}_i^\varepsilon - \boldsymbol{\Upsilon}_{i+1}^\varepsilon\|_A^2 \leq 2\,\langle \boldsymbol{\Upsilon}_i^\varepsilon - \boldsymbol{\Upsilon}_{i+1}^\varepsilon, \boldsymbol{\Sigma}_{B\boldsymbol{\Upsilon}_i^\varepsilon}\rangle_A.$$

Summing over $i \in \{k, k+1, \ldots, N\}$ and using $\boldsymbol{\Upsilon}_{N+1} = \mathbf{0}$ yields

$$\begin{aligned} \|\boldsymbol{\Upsilon}_k^\varepsilon\|_A^2 + \sum_{i=k}^N \|\boldsymbol{\Upsilon}_i^\varepsilon - \boldsymbol{\Upsilon}_{i+1}^\varepsilon\|_A^2 &\leq 2\sum_{i=k}^N \langle \boldsymbol{\Upsilon}_i^\varepsilon - \boldsymbol{\Upsilon}_{i+1}^\varepsilon, \boldsymbol{\Sigma}_{B\boldsymbol{\Upsilon}_i^\varepsilon}\rangle_A \\ &\leq 2\,\langle \boldsymbol{\Upsilon}_k^\varepsilon, \boldsymbol{\Sigma}_{B\boldsymbol{\Upsilon}_k^\varepsilon}\rangle_A - 2\sum_{i=k}^N \langle \boldsymbol{\Upsilon}_{i+1}^\varepsilon, \boldsymbol{\Sigma}_{B\boldsymbol{\Upsilon}_i^\varepsilon - B\boldsymbol{\Upsilon}_{i+1}^\varepsilon}\rangle_A \\ &\leq C \max_{i=k,\ldots,N} \|\boldsymbol{\Upsilon}_i^\varepsilon\|_A \Big(\|B\boldsymbol{\Upsilon}_k^\varepsilon\|_{V'} + \sum_{i=k}^N \|\psi_i^\tau(\boldsymbol{u}^\varepsilon)\|_{V'}\Big), \end{aligned}$$

where C depends only on A and B. By $B\boldsymbol{\Upsilon}_k^\varepsilon = \sum_{i=k}^N \psi_i^\tau(\boldsymbol{u}^\varepsilon)$ we obtain

$$\|\boldsymbol{\Upsilon}_k^\varepsilon\|_A^2 \leq C \max_{i=k,\ldots,N} \|\boldsymbol{\Upsilon}_i^\varepsilon\|_A \sum_{i=k}^N \|\psi_i^\tau(\boldsymbol{u}^\varepsilon)\|_{V'} \leq C \max_{i=1,\ldots,N} \|\boldsymbol{\Upsilon}_i^\varepsilon\|_A \sum_{i=1}^N \|\psi_i^\tau(\boldsymbol{u}^\varepsilon)\|_{V'}.$$

Taking the maximum over $k = 1, \ldots, N$ on the left hand side yields

$$\max_{i=1,\ldots,N} \|\boldsymbol{\Upsilon}_i^\varepsilon\|_A \leq C \sum_{i=1}^N \|\psi_i^\tau(\boldsymbol{u}^\varepsilon)\|_{V'},$$

where C depends only on A and B. The estimate for $\boldsymbol{w}_i^\varepsilon$ follows by taking the difference of (6.6a) and (6.6b) and using the inf-sup condition of $B^\star$, see (1.13).

For convenience we define the abbreviation $\boldsymbol{Q}_i^\varepsilon = -A\boldsymbol{\Upsilon}_i^\varepsilon - B^\star \boldsymbol{w}_i^\varepsilon$. The adjoint system (6.18a) becomes

$$\frac{1}{\tau}\,(\boldsymbol{Q}_i^\varepsilon - \boldsymbol{Q}_{i+1}^\varepsilon) = \lambda_i^\varepsilon\, \mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon + \theta_i^\varepsilon\, \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon. \tag{6.27}$$

Using Lemma 6.11 we obtain the boundedness of $\boldsymbol{Q}^\varepsilon$

$$\|\boldsymbol{Q}_i^\varepsilon\|_{S^2} \leq C \sum_{i=1}^N \|\psi_i^\tau(\boldsymbol{u}^\varepsilon)\|_{V'}. \tag{6.28}$$

As a consequence, we obtain an estimate of the bilinear terms in (6.27) and of the multiplier θ_i^ε in $L^2(\Omega)$.

Lemma 6.12. The estimates

$$\|\theta_i^\varepsilon\, \mathcal{D}^\star\mathcal{D}\mathbf{\Sigma}_i^\varepsilon\|_{S^2}^2 + \|\lambda_i^\varepsilon\, \mathcal{D}^\star\mathcal{D}\mathbf{\Upsilon}_i^\varepsilon\|_{S^2}^2 \le \frac{1}{\tau^2}\,\|\boldsymbol{Q}_i^\varepsilon - \boldsymbol{Q}_{i+1}^\varepsilon\|_{S^2}^2 \tag{6.29a}$$

$$\|\theta_i^\varepsilon\|_{L^2(\Omega)} \le \frac{1+\varepsilon}{(1-\varepsilon)\,\sqrt{2}\,\tilde\sigma_0}\cdot\frac{1}{\tau}\,\|\boldsymbol{Q}_i^\varepsilon - \boldsymbol{Q}_{i+1}^\varepsilon\|_{S^2} \tag{6.29b}$$

hold for $i \in \{1, \ldots, N\}$.

Proof. Taking norms on both sides of (6.27) yields

$$\|\theta_i^\varepsilon\, \mathcal{D}^\star\mathcal{D}\mathbf{\Sigma}_i^\varepsilon\|_{S^2}^2 + \langle\theta_i^\varepsilon\, \mathcal{D}^\star\mathcal{D}\mathbf{\Sigma}_i^\varepsilon,\, \lambda_i^\varepsilon\, \mathcal{D}^\star\mathcal{D}\mathbf{\Upsilon}_i^\varepsilon\rangle_{S^2} + \|\lambda_i^\varepsilon\, \mathcal{D}^\star\mathcal{D}\mathbf{\Upsilon}_i^\varepsilon\|_{S^2}^2 = \frac{1}{\tau^2}\,\|\boldsymbol{Q}_i^\varepsilon - \boldsymbol{Q}_{i+1}^\varepsilon\|_{S^2}^2.$$

Due to Lemma 6.10, the definition of μ_i^ε, see (6.19), and $\lambda_i^\varepsilon \ge 0$ the scalar product is non-negative. Indeed, we have

$$\langle\theta_i^\varepsilon\, \mathcal{D}^\star\mathcal{D}\mathbf{\Sigma}_i^\varepsilon,\, \lambda_i^\varepsilon\, \mathcal{D}^\star\mathcal{D}\mathbf{\Upsilon}_i^\varepsilon\rangle_{S^2} = 2\int_\Omega \lambda_i^\varepsilon\,\theta_i^\varepsilon\,\mathcal{D}\mathbf{\Sigma}_i^\varepsilon : \mathcal{D}\mathbf{\Upsilon}_i^\varepsilon\,\mathrm{d}x = 2\int_\Omega \lambda_i^\varepsilon\,\theta_i^\varepsilon\,\mu_i^\varepsilon\,\mathrm{d}x \ge 0.$$

This yields (6.29a).

Due to $\beta_i^\varepsilon = 0$ on $A_i^{\varepsilon,-}$, we have $\theta_i^\varepsilon = 0$ on $A_i^{\varepsilon,-}$, see (6.17). By using (6.11) we obtain $|\mathcal{D}\mathbf{\Sigma}_i^\varepsilon| = (1-\alpha_i^\varepsilon)\,|\tilde{\mathcal{D}}\mathbf{\Sigma}_i^\varepsilon| \ge \frac{1-\varepsilon}{1+\varepsilon}\,\tilde\sigma_0$ on $A_i^{\varepsilon,0} \cup A_i^{\varepsilon,+}$. Hence, the estimate (6.29b) follows by (6.29a).

Unfortunately, these estimates are not uniform w.r.t. the time step size τ. This will cause severe issues when passing to the limit $\tau \searrow 0$, see in particular Lemmas 7.8 and 7.9.

Finally, we prove the regularized counterparts of the complementarity conditions (6.4b) and (6.4c).

Lemma 6.13. The plastic multiplier λ^ε and the multiplier μ^ε satisfy

$$\|\lambda_i^\varepsilon\,\mu_i^\varepsilon\|_{L^1(\Omega)} \le k_1^{-1}\,\tau^{-1}\,\frac{\varepsilon}{1-\varepsilon}\,\|\mathcal{D}\mathbf{\Sigma}_i^\varepsilon\|_S\,\|\mathcal{D}\mathbf{\Upsilon}_i^\varepsilon\|_S \qquad \text{for all } i \in \{1, \ldots, N\}. \tag{6.30}$$

Proof. By (6.12) and (6.10) we obtain $\lambda_i^\varepsilon = 0$ on $A_i^{\varepsilon,-}$. Using (6.20) and (6.25) we infer $\mu_i^\varepsilon = 0$ on $A_i^{\varepsilon,+}$.

On $A_i^{\varepsilon,0}$ we have $\lambda_i^\varepsilon = k_1^{-1}\,\tau^{-1}\,\frac{\alpha_i^\varepsilon}{1-\alpha_i^\varepsilon} \le k_1^{-1}\,\tau^{-1}\,\frac{\varepsilon}{1-\varepsilon}$. Hence,

$$\|\lambda_i^\varepsilon\,\mu_i^\varepsilon\|_{L^1(\Omega)} \le k_1^{-1}\,\tau^{-1}\,\frac{\varepsilon}{1-\varepsilon}\,\|\mu_i^\varepsilon\|_{L^1(A_i^{\varepsilon,0})} \le k_1^{-1}\,\tau^{-1}\,\frac{\varepsilon}{1-\varepsilon}\,\|\mathcal{D}\mathbf{\Sigma}_i^\varepsilon\|_S\,\|\mathcal{D}\mathbf{\Upsilon}_i^\varepsilon\|_S.$$

Lemma 6.14. The generalized stresses $\mathbf{\Sigma}^\varepsilon$ and the multiplier θ^ε satisfy

$$\|\theta_i^\varepsilon\,\phi(\mathbf{\Sigma}_i^\varepsilon)\|_{L^1(\Omega)} \le \frac{2\,\varepsilon}{(1+\varepsilon)^2}\,\tilde\sigma_0^2\,\|\theta_i^\varepsilon\|_{L^1(A_i^{\varepsilon,0})} \qquad \text{for all } i \in \{1, \ldots, N\}. \tag{6.31}$$

Proof. We have $\beta_i^\varepsilon = 0$ and hence by (6.17) $\theta_i^\varepsilon = 0$ on $A_i^{\varepsilon,-}$. From (6.25) and (6.11) we infer $\phi(\mathbf{\Sigma}_i^\varepsilon) = 0$ on $A_i^{\varepsilon,+}$.

On $A_i^{\varepsilon,0}$ we have $|\mathcal{D}\mathbf{\Sigma}_i^\varepsilon| \in \left[\frac{1-\varepsilon}{1+\varepsilon}, 1\right]\,\tilde\sigma_0$ by (6.10) and (6.11). Hence, we obtain

$$|\phi(\mathbf{\Sigma}_i^\varepsilon)| = \frac{\tilde\sigma_0^2 - |\mathcal{D}\mathbf{\Sigma}_i^\varepsilon|^2}{2} \le \frac{2\,\varepsilon}{(1+\varepsilon)^2}\,\tilde\sigma_0^2 \quad \text{a.e. on } A_i^{\varepsilon,0}.$$

Using Hölder's inequality concludes the proof.

Using the results above we prove that the system (6.1)–(6.4) is a necessary optimality condition for the time-discrete control problem $(\mathbf{P}^\tau)$.

Theorem 6.15. Let $\boldsymbol{g}^\tau$ be a local solution of $(\mathbf{P}^\tau)$. We denote by $(\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau, \lambda^\tau) \in (S \times V \times L^2(\Omega))^N$ the stress, displacement and plastic multiplier, which are associated to $\boldsymbol{g}^\tau$ by (3.8). Then, there are adjoint states $(\boldsymbol{\Upsilon}^\tau, \boldsymbol{w}^\tau) \in (S \times V)^N$ and multipliers $\mu^\tau, \theta^\tau \in L^2(\Omega)^N$, such that (6.1)–(6.4) is fulfilled.

Proof. Corollary 6.6 implies the existence of a sequence of local solutions $\{\boldsymbol{g}^\varepsilon\}$ of $(\mathbf{P}^\varepsilon_{\boldsymbol{g}^\tau})$, such that $\boldsymbol{g}^\varepsilon \to \boldsymbol{g}^\tau$ as $\varepsilon \searrow 0$.

Let us denote by $(\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon, \lambda^\varepsilon)$ the regularized stresses, displacements and plastic multipliers, which are associated to $\boldsymbol{g}^\varepsilon$ by (5.4) and (6.12). From Corollary 5.12 and Corollary 6.7 we infer

$$(\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon, \lambda^\varepsilon) \to (\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau, \lambda^\tau) \quad \text{in } (S^2 \times V \times L^2(\Omega))^N \quad \text{as } \varepsilon \searrow 0,$$

where $(\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau, \lambda^\tau) \in (S \times V \times L^2(\Omega))^N$ are the unregularized stresses, displacements and plastic multipliers associated to $\boldsymbol{g}^\tau$, see (3.8). This shows the forward system (6.1).

Let us define the adjoint states $(\boldsymbol{\Upsilon}^\varepsilon, \boldsymbol{w}^\varepsilon)$ and the multipliers $(\theta^\varepsilon, \mu^\varepsilon)$ associated to $\boldsymbol{g}^\varepsilon$ by (6.18), (6.17) and (6.20). By Lemma 6.11 and Lemma 6.12 the adjoint states $(\boldsymbol{\Upsilon}^\varepsilon, \boldsymbol{w}^\varepsilon)$ and the multipliers $(\theta^\varepsilon, \mu^\varepsilon)$ are bounded in $(S^2 \times V)^N$ and $L^2(\Omega)^N \times L^2(\Omega)$, respectively. Hence, there is a subsequence of ε denoted by the same symbol and an element $(\boldsymbol{\Upsilon}^\tau, \boldsymbol{w}^\tau, \theta^\tau, \mu^\tau) \in (S^2 \times V \times L^2(\Omega) \times L^2(\Omega))^N$, such that

$$(\boldsymbol{\Upsilon}^\varepsilon, \boldsymbol{w}^\varepsilon, \theta^\varepsilon, \mu^\varepsilon) \rightharpoonup (\boldsymbol{\Upsilon}^\tau, \boldsymbol{w}^\tau, \theta^\tau, \mu^\tau) \quad \text{in } (S^2 \times V \times L^2(\Omega) \times L^2(\Omega))^N \quad \text{as } \varepsilon \searrow 0.$$

Therefore, we can pass to the limit in the necessary optimality condition (6.8) of the modified regularized problem $(\mathbf{P}^\varepsilon_{\boldsymbol{g}^\tau})$ and obtain (6.3).

By $\lambda^\varepsilon \to \lambda^\tau$ in $L^2(\Omega)^N$ and by $\boldsymbol{\Upsilon}^\varepsilon \rightharpoonup \boldsymbol{\Upsilon}^\tau$ in $(S^2)^N$, we infer $\lambda^\varepsilon \, \mathcal{D}\boldsymbol{\Upsilon}^\varepsilon \rightharpoonup \lambda^\tau \, \mathcal{D}\boldsymbol{\Upsilon}^\tau$ in $L^1(\Omega; \mathbb{S})^N$. Using (6.29a) we obtain $\lambda^\varepsilon \, \mathcal{D}\boldsymbol{\Upsilon}^\varepsilon \rightharpoonup \lambda^\tau \, \mathcal{D}\boldsymbol{\Upsilon}^\tau$ in S^N for fixed $\tau > 0$, since $\boldsymbol{Q}^\varepsilon$ is bounded, see (6.28). Similarly, we infer $\theta^\varepsilon \, \mathcal{D}\boldsymbol{\Sigma}^\varepsilon \rightharpoonup \theta^\tau \, \mathcal{D}\boldsymbol{\Sigma}^\tau$ in S^N. Therefore, we can pass to the limit in the regularized adjoint equation (6.18) and obtain (6.2).

It remains to check the relations (6.4). Using the definition (6.20) of μ^ε we infer (6.4a). Now we address the complementarity conditions (6.4b) and (6.4c). In view of Lemma 6.13 and Lemma 6.14 it would suffice to prove the weak convergence of $\lambda_i^\varepsilon \, \mu_i^\varepsilon$ and $\theta_i^\varepsilon \, \phi(\boldsymbol{\Sigma}_i^\varepsilon)$ in $L^1(\Omega)$, since the $L^1(\Omega)$-norm is weakly lower semicontinuous. By $\boldsymbol{\Sigma}_i^\varepsilon \to \boldsymbol{\Sigma}_i^\tau$ in S^2 and $\lambda_i^\varepsilon \, \mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon \rightharpoonup \lambda_i^\tau \, \mathcal{D}\boldsymbol{\Upsilon}_i^\tau$ in S we infer $\lambda_i^\varepsilon \, \mu_i^\varepsilon = \lambda_i^\varepsilon \, \mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon : \mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon \rightharpoonup \lambda_i^\tau \, \mu_i^\tau$ in $L^1(\Omega)$. Similar, using $\theta_i^\varepsilon \, \mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon \rightharpoonup \theta_i^\tau \, \mathcal{D}\boldsymbol{\Sigma}_i^\tau$ in S and $\boldsymbol{\Sigma}_i^\varepsilon \to \boldsymbol{\Sigma}_i^\tau$ in S^2 shows $\theta_i^\varepsilon \, \phi(\boldsymbol{\Sigma}_i^\varepsilon) = \theta_i^\varepsilon \, \frac{1}{2} (\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon : \mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon - \tilde{\sigma}_0)^2 \rightharpoonup \theta_i^\tau \, \phi(\boldsymbol{\Sigma}_i^\tau)$ in $L^1(\Omega)$. Here, we used the definition (1.5) of ϕ. This shows the complementarity conditions (6.4b) and (6.4c).

Last we address (6.4d). We will use [Herzog et al., 2010, Proposition 3.15]. To this end, we test (6.18a) with $\varphi \, \boldsymbol{\Upsilon}_i^\varepsilon$, where $\varphi \in C_0^\infty(\Omega)$, $\varphi \geq 0$. Using $\theta_i^\varepsilon \, \mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon : \mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon = \theta_i^\varepsilon \, \mu_i^\varepsilon \geq 0$ by Lemma 6.10, we obtain

$$\langle A(\boldsymbol{\Upsilon}_i^\varepsilon - \boldsymbol{\Upsilon}_{i+1}^\varepsilon) + B^\star(\boldsymbol{w}_i^\varepsilon - \boldsymbol{w}_{i+1}^\varepsilon) + \lambda_i^\varepsilon \, \mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon, \, \varphi \, \boldsymbol{\Upsilon}_i^\varepsilon \rangle_{S^2} \leq 0.$$

Applying [Herzog et al., 2010, Proposition 3.15] yields

$$\langle A(\boldsymbol{\Upsilon}_i^\tau - \boldsymbol{\Upsilon}_{i+1}^\tau) + B^\star(\boldsymbol{w}_i^\tau - \boldsymbol{w}_{i+1}^\tau) + \lambda_i^\tau \, \mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon}_i^\tau, \, \varphi \, \boldsymbol{\Upsilon}_i^\tau \rangle_{S^2} \leq 0.$$

Testing (6.2a) with $\varphi \, \boldsymbol{\Upsilon}_i^\tau$ yields

$$\int_\Omega \varphi \, \theta_i^\tau \, \mu_i^\tau \, \mathrm{d}x \geq 0 \quad \text{for all } \varphi \in C_0^\infty(\Omega) \text{ satisfying } \varphi \geq 0.$$

Hence, $\theta_i^\tau \, \mu_i^\tau \geq 0$ almost everywhere. This shows (6.4d).

Remark 6.16.

(1) Similarly to Theorem 6.15 a necessary optimality condition for the modified time-discrete problem ($\mathbf{P}^\tau_{\boldsymbol{g}}$) can be proven. In the optimality system (6.1)–(6.4) we have to replace the gradient equation (6.3) by

$$\sum_{i=1}^{N} \big\langle E^\star \boldsymbol{w}_i^\tau, \tilde{\boldsymbol{g}}_i^\tau - \tilde{\boldsymbol{g}}_{i-1}^\tau - (\boldsymbol{g}_i^\tau - \boldsymbol{g}_{i-1}^\tau)\big\rangle_{L^2(\Gamma_N;\mathbb{R}^d)} + \langle (\boldsymbol{g}^\tau - \boldsymbol{g}) + \nu\, \boldsymbol{g}^\tau, \tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\tau \rangle_{H^1(0,T;U)} \geq 0, \tag{6.3'}$$

for all $\tilde{\boldsymbol{g}}^\tau \in U^\tau_{\text{ad}}$. As an optimality system for the modified time-discrete problem ($\mathbf{P}^\tau_{\boldsymbol{g}}$) we obtain (6.1), (6.2), (6.3') and (6.4).

(2) In case $N = 1$ (only one time step) we obtain an optimality system for the optimal control of *static* plasticity. The system (6.1)–(6.4) equals the system [Herzog et al., 2010, (3.3)–(3.6)] up to minor differences: in the current paper we neglected volume forces $\boldsymbol{f}$, but considered additionally control constraints.

(3) Using the technique of [Herzog et al., 2011c, Section 3] one may derive a system of B-stationary type for the time-discrete problem ($\mathbf{P}^\tau$).

7 Weak stationarity for the quasistatic problem

In this section we derive an optimality system for the continuous problem

$$\left.\begin{aligned} \text{Minimize}\quad & F(\boldsymbol{u},\boldsymbol{g}) = \psi(\boldsymbol{u}) + \frac{\nu}{2}\|\boldsymbol{g}\|^2_{H^1(0,T;U)} \\ \text{such that}\quad & (\boldsymbol{\Sigma},\boldsymbol{u}) = \mathcal{G}(E\boldsymbol{g}) \\ \text{and}\quad & \boldsymbol{g} \in U_{\text{ad}}. \end{aligned}\right\} \tag{$\mathbf{P}$}$$

Here, $\mathcal{G}$ is the solution map of (**VI**). We use arguments similar to those in the proof of Theorem 6.15. Throughout this section, $\boldsymbol{g}$ denotes a local optimum of ($\mathbf{P}$). Theorem 3.10 yields the existence of a sequence $\boldsymbol{g}^\tau$ of local optima of the time-discrete and modified problems ($\mathbf{P}^\tau_{\boldsymbol{g}}$) which converges to $\boldsymbol{g}$ in the strong topology of $H^1(0,T;U)$. The convergence of the states $(\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau, \lambda^\tau)$ towards $(\boldsymbol{\Sigma}, \boldsymbol{u}, \lambda)$ in $H^1(0,T;S^2\times V)\times L^2(0,T;L^2(\Omega))$ was shown in Theorems 3.3 and 3.4. In this section, we study the convergence properties of the dual quantities $(\boldsymbol{\Upsilon}^\tau, \boldsymbol{w}^\tau, \mu^\tau, \theta^\tau)$ and pass to the limit in the optimality system (6.1)–(6.4) as $\tau \searrow 0$.

Unfortunately, one cannot show the boundedness of $\boldsymbol{\Upsilon}^\tau$ and $\boldsymbol{w}^\tau$ in $H^1(0,T;S^2)$ and $H^1(0,T;V)$, but only in $L^\infty(0,T;S^2)$ and $L^\infty(0,T;V)$, respectively, which was already proven in Lemma 6.11. Due to this lack of regularity, the derivatives in time of $\boldsymbol{\Upsilon}^\tau$ and $\boldsymbol{w}^\tau$ in the adjoint equation (6.2) have to be formulated in a weak sense in order to pass to the limit $\tau \searrow 0$. Hence, the adjoint equation (7.36) of the continuous problem can be stated only in a weak sense. As mentioned in the introduction, this lack of regularity also occurs in the optimal control of the parabolic obstacle problem, see [Ito and Kunisch, 2010, Theorem 6.2] and for the optimal control of ODES involving hysteresis, see [Brokate, 1987, Satz 8.12].

7.1. Additional notations and assumptions. In this section we set up some additional notation required in the sequel.

As in (3.1), let $f^\tau \in X^N$ be given, where X is some Banach space. We define for $t \in [(i-1)\,\tau, i\,\tau)$ the piecewise linear and *continuous* interpolant

$$f^\tau_{\text{c}}(t) = f^\tau_{i-1} + \Big(\frac{t}{\tau} - (i-1)\Big)\,(f^\tau_i - f^\tau_{i-1}), \tag{7.1}$$

with $f_0^\tau = f_1^\tau$ by convention. Compared with definition (3.1), only the fictitious value of f_0^τ was changed. Due to this choice, the adjoint equation (7.22a) is not only satisfied in the interval (τ, T) but also in $(0, \tau)$.

Similarly to (3.3), we define for $f^\tau \in X^N$ and $t \in [(i-1)\,\tau, i\,\tau)$ the piecewise constant and *discontinuous* interpolant

$$f_\mathrm{d}^\tau(t) = f_{i-1}^\tau \tag{7.2}$$

with the convention $f_0^\tau = \mathbf{0}$.

Obviously, we have $f_\mathrm{c}^\tau \in H^1(0,T;S^2)$ and $f_\mathrm{d}^\tau \in L^2(0,T;S^2)$. For later reference we mention that

$$\dot{f}_\mathrm{c}^\tau(t) = \frac{1}{\tau}\,(f_i^\tau - f_{i-1}^\tau) \tag{7.3}$$

holds for all $t \in ((i-1)\,\tau, i\,\tau)$.

We also require an additional structural assumption on the objective.

Assumption 7.1. In addition to Assumption 3.6, we assume that $\psi : H^1(0,T;V) \to \mathbb{R}$ can be decomposed into $\psi_\mathrm{c} : L^2(0,T;V) \to \mathbb{R}$ and $\psi_T : V \to \mathbb{R}$, such that

$$\psi(\boldsymbol{u}) = \psi_\mathrm{c}(\boldsymbol{u}) + \psi_T(\boldsymbol{u}(T)) \tag{7.4}$$

holds for all $\boldsymbol{u} \in H^1(0,T;V)$. Both, ψ_c and ψ_T are assumed to be continuously Fréchet differentiable.

Note that due to $H^1(0,T;V) \hookrightarrow C([0,T];V)$ the value $\boldsymbol{u}(T)$ is well-defined. We remark that Assumption 7.1 is satisfied by all examples of ψ given after Assumption 2.8.

Given $\boldsymbol{u}, \delta\boldsymbol{u} \in H^1(0,T;V)$, the derivative of ψ at $\boldsymbol{u}$ in direction $\delta\boldsymbol{u}$ is given by

$$\psi'(\boldsymbol{u})\,\delta\boldsymbol{u} = \int_0^T \langle \delta\boldsymbol{u},\, \nabla\psi_\mathrm{c}(\boldsymbol{u})\rangle_{V,V'}\,\mathrm{d}t + \psi_T'(\boldsymbol{u}(T))\,\delta\boldsymbol{u}(T), \tag{7.5}$$

where $\nabla\psi_\mathrm{c}(\boldsymbol{u}) \in L^2(0,T;V')$ denotes the gradient of ψ_c at $\boldsymbol{u}$ and $\psi_T' \in V'$ is the Fréchet derivative of ψ_T.

Let us check that Assumption 7.1 implies Assumption 6.1. If we denote by $v_i : [0,T] \to \mathbb{R}$ the usual hat function associated with the node $t = i\,\tau$ (piecewise linear, continuous, 0 at $j\,\tau$ for $j \ne i$ and 1 at $i\,\tau$), we obtain

$$\psi_i^\tau(\boldsymbol{u}^\tau) = \int_0^T v_i\,\nabla\psi_\mathrm{c}(\boldsymbol{u}^\tau)\,\mathrm{d}t \qquad \text{for } i = 1,\ldots,N-1, \tag{7.6a}$$

$$\psi_i^\tau(\boldsymbol{u}^\tau) = \int_0^T v_i\,\nabla\psi_\mathrm{c}(\boldsymbol{u}^\tau)\,\mathrm{d}t + \psi_T'(\boldsymbol{u}^\tau(T)) \qquad \text{for } i = N. \tag{7.6b}$$

This shows that $\psi(\boldsymbol{u}^\tau)$ is Fréchet differentiable w.r.t. the components $\boldsymbol{u}_i^\tau$ of $\boldsymbol{u}^\tau$.

Moreover, (7.6) implies that the right hand side in (6.26) is bounded. Indeed, we obtain

$$\begin{aligned}\sum_{i=1}^N \|\psi_i^\tau(\boldsymbol{u}^\tau)\|_{V'} &\le \sum_{i=1}^N \Big\| \int_0^T v_i\,\nabla\psi_\mathrm{c}(\boldsymbol{u}^\tau)\,\mathrm{d}t \Big\|_{V'} + \|\psi_T'(\boldsymbol{u}^\tau(T))\|_{V'} \\ &\le \int_0^T \|\nabla\psi_\mathrm{c}(\boldsymbol{u}^\tau)\|_{V'}\,\mathrm{d}t + \|\psi_T'(\boldsymbol{u}^\tau(T))\|_{V'},\end{aligned} \tag{7.7}$$

where we used that the v_i form a partition of unity. Together with Lemma 6.11 this estimate is used to derive the boundedness of $(\boldsymbol{\Upsilon}_\mathrm{c}^\tau, \boldsymbol{w}_\mathrm{c}^\tau)$ in $L^\infty(0,T;S^2 \times V)$, see (7.10).

7.2. Basic convergence properties. Let us denote by $\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau, \lambda^\tau, \boldsymbol{\Upsilon}^\tau, \boldsymbol{w}^\tau, \mu^\tau, \theta^\tau$ the states, adjoint states and multipliers, such that the optimality system (6.1), (6.2), (6.3') and (6.4) of the modified problem ($\mathbf{P}^\tau_{\boldsymbol{g}}$) is satisfied, see Remark 6.16 (1). As mentioned in the introduction of this section, Theorems 3.3 and 3.4 imply

$$\left.\begin{aligned} \boldsymbol{\Sigma}^\tau &\to \boldsymbol{\Sigma} \quad \text{in } H^1(0,T;S^2), \qquad \boldsymbol{u}^\tau \to \boldsymbol{u} \quad \text{in } H^1(0,T;V), \\ \lambda^\tau &\to \lambda \quad \text{in } L^2(0,T;L^2(\Omega)), \end{aligned}\right\} \tag{7.8}$$

where $(\boldsymbol{\Sigma}, \boldsymbol{u}, \lambda)$ is the solution of the continuous problem associated to $\boldsymbol{g}$, see (1.22) with $\ell = E\boldsymbol{g}$. Using the definitions introduced in Section 7.1, it is easy to see that

$$\left.\begin{aligned} \boldsymbol{\Sigma}^\tau_{\mathrm{d}} &\to \boldsymbol{\Sigma} \quad \text{in } L^\infty(0,T;S^2), \qquad \boldsymbol{u}^\tau_{\mathrm{d}} \to \boldsymbol{u} \quad \text{in } L^\infty(0,T;S^2), \\ \lambda^\tau_{\mathrm{d}} &\to \lambda \quad \text{in } L^2(0,T;L^2(\Omega)). \end{aligned}\right\} \tag{7.9}$$

Using Lemma 6.11 and (7.7), the continuities of $\nabla\psi_{\mathrm{c}} : L^2(0,T;V) \to L^2(0,T;V')$ and $\psi_T' : V \to V'$ imply

$$\|(\boldsymbol{\Upsilon}^\tau_{\mathrm{c}}, \boldsymbol{w}^\tau_{\mathrm{c}})\|_{L^\infty(0,T;S^2\times V)} + \|(\boldsymbol{\Upsilon}^\tau_{\mathrm{d}}, \boldsymbol{w}^\tau_{\mathrm{d}})\|_{L^\infty(0,T;S^2\times V)} \le C, \tag{7.10}$$

where C does not depend on τ. Since $L^\infty(0,T;S^2\times V)$ is the dual of $L^1(0,T;S^2\times V')$, see [Diestel and Uhl, 1977, Theorem 4.1] or [Edwards, 1995, Theorem 8.18.3], there are subsequences of $(\boldsymbol{\Upsilon}^\tau_{\mathrm{c}}, \boldsymbol{w}^\tau_{\mathrm{c}})$ and of $(\boldsymbol{\Upsilon}^\tau_{\mathrm{d}}, \boldsymbol{w}^\tau_{\mathrm{d}})$ (denoted by the same symbol) which converge in the weak-$\star$ topology of $L^\infty(0,T;S^2\times V)$. Testing $(\boldsymbol{\Upsilon}^\tau_{\mathrm{c}}, \boldsymbol{w}^\tau_{\mathrm{c}})$ and $(\boldsymbol{\Upsilon}^\tau_{\mathrm{d}}, \boldsymbol{w}^\tau_{\mathrm{d}})$ with $\chi_{[0,t]}\,(\boldsymbol{T}, \ell) \in L^1(0,T;S^2\times V')$, where $\boldsymbol{T} \in S^2$, $\ell \in V'$ and $t \in [0,T]$ are arbitrary, shows that the weak-$\star$ limits coincide. Denoting the weak-$\star$ limit by $(\boldsymbol{\Upsilon}, \boldsymbol{w}) \in L^\infty(0,T;S^2\times V)$, we obtain

$$\left.\begin{aligned} \boldsymbol{\Upsilon}^\tau_{\mathrm{c}} &\overset{\star}{\rightharpoonup} \boldsymbol{\Upsilon} \quad \text{in } L^\infty(0,T;S^2), \qquad \boldsymbol{\Upsilon}^\tau_{\mathrm{d}} \overset{\star}{\rightharpoonup} \boldsymbol{\Upsilon} \quad \text{in } L^\infty(0,T;S^2), \\ \boldsymbol{w}^\tau_{\mathrm{c}} &\overset{\star}{\rightharpoonup} \boldsymbol{w} \quad \text{in } L^\infty(0,T;V), \qquad \boldsymbol{w}^\tau_{\mathrm{d}} \overset{\star}{\rightharpoonup} \boldsymbol{w} \quad \text{in } L^\infty(0,T;V). \end{aligned}\right\} \tag{7.11}$$

Using the weak convergence of $\boldsymbol{w}^\tau_{\mathrm{d}}$ we infer the convergence of the gradient equation. Employing the notations (3.1) and (7.2), the gradient equation (6.3') reads

$$\langle E^\star \boldsymbol{w}^\tau_{\mathrm{d}}, \dot{\tilde{\boldsymbol{g}}}^\tau - \dot{\boldsymbol{g}}^\tau\rangle_{L^2(0,T;U)} + \langle(\boldsymbol{g}^\tau - \boldsymbol{g}) + \nu\,\boldsymbol{g}^\tau, \tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\tau\rangle_{H^1(0,T;U)} \ge 0 \tag{7.12}$$

for all $\tilde{\boldsymbol{g}}^\tau \in U^\tau_{\mathrm{ad}}$. Due to Assumption 3.6 every $\tilde{\boldsymbol{g}} \in U_{\mathrm{ad}}$ can be approximated by a sequence $\tilde{\boldsymbol{g}}^\tau \in U^\tau_{\mathrm{ad}}$. Passing to the limit $\tau \searrow 0$ implies

$$\langle E^\star \boldsymbol{w}, \dot{\tilde{\boldsymbol{g}}} - \dot{\boldsymbol{g}}\rangle_{L^2(0,T;U)} + \nu\,\langle \boldsymbol{g}, \tilde{\boldsymbol{g}} - \boldsymbol{g}\rangle_{H^1(0,T;U)} \ge 0 \tag{7.13}$$

for all $\tilde{\boldsymbol{g}} \in U_{\mathrm{ad}}$.

7.3. Preliminary considerations. In this section we provide some results needed several times in the sequel.

We show a relationship between the piecewise linear interpolant $\boldsymbol{\Sigma}^\tau$, see (3.1), and the piecewise constant interpolant $\boldsymbol{\Sigma}^\tau_{\mathrm{d}}$, see (7.2).

Similarly to the relation (3.2), a simple calculation shows

$$\boldsymbol{\Sigma}^\tau_{\mathrm{d}}(t) = \boldsymbol{\Sigma}^\tau(t) - (t - n^\tau(t)\,\tau)\,\dot{\boldsymbol{\Sigma}}^\tau(t) \quad \text{f.a.a. } t \in [0,T], \tag{7.14}$$

where n^τ is given by

$$n^\tau(t) = \max\{n \in \mathbb{N} : t \ge (n-1)\,\tau\}.$$

This definition implies

$$t \in [(n^\tau(t) - 1)\,\tau, n^\tau(t)\,\tau) \quad \text{for all } t \in [0,T].$$

The relation (7.14) gives rise to the definition

$$\kappa^\tau(t) = (t - n^\tau(t)\,\tau) \quad \text{f.a.a. } t \in [0,T]. \tag{7.15}$$

Obviously, we have $\kappa^\tau \in L^\infty(0,T)$ and

$$\kappa^\tau(t) \in [-\tau, 0] \quad \text{f.a.a. } t \in [0,T]. \tag{7.16}$$

Due to (7.14) the term $\kappa^\tau\,(A\dot{\boldsymbol{\Upsilon}}_c^\tau + B^\star\dot{\boldsymbol{w}}_c^\tau)$ appears frequently in Lemmas 7.8 and 7.9. Using the estimates (6.28) and (7.7) we can prove that it converges to zero w.r.t. the weak-$\star$ topology of $L^\infty(0,T;S^2)$.

Theorem 7.2. We have

$$\kappa^\tau\,(A\dot{\boldsymbol{\Upsilon}}_c^\tau + B^\star\dot{\boldsymbol{w}}_c^\tau) \stackrel{\star}{\rightharpoonup} 0 \quad \text{in } L^\infty(0,T;S^2).$$

Proof. Let us use the abbreviation $\boldsymbol{Q}_c^\tau = -A\boldsymbol{\Upsilon}_c^\tau - B^\star\boldsymbol{w}_c^\tau$. Testing $\kappa^\tau\,\dot{\boldsymbol{Q}}_c^\tau$ with $\chi_{(t,T)}\,\boldsymbol{T}$, where $t \in [0,T]$ and $\boldsymbol{T} \in S^2$, yields

$$\langle \chi_{(t,T)}\,\boldsymbol{T},\, \kappa^\tau\,\dot{\boldsymbol{Q}}_c^\tau\rangle_{L^2(0,T;S^2)} = \int_t^{n^\tau(t)\,\tau} \langle \boldsymbol{T},\, \kappa^\tau\,\dot{\boldsymbol{Q}}_c^\tau\rangle_{S^2}\,\mathrm{d}s + \int_{n^\tau(t)\,\tau}^T \langle \boldsymbol{T},\, \kappa^\tau\,\dot{\boldsymbol{Q}}_c^\tau\rangle_{S^2}\,\mathrm{d}s.$$

Let us estimate these two terms. For the first one we have

$$\left|\int_t^{n^\tau(t)\,\tau} \langle \boldsymbol{T},\, \kappa^\tau\,\dot{\boldsymbol{Q}}_c^\tau\rangle_{S^2}\,\mathrm{d}s\right| \le \tau\,\|\boldsymbol{T}\|_{S^2}\,\|\tau\,\dot{\boldsymbol{Q}}_c^\tau\|_{L^\infty(0,T;S^2)},$$

since $|n^\tau(t)\,\tau - t| \le \tau$ and $|\kappa^\tau| \le \tau$ a.e. in $[0,T]$, see (7.16).

Using $\int_{(i-1)\,\tau}^{i\,\tau} \kappa^\tau\,\mathrm{d}s = -\tau/2$ for all $i \in \{1,\ldots,N\}$ and the constantness of $\dot{\boldsymbol{Q}}_c^\tau$ on $((i-1)\,\tau, i\,\tau)$ for all $i \in \{1,\ldots,N\}$, we obtain for the second term

$$\begin{aligned}\left|\int_{n^\tau(t)\,\tau}^T \langle \boldsymbol{T},\, \kappa^\tau\,\dot{\boldsymbol{Q}}_c^\tau\rangle_{S^2}\,\mathrm{d}s\right| &= \left|\frac{\tau}{2}\int_{n^\tau(t)\,\tau}^T \langle \boldsymbol{T},\, \dot{\boldsymbol{Q}}_c^\tau\rangle_{S^2}\,\mathrm{d}s\right| \\ &\le \left|\frac{\tau}{2}\,\langle \boldsymbol{T},\, \boldsymbol{Q}_c^\tau(T) - \boldsymbol{Q}_c^\tau(n^\tau(t)\,\tau)\rangle_{S^2}\right| \\ &\le \tau\,\|\boldsymbol{T}\|_{S^2}\,\|\boldsymbol{Q}_c^\tau\|_{L^\infty(0,T;S^2)}.\end{aligned}$$

Hence, by (6.28) and (7.7) we obtain

$$\left|\langle \chi_{(t,T)}\,\boldsymbol{T},\, \kappa^\tau\,\dot{\boldsymbol{Q}}_c^\tau\rangle_{L^2(0,T;S^2)}\right| \le \tau\,C.$$

Using the boundedness of $\kappa^\tau\,\dot{\boldsymbol{Q}}_c^\tau$ in $L^\infty(0,T;S^2)$, see (6.28), the density of the linear hull of $\{\chi_{[t,T]}\,\boldsymbol{T}\}$ in $L^1(0,T;S^2)$ finishes the proof, see Lemma A.2.

Another term which appears frequently is $\mathcal{D}\boldsymbol{T}^\tau : \mathcal{D}\boldsymbol{\Sigma}^\tau$, where $\boldsymbol{T}^\tau \in L^\infty(0,T;S^2)$ is a sequence which converges in the weak-$\star$ topology. Using the boundedness of $\mathcal{D}\boldsymbol{\Sigma}^\tau$ in $L^\infty((0,T)\times\Omega;\mathbb{S})$, see (1.9), and the convergence $\mathcal{D}\boldsymbol{\Sigma}^\tau \to \mathcal{D}\boldsymbol{\Sigma}$ in $L^\infty(0,T;S)$, an interpolation argument, see [Tartar, 2007, Lemma 8.2], yields $\mathcal{D}\boldsymbol{\Sigma}^\tau \to \mathcal{D}\boldsymbol{\Sigma}$ in $L^\infty(0,T;L^p(\Omega;\mathbb{S}))$ for all $p < \infty$. This gives in turn $\mathcal{D}\boldsymbol{T}^\tau : \mathcal{D}\boldsymbol{\Sigma}^\tau \stackrel{\star}{\rightharpoonup} \mathcal{D}\boldsymbol{T} : \mathcal{D}\boldsymbol{\Sigma}$ in $L^\infty(0,T;L^q(\Omega))$ for all $q < 2$. Using the boundedness of $\mathcal{D}\boldsymbol{T}^\tau : \mathcal{D}\boldsymbol{\Sigma}^\tau$ in $L^\infty(0,T;L^2(\Omega))$ we infer even the weak-$\star$ convergence of the product in $L^\infty(0,T;L^2(\Omega))$.

Lemma 7.3. Let $\boldsymbol{T}^\tau \stackrel{\star}{\rightharpoonup} \boldsymbol{T}$ in $L^\infty(0,T;S^2)$. Then

$$\mathcal{D}\boldsymbol{\Sigma}^\tau : \mathcal{D}\boldsymbol{T}^\tau \stackrel{\star}{\rightharpoonup} \mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T} \quad \text{in } L^\infty(0,T;L^2(\Omega)).$$

Let $f^\tau \rightharpoonup f$ in $L^2(0,T;L^1(\Omega))$. Then

$$f^\tau\,\mathcal{D}\boldsymbol{\Sigma}^\tau \rightharpoonup f\,\mathcal{D}\boldsymbol{\Sigma} \quad \text{in } L^2(0,T;L^1(\Omega;\mathbb{S})).$$

Let $g^\tau \overset{\star}{\rightharpoonup} g$ in $L^\infty(0,T;L^2(\Omega))$. Then

$$g^\tau\, \mathcal{D}\boldsymbol{\Sigma}^\tau \overset{\star}{\rightharpoonup} g\, \mathcal{D}\boldsymbol{\Sigma} \quad \text{in } L^\infty(0,T;S).$$

The statements remain valid if $\boldsymbol{\Sigma}^\tau$ is replaced by $\boldsymbol{\Sigma}_{\mathrm{d}}^\tau$.

Proof. Let us prove the first statement. Since $\mathcal{D}\boldsymbol{\Sigma}^\tau \to \mathcal{D}\boldsymbol{\Sigma}$ in $L^\infty(0,T;S)$ and $\boldsymbol{T}^\tau \overset{\star}{\rightharpoonup} \boldsymbol{T}$ in $L^\infty(0,T;S^2)$, we obtain

$$\mathcal{D}\boldsymbol{T}^\tau : \mathcal{D}\boldsymbol{\Sigma}^\tau \rightharpoonup \mathcal{D}\boldsymbol{T} : \mathcal{D}\boldsymbol{\Sigma} \quad \text{in } L^1((0,T)\times\Omega). \tag{7.17}$$

Since $\mathcal{D}\boldsymbol{\Sigma}^\tau$ is bounded in the space $L^\infty((0,T)\times\Omega;\mathbb{S})$, $\mathcal{D}\boldsymbol{T}^\tau : \mathcal{D}\boldsymbol{\Sigma}^\tau$ is bounded in $L^\infty(0,T;L^2(\Omega))$. Due to this boundedness, there exists a subsequence which converges with respect to the weak-$\star$ topology of $L^\infty(0,T;L^2(\Omega))$. Due to (7.17), the limit is unique and hence we obtain the convergence of the whole sequence.

The statement involving $g \in L^\infty(0,T;L^2(\Omega))$ proves completely analogously. To prove the statement involving $f \in L^2(0,T;L^1(\Omega))$, one has to use additionally Egorov's theorem.

7.4. Passing to the limit in the adjoint equation. We show the weak convergence of the terminal values of the adjoint states $\boldsymbol{\Upsilon}_{\mathrm{c}}^\tau$ and $\boldsymbol{w}_{\mathrm{c}}^\tau$. Note that this does not simply follow from the weak-$\star$ convergence in $L^\infty(0,T;S^2\times V)$.

Using $\boldsymbol{\Upsilon}_{\mathrm{c}}^\tau(T) = \boldsymbol{\Upsilon}_N^\tau$ and $\boldsymbol{w}_{\mathrm{c}}^\tau(T) = \boldsymbol{w}_N^\tau$ the adjoint equation (6.2) with $i = N$ implies

$$A\boldsymbol{\Upsilon}_N^\tau + \tau\,\lambda_N^\tau\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon}_N^\tau + \tau\,\theta_N^\tau\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_N^\tau + B^\star\boldsymbol{w}_N^\tau = \boldsymbol{0}, \tag{7.18a}$$

$$B\boldsymbol{\Upsilon}_N^\tau = \psi_N^\tau(\boldsymbol{u}^\tau). \tag{7.18b}$$

By standard saddle-point arguments, we obtain the boundedness of $\boldsymbol{\Upsilon}_N^\tau$, $\boldsymbol{w}_N^\tau$ and $\tau\,\theta_N^\tau$. Hence, there exists a subsequence, denoted by the same symbol, such that

$$(\boldsymbol{\Upsilon}_N^\tau, \boldsymbol{w}_N^\tau, \tau\,\theta_N^\tau) \rightharpoonup (\boldsymbol{\Upsilon}_T, \boldsymbol{w}_T, \theta_T) \quad \text{in } S^2\times V\times L^2(\Omega). \tag{7.19}$$

Moreover, the boundedness of $\lambda_{\mathrm{d}}^\tau\,\mathcal{D}\boldsymbol{\Upsilon}_{\mathrm{d}}^\tau$ in $L^2(0,T;L^1(\Omega;\mathbb{S}))$ implies $\tau\,\lambda_N^\tau\,\mathcal{D}\boldsymbol{\Upsilon}_N^\tau \to 0$ in $L^1(\Omega;\mathbb{S})$. Hence, we obtain from (7.18)

$$A\boldsymbol{\Upsilon}_T + \theta_T\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}(T) + B^\star\boldsymbol{w}_T = \boldsymbol{0}, \tag{7.20a}$$

$$B\boldsymbol{\Upsilon}_T = \psi_T'(\boldsymbol{u}(T)). \tag{7.20b}$$

Similarly to the derivation of (6.4c) and (6.4d) we obtain

$$\theta_T\,\phi(\boldsymbol{\Sigma}(T)) = 0 \qquad \text{and} \qquad \theta_T\,\mathcal{D}\boldsymbol{\Sigma}(T) : \mathcal{D}\boldsymbol{\Upsilon}_T \geq 0. \tag{7.21}$$

Under an additional regularity assumption we obtain $(\boldsymbol{\Upsilon}, \boldsymbol{w}) \in H^1(0,T;S^2\times V)$ and $(\boldsymbol{\Upsilon}_T, \boldsymbol{w}_T)$ coincides with $(\boldsymbol{\Upsilon}(T), \boldsymbol{w}(T))$, see Remark 7.12 (5).

Due to the choice of the interpolations (7.1)–(7.3), the adjoint equation (6.2) reads

$$-A\dot{\boldsymbol{\Upsilon}}_{\mathrm{c}}^\tau - B^\star\dot{\boldsymbol{w}}_{\mathrm{c}}^\tau + \lambda_{\mathrm{d}}^\tau\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon}_{\mathrm{d}}^\tau + \theta_{\mathrm{d}}^\tau\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_{\mathrm{d}}^\tau = \boldsymbol{0}, \tag{7.22a}$$

$$-B\dot{\boldsymbol{\Upsilon}}_{\mathrm{c}}^\tau = \psi_{\mathrm{d}}^\tau(\boldsymbol{u}^\tau). \tag{7.22b}$$

Here we used the notation

$$\begin{aligned} \psi_{\mathrm{d}}^\tau(\boldsymbol{u}^\tau)(t) &= \psi_{i-1}^\tau(\boldsymbol{u}^\tau) && \text{for } t \in [(i-1)\,\tau, i\,\tau),\ i \in \{2,\ldots,N\},\\ \psi_{\mathrm{d}}^\tau(\boldsymbol{u}^\tau)(t) &= \boldsymbol{0} && \text{for } t \in [0,\tau), \end{aligned}$$

similarly to (7.2). Let us pass to the limit in (7.22b). Integration over $[t,T]$ implies

$$B(\boldsymbol{\Upsilon}_{\mathrm{c}}^\tau(t) - \boldsymbol{\Upsilon}_{\mathrm{c}}^\tau(T)) = \int_t^T \psi_{\mathrm{d}}^\tau(\boldsymbol{u}^\tau)\,\mathrm{d}t \quad \text{for all } t \in [0,T].$$

Hence, $\tau \searrow 0$, (7.19) and (7.6a) yield

$$B(\boldsymbol{\Upsilon}(t) - \boldsymbol{\Upsilon}_T) = \int_t^T \nabla\psi_c(\boldsymbol{u})\,\mathrm{d}t \quad \text{f.a.a. } t \in [0,T].$$

We show that the first three addends in (7.22a) converge weakly in adequate spaces. The convergence of the fourth addend $\theta_d^\tau\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_d^\tau$ is more delicate and is addressed afterwards.

Since $(\boldsymbol{\Upsilon}_c^\tau, \boldsymbol{w}_c^\tau)$ is bounded only in $L^\infty(0,T;S^2 \times V)$, we have to test the first two addends in (7.22b) with a differentiable test function. Let

$$\boldsymbol{T} \in W_0^{1,1}(0,T;S^2) := \{\boldsymbol{T} \in W^{1,1}(0,T;S^2) : \boldsymbol{T}(0) = \boldsymbol{0}\}$$

be given. Integration by parts implies

$$\begin{aligned}
&\int_0^T \langle A\dot{\boldsymbol{\Upsilon}}_c^\tau + B^\star\dot{\boldsymbol{w}}_c^\tau,\, \boldsymbol{T}\rangle_{S^2}\,\mathrm{d}t \\
&\qquad = -\int_0^T \langle A\boldsymbol{\Upsilon}_c^\tau + B^\star\boldsymbol{w}_c^\tau,\, \dot{\boldsymbol{T}}\rangle_{S^2}\,\mathrm{d}t + \langle A\boldsymbol{\Upsilon}_c^\tau(T) + B^\star\boldsymbol{w}_c^\tau(T),\, \boldsymbol{T}(T)\rangle_{S^2} \\
&\qquad \to -\int_0^T \langle A\boldsymbol{\Upsilon} + B^\star\boldsymbol{w},\, \dot{\boldsymbol{T}}\rangle_{S^2}\,\mathrm{d}t + \langle A\boldsymbol{\Upsilon}_T + B^\star\boldsymbol{w}_T,\, \boldsymbol{T}(T)\rangle_{S^2}.
\end{aligned}$$

In order to study the third addend in (7.22a), let us define the space

$$S_1^2 = S^2 + \{(\boldsymbol{\eta},\boldsymbol{\eta}) \in L^1(\Omega;\mathbb{S}^2) : \operatorname{trace}(\boldsymbol{\eta}) = 0\} \tag{7.23a}$$

equipped with the norm

$$\|\boldsymbol{T}\|_{S_1^2} = \inf_{\boldsymbol{T}=(\boldsymbol{\tau}+\boldsymbol{\eta},\boldsymbol{\mu}+\boldsymbol{\eta})} \|(\boldsymbol{\tau},\boldsymbol{\mu})\|_{S^2} + \|\boldsymbol{\eta}\|_{L^1(\Omega;\mathbb{S})}, \tag{7.23b}$$

where the infimum is taken over $(\boldsymbol{\tau},\boldsymbol{\mu}) \in S^2$ and $\boldsymbol{\eta} \in L^1(\Omega;\mathbb{S})$ such that $\operatorname{trace}(\boldsymbol{\eta}) = 0$. A simple calculation shows that the dual of S_1^2 is

$$S_\infty^2 = \{\boldsymbol{T} \in S^2 : \mathcal{D}\boldsymbol{T} \in L^\infty(\Omega;\mathbb{S})\} \tag{7.24a}$$

with the norm given by

$$\|\boldsymbol{T}\|_{S_\infty^2} = \max\{\|\boldsymbol{T}\|_{S^2}, \|\mathcal{D}\boldsymbol{T}\|_{L^\infty(\Omega;\mathbb{S})}\}, \tag{7.24b}$$

see also [Tartar, 2007, Lemma 41.2].

Using the convergence properties of λ_d^τ and $\boldsymbol{\Upsilon}_d^\tau$, see (7.9) and (7.11), we are led to expect $\lambda_d^\tau\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon}_d^\tau \rightharpoonup \lambda\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon}$ in $L^2(0,T;S_1^2)$. In order to prove this, we have to determine the dual of this space.

Theorem 7.4 ([Edwards, 1995, Theorem 8.20.3])**.** Let $\boldsymbol{T} \in L^2(0,T;S_1^2)'$ be given. It can be identified with a function $\boldsymbol{T} : [0,T] \to S_\infty^2$, which is weakly measurable and

$$\|\boldsymbol{T}\|_{L^2(0,T;S_1^2)'} = \Big(\int_0^T \|\boldsymbol{T}(t)\|_{S_\infty^2}^2\,\mathrm{d}t\Big)^{1/2}.$$

Note, that the measurability of $\|\boldsymbol{T}(\cdot)\|_{S_\infty^2}$ is ensured by [Edwards, 1995, Proposition 8.15.3]. The duality pairing is given by

$$\langle \boldsymbol{\Upsilon},\, \boldsymbol{T}\rangle_{L^2(0,T;S_1^2),L^2(0,T;S_1^2)'} = \int_0^T \langle \boldsymbol{\Upsilon}(t),\, \boldsymbol{T}(t)\rangle_{S_1^2,S_\infty^2}\,\mathrm{d}t$$

for all $\boldsymbol{\Upsilon} \in L^2(0,T;S_1^2)$.

Remark 7.5. In [Edwards, 1995, page 558] a function $\boldsymbol{T} : [0,T] \to S^2_\infty$ is defined to be weakly measurable if for every $\varepsilon > 0$, there is a compact set $K \subset [0,T]$ such that $\mu([0,T] \setminus K) < \varepsilon$ and $\boldsymbol{T}_{|K} : K \to S^2_\infty$ is continuous w.r.t. the weak topology of S^2_∞.

This is different from the more commonly used definition of weak measurability, which requires only $\langle f, \boldsymbol{T}(\cdot)\rangle$ to be measurable for all f in the dual of S^2_∞. Nevertheless, both concepts coincide in our situation, see [Edwards, 1995, Proposition 8.15.3].

The key issue for proving that $\langle \lambda^\tau_{\mathrm{d}}\, \mathcal{D}\boldsymbol{\Upsilon}^\tau_{\mathrm{d}},\, \mathcal{D}\boldsymbol{T}\rangle \to \langle \lambda\, \mathcal{D}\boldsymbol{\Upsilon},\, \mathcal{D}\boldsymbol{T}\rangle$ for all $\boldsymbol{T} \in L^2(0,T;S^2_1)'$ is resolved by the following lemma.

Lemma 7.6. For all $\boldsymbol{T} \in L^2(0,T;S^2_1)'$ we have $\lambda^\tau_{\mathrm{d}}\, \mathcal{D}\boldsymbol{T}, \lambda\, \mathcal{D}\boldsymbol{T} \in L^1(0,T;S)$. Moreover, $\lambda^\tau_{\mathrm{d}}\, \mathcal{D}\boldsymbol{T} \to \lambda\, \mathcal{D}\boldsymbol{T}$ in $L^1(0,T;S)$.

Proof. We sketch the main steps of the proof.

Step (1): Using the definition of weak measurability (see Remark 7.5), we infer the weak measurability of $\lambda\, \mathcal{D}\boldsymbol{T}$.

Step (2): By [Edwards, 1995, Theorem 8.15.2] (using the separability of S) we obtain the measurability of $\lambda\, \mathcal{D}\boldsymbol{T}$.

Step (3): The integrability of $\lambda\, \mathcal{D}\boldsymbol{T}$ (and hence $\lambda\, \mathcal{D}\boldsymbol{T} \in L^1(0,T;S)$) follows by the simple estimate

$$\int_0^T \|\lambda\, \mathcal{D}\boldsymbol{T}\|_S \,\mathrm{d}t \le \int_0^T \|\lambda\|_{L^2(\Omega)}\, \|\boldsymbol{T}\|_{S^2_\infty} \,\mathrm{d}t \le \|\lambda\|_{L^2(0,T;L^2(\Omega))}\, \|\boldsymbol{T}\|_{L^2(0,T;S^2_1)'}.$$

Step (4): Similar to the estimate in Step (3) we show the convergence $\lambda^\tau_{\mathrm{d}}\, \mathcal{D}\boldsymbol{T} \to \lambda\, \mathcal{D}\boldsymbol{T}$ in $L^1(0,T;S)$.

Using that the dual of $L^1(0,T;S)$ is $L^\infty(0,T;S)$, see [Diestel and Uhl, 1977, Theorem 4.1] or [Edwards, 1995, Theorem 8.18.3], and $\mathcal{D}\boldsymbol{\Upsilon}^\tau_{\mathrm{d}} \overset{\star}{\rightharpoonup} \mathcal{D}\boldsymbol{\Upsilon}$ in $L^\infty(0,T;S)$, we infer the expected weak convergence result.

Corollary 7.7. For all $\boldsymbol{T} \in L^2(0,T;S^2_1)'$ we obtain

$$\langle \lambda^\tau_{\mathrm{d}}\, \mathcal{D}\boldsymbol{\Upsilon}^\tau_{\mathrm{d}},\, \mathcal{D}\boldsymbol{T}\rangle_{L^2(0,T;S)} \to \langle \lambda\, \mathcal{D}\boldsymbol{\Upsilon},\, \mathcal{D}\boldsymbol{T}\rangle_{L^2(0,T;L^1(\Omega;\mathbb{S})),L^2(0,T;L^1(\Omega;\mathbb{S}))'}.$$

If we choose a test function $\boldsymbol{T} \in W^{1,1}_0(0,T;S^2) \cap L^2(0,T;S^2_1)'$ we can pass to the limit with the first three terms in the adjoint system (7.22a). For brevity, we define the spaces

$$\mathcal{X}(0,T) = W^{1,1}(0,T;L^2(\Omega)) \cap L^2(0,T;L^1(\Omega))',$$
$$\mathcal{X}_{S^2,0}(0,T) = W^{1,1}_0(0,T;S^2) \cap L^2(0,T;S^2_1)'.$$

The dual space of $\mathcal{X}(0,T)$ can be determined similarly to Theorem 7.4. We obtain

$$\begin{aligned}\langle -A\dot{\boldsymbol{\Upsilon}}^\tau_{\mathrm{c}} - B^\star \dot{\boldsymbol{w}}^\tau_{\mathrm{c}} + \lambda^\tau_{\mathrm{d}}\, \mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon}^\tau_{\mathrm{d}},\, \boldsymbol{T}\rangle_{L^2(0,T;S^2)} \to \\ \langle A\boldsymbol{\Upsilon} + B^\star\boldsymbol{w},\, \dot{\boldsymbol{T}}\rangle_{L^\infty(0,T;S^2),L^1(0,T;S^2)} - \langle A\boldsymbol{\Upsilon}_T + B^\star\boldsymbol{w}_T,\, \boldsymbol{T}(T)\rangle_{S^2} \\ + \langle \lambda\, \mathcal{D}\boldsymbol{\Upsilon},\, \mathcal{D}\boldsymbol{T}\rangle_{L^2(0,T;S^2_1),L^2(0,T;S^2_1)'}\end{aligned} \tag{7.25}$$

for all $\boldsymbol{T} \in \mathcal{X}_{S^2,0}(0,T)$. As an immediate consequence of (7.22a), there exists a functional $\Theta \in \mathcal{X}_{S^2,0}(0,T)'$ such that

$$\langle \theta^\tau_{\mathrm{d}}\, \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{d}},\, \boldsymbol{T}\rangle_{L^2(0,T;S^2)} \to \Theta(\boldsymbol{T}) \tag{7.26}$$

for all $\boldsymbol{T} \in \mathcal{X}_{S^2,0}(0,T)$. The next two lemmas show that $\Theta = \theta\, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}$, where θ is the weak-$\star$ limit of θ_{d}^τ in $\mathcal{X}(0,T)'$.

Lemma 7.8. Define $\theta \in \mathcal{X}(0,T)'$ by

$$2\,\tilde{\sigma}_0^2\,\theta(v) := -\Big\langle \frac{\mathrm{d}}{\mathrm{d}t}(v\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}),\, \boldsymbol{Q}\Big\rangle_{L^2(0,T;S^2)} + \big\langle v(T)\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}(T),\, \boldsymbol{Q}_T\big\rangle_{S^2} \tag{7.27}$$

for all $v \in \mathcal{X}(0,T)$. Then $\theta_{\mathrm{d}}^\tau \stackrel{\star}{\rightharpoonup} \theta$ in $\mathcal{X}(0,T)'$.

Proof. Multiplying (7.22a) with $\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_{\mathrm{d}}^\tau$ and using

$$\begin{aligned} \lambda_{\mathrm{d}}^\tau\,\mathcal{D}\boldsymbol{\Upsilon}_{\mathrm{d}}^\tau : \mathcal{D}\boldsymbol{\Sigma}_{\mathrm{d}}^\tau &= 0, && \text{by (6.4b)},\\ \theta_{\mathrm{d}}^\tau\,\mathcal{D}\boldsymbol{\Sigma}_{\mathrm{d}}^\tau : \mathcal{D}\boldsymbol{\Sigma}_{\mathrm{d}}^\tau &= \tilde{\sigma}_0^2\,\theta_{\mathrm{d}}^\tau, && \text{by (6.4c)},\\ (\mathcal{D}^\star\mathcal{D})^2 &= 2\,\mathcal{D}^\star\mathcal{D}, && \text{by definition of } \mathcal{D} \text{ and } \mathcal{D}^\star, \end{aligned}$$

yields

$$\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_{\mathrm{d}}^\tau : \dot{\boldsymbol{Q}}_{\mathrm{c}}^\tau + 2\,\tilde{\sigma}_0^2\,\theta_{\mathrm{d}}^\tau = 0 \quad \text{a.e. in } \Omega, \tag{7.28}$$

where $\boldsymbol{Q}_{\mathrm{c}}^\tau = -A\boldsymbol{\Upsilon}_{\mathrm{c}}^\tau - B^\star\boldsymbol{w}_{\mathrm{c}}^\tau$. Let $v \in \mathcal{X}(0,T)$ be given. Multiplying (7.28) with v, integrating over $(0,T)\times\Omega$ and using (7.14) we obtain

$$2\,\tilde{\sigma}_0^2\,\langle \theta_{\mathrm{d}}^\tau,\, v\rangle_{L^2((0,T)\times\Omega)} = -\big\langle v\,(\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau - \kappa^\tau\,\mathcal{D}^\star\mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau),\, \dot{\boldsymbol{Q}}_{\mathrm{c}}^\tau\big\rangle_{L^2(0,T;S^2)}, \tag{7.29}$$

where κ^τ is given by (7.15). Due to the regularity of v and using the convergence of $\boldsymbol{\Sigma}^\tau$, see (7.8), we obtain similarly to Lemma 7.6

$$v\,\mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau \to v\,\mathcal{D}\dot{\boldsymbol{\Sigma}} \quad \text{in } L^1(0,T;S^2). \tag{7.30}$$

Together with Theorem 7.2 we infer

$$\langle v\,\kappa^\tau\,\mathcal{D}^\star\mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau,\, \dot{\boldsymbol{Q}}_{\mathrm{c}}^\tau\rangle_{L^2(0,T;S^2)} \to 0$$

as $\tau \searrow 0$.

It remains to study the first addend on the right hand side of (7.29). Integration by parts yields

$$\begin{aligned} \langle v\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau,\, \dot{\boldsymbol{Q}}_{\mathrm{c}}^\tau\rangle_{L^2(0,T;S^2)} = &-\langle \dot{v}\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau + v\,\mathcal{D}^\star\mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau,\, \boldsymbol{Q}_{\mathrm{c}}^\tau\rangle_{L^2(0,T;S^2)}\\ &+ \langle v(T)\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau(T),\, \boldsymbol{Q}_{\mathrm{c}}^\tau(T)\rangle_{S^2}. \end{aligned}$$

Using $\dot{v} \in L^1(0,T;L^2(\Omega))$ and Lemma 7.3 with $\boldsymbol{T}^\tau = \boldsymbol{Q}_{\mathrm{c}}^\tau$, we obtain the convergence of the first addend of the right hand side. By (7.30) and $\boldsymbol{Q}_{\mathrm{c}}^\tau \stackrel{\star}{\rightharpoonup} \boldsymbol{Q} = -A\boldsymbol{\Upsilon} - B^\star\boldsymbol{w}$ in $L^\infty(0,T;S^2)$, we infer the convergence of the second addend. Hence,

$$\begin{aligned} \langle v\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau,\, \dot{\boldsymbol{Q}}_{\mathrm{c}}^\tau\rangle_{L^2(0,T;S^2)} \to &-\langle \dot{v}\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma} + v\,\mathcal{D}^\star\mathcal{D}\dot{\boldsymbol{\Sigma}},\, \boldsymbol{Q}\rangle_{L^2(0,T;S^2)}\\ &+ \langle v(T)\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}(T),\, \boldsymbol{Q}_T\rangle_{S^2}, \end{aligned}$$

where $\boldsymbol{Q}_T = -A\boldsymbol{\Upsilon}_T - B^\star\boldsymbol{w}_T$.

Altogether, we obtain

$$\begin{aligned} 2\,\tilde{\sigma}_0^2\,\langle \theta_{\mathrm{d}}^\tau,\, v\rangle_{L^2((0,T)\times\Omega)} \to &-\Big\langle \frac{\mathrm{d}}{\mathrm{d}t}(v\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}),\, \boldsymbol{Q}\Big\rangle_{L^2(0,T;S^2)}\\ &+ \langle v(T)\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}(T),\, \boldsymbol{Q}_T\rangle_{S^2}, \end{aligned} \tag{7.31}$$

for all $v \in \mathcal{X}(0,T)$. This shows the claim.

Lemma 7.9. For all $\boldsymbol{T} \in \mathcal{X}_{S^2,0}(0,T)$ we have

$$\theta(\mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T}) = \Theta(\boldsymbol{T}).$$

Consequently,

$$\langle A\boldsymbol{\Upsilon} + B^\star \boldsymbol{w},\, \dot{\boldsymbol{T}}\rangle_{L^\infty(0,T;S^2),L^1(0,T;S^2)} - \langle A\boldsymbol{\Upsilon}_T + B^\star \boldsymbol{w}_T,\, \boldsymbol{T}(T)\rangle_{S^2} + \langle \lambda\, \mathcal{D}\boldsymbol{\Upsilon},\, \mathcal{D}\boldsymbol{T}\rangle_{L^2(0,T;S^2_1),L^2(0,T;S^2_1)'} + \theta(\mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T}) = 0 \tag{7.32}$$

for all $\boldsymbol{T} \in \mathcal{X}_{S^2,0}(0,T)$.

Proof. Let a test function $\boldsymbol{T} \in \mathcal{X}_{S^2,0}(0,T)$ be given.

By (7.14) we have

$$\langle \theta^\tau_{\mathrm{d}}\, \mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{d}},\, \mathcal{D}\boldsymbol{T}\rangle_{L^2(0,T;S)} = \big\langle \theta^\tau_{\mathrm{d}}\, \big(\mathcal{D}\boldsymbol{\Sigma}^\tau - \kappa^\tau\, \mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau\big),\, \mathcal{D}\boldsymbol{T}\big\rangle_{L^2(0,T;S)}.$$

Using (7.28), Theorem 7.2, and Lemma 7.3 with $\boldsymbol{T}^\tau = \dot{\boldsymbol{Q}}^\tau_{\mathrm{c}}$, we obtain

$$\kappa^\tau\, \theta^\tau_{\mathrm{d}} \overset{\star}{\rightharpoonup} 0 \quad \text{in } L^\infty(0,T;L^2(\Omega)). \tag{7.33}$$

Due to $\dot{\boldsymbol{\Sigma}}^\tau \to \dot{\boldsymbol{\Sigma}}$ in $L^2(0,T;S^2)$ and $\boldsymbol{T} \in L^2(0,T;S^2_1)'$, we infer the convergence $\langle \theta^\tau_{\mathrm{d}}\, \kappa^\tau\, \mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau,\, \mathcal{D}\boldsymbol{T}\rangle_{L^2(0,T;S^2)} \to 0$ as $\tau \searrow 0$.

Using (7.28) and (7.14) we obtain

$$2\, \tilde\sigma_0^2\, \langle \theta^\tau_{\mathrm{d}}\, \mathcal{D}\boldsymbol{\Sigma}^\tau,\, \mathcal{D}\boldsymbol{T}\rangle_{L^2(0,T;S)} = -\langle (\mathcal{D}\boldsymbol{\Sigma}^\tau : \mathcal{D}\boldsymbol{T})\, (\mathcal{D}\boldsymbol{\Sigma}^\tau - \kappa^\tau\, \mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau),\, \mathcal{D}\dot{\boldsymbol{Q}}^\tau_{\mathrm{c}}\rangle_{L^2(0,T;S)}.$$

By Theorem 7.2 and Lemma 7.3 with $f^\tau = \kappa^\tau\, \mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau : \mathcal{D}\dot{\boldsymbol{Q}}^\tau_{\mathrm{c}}$ we obtain

$$\langle (\mathcal{D}\boldsymbol{\Sigma}^\tau : \mathcal{D}\boldsymbol{T})\, \kappa^\tau\, \mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau,\, \mathcal{D}\dot{\boldsymbol{Q}}^\tau_{\mathrm{c}}\rangle_{L^2(0,T;S)} \to 0$$

as $\tau \searrow 0$. Integration by parts implies

$$-\langle (\mathcal{D}\boldsymbol{\Sigma}^\tau : \mathcal{D}\boldsymbol{T})\, \mathcal{D}\boldsymbol{\Sigma}^\tau,\, \mathcal{D}\dot{\boldsymbol{Q}}^\tau_{\mathrm{c}}\rangle_{L^2(0,T;S)} = \Big\langle \frac{\mathrm{d}}{\mathrm{d}t}\big((\mathcal{D}\boldsymbol{\Sigma}^\tau : \mathcal{D}\boldsymbol{T})\, \mathcal{D}\boldsymbol{\Sigma}^\tau\big),\, \mathcal{D}\boldsymbol{Q}^\tau_{\mathrm{c}}\Big\rangle_{L^2(0,T;S)} - \big\langle (\mathcal{D}\boldsymbol{\Sigma}^\tau(T) : \mathcal{D}\boldsymbol{T}(T))\, \mathcal{D}\boldsymbol{\Sigma}^\tau(T),\, \mathcal{D}\boldsymbol{Q}^\tau_{\mathrm{c}}(T)\big\rangle_S.$$

Using the chain rule and applying Lemma 7.3 thrice (with $\boldsymbol{T}^\tau = \boldsymbol{Q}^\tau_{\mathrm{c}}$, $g^\tau = \mathcal{D}\boldsymbol{\Sigma}^\tau : \mathcal{D}\boldsymbol{Q}^\tau_{\mathrm{c}}$ and $f^\tau = \mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau : \mathcal{D}\boldsymbol{Q}^\tau_{\mathrm{c}}$) we obtain

$$\Big\langle \frac{\mathrm{d}}{\mathrm{d}t}\big((\mathcal{D}\boldsymbol{\Sigma}^\tau : \mathcal{D}\boldsymbol{T})\, \mathcal{D}\boldsymbol{\Sigma}^\tau\big),\, \mathcal{D}\boldsymbol{Q}^\tau_{\mathrm{c}}\Big\rangle_{L^2(0,T;S)} \to \Big\langle \frac{\mathrm{d}}{\mathrm{d}t}\big((\mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T})\, \mathcal{D}\boldsymbol{\Sigma}\big),\, \mathcal{D}\boldsymbol{Q}\Big\rangle_{L^2(0,T;S)}.$$

Putting everything together, we obtain

$$\begin{aligned} 2\, \tilde\sigma_0^2\, \langle \theta^\tau_{\mathrm{d}}\, \mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{d}},\, \mathcal{D}\boldsymbol{T}\rangle_{L^2(0,T;S^2)} \to{}& \Big\langle \frac{\mathrm{d}}{\mathrm{d}t}\big((\mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T})\, \mathcal{D}\boldsymbol{\Sigma}\big),\, \mathcal{D}\boldsymbol{Q}\Big\rangle_{L^2(0,T;S^2)} \\ &- \Big\langle \big((\mathcal{D}\boldsymbol{\Sigma}(T) : \mathcal{D}\boldsymbol{T}(T))\, \mathcal{D}\boldsymbol{\Sigma}(T)\big),\, \mathcal{D}\boldsymbol{Q}_T\Big\rangle_{S^2} \\ ={}& 2\, \tilde\sigma_0^2\, \theta(\mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T}) \end{aligned}$$

for all $\boldsymbol{T} \in \mathcal{X}_{S^2,0}(0,T)$, see (7.27). Together with (7.25), (7.26), (7.31), (7.27) this shows the claim.

7.5. Weak stationarity. First, we pass to the limit in the complementarity conditions (6.4b) and (6.4c).

Lemma 7.10. We obtain

$$\lambda\, \mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{\Upsilon} = 0 \quad \text{a.e. in } (0,T) \times \Omega, \tag{7.34a}$$

$$\theta(v\, \phi(\boldsymbol{\Sigma})) = 0 \quad \text{for all } v \in \mathcal{X}(0,T). \tag{7.34b}$$

Proof. Similarly as in Lemma 7.3, we obtain

$$\lambda^\tau_{\mathrm{d}}\, \mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{d}} : \mathcal{D}\boldsymbol{\Upsilon}^\tau_{\mathrm{d}} \rightharpoonup \lambda\, \mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{\Upsilon} \quad \text{in } L^2((0,T);L^1(\Omega)).$$

Since the set $\{\mathbf{0}\}$ is weakly closed in $L^2((0,T);L^1(\Omega))$, (6.4b) implies (7.34a).

Testing (6.4c) with $v \in C^1((0,T)\times\Omega)$, we obtain

$$\langle \theta_{\mathrm{d}}^\tau\, \phi(\boldsymbol{\Sigma}_{\mathrm{d}}^\tau),\, v\rangle_{L^2((0,T)\times\Omega)} = 0.$$

We have by (7.14)

$$\mathcal{D}\boldsymbol{\Sigma}_{\mathrm{d}}^\tau : \mathcal{D}\boldsymbol{\Sigma}_{\mathrm{d}}^\tau = \mathcal{D}\boldsymbol{\Sigma}^\tau : \mathcal{D}\boldsymbol{\Sigma}^\tau - \kappa^\tau\, \mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau : (\mathcal{D}\boldsymbol{\Sigma}^\tau + \mathcal{D}\boldsymbol{\Sigma}_{\mathrm{d}}^\tau).$$

Using (7.33) and Lemma 7.3 with $g^\tau = v\,\kappa^\tau\,\theta_{\mathrm{d}}^\tau$, we obtain

$$\langle \mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau : (\mathcal{D}\boldsymbol{\Sigma}^\tau + \mathcal{D}\boldsymbol{\Sigma}_{\mathrm{d}}^\tau),\, v\,\kappa^\tau\,\theta_{\mathrm{d}}^\tau\rangle_{L^2((0,T)\times\Omega)} \to 0.$$

Hence, it remains to study the convergence of

$$\langle \theta_{\mathrm{d}}^\tau\, \phi(\boldsymbol{\Sigma}^\tau),\, v\rangle_{L^2((0,T)\times\Omega)}.$$

Using an argument similar to those in the proof of Lemma 7.9, we obtain

$$\langle \theta_{\mathrm{d}}^\tau\, \phi(\boldsymbol{\Sigma}^\tau),\, v\rangle_{L^2((0,T)\times\Omega)} \to \theta(v\,\phi(\boldsymbol{\Sigma})).$$

Hence, (7.34b) holds for all $v \in C^1((0,T)\times\Omega)$. A density argument finishes the proof.

The following theorem summarizes the results of this section.

Theorem 7.11. Let $\boldsymbol{g} \in H^1(0,T;U)$ be a local minimum of the optimal control problem **(P)**. There are

$$(\boldsymbol{\Sigma}, \boldsymbol{u}) \in H^1(0,T;S^2\times V) \qquad \lambda \in L^2(0,T;L^2(\Omega))$$
$$(\boldsymbol{\Upsilon}, \boldsymbol{w}) \in L^\infty(0,T;S^2\times V) \qquad (\theta,\mu) \in \mathcal{X}(0,T)' \times L^\infty(0,T;L^2(\Omega))$$
$$(\boldsymbol{\Upsilon}_T, \boldsymbol{w}_T) \in S^2\times V \qquad (\theta_T,\mu_T) \in L^2(\Omega)\times L^2(\Omega)$$

satisfying

$$A\dot{\boldsymbol{\Sigma}} + B^\star\dot{\boldsymbol{u}} + \lambda\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma} = \mathbf{0}, \tag{7.35a}$$
$$B\dot{\boldsymbol{\Sigma}} = E\dot{\boldsymbol{g}}, \tag{7.35b}$$
$$0 \le \lambda \quad \perp \quad \phi(\boldsymbol{\Sigma}) \le 0, \tag{7.35c}$$

$$\begin{aligned}\langle A\boldsymbol{\Upsilon} + B^\star\boldsymbol{w},\, \dot{\boldsymbol{T}}\rangle_{L^\infty(0,T;S^2),L^1(0,T;S^2)} &- \langle A\boldsymbol{\Upsilon}_T + B^\star\boldsymbol{w}_T,\, \boldsymbol{T}(T)\rangle_{S^2}\\ +\langle \lambda\,\mathcal{D}\boldsymbol{\Upsilon},\, \mathcal{D}\boldsymbol{T}\rangle_{L^2(0,T;S^2_1),L^2(0,T;S^2_1)'} &+ \theta(\mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T}) = \mathbf{0},\end{aligned} \tag{7.36a}$$
$$B(\boldsymbol{\Upsilon} - \boldsymbol{\Upsilon}_T) - \int_\cdot^T \nabla\psi_{\mathrm{c}}(\boldsymbol{u})\,\mathrm{d}s = \mathbf{0}, \tag{7.36b}$$

$$\big\langle E^\star\boldsymbol{w},\, \dot{\tilde{\boldsymbol{g}}} - \dot{\boldsymbol{g}}\big\rangle_{L^2(0,T;U)} + \langle \nu\,\boldsymbol{g},\, \tilde{\boldsymbol{g}} - \boldsymbol{g}\rangle_{H^1(0,T;U)} \ge 0, \tag{7.37}$$

$$\mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{\Upsilon} - \mu = 0, \tag{7.38a}$$
$$\mu\,\lambda = 0, \tag{7.38b}$$
$$\theta(v\,\phi(\boldsymbol{\Sigma})) = 0, \tag{7.38c}$$

$$A\boldsymbol{\Upsilon}_T + \theta_T\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}(T) + B^\star\boldsymbol{w}_T = \mathbf{0}, \tag{7.39a}$$
$$B\boldsymbol{\Upsilon}_T - \psi_T'(\boldsymbol{u}(T)) = \mathbf{0}, \tag{7.39b}$$

$$\mathcal{D}\boldsymbol{\Sigma}(T) : \mathcal{D}\boldsymbol{\Upsilon}_T - \mu_T = 0, \tag{7.40a}$$
$$\theta_T\,\phi(\boldsymbol{\Sigma}(T)) = 0, \tag{7.40b}$$
$$\theta_T\,\mu_T \ge 0, \tag{7.40c}$$

for all $\boldsymbol{T} \in \mathcal{X}_{S^2,0}(0,T)$, $\tilde{\boldsymbol{g}} \in U_{\mathrm{ad}}$ and $v \in \mathcal{X}(0,T)$.

Remark 7.12.

(1) Following the notation for finite-dimensional MPECs, see Scheel and Scholtes [2000], Kanzow and Schwartz [2010], the optimality system (7.35)–(7.40) is of weak-stationary type.
(2) In a system of C-stationary type, the product of the multipliers μ and θ is required to be non-negative. Due to the low regularity of θ, however, the product $\theta\,\mu$ cannot be defined.
(3) Similarly to optimal control problems involving state constraints, one can construct examples where the multiplier θ is not a function. Its low regularity is induced by the constraint $\phi(\boldsymbol{\Sigma}) \le 0$.
(4) The remarks (2) and (3) above also apply to optimal control of parabolic VIs, see e.g. the optimality system in [Ito and Kunisch, 2010, Theorem 6.2].
(5) Using (7.27) and (7.36), it is easy to prove that $\theta \in L^2(0,T;L^2(\Omega))$ if and only if $(\boldsymbol{\Upsilon}, \boldsymbol{w}) \in H^1(0,T;S^2 \times V)$, $(\boldsymbol{\Upsilon}(T), \boldsymbol{w}(T)) = (\boldsymbol{\Upsilon}_T, \boldsymbol{w}_T)$, and $\lambda\,\mathcal{D}\boldsymbol{\Upsilon} \in L^2(0,T;S)$. Hence, the low regularity of θ is directly related to the non-differentiability of $(\boldsymbol{\Upsilon}, \boldsymbol{w})$ as functions of time.
(6) The equations (7.36) and (7.39) can be stated equivalently as

$$\langle A\boldsymbol{\Upsilon} + B^\star\boldsymbol{w},\, \dot{\boldsymbol{T}}\rangle_{L^\infty(0,T;S^2),L^1(0,T;S^2)} + \langle\lambda\,\mathcal{D}\boldsymbol{\Upsilon},\, \mathcal{D}\boldsymbol{T}\rangle_{L^2(0,T;S^2_1),L^2(0,T;S^2_1)'} + \tilde\theta(\mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T}) = \mathbf{0}, \tag{7.41a}$$

$$B\boldsymbol{\Upsilon} - \psi_T'(\boldsymbol{u}(T)) - \int_\cdot^T \nabla\psi_c(\boldsymbol{u})\,\mathrm{d}s = \mathbf{0}, \tag{7.41b}$$

with $\tilde\theta(v) = \theta(v) + \langle\theta_T,\, v(T)\rangle_{L^2(\Omega)}$ for all functions $v \in \mathcal{X}(0,T)$.

We prefer (7.36) with terminal conditions (7.39) over (7.41) since the former more clearly show the conditions at time T.
(7) There are two contributions to the terminal condition (7.39). The term $\psi_T'(\boldsymbol{u}(T))$ is induced by the observation $\psi_T(\boldsymbol{u}(T))$ at final time in the objective. This is typical for optimal control problems with differential equations.

The term $\theta_T\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}(T)$ can be understood as a Lagrange multiplier to the constraint $\phi(\boldsymbol{\Sigma}(T)) \le 0$ in the state equation (7.35). In fact, similar terms appear also in the adjoint equation (7.36a) at times $t \in (0,T)$ where θ has Dirac contributions.

B 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 [Herzog et al., 2010, Section A]. Throughout this section we consider the abstract system

$$A\,\boldsymbol{\sigma} + J(\boldsymbol{\sigma} + \mathscr{L}_2) + B^\star\boldsymbol{u} = \mathscr{L}_1, \tag{B.1a}$$

$$B\,\boldsymbol{\sigma} = \ell, \tag{B.1b}$$

where $(\mathscr{L}_1, \mathscr{L}_2, \ell)$ is the data and $(\boldsymbol{\sigma}, \boldsymbol{u})$ is the solution. We will show in Theorem B.6 that the solution map of (B.1) is Fréchet differentiable under certain assumptions. The functional analytic setting is made precise in the following assumption.

Assumption B.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.

$$\|B^\star \boldsymbol{u}\|_X \ge \beta \, \|\boldsymbol{u}\|_V \quad \text{for all } \boldsymbol{u} \in V. \tag{B.2}$$

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

It is a standard result that the system (B.1) 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 B.1 is satisfied. A proof can be found in [Herzog et al., 2010, Lemma A.1].

Lemma B.2 (Nonlinear saddle point problem)**.** Let Assumption B.1 be satisfied. Then, for all $(\mathscr{L}_1, \mathscr{L}_2, \ell) \in X \times X \times V'$, the system (B.1) 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 (B.1). 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 Nemytzki operator, see [Tröltzsch, 2010, Section 4.3.2] or Krasnoselskii et al. [1976].

Assumption B.3 (Fréchet differentiability of J)**.** In addition to Assumption B.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 B.3, we state the linearization of (B.1),

$$(A + J'(\boldsymbol{\sigma} + \mathscr{L}_2)) \, \delta\boldsymbol{\sigma} + B^\star \delta\boldsymbol{u} = \delta\mathscr{L}, \tag{B.3a}$$

$$B \, \delta\boldsymbol{\sigma} = \delta\ell. \tag{B.3b}$$

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

Lemma B.4 (Solvability of the linearized problem)**.** Let Assumption B.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 (B.3) 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 B.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 B.5 (Lipschitz continuity in stronger norms)**.** In addition to Assumption B.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 (B.1) 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 (B.3) 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 (B.1) 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 (B.3) with $\delta\mathscr{L} = \delta\mathscr{L}_1 - J'(\boldsymbol{\sigma} + \mathscr{L}_2) \, \delta\mathscr{L}_2$.

Theorem B.6 (Differentiability)**.** Let Assumption B.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 (B.3), 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

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

By definition of $\tilde{G}(\boldsymbol{\sigma}, \mathscr{L}_2)$, see Lemma B.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 B.6 does not simply follow from the implicit function theorem. In order to apply the implicit function theorem to (B.1), 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 B.6 in Section 5.2.

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

D. R. Adams and S. Lenhart. Optimal control of the obstacle for a parabolic variational inequality. *Journal of Mathematical Analysis and Applications*, 268 (2):602 – 614, 2002. doi: 10.1006/jmaa.2001.7833.

V. Barbu. Necessary conditions for distributed control problems governed by parabolic variational inequalities. *SIAM Journal on Control and Optimization*, 19 (1):64–86, 1981. ISSN 0363-0129. doi: 10.1137/0319006.

M. Brokate. *Optimale Steuerung von gewöhnlichen Differentialgleichungen mit Nichtlinearitäten vom Hysteresis-Typ*. Number 35 in Methoden und Verfahren der mathematischen Physik. Verlag Peter Lang, Frankfurt, 1987.

E. Casas and F. Tröltzsch. Second order necessary and sufficient conditions for optimization problems and applications to control theory. *SIAM Journal on Optimization*, 13:406–431, 2002.

J. Diestel and J. J. Uhl. *Vector Measures*. Mathematical Surveys and Monographs. American Mathematical Society, Providence, 1977.

R. E. Edwards. *Functional Analysis*. Dover Publications, 1995.

M. H. Farshbaf-Shaker. A penalty approach to optimal control of allen-cahn variational inequalities: MPEC-view. *submitted*, 2011.

R. Herzog, C. Meyer, and G. Wachsmuth. C-stationarity for optimal control of static plasticity with linear kinematic hardening. *submitted*, 2010.

R. Herzog, C. Meyer, and G. Wachsmuth. Integrability of displacement and stresses in linear and nonlinear elasticity with mixed boundary conditions. *Journal of Mathematical Analysis and Applications*, 382(2):802–813, 2011a. doi: 10.1016/j.jmaa.2011.04.074.

R. Herzog, C. Meyer, and G. Wachsmuth. Existence and regularity of the plastic multiplier in static and quasistatic plasticity. *GAMM Reports*, 34(1):39–44, 2011b. doi: 10.1002/gamm.201110006.

R. Herzog, C. Meyer, and G. Wachsmuth. B- and strong stationarity for optimal control of static plasticity with hardening. *in revision*, 2011c.

M. Hintermüller and D. Wegner. Distributed optimal control of the Cahn-Hilliard system including the case of a double-obstacle homogeneous free energy density. IFB-Report No. 43, Institute of Mathematics and Scientific Computing, University of Graz, 2011.

K. Ito and K. Kunisch. Optimal control of parabolic variational inequalities. *Journal de Mathématiques Pures et Appliqués*, 93(4):329–360, 2010. doi: 10.1016/j.matpur.2009.10.005.

C. Kanzow and A. Schwartz. Mathematical programs with equilibrium constraints: enhanced Fritz John-conditions, new constraint qualifications, and improved exact penalty results. *SIAM Journal on Optimization*, 20(5):2730–2753, 2010. ISSN 1052-6234. doi: 10.1137/090774975.

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

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

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

S. Steffensen and M. Ulbrich. A new relaxation scheme for mathematical programs with equilibrium constraints. *SIAM Journal on Optimization*, 20(5):2504–2539, 2010. doi: 10.1137/090748883.

L. Tartar. *An Introduction to Sobolev Spaces and Interpolation Spaces.* Lecture Notes of the Unione Matematica Italiana. Springer, 2007.

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

Chemnitz University of Technology, Faculty of Mathematics, D–09107 Chemnitz, Germany

E-mail address: gerd.wachsmuth@mathematik.tu-chemnitz.de

URL: http://www.tu-chemnitz.de/mathematik/part_dgl/people/wachsmuth