Zeitschrift für Analysis und ihre Anwendungen
Journal of Analysis and its Applications
Volume 35 (2016), 81–118
DOI: 10.4171/ZAA/1556

Optimal Control of Quasistatic Plasticity with Linear Kinematic Hardening III: Optimality Conditions

Gerd Wachsmuth

Abstract. In this paper we consider an optimal control problem governed by a rate-independent variational inequality arising in quasistatic plasticity with linear kinematic hardening. Since the solution operator of a variational inequality is not differentiable, the Karush-Kuhn-Tucker system is not a necessary optimality condition. We show a system of weakly stationary type by passing to the limit with the optimality system of a regularized and time-discretized problem.

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

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

1. Introduction

In this paper we prove a necessary optimality system for an optimal control problem governed by the quasistatic forward problem of small-strain elastoplasticity. 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.

As a particular problem, we mention

$$\begin{aligned} \text{Minimize} \quad & F(\boldsymbol{u}, \boldsymbol{g}) = \|\boldsymbol{u}(T) - \boldsymbol{u}_d\|_{L^2(\Omega;\mathbb{R}^d)} + \frac{\nu}{2}\|\boldsymbol{g}\|^2_{H^1(0,T;L^2(\Gamma_N;\mathbb{R}^d))}, \\ \text{with respect to} \quad & \boldsymbol{\Sigma}, \boldsymbol{u}, \boldsymbol{g} \\ \text{such that} \quad & (\boldsymbol{\Sigma}, \boldsymbol{u}) = \mathcal{G}(E\boldsymbol{g}) \quad \text{and} \quad \boldsymbol{g}(0) = \boldsymbol{g}(T) = \mathbf{0}. \end{aligned}$$

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

Here, $\mathcal{G}$ is the solution map of the quasistatic forward problem and E is the control operator. The definition of the forward problem needs some notation and is done in Section 1.2. The constraint $\boldsymbol{g}(T) = \mathbf{0}$ implies that the body at the final time T is unloaded. Due to the observation of the final displacement $\boldsymbol{u}(T)$ in the objective, this combination of objective and control constraints corresponds to controlling the springback of the solid body.

The forward system in the stress-based (so-called dual) form is represented by a time-dependent, rate-independent variational inequality (VI) of mixed type, see Section 1.2. Hence, the control-to-state map is not, in general, differentiable. Moreover, it is already know from finite-dimensional problems, that the associated Karush-Kuhn-Tucker (KKT) system is not a necessary optimality system for optimization problems constrained by a VI. Therefore, one considers regularizations of VIs, see [3]. The main contribution of this paper is Theorem 3.1, in which we provide an optimality system for the optimal control problem under consideration. To our knowledge, necessary optimality systems for the control of rate-independent VIs in function space are not known up to now.

A regularized, time-discrete approximation of the forward problem is the subject of [29]. There, the author proved the Fréchet differentiability of the solution map of the regularized forward problem which implies a first order necessary optimality condition for the regularized optimal control problem. Based on this result, we are going to prove an optimality system (of weakly stationary type) for the unregularized optimal control problem by passing to the limit in the optimality system of the regularized optimal control problem. In particular, passing to the limit with the time discretization parameter τ requires some new and subtle arguments, see Section 3.

For the notions of the various optimality systems, we refer to [25, Section 2]. Let us put our work into perspective. We give some references for optimal control of *time-dependent* VIs. We mention [1,2,19,21], which deal with optimal control of a *parabolic obstacle problem.* Moreover, [8] and [18] 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. We also mention [22, 23] who studies the optimal control of rate-independent evolution processes in a general setting. The existence of an optimal control and the approximability by solutions of discretized problems is shown, but no optimality conditions are given.

Let us briefly highlight the main contributions of some of these references. In [21] 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 [17, Section 4]. To our knowledge, there are no results on optimality systems of C-stationary type for optimal control problems governed by time-dependent VIs. In [19] 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.

We also mention [16, 17], which considered the optimal control of *static* plasticity. For locally optimal controls, systems of B- and C-stationary type were obtained.

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 proper function, but a measure in time. Moreover, in both problems the adjoint states (in our notation $\boldsymbol{\Upsilon}$ and $\boldsymbol{w}$) 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 [19] for control of the parabolic obstacle problem. Moreover, the analysis is rendered more challenging due to the nonlinearity in the set $\mathcal{K}$, see (2), and due to the constraint equation (equilibrium of forces) in the VI. Another difficulty arises from the fact that there seem to be no existence results for regularized versions of the time-dependent variational inequality. Therefore, it is more convenient to regularize the discretization in time 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, see [29, Section 3].

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 [2, 21] derive necessary conditions in this sense for time-dependent VIs. We utilize the approximation results of [28, Section 3.4] and [29, Section 4.2] in order to show that the derived optimality system (33)–(38) holds for all local minimizers.

Let us sketch the outline of the paper. In the remainder of the introduction, we fix the notation (Section 1.1), and state the forward and optimal control problems together with their regularizations (Sections 1.2 and 1.3). Section 2 is devoted to showing an optimality system for the time-discrete optimal control problem ($\mathbf{P}^\tau$) by passing to the limit with the optimality conditions for the regularization ($\mathbf{P}^\varepsilon$). In Section 3 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 3.8 and 3.9. We finally arrive at the optimality system of weakly stationary type, see Theorem 3.1.

1.1. Notation and assumptions. Our notation follows [13] and [16].

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

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

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

as the spaces for the displacement $\boldsymbol{u}$, stress $\boldsymbol{\sigma}$, and back stress $\boldsymbol{\chi}$, respectively. The control $\boldsymbol{g}$ belongs to the space of boundary forces

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

The control operator $E : U \to V'$, $\boldsymbol{g} \mapsto \ell$, which maps boundary forces (i.e., controls) $\boldsymbol{g} \in U$ to functionals $\ell \in V'$ (i.e., right-hand sides of the weak formulation (9)) is given by

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

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

For a Banach space X and $p \in [1,\infty]$, we define the Bochner-Lebesgue space

$$L^p(0,T;X) = \{u : [0,T] \to X,\ u \text{ is Bochner measurable and } p\text{-integrable}\}.$$

In the case $p = \infty$ one has to replace p-integrability by essential boundedness. The norm in $L^p(0,T;X)$ is given by

$$\|u\|_{L^p(0,T;X)} = \big\|\|u(\cdot)\|_X\big\|_{L^p(0,T)}.$$

By $W^{1,p}(0,T;X)$ we denote the Bochner-Sobolev space consisting of functions $u \in L^p(0,T;X)$ which possess a weak derivative $\dot{u} \in L^p(0,T;X)$. Two equivalent norms on $W^{1,p}(0,T;X)$ are given by

$$\big(\|u\|^p_{L^p(0,T;X)} + \|\dot{u}\|^p_{L^p(0,T;X)}\big)^{\frac{1}{p}} \quad \text{and} \quad \big(\|u(0)\|^p_X + \|\dot{u}\|^p_{L^p(0,T;X)}\big)^{\frac{1}{p}},$$

where the extension to the case $p = \infty$ is clear. We use $H^1(0,T;X) = W^{1,2}(0,T;X)$. Moreover, we define the space of functions in $H^1(0,T;X)$ vanishing at $t = 0$

$$H^1_{\{0\}}(0,T;X) = \{u \in H^1(0,T;X) : u(0) = 0\}.$$

Details on Bochner-Lebesgue and Bochner-Sobolev spaces can be found in [6, 9, 24, 30].

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

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

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

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

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

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

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

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

This property is exploited quite often in the analysis.

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

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

for matrices $\boldsymbol{\Sigma} \in \mathbb{S}^2$ as well as for functions $\boldsymbol{\Sigma} \in S^2$ and $\boldsymbol{\Sigma} \in L^p(0,T;S^2)$. When considered as an operator in function space, $\mathcal{D}$ maps S^2 and $L^p(0,T;S^2)$ continuously into S and $L^p(0,T;S)$, respectively. For later reference, we also remark that

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

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

$$\mathbf{\Sigma} \in \mathcal{K} \quad \Rightarrow \quad \mathcal{D}\mathbf{\Sigma} \in L^\infty(\Omega;\mathbb{S}). \tag{3}$$

Here and in the sequel we denote linear operators, e.g. $\mathcal{D} : S^2 \to S$, and the induced Nemytskii operators, e.g. $\mathcal{D} : H^1(0,T;S^2) \to H^1(0,T;S)$ and $\mathcal{D} : L^2(0,T;S^2) \to L^2(0,T;S)$, with the same symbol. This will cause no confusion, since the meaning will be clear from the context.

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

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

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

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

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

Standing assumptions. Throughout the paper, we require

Assumption 1.1.

(1) The domain $\Omega \subset \mathbb{R}^d$, $d = 3$ is a bounded Lipschitz domain in the sense of [10, Chapter 1.2]. The boundary of Ω, denoted by Γ, consists of two disjoint measurable parts Γ_N and Γ_D such that $\Gamma = \Gamma_N \cup \Gamma_D$. While Γ_N is a relatively open subset, Γ_D is a relatively closed subset of Γ. Furthermore Γ_D is assumed to have positive measure. In addition, the set $\Omega \cup \Gamma_N$ is regular in the sense of Gröger, cf. [11]. A characterization of regular domains for the case $d \in \{2, 3\}$ can be found in [12, Section 5]. This class of domains covers a wide range of geometries.

(2) The yield stress $\tilde\sigma_0$ is assumed to be a positive constant. It equals $\sqrt{\frac{2}{3}}\,\sigma_0$, where σ_0 is the uni-axial yield stress.

(3) $\mathbb{C}^{-1}$ is a uniformly coercive element of $L^\infty(\Omega; \mathcal{L}(\mathbb{S},\mathbb{S}))$, where $\mathcal{L}(\mathbb{S},\mathbb{S})$ denotes the space of linear operators $\mathbb{S} \to \mathbb{S}$. Moreover, we assume that $\mathbb{C}^{-1}(x)$ is symmetric, i.e., $\boldsymbol{\tau} : \mathbb{C}^{-1}(x)\,\boldsymbol{\sigma} = \boldsymbol{\sigma} : \mathbb{C}^{-1}(x)\,\boldsymbol{\tau}$.

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

Assumption 1.1(1) enables us to apply the regularity results in [15] pertaining to systems of nonlinear elasticity. The latter appear in the time-discrete forward problem and its regularizations. Additional regularity leads to a norm gap, which is needed to prove the differentiability of the control-to-state map.

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

$$\|\boldsymbol{u}\|^2_{H^1(\Omega;\mathbb{R}^d)} \le c_K \left(\|\boldsymbol{u}\|^2_{L^2(\Gamma_D;\mathbb{R}^d)} + \|\boldsymbol{\varepsilon}(\boldsymbol{u})\|^2_S\right) \tag{4}$$

for all $\boldsymbol{u} \in H^1(\Omega;\mathbb{R}^d)$, see e.g. [15, Lemma C.1]. Note that (4) entails in particular that $\|\boldsymbol{\varepsilon}(\boldsymbol{u})\|_S$ is a norm on $H^1_D(\Omega;\mathbb{R}^d)$ equivalent to the standard $H^1(\Omega;\mathbb{R}^d)$ norm. A further consequence is that $B^\star$ satisfies the inf-sup condition

$$\|\boldsymbol{u}\|_V \le \sqrt{c_K}\, \|B^\star \boldsymbol{u}\|_{S^2} \quad \text{for all } \boldsymbol{u} \in V. \tag{5}$$

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

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

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

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

Regularization. For the regularized problem, we need a regularization of the function $\max\{0,\cdot\}$.

Assumption 1.2. 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. Since the term appearing inside $\max^\varepsilon$ will be smaller than 1, we will assume $\varepsilon \in (0, 1)$, see (12c).

Interpolation of time-discrete functions. Let $f^\tau \in X^N$ be given, where X is some Banach space and N is the number of time steps. We define certain interpolants of f^τ which will be useful for defining the time-discrete problem as well as for passing to the limit with the time step size.

We define the piecewise linear and *continuous* interpolant $f^\tau_{\mathrm{c,p}}$ which will be used for the *primal* variables $\boldsymbol{\Sigma}$, $\boldsymbol{u}$, and $\boldsymbol{g}$

$$f^\tau_{\mathrm{c,p}}(t) = \frac{t - (i-1)\,\tau}{\tau}\, f^\tau_i + \frac{i\,\tau - t}{\tau}\, f^\tau_{i-1} \quad \text{for } t \in [(i-1)\,\tau, i\,\tau), \tag{6}$$

with $f_0^\tau = 0$. Therefore, we can identify X^N with a subspace of $H^1_{\{0\}}(0,T;X)$. Note that this interpolation coincides with the one given in [28, (3.1)].

The piecewise linear and *continuous* interpolant $f^\tau_{\mathrm{c,d}}$ will be used for the *dual* variables $\boldsymbol{\Upsilon}$ and $\boldsymbol{w}$ and is defined as

$$f^\tau_{\mathrm{c,d}}(t) = \frac{t-(i-1)\,\tau}{\tau}\, f_i^\tau + \frac{i\,\tau - t}{\tau}\, f_{i-1}^\tau \quad \text{for } t \in [(i-1)\,\tau, i\,\tau), \tag{7}$$

with $f_0^\tau = f_1^\tau$. Compared with the definition of $f^\tau_{\mathrm{c,p}}$, only the fictitious value of f_0^τ was changed. Due to this choice of the initial value, the adjoint equation (52a) is not only satisfied in the interval (τ, T) but also in $(0,\tau)$, see Section 3.3.

Moreover, we define the piecewise *constant* interpolations $f^\tau_{\mathrm{d}+}$ and $f^\tau_{\mathrm{d}-}$ by

$$f^\tau_{\mathrm{d}+}(t) = f_i^\tau \qquad \text{for } t \in [(i-1)\,\tau, i\,\tau), \tag{8a}$$

$$f^\tau_{\mathrm{d}-}(t) = f_{i-1}^\tau \quad \text{for } t \in [(i-1)\,\tau, i\,\tau), \tag{8b}$$

with the convention $f_0^\tau = 0$. The interpolant $f^\tau_{\mathrm{d}+}$ will be used for the adjoint displacement $\boldsymbol{w}$ in the gradient equation, see (45), whereas $f^\tau_{\mathrm{d}-}$ will be used for several quantities in the adjoint system, see (52). Note that the interpolant $f^\tau_{\mathrm{d}+}$ coincides with the one given in [28, (3.3)].

1.2. The forward problem. Now, we are in the position to state the forward problem of quasistatic plasticity. Given $\ell \in H^1_{\{0\}}(0,T;V')$, one has to find generalized stresses $\boldsymbol{\Sigma} \in H^1_{\{0\}}(0,T;S^2)$ and displacements $\boldsymbol{u} \in H^1_{\{0\}}(0,T;V)$ which satisfy $\boldsymbol{\Sigma}(t) \in \mathcal{K}$ and

$$\langle A\dot{\boldsymbol{\Sigma}}(t) + B^\star \dot{\boldsymbol{u}}(t), \boldsymbol{T} - \boldsymbol{\Sigma}(t)\rangle_{S^2} \ge 0 \qquad \text{for all } \boldsymbol{T} \in \mathcal{K}, \tag{9a}$$

$$B\boldsymbol{\Sigma}(t) = \ell(t) \quad \text{in } V' \tag{9b}$$

for almost all $t \in (0,T)$. The unique solvability of (9) is shown in [13, Theorem 8.12], see also [28, (1.19)] for the uniqueness of the displacement $\boldsymbol{u}$ in case of linear kinematic hardening. We denote the solution operator which maps $\ell \mapsto (\boldsymbol{\Sigma}, \boldsymbol{u})$ by $\mathcal{G}$. For continuity properties of the solution map $\mathcal{G}$ we refer to [28, Section 2]. Equivalently, by introducing a Lagrange multiplier λ associated with the constraint $\phi(\boldsymbol{\Sigma}) \le 0$, the system (9) can be written as

$$A\dot{\boldsymbol{\Sigma}} + B^\star \dot{\boldsymbol{u}} + \lambda\, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma} = \boldsymbol{0} \quad \text{in } L^2(0,T;S^2), \tag{10a}$$

$$B\boldsymbol{\Sigma} = \ell \quad \text{in } L^2(0,T;V'), \tag{10b}$$

$$0 \le \lambda \quad \perp \quad \phi(\boldsymbol{\Sigma}) \le 0 \quad \text{a.e. in } (0,T)\times\Omega, \tag{10c}$$

see [14]. As usual, $0 \le \lambda \perp \phi(\boldsymbol{\Sigma}) \le 0$ is short for $\lambda \ge 0$, $\phi(\boldsymbol{\Sigma}) \le 0$, and $\lambda\, \phi(\boldsymbol{\Sigma}) = 0$ a.e. in Ω. Note that the derivation of (10) based on its strong formulation is given in [28, Section 1.2], see also [13, Chapter 3].

By replacing the time derivatives by backward differences with time step size $\tau = \frac{T}{N}$, we obtain the discretized problem: given $\ell^\tau \in (V')^N$, find $(\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau, \lambda^\tau) \in (S^2 \times V \times L^2(\Omega))^N$ such that $\boldsymbol{\Sigma}^\tau_i \in \mathcal{K}$ and

$$A(\boldsymbol{\Sigma}^\tau_i - \boldsymbol{\Sigma}^\tau_{i-1}) + B^\star(\boldsymbol{u}^\tau_i - \boldsymbol{u}^\tau_{i-1}) + \tau\, \lambda^\tau_i\, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}^\tau_i = \mathbf{0} \quad \text{in } S^2, \tag{11a}$$

$$B\boldsymbol{\Sigma}^\tau_i = \ell^\tau_i \quad \text{in } V', \tag{11b}$$

$$0 \le \lambda^\tau_i \quad \perp \quad \phi(\boldsymbol{\Sigma}^\tau_i) \le 0 \quad \text{a.e. in } \Omega \tag{11c}$$

is satisfied for all $i \in \{1, \dots, N\}$, where $(\boldsymbol{\Sigma}^\tau_0, \boldsymbol{u}^\tau_0) = \mathbf{0}$. The unique solvability of this incremental problem is shown in [13, Proof of Theorem 8.12, p. 196], for the formulation involving the plastic multiplier, we refer to [14, Theorem 1.4]. We denote the solution operator which maps $\ell^\tau \mapsto (\boldsymbol{\Sigma}^\tau, \boldsymbol{u}^\tau)$ by $\mathcal{G}^\tau$.

A regularization of (11) is given in [29, Section 2]: given the loads $\ell^\varepsilon \in (V')^N$, find $(\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon) \in (S^2 \times V)^N$ satisfying

$$A(\boldsymbol{\Sigma}^\varepsilon_i - \boldsymbol{\Sigma}^\varepsilon_{i-1}) + B^\star(\boldsymbol{u}^\varepsilon_i - \boldsymbol{u}^\varepsilon_{i-1}) + \tau\, \lambda^\varepsilon_i\, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}^\varepsilon_i = \mathbf{0}, \tag{12a}$$

$$B\boldsymbol{\Sigma}^\varepsilon_i = \ell^\varepsilon_i, \tag{12b}$$

$$\lambda^\varepsilon_i = \tau^{-1}\, k_1^{-1} \frac{\alpha^\varepsilon_i}{1 - \alpha^\varepsilon_i}, \quad \text{where } \alpha^\varepsilon_i = \max{}^\varepsilon \left(1 - \frac{\tilde{\sigma}_0}{|(\boldsymbol{\sigma}^\varepsilon_i + \boldsymbol{\chi}^\varepsilon_{i-1})^D|}\right), \tag{12c}$$

for all $i \in \{1, \dots, N\}$, and with the initial condition $(\boldsymbol{\Sigma}^\varepsilon_0, \boldsymbol{u}^\varepsilon_0) = (\mathbf{0}, \mathbf{0})$. The unique solvability of (12) can be proved using the Browder-Minty theorem, see the discussion after [29, Section 2]. We denote the solution operator which maps $\ell^\varepsilon \mapsto (\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon)$ by $\mathcal{G}^\varepsilon$. Note that we suppress the dependence of $\mathcal{G}^\varepsilon$ on τ.

1.3. The optimal control problem. At first, we substantiate the assumptions on the objective ψ. Throughout this paper we assume

Assumption 1.3.

(1) The function $\psi : H^1(0,T;V) \to \mathbb{R}$ is weakly lower semicontinuous, continuous and bounded from below.

(2) 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

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

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

(3) The cost parameter ν is a positive, real number.

The assumptions on the admissible set $U_\mathrm{ad} \subset H^1_{\{0\}}(0,T;U)$ and its time-discretization $U^\tau_\mathrm{ad} = U_\mathrm{ad} \cap U^N$, where U^N is identified with a linear subspace of $H^1_{\{0\}}(0,T;U)$ via the linear interpolation (6), is given by

Assumption 1.4.

(1) The admissible set U_{ad} is nonempty, convex and closed in $H^1_{\{0\}}(0,T;U)$.
(2) We suppose that for all $\boldsymbol{g} \in U_{\mathrm{ad}}$, there exists $\boldsymbol{g}^\tau \in U^\tau_{\mathrm{ad}}$, such that $\boldsymbol{g}^\tau_{\mathrm{c,p}} \to \boldsymbol{g}$ in $H^1(0,T;U)$ as $\tau \searrow 0$.

Some examples of ψ and U_{ad} satisfying Assumption 1.3 are given after [28, Assumption 2.8], see also the problem given in the introduction.

The optimal control problem under consideration is given by

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

Here, $\mathcal{G}$ is the solution map of (9) and E is defined in (1). The existence of an optimal control is shown in [28, Theorem 2.9].

Since the control-to-state map $\mathcal{G} \circ E$ is given by the solution of the VI (9) or, equivalently, by the complementarity system (10), the optimal control problem ($\mathbf{P}$) is a mathematical program with equilibrium constraints (MPEC) or with complementarity constraints (MPCC), respectively. Hence, optimality conditions are not given by the KKT system. In order to prove optimality conditions for ($\mathbf{P}$), we replace the solution map $\mathcal{G}$ of (10) by the solution map $\mathcal{G}^\varepsilon$ of the discretized and regularized problem (12).

By restricting $\boldsymbol{g}$ to U^τ_{ad} and by replacing the control-to-state map $\mathcal{G}$ by its discretization $\mathcal{G}^\tau$, we obtain the time-discrete optimal control problem

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

where the discrete functionals are defined by using the interpolation (6), i.e.,

$$\psi^\tau(\boldsymbol{u}^\tau) = \psi(\boldsymbol{u}^\tau_{\mathrm{c,p}}) \quad \text{and} \quad \|\boldsymbol{g}^\tau\|_{U^N} = \|\boldsymbol{g}^\tau_{\mathrm{c,p}}\|_{H^1(0,T;U)}.$$

We refer to [28, Section 3.4] concerning the existence of an optimal control.

Replacing $\mathcal{G}^\tau$ by $\mathcal{G}^\varepsilon$, we obtain the regularized control problem

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

The existence of an optimal control is proven in [29, Lemma 2.2] and an optimality system is given in [29, Section 3.3], see also (20).

2. 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 ($\mathbf{P}^\tau$) by passing to the limit in the optimality system of its regularization ($\mathbf{P}^\varepsilon$). Note that *all* local solutions $\boldsymbol{g}^\tau$ of ($\mathbf{P}^\tau$) can be approximated by solutions to the following, slightly modified version of ($\mathbf{P}^\varepsilon$), see [29, Section 4.2],

$$\left.\begin{array}{ll}\text{Minimize} & \psi^\tau(\boldsymbol{u}^\varepsilon) + \dfrac{\nu}{2}\|\boldsymbol{g}^\varepsilon\|^2_{U^N} + \dfrac{1}{2}\,\|\boldsymbol{g}^\varepsilon - \boldsymbol{g}^\tau\|^2_{U^N} \\ \text{such that} & (\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon) = \mathcal{G}^\varepsilon(E\boldsymbol{g}^\varepsilon) \quad \text{and} \quad \boldsymbol{g}^\varepsilon \in U^\tau_{\text{ad}}.\end{array}\right\} \qquad (\mathbf{P}^\varepsilon_{\boldsymbol{g}^\tau})$$

In order to derive the time-discrete optimality system formally, we introduce the multipliers

$$\begin{array}{llll}\boldsymbol{\Upsilon}^\tau_i \text{ for (11a)}, & \mu^\tau_i \text{ for } \lambda^\tau_i \ge 0, & \text{see (11c)}, \\ \boldsymbol{w}^\tau_i \text{ for (11b)}, & \theta^\tau_i \text{ for } \phi(\boldsymbol{\Sigma}^\tau_i) \le 0, & \text{see (11c)}.\end{array}$$

As usual in the context of MPCCs, we did not introduce a multiplier for the complementarity condition $\lambda^\tau_i \perp \phi(\boldsymbol{\Sigma}^\tau_i)$. Proceeding, we expect the following system of C-stationarity to hold for local optima of ($\mathbf{P}^\tau$), cf. [25],

$$A(\boldsymbol{\Sigma}^\tau_i - \boldsymbol{\Sigma}^\tau_{i-1}) + B^\star(\boldsymbol{u}^\tau_i - \boldsymbol{u}^\tau_{i-1}) + \tau\lambda^\tau_i \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau_i = \mathbf{0}, \tag{13a}$$

$$B(\boldsymbol{\Sigma}^\tau_i - \boldsymbol{\Sigma}^\tau_{i-1}) = E(\boldsymbol{g}^\tau_i - \boldsymbol{g}^\tau_{i-1}), \tag{13b}$$

$$0 \le \lambda^\tau_i \quad \perp \quad \phi(\boldsymbol{\Sigma}^\tau_i) \le 0, \tag{13c}$$

$$A(\boldsymbol{\Upsilon}^\tau_i - \boldsymbol{\Upsilon}^\tau_{i+1}) + B^\star(\boldsymbol{w}^\tau_i - \boldsymbol{w}^\tau_{i+1}) + \tau\lambda^\tau_i\mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon}^\tau_i + \tau\theta^\tau_i\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau_i = \mathbf{0}, \tag{14a}$$

$$B(\boldsymbol{\Upsilon}^\tau_i - \boldsymbol{\Upsilon}^\tau_{i+1}) = \psi^\tau_i(\boldsymbol{u}^\tau), \tag{14b}$$

$$\sum_{i=1}^N \big\langle E^\star\boldsymbol{w}^\tau_i, \tilde{\boldsymbol{g}}^\tau_i - \tilde{\boldsymbol{g}}^\tau_{i-1} - (\boldsymbol{g}^\tau_i - \boldsymbol{g}^\tau_{i-1})\big\rangle_{L^2(\Gamma_N;\mathbb{R}^d)} + \langle \nu\boldsymbol{g}^\tau, \tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\tau\rangle_{U^N} \ge 0, \tag{15}$$

$$\mathcal{D}\boldsymbol{\Sigma}^\tau_i : \mathcal{D}\boldsymbol{\Upsilon}^\tau_i - \mu^\tau_i = 0, \tag{16a}$$

$$\mu^\tau_i\lambda^\tau_i = 0, \tag{16b}$$

$$\theta^\tau_i\phi(\boldsymbol{\Sigma}^\tau_i) = 0, \tag{16c}$$

$$\theta^\tau_i\,\mu^\tau_i \ge 0, \tag{16d}$$

for $i = 1, \ldots, N$ and for all $\tilde{\boldsymbol{g}}^\tau \in U^\tau_{\text{ad}}$. Here, we used $(\boldsymbol{\Sigma}^\tau_0, \boldsymbol{u}^\tau_0) = (\mathbf{0}, \mathbf{0})$ and $(\boldsymbol{\Upsilon}^\tau_{N+1}, \boldsymbol{w}^\tau_{N+1}) = (\mathbf{0}, \mathbf{0})$. Moreover, $\psi^\tau_i(\boldsymbol{u}^\tau) \in V'$ denotes the partial derivatives of ψ^τ w.r.t. $\boldsymbol{u}^\tau_i$, see also (41).

Here, (13) is the forward system and (14) is the adjoint system. The variational inequality (15) is a relationship between the adjoint state $\boldsymbol{w}^\tau$ and the control $\boldsymbol{g}^\tau$, i.e., it is the gradient equation. The pointwise complementarity conditions on the multipliers (16) complete the system of C-stationary type.

The main result of this section is the following theorem.

Theorem 2.1. *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* (11)*. 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* (13)–(16) *is fulfilled.*

In Section 2.1 we reformulate an optimality system of $(\mathbf{P}^\varepsilon_{\boldsymbol{g}^\tau})$, see [29, Section 3.3], such that it involves the regularized counterparts of the multipliers appearing in the C-stationarity system. By passing to the limit with ε, we prove the C-stationarity result in Section 2.2. The main work is to verify certain estimates for the multipliers and adjoint states, which have to be uniform w.r.t. the regularization parameter ε.

2.1. Alternative formulation of forward and adjoint systems. The aim of this section is to state an optimality system for $(\mathbf{P}^\varepsilon_{\boldsymbol{g}^\tau})$ which resembles the C-stationarity system (13)–(16). To this end, we denote by $(\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon, \boldsymbol{g}^\varepsilon)$ a local minimizer of $(\mathbf{P}^\varepsilon_{\boldsymbol{g}^\tau})$.

According to [29, (34)], the adjoint states $(\boldsymbol{\Upsilon}^\varepsilon, \boldsymbol{w}^\varepsilon) = (\boldsymbol{\upsilon}^\varepsilon, \boldsymbol{\zeta}^\varepsilon, \boldsymbol{w}^\varepsilon) \in (S^2 \times V)^N$ are defined as the solution of the system

$$A(\boldsymbol{\Upsilon}^\varepsilon_i - \boldsymbol{\Upsilon}^\varepsilon_{i+1}) + B^\star(\boldsymbol{w}^\varepsilon_i - \boldsymbol{w}^\varepsilon_{i+1}) + \mathcal{D}^\star J^\varepsilon_i(\boldsymbol{\upsilon}^\varepsilon_i + \boldsymbol{\zeta}^\varepsilon_{i+1}) = \mathbf{0}, \tag{17a}$$

$$B(\boldsymbol{\Upsilon}^\varepsilon_i - \boldsymbol{\Upsilon}^\varepsilon_{i+1}) = \psi^\tau_i(\boldsymbol{u}^\varepsilon), \tag{17b}$$

with $(\boldsymbol{\Upsilon}^\varepsilon_{N+1}, \boldsymbol{w}^\varepsilon_{N+1}) = \mathbf{0}$. Here, $\psi^\tau_i(\boldsymbol{u}^\varepsilon) \in V'$ denotes the partial derivative of ψ^τ w.r.t. $\boldsymbol{u}^\varepsilon_i$. Moreover, J^ε_i is defined by

$$J^\varepsilon_i \,\boldsymbol{\tau} = k_1^{-1} \Big(\alpha^\varepsilon_i \,\boldsymbol{\tau}^D + \beta^\varepsilon_i \,\frac{\mathcal{D}\boldsymbol{\Sigma}^\varepsilon_i : \boldsymbol{\tau}^D}{|\mathcal{D}\boldsymbol{\Sigma}^\varepsilon_i|^2} \,\mathcal{D}\boldsymbol{\Sigma}^\varepsilon_i \Big), \tag{18}$$

where α^ε_i is defined in (12c), and β^ε_i is given by

$$\beta^\varepsilon_i = \big(\max{}^\varepsilon\big)' \Big(1 - \frac{\tilde\sigma_0}{|(\boldsymbol{\sigma}^\varepsilon_i + \boldsymbol{\chi}^\varepsilon_{i-1})^D|}\Big) \frac{\tilde\sigma_0}{|(\boldsymbol{\sigma}^\varepsilon_i + \boldsymbol{\chi}^\varepsilon_{i-1})^D|}. \tag{19}$$

The local optimality of $(\boldsymbol{\Sigma}^\varepsilon, \boldsymbol{u}^\varepsilon, \boldsymbol{g}^\varepsilon)$ implies that

$$\sum_{i=1}^N \big\langle E^\star \boldsymbol{w}^\varepsilon_i, \, \tilde{\boldsymbol{g}}^\tau_i - \tilde{\boldsymbol{g}}^\tau_{i-1} - (\boldsymbol{g}^\varepsilon_i - \boldsymbol{g}^\varepsilon_{i-1}) \big\rangle_{L^2(\Gamma_N;\mathbb{R}^d)} + \langle \nu \,\boldsymbol{g}^\varepsilon + \boldsymbol{g}^\varepsilon - \boldsymbol{g}^\tau, \, \tilde{\boldsymbol{g}}^\tau - \boldsymbol{g}^\varepsilon \rangle_{U^N} \ge 0 \tag{20}$$

holds for all $\tilde{\boldsymbol{g}}^\tau \in U^\tau_{\mathrm{ad}}$, see [29, Section 3.3].

Using (17a) and $\boldsymbol{\zeta}^\varepsilon_{N+1} = \mathbf{0}$ we infer $\boldsymbol{\zeta}^\varepsilon_i = (\boldsymbol{\zeta}^\varepsilon_i)^D$ for all $i = 1, \ldots, N$. We define

$$\tilde{\mathcal{D}}\boldsymbol{\Upsilon}^\varepsilon_i := (\boldsymbol{\upsilon}^\varepsilon_i)^D + \boldsymbol{\zeta}^\varepsilon_{i+1}. \tag{21}$$

Now we are going to manipulate the term $J^\varepsilon_i \,\tilde{\mathcal{D}}\boldsymbol{\Upsilon}^\varepsilon_i$ in order that the adjoint equation (17) resembles its counterpart in the expected C-stationarity system (14).

We obtain from the second component in (17a), (18) and (21) and Assumption 1.1(4)

$$\mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon - \tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon = \boldsymbol{\zeta}_i^\varepsilon - \boldsymbol{\zeta}_{i+1}^\varepsilon = -\Big(\alpha_i^\varepsilon\,\tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon + \beta_i^\varepsilon\,\frac{\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon : \tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon}{|\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon|^2}\,\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon\Big). \tag{22}$$

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 (18) 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

$$\theta_i^\varepsilon := k_1^{-1}\,\tau^{-1}\,\frac{\beta_i^\varepsilon}{1-\alpha_i^\varepsilon}\,\frac{\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon : \tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon}{|\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon|^2}. \tag{23}$$

The $L^2(\Omega)$-regularity of θ_i^ε is shown in Lemma 2.5. 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 (17) becomes

$$A(\boldsymbol{\Upsilon}_i^\varepsilon - \boldsymbol{\Upsilon}_{i+1}^\varepsilon) + B^\star(\boldsymbol{w}_i^\varepsilon - \boldsymbol{w}_{i+1}^\varepsilon) + \tau\lambda_i^\varepsilon\mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon + \tau\theta_i^\varepsilon\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon = \mathbf{0}, \tag{24a}$$

$$B(\boldsymbol{\Upsilon}_i^\varepsilon - \boldsymbol{\Upsilon}_{i+1}^\varepsilon) = \psi_i^\tau(\boldsymbol{u}^\varepsilon). \tag{24b}$$

It remains to define the multiplier μ^ε. According to (16a) we define

$$\mu_i^\varepsilon := \mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon : \mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon \in L^2(\Omega). \tag{25}$$

Testing (22) with $\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon$ implies

$$\mu_i^\varepsilon = \mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon : \mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon = (1-\alpha_i^\varepsilon - \beta_i^\varepsilon)\,\tilde{\mathcal{D}}\boldsymbol{\Upsilon}_i^\varepsilon : \mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon. \tag{26}$$

This equation is the starting point to estimate the multiplier μ_i^ε, see Lemma 2.2.

Now, the optimality system of $(\mathbf{P}^\varepsilon_{\boldsymbol{g}^\tau})$ consists of the state equation (12), the adjoint equation (24) and the gradient equation (20).

2.2. Convergence of the regularization. As a preparation for the proof of Theorem 2.1 we verify estimates for various quantities introduced in Section 2.1.

In Assumption 1.2 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 (12c) and (19) is smaller than $-\varepsilon$, larger than ε or in $(-\varepsilon, \varepsilon)$.

$$\begin{aligned} A_i^{\varepsilon,-} &:= \Big\{x \in \Omega : |\tilde{\mathcal{D}}\mathbf{\Sigma}_i^\varepsilon| \le \frac{\tilde{\sigma}_0}{1+\varepsilon}\Big\} = \Big\{x \in \Omega : 1 - \frac{\tilde{\sigma}_0}{|\tilde{\mathcal{D}}\mathbf{\Sigma}_i^\varepsilon|} \le -\varepsilon\Big\}, \\ A_i^{\varepsilon,+} &:= \Big\{x \in \Omega : |\tilde{\mathcal{D}}\mathbf{\Sigma}_i^\varepsilon| \ge \frac{\tilde{\sigma}_0}{1-\varepsilon}\Big\} = \Big\{x \in \Omega : 1 - \frac{\tilde{\sigma}_0}{|\tilde{\mathcal{D}}\mathbf{\Sigma}_i^\varepsilon|} \ge \varepsilon\Big\}, \\ A_i^{\varepsilon,0} &:= \Omega \setminus (A_i^{\varepsilon,-} \cup A_i^{\varepsilon,+}). \end{aligned}$$

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 (26).

Lemma 2.2. *We have*

$$1 - \alpha_i^\varepsilon - \beta_i^\varepsilon \in \begin{cases} \{1\}, & \textit{on } A_i^{\varepsilon,-}, \\ [0,1], & \textit{on } A_i^{\varepsilon,0}, \\ \{0\}, & \textit{on } A_i^{\varepsilon,+}, \end{cases} \tag{27}$$

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

Proof. By the definition of α_i^ε and β_i^ε in (12c) and (19), we infer immediately $\alpha_i^\varepsilon = \beta_i^\varepsilon = 0$ on $A_i^{\varepsilon,-}$. On $A_i^{\varepsilon,+}$ we have $\alpha_i^\varepsilon = 1 - \frac{\tilde{\sigma}_0}{|\tilde{\mathcal{D}}\mathbf{\Sigma}_i^\varepsilon|}$ and $\beta_i^\varepsilon = \frac{\tilde{\sigma}_0}{|\tilde{\mathcal{D}}\mathbf{\Sigma}_i^\varepsilon|}$. This implies $1 - \alpha_i^\varepsilon - \beta_i^\varepsilon = 0$.

It remains to check the assertion on $A_i^{\varepsilon,0}$. Let us define $\kappa_i^\varepsilon = 1 - \frac{\tilde{\sigma}_0}{|\tilde{\mathcal{D}}\mathbf{\Sigma}_i^\varepsilon|}$. On $A_i^{\varepsilon,0}$ we have $\kappa_i^\varepsilon \in (-\varepsilon, \varepsilon)$. By definition of α_i^ε and β_i^ε in (12c) and (19), 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 1.2 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

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

By $\alpha_i^\varepsilon \geq 0$ and $\beta_i^\varepsilon \geq 0$ 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 (16d).

Lemma 2.3. *For all $i \in \{1,\dots,N\}$, the condition $\theta_i^\varepsilon\,\mu_i^\varepsilon \geq 0$ is satisfied almost everywhere in Ω.*

Proof. By (26) we obtain

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

Now, Lemma 2.2 and the definition of θ_i^ε in (23) imply $\theta_i^\varepsilon\,\mu_i^\varepsilon \geq 0$. □

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

Lemma 2.4. *The adjoint states $(\boldsymbol{\Upsilon}^\varepsilon,\boldsymbol{w}^\varepsilon)$ satisfy*

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

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

Proof. There exists $\boldsymbol{\Sigma}_{B\boldsymbol{\Upsilon}_i^\varepsilon} \in S^2$, such that $B\boldsymbol{\Sigma}_{B\boldsymbol{\Upsilon}_i^\varepsilon}=B\boldsymbol{\Upsilon}_i^\varepsilon$ and $\mathcal{D}\boldsymbol{\Sigma}_{B\boldsymbol{\Upsilon}_i^\varepsilon}=\mathbf{0}$, see [28, (2.4)]. Let us define $\boldsymbol{T}=\boldsymbol{\Upsilon}_i^\varepsilon-\boldsymbol{\Sigma}_{B\boldsymbol{\Upsilon}_i^\varepsilon}$. This implies

$$
\begin{aligned}
B\boldsymbol{T}&=\mathbf{0}, &&\text{by } B\boldsymbol{\Sigma}_{B\boldsymbol{\Upsilon}_i^\varepsilon}=B\boldsymbol{\Upsilon}_i^\varepsilon,\\
\lambda_i^\varepsilon\,\mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon:\mathcal{D}\boldsymbol{T}=\lambda_i^\varepsilon\,\mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon:\mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon&\geq 0, &&\text{by } \lambda_i^\varepsilon\geq 0,\\
\theta_i^\varepsilon\,\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon:\mathcal{D}\boldsymbol{T}=\theta_i^\varepsilon\,\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon:\mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon=\theta_i^\varepsilon\,\mu_i^\varepsilon&\geq 0, &&\text{by Lemma 2.3.}
\end{aligned}
$$

Testing (24a) 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

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

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

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

where C depends only on A and B. Here, we used

$$\|\boldsymbol{\Sigma}_{B\boldsymbol{\Upsilon}_i^\varepsilon}\|_A \le C\,\|B\boldsymbol{\Upsilon}_i^\varepsilon\|_{V'}$$

and

$$\|\boldsymbol{\Sigma}_{B\boldsymbol{\Upsilon}_i^\varepsilon - B\boldsymbol{\Upsilon}_{i+1}^\varepsilon}\|_A \le C\,\|B\boldsymbol{\Upsilon}_i^\varepsilon - B\boldsymbol{\Upsilon}_{i+1}^\varepsilon\|_{V'} = C\,\|\psi_i^\tau(\boldsymbol{u}^\varepsilon)\|_{V'},$$

which follow from the inf-sup condition of $B^\star$, see [28, (2.4)]. By $B\boldsymbol{\Upsilon}_k^\varepsilon = \sum_{i=k}^N \psi_i^\tau(\boldsymbol{u}^\varepsilon)$ we obtain

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

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

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

where C depends only on A and B. The estimate for $\boldsymbol{w}_i^\varepsilon$ follows by (17a) and using the inf-sup condition of $B^\star$, see (5). □

For convenience we define the abbreviation

$$\boldsymbol{Q}_i^\varepsilon = -A\boldsymbol{\Upsilon}_i^\varepsilon - B^\star \boldsymbol{w}_i^\varepsilon, \tag{28}$$

which will be used frequently (also with other sub- and superscripts) in the sequel. The adjoint system (24a) becomes

$$\frac{1}{\tau}\,(\boldsymbol{Q}_i^\varepsilon - \boldsymbol{Q}_{i+1}^\varepsilon) = \lambda_i^\varepsilon\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon}_i^\varepsilon + \theta_i^\varepsilon\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon. \tag{29}$$

Using Lemma 2.4 we obtain the boundedness of $\boldsymbol{Q}^\varepsilon$

$$\max_{i=1,\ldots,N} \|\boldsymbol{Q}_i^\varepsilon\|_{S^2} \le C \sum_{i=1}^N \|\psi_i^\tau(\boldsymbol{u}^\varepsilon)\|_{V'}. \tag{30}$$

As a consequence, we obtain an estimate of the bilinear terms in (29) and of the multiplier θ_i^ε in $L^2(\Omega)$.

Lemma 2.5. *The estimates*

$$\|\theta_i^\varepsilon\, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}_i^\varepsilon\|_{S^2}^2 + \|\lambda_i^\varepsilon\, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Upsilon}_i^\varepsilon\|_{S^2}^2 \le \frac{1}{\tau^2}\, \|\boldsymbol{Q}_i^\varepsilon - \boldsymbol{Q}_{i+1}^\varepsilon\|_{S^2}^2 \tag{31a}$$

$$\|\theta_i^\varepsilon\|_{L^2(\Omega)} \le \frac{1+\varepsilon}{(1-\varepsilon)\sqrt{2}\,\tilde{\sigma}_0} \cdot \frac{1}{\tau}\, \|\boldsymbol{Q}_i^\varepsilon - \boldsymbol{Q}_{i+1}^\varepsilon\|_{S^2} \tag{31b}$$

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

Proof. Taking norms on both sides of (29) yields

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

Due to Lemma 2.3, the definition of μ_i^ε, see (25), and $\lambda_i^\varepsilon \ge 0$ the scalar product is non-negative. Indeed, we have

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

This yields (31a).

Due to $\beta_i^\varepsilon = 0$ on $A_i^{\varepsilon,-}$, we have $\theta_i^\varepsilon = 0$ on $A_i^{\varepsilon,-}$, see (23). By using

$$\mathcal{D} \boldsymbol{\Sigma}_i^\varepsilon = (1 - \alpha_i^\varepsilon)\, (\boldsymbol{\sigma}_i^\varepsilon + \boldsymbol{\chi}_{i-1}^\varepsilon)^D \tag{32}$$

see [29, (16)], we obtain $|\mathcal{D} \boldsymbol{\Sigma}_i^\varepsilon| = (1 - \alpha_i^\varepsilon)\, |\tilde{\mathcal{D}} \boldsymbol{\Sigma}_i^\varepsilon| \ge \frac{1-\varepsilon}{1+\varepsilon}\, \tilde{\sigma}_0$ on $A_i^{\varepsilon,0} \cup A_i^{\varepsilon,+}$. Hence, the estimate (31b) follows by (31a). □

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 3.8 and 3.9.

Finally, we prove the regularized counterparts of the complementarity conditions (16b) and (16c).

Lemma 2.6. *The plastic multiplier* λ^ε *and the multiplier* μ^ε *satisfy*

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

Proof. By (12c) we obtain $\lambda_i^\varepsilon = 0$ on $A_i^{\varepsilon,-}$. Using (26) and (27) 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,

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

Lemma 2.7. *The generalized stresses $\boldsymbol{\Sigma}^\varepsilon$ and the multiplier θ^ε satisfy*

$$\|\theta_i^\varepsilon\,\phi(\boldsymbol{\Sigma}_i^\varepsilon)\|_{L^1(\Omega)} \le \frac{2\,\varepsilon}{(1+\varepsilon)^2}\,\tilde\sigma_0^2\,\|\theta_i^\varepsilon\|_{L^1(A_i^{\varepsilon,0})} \quad \textit{for all } i \in \{1,\dots,N\}.$$

Proof. We have $\beta_i^\varepsilon = 0$ and hence by (23) $\theta_i^\varepsilon = 0$ on $A_i^{\varepsilon,-}$. From (27) and (32) we infer $\phi(\boldsymbol{\Sigma}_i^\varepsilon) = 0$ on $A_i^{\varepsilon,+}$.

On $A_i^{\varepsilon,0}$ we have $|\mathcal{D}\boldsymbol{\Sigma}_i^\varepsilon| \in \left[\frac{1-\varepsilon}{1+\varepsilon}, 1\right]\tilde\sigma_0$ by (12c) and (32). Hence, we obtain

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

Using Hölder's inequality concludes the proof. $\square$

Using the results above we prove that the system (13)–(16) is a necessary optimality condition for the time-discrete control problem $(\mathbf{P}^\tau)$.

Proof of Theorem 2.1. [29, Corollary 4.7] 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 (12). From [29, Theorem 4.3 and Corollary 4.4] 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 (11). This shows the forward system (13).

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 (23), (24) and (26). By Lemma 2.4 and Lemma 2.5 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 $(L^2(\Omega) \times L^2(\Omega))^N$, 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 (20) of the modified regularized problem $(\mathbf{P}^\varepsilon_{\boldsymbol{g}^\tau})$ and obtain (15).

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 (31a) 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 (30). 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 (24) and obtain (14).

It remains to check the relations (16). Using the definition (26) of μ^ε we infer (16a). Now we address the complementarity conditions (16b) and (16c). In view of Lemma 2.6 and Lemma 2.7 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 (2) of ϕ. This shows the complementarity conditions (16b) and (16c).

Last we address (16d). We will use [16, Proposition 3.15]. To this end, we test (24a) 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 2.3, 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 [16, 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 (14a) 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 (16d). $\square$

Remark 2.8. (1) Similarly to Theorem 2.1 a necessary optimality condition for the modified time-discrete problem $(\mathbf{P}_{\boldsymbol{g}}^\tau)$ can be proven, see the introduction of Section 3 for the definition of the modified problem. In the optimality system (13)–(16) we have to replace the gradient equation (15) by

$$\begin{aligned} &\sum_{i=1}^N \big\langle E^\star \boldsymbol{w}_i^\tau, \, \tilde{\boldsymbol{g}}_i^\tau - \tilde{\boldsymbol{g}}_{i-1}^\tau - (\boldsymbol{g}_i^\tau - \boldsymbol{g}_{i-1}^\tau) \big\rangle_{L^2(\Gamma_N;\mathbb{R}^d)} \\ &+ \langle (\boldsymbol{g}_{\mathrm{c,p}}^\tau - \boldsymbol{g}) + \nu \, \boldsymbol{g}_{\mathrm{c,p}}^\tau, \, \tilde{\boldsymbol{g}}_{\mathrm{c,p}}^\tau - \boldsymbol{g}_{\mathrm{c,p}}^\tau \rangle_{H^1(0,T;U)} \\ &\geq 0, \end{aligned} \tag{15'}$$

for all $\tilde{\boldsymbol{g}}^\tau \in U_{\mathrm{ad}}^\tau$. Here, we used the linear interpolation (6). As an optimality system for the modified time-discrete problem $(\mathbf{P}_{\boldsymbol{g}}^\tau)$ we obtain (13), (14), (15') and (16).

(2) In case $N = 1$ (only one time step) we obtain an optimality system for the optimal control of *static* plasticity. The system (13)–(16) equals the system [16, (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 [17, Section 3] one may derive a system of B-stationary type for the time-discrete problem $(\mathbf{P}^\tau)$.

3. Weak stationarity for the quasistatic problem

In this section we derive an optimality system for the continuous problem $(\mathbf{P})$. We use arguments similar to those in the proof of Theorem 2.1. Throughout this section, $\boldsymbol{g}$ denotes a fixed local optimum of $(\mathbf{P})$. [28, Theorem 3.10] yields the existence of a sequence $\{\boldsymbol{g}^\tau\}_{\tau>0}$ of local optima of the time-discrete and modified problems

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

such that their interpolations $\boldsymbol{g}^\tau_{\text{c,p}}$, see (6), converge to $\boldsymbol{g}$ in the strong topology of $H^1(0,T;U)$. This sequence $\{\boldsymbol{g}^\tau\}_{\tau>0}$ is fixed throughout this section. The convergence of (the interpolations of) the states $(\boldsymbol{\Sigma}^\tau_{\text{c,p}}, \boldsymbol{u}^\tau_{\text{c,p}}, \lambda^\tau_{\text{c,p}})$ 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 [28, 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 (13)–(16) as $\tau \searrow 0$.

Unfortunately, one cannot show the boundedness of the adjoints $\boldsymbol{\Upsilon}^\tau_{\text{c,d}}$ and $\boldsymbol{w}^\tau_{\text{c,d}}$ 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 2.4. Due to this lack of regularity, the derivatives in time of $\boldsymbol{\Upsilon}^\tau_{\text{c,d}}$ and $\boldsymbol{w}^\tau_{\text{c,d}}$ in the adjoint equation (14) have to be formulated in a weak sense in order to pass to the limit $\tau \searrow 0$. Hence, the adjoint equation (34) 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 [19, Theorem 6.2] and for the optimal control of ODEs involving hysteresis, see [5, Satz 8.12].

The main result of this paper is the following theorem.

Theorem 3.1. *Let* $\boldsymbol{g} \in H^1(0,T;U)$ *be a local minimum of the optimal control problem* $(\mathbf{P})$*. Then there exist*

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

satisfying

$$A\dot{\boldsymbol{\Sigma}} + B^\star \dot{\boldsymbol{u}} + \lambda\, \mathcal{D}^\star \mathcal{D}\boldsymbol{\Sigma} = \mathbf{0}, \tag{33a}$$

$$B\dot{\boldsymbol{\Sigma}} = E\dot{\boldsymbol{g}}, \tag{33b}$$

$$0 \le \lambda \quad \perp \quad \phi(\boldsymbol{\Sigma}) \le 0, \tag{33c}$$

$$\begin{aligned}\langle A\boldsymbol{\Upsilon} + B^\star \boldsymbol{w},\, \dot{\boldsymbol{T}}\rangle_{L^\infty(0,T;S^2),L^1(0,T;S^2)} - \langle A\boldsymbol{\Upsilon}_T + B^\star \boldsymbol{w}_T,\, \boldsymbol{T}(T)\rangle_{S^2}&\\ +\langle \lambda\, \mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon},\, \boldsymbol{T}\rangle_{L^2(0,T;S^2_1),L^2(0,T;S^2_1)'} + \theta(\mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T}) = \mathbf{0},&\end{aligned} \tag{34a}$$

$$B(\boldsymbol{\Upsilon} - \boldsymbol{\Upsilon}_T) - \int_\cdot^T \nabla\psi_c(\boldsymbol{u})\,\mathrm{d}s = \mathbf{0}, \tag{34b}$$

$$\big\langle E^\star \boldsymbol{w},\, \dot{\tilde{\boldsymbol{g}}} - \dot{\boldsymbol{g}}\big\rangle_{L^2(0,T;U)} + \langle \nu\, \boldsymbol{g},\, \tilde{\boldsymbol{g}} - \boldsymbol{g}\rangle_{H^1(0,T;U)} \ge 0, \tag{35}$$

$$\mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{\Upsilon} - \mu = 0, \tag{36a}$$

$$\mu\,\lambda = 0, \tag{36b}$$

$$\theta(v\,\phi(\boldsymbol{\Sigma})) = 0, \tag{36c}$$

$$A\boldsymbol{\Upsilon}_T + \theta_T\, \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}(T) + B^\star \boldsymbol{w}_T = \mathbf{0}, \tag{37a}$$

$$B\boldsymbol{\Upsilon}_T - \psi_T'(\boldsymbol{u}(T)) = \mathbf{0}, \tag{37b}$$

$$\mathcal{D}\boldsymbol{\Sigma}(T) : \mathcal{D}\boldsymbol{\Upsilon}_T - \mu_T = 0, \tag{38a}$$

$$\theta_T\, \phi(\boldsymbol{\Sigma}(T)) = 0, \tag{38b}$$

$$\theta_T\, \mu_T \ge 0, \tag{38c}$$

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

For definition of the spaces $\mathcal{X}(0,T)$ and $\mathcal{X}_{S^2,0}(0,T)$, we refer to (53).

The remainder of this section is devoted to the proof of Theorem 3.1 and is organized as follows. In Section 3.1, we use some basic convergence results in order to establish the state equation (33) and the gradient inequality (35). After deriving some auxiliary results in Section 3.2, we obtain the adjoint equation (34) and the terminal conditions (37), (38). Finally, the complementarity conditions (36) are verified in Section 3.4.

Throughout this section, we 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 (13), (14), (15') and (16) of the modified problem $(\mathbf{P}^\tau_{\boldsymbol{g}})$ is satisfied, see Remark 2.8(1). Moreover, we denote by $\boldsymbol{\Sigma}, \boldsymbol{u}, \lambda$ the states associated with $\boldsymbol{g}$ by (10) with $\ell = E\boldsymbol{g}$.

3.1. Basic convergence results. As already mentioned in the introduction of this section, [28, Theorems 3.3 and 3.4] imply

$$\left.\begin{aligned}&\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} \to \boldsymbol{\Sigma} \quad \text{in } H^1(0,T;S^2), \qquad \boldsymbol{u}^\tau_{\mathrm{c,p}} \to \boldsymbol{u} \quad \text{in } H^1(0,T;V),\\ &\lambda^\tau_{\mathrm{d+}} \to \lambda \quad \text{in } L^2(0,T;L^2(\Omega)),\end{aligned}\right\} \tag{39}$$

where $(\boldsymbol{\Sigma}, \boldsymbol{u}, \lambda)$ is the solution of the continuous problem associated to $\boldsymbol{g}$, see (10) with $\ell = E\boldsymbol{g}$. We refer to (6)–(8) for the definitions of $f^\tau_{\mathrm{c,p}}, f^\tau_{\mathrm{c,d}}, f^\tau_{\mathrm{d+}}$ and $f^\tau_{\mathrm{d-}}$. For later use, we mention that this implies

$$\left.\begin{aligned} &\boldsymbol{\Sigma}^\tau_{\mathrm{d-}} \to \boldsymbol{\Sigma} \quad \text{in } L^\infty(0,T;S^2), \qquad \boldsymbol{u}^\tau_{\mathrm{d-}} \to \boldsymbol{u} \quad \text{in } L^\infty(0,T;S^2),\\ &\lambda^\tau_{\mathrm{d-}} \to \lambda \quad \text{in } L^2(0,T;L^2(\Omega)). \end{aligned}\right\} \tag{40}$$

This shows the satisfaction of the state equation (33).

Now, we will give a formula for the partial derivatives ψ^τ_i of ψ^τ in order to show that the right hand side in the estimate of Lemma 2.4 is uniform w.r.t. the time step size τ. 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 $j \neq i$ and 1 at $i\,\tau$), we obtain

$$\psi^\tau_i(\boldsymbol{u}^\tau_{\mathrm{c,p}}) = \int_0^T v_i\, \nabla\psi_{\mathrm{c}}(\boldsymbol{u}^\tau_{\mathrm{c,p}})\,\mathrm{d}t \qquad \text{for } i = 1,\ldots,N-1, \tag{41a}$$

$$\psi^\tau_i(\boldsymbol{u}^\tau_{\mathrm{c,p}}) = \int_0^T v_i\, \nabla\psi_{\mathrm{c}}(\boldsymbol{u}^\tau_{\mathrm{c,p}})\,\mathrm{d}t + \psi'_T(\boldsymbol{u}^\tau_{\mathrm{c,p}}(T)) \quad \text{for } i = N, \tag{41b}$$

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. This shows

$$\begin{aligned} \sum_{i=1}^N \|\psi^\tau_i(\boldsymbol{u}^\tau_{\mathrm{c,p}})\|_{V'} &\le \sum_{i=1}^N \Big\| \int_0^T v_i\, \nabla\psi_{\mathrm{c}}(\boldsymbol{u}^\tau_{\mathrm{c,p}})\,\mathrm{d}t \Big\|_{V'} + \|\psi'_T(\boldsymbol{u}^\tau_{\mathrm{c,p}}(T))\|_{V'} \\ &\le \int_0^T \|\nabla\psi_{\mathrm{c}}(\boldsymbol{u}^\tau_{\mathrm{c,p}})\|_{V'}\,\mathrm{d}t + \|\psi'_T(\boldsymbol{u}^\tau_{\mathrm{c,p}}(T))\|_{V'}, \end{aligned} \tag{42}$$

where we used that the v_i form a partition of unity. Using Lemma 2.4, the continuities of $\nabla\psi_{\mathrm{c}} : L^2(0,T;V) \to L^2(0,T;V')$ and $\psi'_T : V \to V'$, and (39), we obtain

$$\|(\boldsymbol{\Upsilon}^\tau_{\mathrm{c,d}}, \boldsymbol{w}^\tau_{\mathrm{c,d}})\|_{L^\infty(0,T;S^2\times V)} + \|(\boldsymbol{\Upsilon}^\tau_{\mathrm{d-}}, \boldsymbol{w}^\tau_{\mathrm{d+}})\|_{L^\infty(0,T;S^2\times V)} \le C, \tag{43}$$

where $C > 0$ does not depend on τ. Since $L^\infty(0,T;S^2 \times V)$ is the dual of the separable space $L^1(0,T;S^2 \times V')$, see [6, Theorem 4.1] or [7, Theorem 8.18.3], there are subsequences of $\{(\boldsymbol{\Upsilon}^\tau_{\mathrm{c,d}}, \boldsymbol{w}^\tau_{\mathrm{c,d}})\}_{\tau>0}$ and of $\{(\boldsymbol{\Upsilon}^\tau_{\mathrm{d-}}, \boldsymbol{w}^\tau_{\mathrm{d+}})\}_{\tau>0}$ (denoted by the same symbol) which converge in the weak-$\star$ topology of $L^\infty(0,T;S^2 \times V)$. Testing $(\boldsymbol{\Upsilon}^\tau_{\mathrm{c,d}}, \boldsymbol{w}^\tau_{\mathrm{c,d}})$ 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

$$\left.\begin{aligned} &\boldsymbol{\Upsilon}^\tau_{\mathrm{c,d}} \stackrel{\star}{\rightharpoonup} \boldsymbol{\Upsilon} \quad \text{in } L^\infty(0,T;S^2), \qquad \boldsymbol{\Upsilon}^\tau_{\mathrm{d-}} \stackrel{\star}{\rightharpoonup} \boldsymbol{\Upsilon} \quad \text{in } L^\infty(0,T;S^2),\\ &\boldsymbol{w}^\tau_{\mathrm{c,d}} \stackrel{\star}{\rightharpoonup} \boldsymbol{w} \quad \text{in } L^\infty(0,T;V), \qquad \boldsymbol{w}^\tau_{\mathrm{d+}} \stackrel{\star}{\rightharpoonup} \boldsymbol{w} \quad \text{in } L^\infty(0,T;V). \end{aligned}\right\} \tag{44}$$

We will use the weak convergence of $\boldsymbol{w}^\tau_{\mathrm{d}+}$ in order to pass to the limit in the gradient equation. Using the interpolations (6) and (8b), the time-discrete gradient equation (15') reads

$$\langle E^\star \boldsymbol{w}^\tau_{\mathrm{d}+},\, \dot{\tilde{\boldsymbol{g}}}^\tau_{\mathrm{c,p}} - \dot{\boldsymbol{g}}^\tau_{\mathrm{c,p}} \rangle_{L^2(0,T;U)} + \langle (\boldsymbol{g}^\tau_{\mathrm{c,p}} - \boldsymbol{g}) + \nu\, \boldsymbol{g}^\tau_{\mathrm{c,p}},\, \tilde{\boldsymbol{g}}^\tau_{\mathrm{c,p}} - \boldsymbol{g}^\tau_{\mathrm{c,p}} \rangle_{H^1(0,T;U)} \geq 0 \quad (45)$$

for all $\tilde{\boldsymbol{g}}^\tau \in U^\tau_{\mathrm{ad}}$. Due to Assumption 1.4 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

$$\langle E^\star \boldsymbol{w},\, \dot{\tilde{\boldsymbol{g}}} - \dot{\boldsymbol{g}} \rangle_{L^2(0,T;U)} + \nu\, \langle \boldsymbol{g},\, \tilde{\boldsymbol{g}} - \boldsymbol{g} \rangle_{H^1(0,T;U)} \geq 0 \quad \text{for all } \tilde{\boldsymbol{g}} \in U_{\mathrm{ad}}.$$

This proves (35).

3.2. Auxiliary results. In this section we provide some results needed several times in the sequel.

We derive a relationship between the piecewise linear interpolant $\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}$, see (6), and the piecewise constant interpolant $\boldsymbol{\Sigma}^\tau_{\mathrm{d}-}$, see (8b). A simple calculation shows

$$\boldsymbol{\Sigma}^\tau_{\mathrm{d}-}(t) = \boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}(t) - (t - n^\tau(t)\,\tau)\, \dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,p}}(t) \quad \text{for a.a. } t \in [0,T], \tag{46}$$

where n^τ is given by

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

This definition implies $t \in [(n^\tau(t) - 1)\,\tau, n^\tau(t)\,\tau)$ for all $t \in [0,T]$. The relation (46) gives rise to the definition

$$\kappa^\tau(t) = (t - n^\tau(t)\,\tau) \quad \text{for a.a. } t \in [0,T]. \tag{47}$$

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

$$\kappa^\tau(t) \in [-\tau, 0] \quad \text{for a.a. } t \in [0,T]. \tag{48}$$

Due to (46) the term $\kappa^\tau\,(A\dot{\boldsymbol{\Upsilon}}^\tau_{\mathrm{c,d}} + B^\star \dot{\boldsymbol{w}}^\tau_{\mathrm{c,d}})$ appears frequently in Lemmas 3.8 and 3.9. Using the estimates (30) and (42) we can prove that it converges to zero w.r.t. the weak-$\star$ topology of $L^\infty(0,T;S^2)$. Similar to (28), we define

$$\boldsymbol{Q}^\tau_{\mathrm{c,d}} = -A\boldsymbol{\Upsilon}^\tau_{\mathrm{c,d}} - B^\star \boldsymbol{w}^\tau_{\mathrm{c,d}}.$$

Lemma 3.2. *The function $\kappa^\tau\, \dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}}$ converges towards $\mathbf{0}$ w.r.t. the weak-$\star$ topology of $L^\infty(0,T;S^2)$.*

Proof. Testing $\kappa^\tau\, \dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}}$ with $\chi_{(t,T)}\, \boldsymbol{T}$, where $t \in [0,T]$ and $\boldsymbol{T} \in S^2$, yields

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

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

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

since $|n^\tau(t)\,\tau - t| \le \tau$ and $|\kappa^\tau| \le \tau$ a.e. in $[0,T]$, see (48).

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

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

Hence, by (30) and (42) we obtain $\left| \langle \chi_{(t,T)}\, \boldsymbol{T},\, \kappa^\tau\, \dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}} \rangle_{L^2(0,T;S^2)} \right| \le \tau\, C$.

Using the boundedness of $\kappa^\tau\, \dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}}$ in $L^\infty(0,T;S^2)$, see (43) and (47), the density of the linear hull of $\{\chi_{[t,T]}\, \boldsymbol{T}\}$ in $L^1(0,T;S^2)$ finishes the proof. This density can be found in [9, Lemma IV.1.3]. □

Another term which will appear frequently is $\mathcal{D}\boldsymbol{T}^\tau : \mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}$, 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 $\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}$ in $L^\infty((0,T)\times\Omega;\mathbb{S})$, see (3), and the convergence $\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} \to \mathcal{D}\boldsymbol{\Sigma}$ in $L^\infty(0,T;S)$, an interpolation argument, see [26, Lemma 8.2], yields $\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} \to \mathcal{D}\boldsymbol{\Sigma}$ in $L^\infty(0,T;L^p(\Omega;\mathbb{S}))$ for all $p < \infty$. This gives in turn $\mathcal{D}\boldsymbol{T}^\tau : \mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} \overset{\star}{\rightharpoonup} \mathcal{D}\boldsymbol{T} : \mathcal{D}\boldsymbol{\Sigma}$ in $L^\infty(0,T;L^q(\Omega))$ for all $q < 2$. Using the boundedness of $\mathcal{D}\boldsymbol{T}^\tau : \mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}$ 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 3.3. *Let $\boldsymbol{T}^\tau \overset{\star}{\rightharpoonup} \boldsymbol{T}$ in $L^\infty(0,T;S^2)$. Then*

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

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

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

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

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

The statements remain valid if $\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}$ is replaced by $\boldsymbol{\Sigma}^\tau_{\mathrm{d-}}$.

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

$$\mathcal{D}\boldsymbol{T}^\tau : \mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} \rightharpoonup \mathcal{D}\boldsymbol{T} : \mathcal{D}\boldsymbol{\Sigma} \quad \text{in } L^1((0,T)\times\Omega). \tag{49}$$

Since $\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}$ is bounded in the space $L^\infty((0,T)\times\Omega;\mathbb{S})$, $\mathcal{D}\boldsymbol{T}^\tau : \mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}$ 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 (49), 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.

It remains to prove the statement involving f. Since $f^\tau\,\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}$ is bounded in $L^2(0,T;L^1(\Omega;\mathbb{S}))$, it is sufficient to show the convergence

$$\langle f^\tau\,\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}},\, \boldsymbol{T}\rangle \to \langle f\,\mathcal{D}\boldsymbol{\Sigma},\, \boldsymbol{T}\rangle$$

for all $\boldsymbol{T}$ from a dense subset of the dual space $L^2(0,T;L^1(\Omega;\mathbb{S}))'$. This dual space $L^2(0,T;L^1(\Omega;\mathbb{S}))'$ consists of (equivalence classes) of measurable functions $\boldsymbol{T} : (0,T)\times\Omega \to \mathbb{S}$ for which the norm

$$\|\boldsymbol{T}\|_{L^2(0,T;L^1(\Omega;\mathbb{S}))'} = \int_0^T \|\boldsymbol{T}(t)\|_{L^\infty(\Omega;\mathbb{S})}\,\mathrm{d}t$$

is finite, see [7, Theorem 8.20.3]. Hence, the space $L^\infty((0,T)\times\Omega;\mathbb{S})$ is dense in $L^2(0,T;L^1(\Omega;\mathbb{S}))'$. Therefore it remains to show that

$$f^\tau\,\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} \rightharpoonup f\,\mathcal{D}\boldsymbol{\Sigma} \quad \text{in } L^1(0,T;L^1(\Omega;\mathbb{S})) = L^1((0,T)\times\Omega;\mathbb{S}).$$

This follows by applying Theorem A.2 to the components of $f^\tau\,\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}$, since we have $f^\tau \rightharpoonup f$ in $L^1((0,T)\times\Omega)$ and the components of $\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}$ converges in $L^2((0,T)\times\Omega)$ and are bounded in $L^\infty((0,T)\times\Omega)$. □

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

Using $\boldsymbol{\Upsilon}^\tau_{\mathrm{c,d}}(T) = \boldsymbol{\Upsilon}^\tau_N$ and $\boldsymbol{w}^\tau_{\mathrm{c,d}}(T) = \boldsymbol{w}^\tau_N$ the adjoint equation (14) with $i = N$ implies

$$A\boldsymbol{\Upsilon}^\tau_N + \tau\,\lambda^\tau_N\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon}^\tau_N + \tau\,\theta^\tau_N\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau_N + B^\star\boldsymbol{w}^\tau_N = \mathbf{0}, \tag{50a}$$

$$B\boldsymbol{\Upsilon}^\tau_N = \psi^\tau_N(\boldsymbol{u}^\tau). \tag{50b}$$

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

$$(\boldsymbol{\Upsilon}^\tau_N, \boldsymbol{w}^\tau_N, \tau\,\theta^\tau_N) \rightharpoonup (\boldsymbol{\Upsilon}_T, \boldsymbol{w}_T, \theta_T) \quad \text{in } S^2\times V\times L^2(\Omega). \tag{51}$$

Moreover, the boundedness of $\lambda^\tau_{\mathrm{d}-}\,\mathcal{D}\boldsymbol{\Upsilon}^\tau_{\mathrm{d}-}$ in $L^2(0,T;L^1(\Omega;\mathbb{S}))$ implies the convergence $\tau\,\lambda^\tau_N\,\mathcal{D}\boldsymbol{\Upsilon}^\tau_N \to 0$ in $L^1(\Omega;\mathbb{S})$. Hence, we obtain from (50)

$$\begin{aligned} A\boldsymbol{\Upsilon}_T + \theta_T\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}(T) + B^\star\boldsymbol{w}_T &= \mathbf{0}, \\ B\boldsymbol{\Upsilon}_T &= \psi_T'(\boldsymbol{u}(T)). \end{aligned}$$

Similarly to the derivation of (16c) and (16d) we obtain

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

Note that under an additional regularity assumption we would also obtain that $(\boldsymbol{\Upsilon},\boldsymbol{w}) \in H^1(0,T;S^2\times V)$ and $(\boldsymbol{\Upsilon}_T,\boldsymbol{w}_T)$ coincides with $(\boldsymbol{\Upsilon}(T),\boldsymbol{w}(T))$, see Remark 3.12(5). This shows the terminal conditions (37), (38).

Due to the choice of the interpolations (7) and (8b), the discrete adjoint equation (14) reads

$$-A\dot{\boldsymbol{\Upsilon}}^\tau_{\mathrm{c,d}} - B^\star\dot{\boldsymbol{w}}^\tau_{\mathrm{c,d}} + \lambda^\tau_{\mathrm{d}-}\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon}^\tau_{\mathrm{d}-} + \theta^\tau_{\mathrm{d}-}\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{d}-} = \mathbf{0}, \tag{52a}$$

$$-B\dot{\boldsymbol{\Upsilon}}^\tau_{\mathrm{c,d}} = \psi^\tau_{\mathrm{d}-}(\boldsymbol{u}^\tau). \tag{52b}$$

Here we used the notation

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

similarly to (8b). Let us pass to the limit in (52b). Integration over $[t,T]$ implies

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

Hence, $\tau \searrow 0$, (51) and (41a) yield

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

Now, we turn to (52a). We show that the first three addends in (52a) converge weakly in adequate spaces. The convergence of the fourth addend $\theta^\tau_{\mathrm{d}-}\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{d}-}$ is more delicate and is addressed afterwards.

Since $(\boldsymbol{\Upsilon}^\tau_{\mathrm{c,d}},\boldsymbol{w}^\tau_{\mathrm{c,d}})$ is bounded only in $L^\infty(0,T;S^2\times V)$, we have to test the first two addends in (52a) with a differentiable test function. Let

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

be given. Integration by parts implies

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

In order to study the third addend in (52a), let us define the space

$$S^2_1 = S^2 + \{(\boldsymbol{\eta}, \boldsymbol{\eta}) \in L^1(\Omega; \mathbb{S}^2) : \mathrm{trace}(\boldsymbol{\eta}) = 0\}$$

equipped with the norm

$$\|\boldsymbol{T}\|_{S^2_1} = \inf_{\boldsymbol{T}=(\boldsymbol{\tau}+\boldsymbol{\eta}, \boldsymbol{\mu}+\boldsymbol{\eta})} \|(\boldsymbol{\tau}, \boldsymbol{\mu})\|_{S^2} + \|\boldsymbol{\eta}\|_{L^1(\Omega;\mathbb{S})},$$

where the infimum is taken over $(\boldsymbol{\tau}, \boldsymbol{\mu}) \in S^2$ and $\boldsymbol{\eta} \in L^1(\Omega; \mathbb{S})$ such that $\mathrm{trace}(\boldsymbol{\eta}) = 0$. A simple calculation shows that the dual of S^2_1 is

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

with the norm given by

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

see also [26, Lemma 41.2].

Using the convergence properties of $\lambda^\tau_{\mathrm{d}-}$ and $\boldsymbol{\Upsilon}^\tau_{\mathrm{d}-}$, see (40) and (44), we are led to expect $\lambda^\tau_{\mathrm{d}-}\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon}^\tau_{\mathrm{d}-} \rightharpoonup \lambda\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon}$ in $L^2(0,T; S^2_1)$. In order to prove this, we have to determine the dual of this space. Since S^2_∞ does not have the Radon-Nikodým property, we do not have $L^2(0,T;S^2_1)' = L^2(0,T;S^2_\infty)$, see [6, Theorem IV.1.1].

Theorem 3.4 ([7, Theorem 8.20.3]). *Let $\boldsymbol{T} \in L^2(0,T;S^2_1)'$ be given. It can be identified with a function $\boldsymbol{T} : [0,T] \to S^2_\infty$, which is weakly measurable and*

$$\|\boldsymbol{T}\|_{L^2(0,T;S^2_1)'} = \Big(\int_0^T \|\boldsymbol{T}(t)\|^2_{S^2_\infty}\,\mathrm{d}t\Big)^{\frac{1}{2}}.$$

Note, that the measurability of $\|\boldsymbol{T}(\cdot)\|_{S^2_\infty}$ is ensured by [7, Proposition 8.15.3]. *The duality pairing is given by*

$$\langle \boldsymbol{\Upsilon}, \boldsymbol{T}\rangle_{L^2(0,T;S^2_1),L^2(0,T;S^2_1)'} = \int_0^T \langle \boldsymbol{\Upsilon}(t), \boldsymbol{T}(t)\rangle_{S^2_1,S^2_\infty}\,\mathrm{d}t$$

for all $\boldsymbol{\Upsilon} \in L^2(0,T;S^2_1)$.

Remark 3.5. In [7, p. 558] a function $\boldsymbol{T}\colon [0,T] \to S^2_\infty$ 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 $\boldsymbol{T}_{|K}\colon K \to S^2_\infty$ 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 [7, 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 3.6. *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. **Step (1):** We show the weak measurability of $\lambda\,\mathcal{D}\boldsymbol{T} : [0,T] \to S$. Let $\varepsilon > 0$. By the definition of weak measurability, see Remark 3.5, we infer the existence of a compact set $K_1 \subset [0,T]$ such that $\mu([0,T]\setminus K_1) \le \varepsilon$ and $\boldsymbol{T}_{|K_1} : K_1 \to S^2_\infty$ is continuous w.r.t. the weak topology of S^2_∞. By Lusin's theorem, see [7, Corollary 4.8.5], we infer the existence of a compact set $K_2 \subset [0,T]$ such that $\mu([0,T]\setminus K_2) \le \varepsilon$ and $\lambda_{|K_2} : K_2 \to L^2(\Omega)$ is a continuous function. We set $K = K_1 \cap K_2$. We have $\mu([0,T]\setminus K) \le 2\,\varepsilon$. Let a sequence $\{t_i\}$ in K be given such that $t_i \to t \in K$. Then $\boldsymbol{T}(t_n) \rightharpoonup \boldsymbol{T}(t)$ in S^2_∞ and $\lambda(t_n) \to \lambda(t)$ in $L^2(\Omega)$. Hence, $(\lambda\,\mathcal{D}\boldsymbol{T})(t_n) \rightharpoonup (\lambda\,\mathcal{D}\boldsymbol{T})(t)$ in S. This shows that $(\lambda\,\mathcal{D}\boldsymbol{T})_{|K} : K \to S$ is weakly continuous. Hence, $\lambda\,\mathcal{D}\boldsymbol{T} : [0,T] \to S$ is weakly measurable.

Step (2): Since S is separable, [7, Theorem 8.15.2] implies the measurability of $\lambda\,\mathcal{D}\boldsymbol{T} : [0,T] \to S$.

Step (3): The integrability of $\lambda\,\mathcal{D}\boldsymbol{T}$: The simple estimate

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

implies the integrability of $\lambda\,\mathcal{D}\boldsymbol{T}$. Hence $\lambda\,\mathcal{D}\boldsymbol{T} \in L^1(0,T;S)$. Analogously to Steps (1)–(3), we show $\lambda^\tau_{\mathrm{d}}\,\mathcal{D}\boldsymbol{T} \in L^1(0,T;S)$.

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

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

This shows $\lambda^\tau_{\mathrm{d}}\,\mathcal{D}\boldsymbol{T} \to \lambda\,\mathcal{D}\boldsymbol{T}$ in $L^1(0,T;S)$. □

Using that the dual of $L^1(0,T;S)$ is $L^\infty(0,T;S)$, see [6, Theorem IV.1.1] or [7, 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 3.7. *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 (52a). For brevity, we define the spaces

$$\mathcal{X}(0,T) = W^{1,1}(0,T;L^2(\Omega)) \cap L^2(0,T;L^1(\Omega))', \tag{53a}$$

$$\mathcal{X}_{S^2,0}(0,T) = W^{1,1}_{\{0\}}(0,T;S^2) \cap L^2(0,T;S^2_1)'. \tag{53b}$$

The dual space of $\mathcal{X}(0,T)$ can be determined similarly to Theorem 3.4. We obtain

$$\begin{aligned}&\langle -A\dot{\boldsymbol{\Upsilon}}^\tau_{\text{c,d}} - B^\star \dot{\boldsymbol{w}}^\tau_{\text{c,d}} + \lambda^\tau_{\text{d}-}\, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Upsilon}^\tau_{\text{d}-},\, \boldsymbol{T}\rangle_{L^2(0,T;S^2)} \to \\ &\langle A\boldsymbol{\Upsilon} + B^\star \boldsymbol{w},\, \dot{\boldsymbol{T}}\rangle_{L^\infty(0,T;S^2),L^1(0,T;S^2)} - \langle A\boldsymbol{\Upsilon}_T + B^\star \boldsymbol{w}_T,\, \boldsymbol{T}(T)\rangle_{S^2} \\ &+ \langle \lambda\, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Upsilon},\, \boldsymbol{T}\rangle_{L^2(0,T;S^2_1),L^2(0,T;S^2_1)'}\end{aligned} \tag{54}$$

for all $\boldsymbol{T} \in \mathcal{X}_{S^2,0}(0,T)$. As an immediate consequence of (52a), there exists a functional $\Theta \in \mathcal{X}_{S^2,0}(0,T)'$ such that

$$\langle \theta^\tau_{\text{d}-}\, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}^\tau_{\text{d}-},\, \boldsymbol{T}\rangle_{L^2(0,T;S^2)} \to \Theta(\boldsymbol{T}) \tag{55}$$

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 $\theta^\tau_{\text{d}-}$ in $\mathcal{X}(0,T)'$.

For brevity, we denote $\boldsymbol{Q} = -A\boldsymbol{\Upsilon} - B^\star \boldsymbol{w}$ and $\boldsymbol{Q}_T = -A\boldsymbol{\Upsilon}_T - B^\star \boldsymbol{w}_T$.

Lemma 3.8. *Define* $\theta \in \mathcal{X}(0,T)'$ *by*

$$2\tilde{\sigma}_0^2 \theta(v) := \Big\langle \frac{\mathrm{d}}{\mathrm{d}t}(v \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}),\, \boldsymbol{Q} \Big\rangle_{L^1(0,T;S^2),L^\infty(0,T;S^2)} - \big\langle v(T) \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}(T),\, \boldsymbol{Q}_T \big\rangle_{S^2} \tag{56}$$

for all $v \in \mathcal{X}(0,T)$. *Then* $\theta^\tau_{\text{d}-} \overset{\star}{\rightharpoonup} \theta$ *in* $\mathcal{X}(0,T)'$.

Proof. Multiplying (52a) with $\mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}^\tau_{\text{d}-}$ and using

$$\begin{aligned} \lambda^\tau_{\text{d}-}\, \mathcal{D} \boldsymbol{\Upsilon}^\tau_{\text{d}-} : \mathcal{D} \boldsymbol{\Sigma}^\tau_{\text{d}-} &= 0, && \text{by (16b)}, \\ \theta^\tau_{\text{d}-}\, \mathcal{D} \boldsymbol{\Sigma}^\tau_{\text{d}-} : \mathcal{D} \boldsymbol{\Sigma}^\tau_{\text{d}-} &= \tilde{\sigma}_0^2\, \theta^\tau_{\text{d}-}, && \text{by (16c)}, \\ (\mathcal{D}^\star \mathcal{D})^2 &= 2\, \mathcal{D}^\star \mathcal{D}, && \text{by definition of } \mathcal{D} \text{ and } \mathcal{D}^\star, \end{aligned}$$

yields

$$\mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}^\tau_{\text{d}-} : \dot{\boldsymbol{Q}}^\tau_{\text{c,d}} + 2\, \tilde{\sigma}_0^2\, \theta^\tau_{\text{d}-} = 0 \quad \text{a.e. in } \Omega, \tag{57}$$

where $\boldsymbol{Q}^\tau_{\text{c,d}} = -A\boldsymbol{\Upsilon}^\tau_{\text{c,d}} - B^\star \boldsymbol{w}^\tau_{\text{c,d}}$. Let $v \in \mathcal{X}(0,T)$ be given. Multiplying (57) with v, integrating over $(0,T) \times \Omega$ and using (46) we obtain

$$2\, \tilde{\sigma}_0^2\, \langle \theta^\tau_{\text{d}-},\, v\rangle_{L^2((0,T)\times\Omega)} = -\big\langle v\, (\mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}^\tau_{\text{c,p}} - \kappa^\tau\, \mathcal{D}^\star \mathcal{D} \dot{\boldsymbol{\Sigma}}^\tau_{\text{c,p}}),\, \dot{\boldsymbol{Q}}^\tau_{\text{c,d}} \big\rangle_{L^2(0,T;S^2)}, \tag{58}$$

where κ^τ is given by (47). Due to the regularity of v and using the convergence of $\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}$, see (39), we obtain similarly to Lemma 3.6

$$v\,\mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,p}} \to v\,\mathcal{D}\dot{\boldsymbol{\Sigma}} \quad \text{in } L^1(0,T;S). \tag{59}$$

Together with Lemma 3.2 we infer

$$\langle v\,\kappa^\tau\,\mathcal{D}^\star\mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,p}},\, \dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}}\rangle_{L^2(0,T;S^2)} \to 0 \quad \text{as } \tau \searrow 0.$$

It remains to study the first addend on the right hand side of (58). Integration by parts yields

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

Using $\dot v \in L^1(0,T;L^2(\Omega))$ and Lemma 3.3 with $\boldsymbol{T}^\tau = \boldsymbol{Q}^\tau_{\mathrm{c,d}}$, we obtain the convergence of the first addend of the right hand side. By (59) and $\boldsymbol{Q}^\tau_{\mathrm{c,d}} \overset{\star}{\rightharpoonup} \boldsymbol{Q} = -A\boldsymbol{\Upsilon} - B^\star\boldsymbol{w}$ in $L^\infty(0,T;S^2)$, we infer the convergence of the second addend. It remains to study the convergence of the third addend. We have $\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}(T) \to \mathcal{D}\boldsymbol{\Sigma}(T)$ in S^2, see (39), and that sequence is bounded in $L^\infty(\Omega;\mathbb{S})$. Moreover, we have $\boldsymbol{Q}^\tau_{\mathrm{c,d}}(T) \rightharpoonup \boldsymbol{Q}_T$, see (51). This yields $\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}(T) : \boldsymbol{Q}^\tau_{\mathrm{c,d}}(T) \rightharpoonup \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}(T) : \boldsymbol{Q}_T$ in $L^2(\Omega)$. Since $v(T) \in L^2(\Omega)$, this implies the convergence of the third addend. Hence,

$$\begin{aligned}&\langle v\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}},\, \dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}}\rangle_{L^2(0,T;S^2)}\\ &\to -\langle \dot v\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma} + v\,\mathcal{D}^\star\mathcal{D}\dot{\boldsymbol{\Sigma}},\, \boldsymbol{Q}\rangle_{L^1(0,T;S^2),L^\infty(0,T;S^2)} + \langle v(T)\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}(T),\, \boldsymbol{Q}_T\rangle_{S^2},\end{aligned}$$

with $\boldsymbol{Q}_T = -A\boldsymbol{\Upsilon}_T - B^\star\boldsymbol{w}_T$. Altogether, we obtain

$$\begin{aligned}&2\,\tilde\sigma_0^2\,\langle \theta^\tau_{\mathrm{d}-},\, v\rangle_{L^2((0,T)\times\Omega)}\\ &\to \Big\langle \frac{\mathrm{d}}{\mathrm{d}t}(v\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}),\, \boldsymbol{Q}\Big\rangle_{L^1(0,T;S^2),L^\infty(0,T;S^2)} - \langle v(T)\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}(T),\, \boldsymbol{Q}_T\rangle_{S^2},\end{aligned} \tag{60}$$

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

Lemma 3.9. *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,

$$\begin{aligned}&\langle A\boldsymbol{\Upsilon} + B^\star\boldsymbol{w},\, \dot{\boldsymbol{T}}\rangle_{L^\infty(0,T;S^2),L^1(0,T;S^2)} - \langle A\boldsymbol{\Upsilon}_T + B^\star\boldsymbol{w}_T,\, \boldsymbol{T}(T)\rangle_{S^2}\\ &+ \langle \lambda\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon},\, \boldsymbol{T}\rangle_{L^2(0,T;S^2_1),L^2(0,T;S^2_1)'} + \theta(\mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T})\\ &= 0 \quad \text{for all } \boldsymbol{T} \in \mathcal{X}_{S^2,0}(0,T).\end{aligned}$$

Proof. Let a test function $\boldsymbol{T} \in \mathcal{X}_{S^2,0}(0,T)$ be given. Note that $\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{d}-} : \mathcal{D}\boldsymbol{T}$ does not belong to $\mathcal{X}(0,T)$ due to the discontinuities of $\boldsymbol{\Sigma}^\tau_{\mathrm{d}-}$. Hence, we cannot simply apply Lemma 3.8.

By (46) we have

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

Using (57), Lemma 3.2, and Lemma 3.3 with $\boldsymbol{T}^\tau = \dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}}$, we obtain

$$\kappa^\tau\, \theta^\tau_{\mathrm{d}-} \overset{\star}{\rightharpoonup} 0 \quad \text{in } L^\infty(0,T;L^2(\Omega)). \tag{61}$$

Due to $\dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,p}} \to \dot{\boldsymbol{\Sigma}}$ in $L^2(0,T;S^2)$ and $\boldsymbol{T} \in L^2(0,T;S^2_1)'$, we infer the convergence $\langle \theta^\tau_{\mathrm{d}-}\, \kappa^\tau\, \mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,p}},\, \mathcal{D}\boldsymbol{T}\rangle_{L^2(0,T;S)} \to 0$ as $\tau \searrow 0$. Hence, it remains to study the convergence of $\langle \theta^\tau_{\mathrm{d}-}\, \mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}},\, \mathcal{D}\boldsymbol{T}\rangle_{L^2(0,T;S)}$. However, $\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} : \mathcal{D}\boldsymbol{T}$ does not converge in $\mathcal{X}(0,T)$. Again, we cannot simply apply Lemma 3.8.

Using (57) and (46) we obtain

$$2\tilde{\sigma}_0^2 \langle \theta^\tau_{\mathrm{d}-} \mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}, \mathcal{D}\boldsymbol{T}\rangle_{L^2(0,T;S)} = -\langle (\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} : \mathcal{D}\boldsymbol{T})(\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} - \kappa^\tau \mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,p}}), \mathcal{D}\dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}}\rangle_{L^2(0,T;S)}.$$

By Lemma 3.2 and Lemma 3.3 with $f^\tau = \kappa^\tau\, \mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,p}} : \mathcal{D}\dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}}$ we obtain

$$\kappa^\tau\, (\mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,d}} : \mathcal{D}\dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}})\, \mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,d}} \rightharpoonup 0 \quad \text{in } L^2(0,T;L^1(\Omega;\mathbb{S})).$$

This directly yields $\kappa^\tau\, (\mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,d}} : \mathcal{D}\dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}})\, \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,d}} \rightharpoonup 0$ in $L^2(0,T;S^2_1)$. By $\boldsymbol{T} \in L^2(0,T;S^2_1)'$, this yields

$$\begin{aligned}&\langle (\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} : \mathcal{D}\boldsymbol{T}) \kappa^\tau \mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,p}}, \mathcal{D}\dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}}\rangle_{L^2(0,T;S)} = \langle \kappa^\tau (\mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,d}} : \mathcal{D}\dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}}) \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,d}}, \boldsymbol{T}\rangle_{L^2(0,T;S^2_1)}\\ &\to 0 \quad \text{as } \tau \searrow 0.\end{aligned}$$

Integration by parts implies

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

Using the chain rule and applying Lemma 3.3 thrice (with $\boldsymbol{T}^\tau = \boldsymbol{Q}^\tau_{\mathrm{c,d}}$, $g^\tau = \mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} : \mathcal{D}\boldsymbol{Q}^\tau_{\mathrm{c,d}}$ and $f^\tau = \mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,p}} : \mathcal{D}\boldsymbol{Q}^\tau_{\mathrm{c,d}}$) we obtain

$$\begin{aligned}&\Big\langle \frac{\mathrm{d}}{\mathrm{d}t}\big((\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} : \mathcal{D}\boldsymbol{T})\, \mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}\big),\, \mathcal{D}\boldsymbol{Q}^\tau_{\mathrm{c,d}}\Big\rangle_{L^2(0,T;S)}\\ &\quad \to \Big\langle \frac{\mathrm{d}}{\mathrm{d}t}\big((\mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T})\, \mathcal{D}\boldsymbol{\Sigma}\big),\, \mathcal{D}\boldsymbol{Q}\Big\rangle_{L^1(0,T;S),L^\infty(0,T;S)}.\end{aligned}$$

Putting everything together, we obtain

$$\begin{aligned} 2\,\tilde\sigma_0^2\,\langle \theta_{\mathrm{d}-}^\tau\,\mathcal{D}\boldsymbol{\Sigma}_{\mathrm{d}-}^\tau,\,\mathcal{D}\boldsymbol{T}\rangle_{L^2(0,T;S^2)} \to & \Big\langle \frac{\mathrm{d}}{\mathrm{d}t}\big((\mathcal{D}\boldsymbol{\Sigma}:\mathcal{D}\boldsymbol{T})\,\mathcal{D}\boldsymbol{\Sigma}\big),\,\mathcal{D}\boldsymbol{Q}\Big\rangle_{L^1(0,T;S),L^\infty(0,T;S)} \\ & - \Big\langle \big((\mathcal{D}\boldsymbol{\Sigma}(T):\mathcal{D}\boldsymbol{T}(T))\,\mathcal{D}\boldsymbol{\Sigma}(T)\big),\,\mathcal{D}\boldsymbol{Q}_T\Big\rangle_{S^2} \\ & = 2\,\tilde\sigma_0^2\,\theta(\mathcal{D}\boldsymbol{\Sigma}:\mathcal{D}\boldsymbol{T}) \end{aligned}$$

for all $\boldsymbol{T}\in\mathcal{X}_{S^2,0}(0,T)$, see (56). Together with (54)–(56), (60) this shows the claim. □

This shows the satisfaction of the adjoint equation (34).

3.4. Complementarity conditions. To complete the proof of Theorem 3.1, it remains to show the complementarity conditions (36). This is obtained by passing to the limit in the complementarity conditions (16a)–(16c).

In order to satisfy (36a), we define

$$\mu = \mathcal{D}\boldsymbol{\Sigma}:\boldsymbol{\Upsilon}.$$

Lemma 3.10. *We have*

$$\mu\,\lambda = 0 \quad \text{a.e. in } (0,T)\times\Omega.$$

Proof. From (16a) and (16b), we infer

$$\lambda_{\mathrm{d}-}^\tau\,\mathcal{D}\boldsymbol{\Sigma}_{\mathrm{d}-}^\tau:\mathcal{D}\boldsymbol{\Upsilon}_{\mathrm{d}-}^\tau = 0 \quad \text{a.e. in } (0,T)\times\Omega.$$

Hence, we have to establish the (weak) convergence of $\lambda_{\mathrm{d}-}^\tau\,\mathcal{D}\boldsymbol{\Sigma}_{\mathrm{d}-}^\tau:\mathcal{D}\boldsymbol{\Upsilon}_{\mathrm{d}-}^\tau$ towards $\lambda\,\mathcal{D}\boldsymbol{\Sigma}:\mathcal{D}\boldsymbol{\Upsilon}$.

We already know, see (40) and (44),

$$\begin{aligned} &\mathcal{D}\boldsymbol{\Sigma}_{\mathrm{d}-}^\tau \to \mathcal{D}\boldsymbol{\Sigma} \quad \text{in } L^2(0,T;S), \qquad \lambda_{\mathrm{d}-}^\tau \to \lambda \quad \text{in } L^2(0,T;L^2(\Omega)), \\ &\mathcal{D}\boldsymbol{\Upsilon}_{\mathrm{d}-}^\tau \stackrel{\star}{\rightharpoonup} \mathcal{D}\boldsymbol{\Upsilon} \quad \text{in } L^\infty(0,T;S). \end{aligned}$$

Moreover, $\mathcal{D}\boldsymbol{\Sigma}_{\mathrm{d}-}^\tau$ is bounded in $L^\infty((0,T)\times\Omega;\mathbb{S})$. Hence, we obtain

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

but that sequence is even bounded in $L^\infty(0,T;L^2(\Omega))$. Thus

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

see also the proof of Lemma 3.3 for similar arguments. This yields

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

The claim follows since set $\{\mathbf{0}\}$ is weakly closed in $L^2((0,T);L^1(\Omega))$. □

It remains to show (36c).

Lemma 3.11. *We have*

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

Proof. Let $v \in \mathcal{X}(0,T)$ be given. Testing (16c) with v, we obtain

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

We show the convergence of the left-hand side. By (46) we have

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

Using (39), (40), (61) we find $\kappa^\tau\, \theta^\tau_{\mathrm{d}-}\, (\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} + \mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{d}-}) \overset{\star}{\rightharpoonup} 0$ in $L^\infty(0,T;S)$. Together with (39), we have

$$\langle \kappa^\tau\, \theta^\tau_{\mathrm{d}-}\, (\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} + \mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{d}-}) : \mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,p}},\, v\rangle_{L^2(0,T;L^1(\Omega)),L^2(0,T;L^1(\Omega))'} \to 0.$$

Hence, it remains to show

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

Unfortunately, $\phi(\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}})\, v$ does not converges in $\mathcal{X}(0,T)$, and we cannot simply apply Lemma 3.8. By definition of ϕ, see (2), we have $2\,\langle \theta^\tau_{\mathrm{d}-}\, \phi(\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}),\, v\rangle = \langle \theta^\tau_{\mathrm{d}-}\, |\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}|^2,\, v\rangle - \tilde{\sigma}_0^2 \langle \theta^\tau_{\mathrm{d}-},\, v\rangle$. Due to Lemma 3.8, the second addend converges to $\tilde{\sigma}_0^2\, \theta(v)$. By (57), we obtain

$$-2\,\tilde{\sigma}_0^2\, \langle \theta^\tau_{\mathrm{d}-}\, |\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}|^2,\, v\rangle = \langle \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{d}-} : \dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}}\, |\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}|^2,\, v\rangle.$$

By (46), $\langle \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{d}-} : \dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}}\, |\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}|^2,\, v\rangle = \big\langle \mathcal{D}^\star\mathcal{D}(\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} - \kappa^\tau\, \dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,p}}) : \dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}}\, |\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}|^2,\, v\big\rangle$ follows. By using Lemma 3.2 and (39), we obtain

$$\mathcal{D}^\star\mathcal{D}\kappa^\tau\, \dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,p}} : \dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}} \rightharpoonup 0 \quad \text{in } L^2(0,T;L^1(\Omega)).$$

Similar arguments as in Lemma 3.3 show

$$\mathcal{D}^\star\mathcal{D}\kappa^\tau\, \dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,p}} : \dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}}\, |\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}|^2 \rightharpoonup 0 \quad \text{in } L^2(0,T;L^1(\Omega)).$$

Hence, we have

$$\big\langle \mathcal{D}^\star\mathcal{D}\kappa^\tau\, \dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,p}} : \dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}}\, |\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}|^2,\, v\big\rangle \to 0.$$

Hence, it remains to study

$$\big\langle \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} : \dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}}\, |\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}|^2,\, v\big\rangle.$$

Integration by parts yields

$$\begin{aligned}
&\big\langle \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} : \dot{\boldsymbol{Q}}^\tau_{\mathrm{c,d}} |\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}|^2, v\big\rangle \\
&= -\big\langle \mathcal{D}^\star\mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,p}} : \boldsymbol{Q}^\tau_{\mathrm{c,d}} |\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}|^2, v\big\rangle - 2\big\langle \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} : \boldsymbol{Q}^\tau_{\mathrm{c,d}} \mathcal{D}\dot{\boldsymbol{\Sigma}}^\tau_{\mathrm{c,p}} : \mathcal{D}\boldsymbol{\Sigma}, v\big\rangle \\
&\quad - \big\langle \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} : \boldsymbol{Q}^\tau_{\mathrm{c,d}} \mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}} : \mathcal{D}\boldsymbol{\Sigma}, \dot{v}\big\rangle + \big\langle \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}(T) : \boldsymbol{Q}^\tau_{\mathrm{c,d}}(T) |\mathcal{D}\boldsymbol{\Sigma}^\tau_{\mathrm{c,p}}(T)|^2, v(T)\big\rangle
\end{aligned}$$

Using Lemma 3.3, this converges towards

$$-\Big\langle \frac{\mathrm{d}}{\mathrm{d}t}\big(v\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}\,|\mathcal{D}\boldsymbol{\Sigma}|^2\big),\, \boldsymbol{Q}\Big\rangle + \big\langle v(T)\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}(T)\,|\mathcal{D}\boldsymbol{\Sigma}(T)|^2,\, \boldsymbol{Q}_T\big\rangle.$$

Due to (56), this term is equal to $-2\,\tilde{\sigma}_0^2\,\theta(v\,|\mathcal{D}\boldsymbol{\Sigma}|^2)$, since $v\,|\mathcal{D}\boldsymbol{\Sigma}|^2 \in \mathcal{X}(0,T)$.

Putting everything together, we find

$$0 = \langle \theta^\tau_{\mathrm{d}-}\,\phi(\boldsymbol{\Sigma}^\tau_{\mathrm{d}-}),\, v\rangle \to \theta(v\,\phi(\boldsymbol{\Sigma})). \qquad \square$$

This finishes the proof of Theorem 3.1. We conclude by giving some remarks on the optimality system obtained in Theorem 3.1.

Remark 3.12. (1) Following the notation for finite-dimensional MPECs, see [20, 25], the optimality system (33)–(38) 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, the low regularity of the multiplier θ is induced by the constraint $\phi(\boldsymbol{\Sigma}) \le 0$. For problems with a state equation much simpler than (9), e.g. with a scalar evolution variational inequality, one can construct examples where the multiplier θ is not a function.
The low regularity of the multiplier θ is also confirmed by numerical experiments, see [27, Chapter 6] or, for an ODE setting, [4].

(4) The remarks (2) and (3) above also apply to optimal control of parabolic VIs, see e.g. the optimality system in [19, Theorem 6.2].

(5) Using (56) and (34), 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 (34) and (37) can be stated equivalently as

$$\begin{aligned}
&\langle A\boldsymbol{\Upsilon} + B^\star\boldsymbol{w},\, \dot{\boldsymbol{T}}\rangle_{L^\infty(0,T;S^2),L^1(0,T;S^2)} \\
&+\langle \lambda\,\mathcal{D}\boldsymbol{\Upsilon},\, \mathcal{D}\boldsymbol{T}\rangle_{L^2(0,T;S^2_1),L^2(0,T;S^2_1)'} + \tilde{\theta}(\mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T}) = \boldsymbol{0}, &&\text{(62a)}\\
&B\boldsymbol{\Upsilon} - \psi_T'(\boldsymbol{u}(T)) - \int_\cdot^T \nabla\psi_{\mathrm{c}}(\boldsymbol{u})\,\mathrm{d}s = \boldsymbol{0}, &&\text{(62b)}
\end{aligned}$$

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 (34) with terminal conditions (37) over (62) since the former more clearly show the conditions at time T.

(7) There are two contributions to the terminal condition (37). 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 (33). In fact, similar terms appear also in the adjoint equation (34a) at times $t \in (0,T)$ where θ has Dirac contributions.

A. Weak convergence of products in Lebesgue spaces

In this appendix, we provide a result about the weak convergence of the product a weakly and a strongly convergent sequence in certain Lebesgue spaces. The result is applied to $\Omega_T = (0,T) \times \Omega$ in the main text.

First, we recall a basic result about weak convergence in L^1.

Lemma A.1. *Let (Ω_T, m) be a finite measure space. Suppose that the sequence $\{g_k\} \subset L^1(\Omega_T)$ converges weakly in $L^1(\Omega_T)$. Then, $\{g_k\}$ is uniformly integrable. That is, for all $\varepsilon > 0$, there is $\delta > 0$, such that*

$$\int_M |g_k| \,\mathrm{d}x \le \varepsilon$$

for all measurable $M \subset \Omega_T$ with $m(M) \le \delta$.

We refer to [6, Theorem IV.2.1] for a proof.

Now, we can prove the main result of the appendix.

Theorem A.2. *Let (Ω_T, m) be a finite measure space. Suppose that the sequences $\{f_k\} \subset L^2(\Omega_T)$, $\{g_k\} \subset L^1(\Omega_T)$ satisfy*

$$f_k \to f \quad \text{in } L^2(\Omega_T), \qquad g_k \rightharpoonup g \quad \text{in } L^1(\Omega_T)$$

for some $f \in L^2(\Omega_T)$ and $g \in L^1(\Omega_T)$. Moreover, assume that the sequence $\{f_k\}$ is bounded in $L^\infty(\Omega_T)$, that is $\|f_k\|_{L^\infty(\Omega_T)} \le K$.

Then,

$$f_k \, g_k \rightharpoonup f \, g \quad \text{in } L^1(\Omega_T).$$

Proof. Let $v \in L^\infty(\Omega_T)$ and $\varepsilon > 0$ be given. By Lemma A.1, there is $\delta > 0$, such that

$$\int_M |g_k| \,\mathrm{d}x \le \varepsilon$$

for all measurable $M \subset \Omega_T$ with $m(M) \le \delta$. By the weak convergence $g_k \rightharpoonup g$, we also obtain $\int_M |g| \,\mathrm{d}x \le \varepsilon$ for all such sets M.

Since $f_k \to f$ in $L^2(\Omega_T)$, there exists a subsequence (denoted by the same symbol), such that $f_k \to f$ a.e. in Ω_T. By Egorov's Theorem, there is a measurable set $O \subset \Omega_T$ with $m(O) \le \delta$ and

$$\|f_k - f\|_{L^\infty(\Omega_T \setminus O)} \to 0.$$

Together with $g_k \rightharpoonup g$ in $L^1(\Omega_T \setminus O)$, this yields

$$f_k\, g_k \rightharpoonup f\, g \quad \text{in } L^1(\Omega_T \setminus O). \tag{63}$$

This yields

$$\begin{aligned}
\left|\int_{\Omega_T} (f_k g_k - f g) v \,\mathrm{d}x\right| &\le \left|\int_{\Omega_T \setminus O} (f_k g_k - f g) v \,\mathrm{d}x\right| + \int_O |(f_k g_k - f g) v| \,\mathrm{d}x \\
&\le \left|\int_{\Omega_T \setminus O} (f_k g_k - f g) v \,\mathrm{d}x\right| + K \|v\|_{L^\infty(\Omega_T)} \int_O (|g_k| + |g|) \,\mathrm{d}x \\
&\le \left|\int_{\Omega_T \setminus O} (f_k g_k - f g) v \,\mathrm{d}x\right| + 2K \|v\|_{L^\infty(\Omega_T)} \varepsilon.
\end{aligned}$$

Together with (63), this yields

$$\int_{\Omega_T} (f_k\, g_k - f\, g)\, v \,\mathrm{d}x \to 0.$$

Since $v \in L^\infty(\Omega_T)$ was arbitrary, this shows the weak convergence of $f_k\, g_k$ in $L^1(\Omega_T)$. A subsequence-subsequence argument shows the convergence of the whole sequence. □

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

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

References

[1] Adams, D. R. and Lenhart, S., Optimal control of the obstacle for a parabolic variational inequality. *J. Math. Anal. Appl.* 268 (2002)(2), 602 – 614.

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

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

[4] Brokate, M., Numerical solution of an optimal control problem with hysteresis. In: *Optimal Control* (Proceedings Oberwolfach 1986; eds.: R. Bulirsch et al.). Lect. Notes Control Inform. Sci. 95. Berlin: Springer 1987, pp. 68–78.

[5] Brokate, M., *Optimale Steuerung von Gewöhnlichen Differentialgleichungen mit Nichtlinearitäten vom Hysteresis-Typ* (in German). Methoden Verfahren Math. Phys. 35. Frankfurt: Peter Lang 1987.

[6] Diestel, J. and Uhl, J. J., *Vector Measures.* Math. Surveys Monogr. 15. Providence (RI): Amer. Math. Soc. 1977.

[7] Edwards, R. E., *Functional Analysis.* New York: Dover Publ. 1995.

[8] Farshbaf-Shaker, M. H., A penalty approach to optimal control of Allen-Cahn variational inequalities: MPEC-view. *Numer. Funct. Anal. Optim.* 33 (2012)(11), 1321 – 1349.

[9] Gajewski, H., Gröger, K. and Zacharias, K., *Nichtlineare Operatorgleichungen und Operatordifferentialgleichungen* (in German). Berlin: Akademie 1974.

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

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

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

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

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

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

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

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

[18] Hintermüller, M. and Wegner, D., Distributed optimal control of the Cahn-Hilliard system including the case of a double-obstacle homogeneous free energy density. *SIAM J. Control Optim.* 50 (2012)(1), 388 – 418.

[19] Ito, K. and Kunisch, K., Optimal control of parabolic variational inequalities. *J. Math. Pures App.*, 93 (2010)(4), 329 – 360.

[20] Kanzow, C. and Schwartz, A., Mathematical programs with equilibrium constraints: enhanced Fritz John-conditions, new constraint qualifications, and improved exact penalty results. *SIAM J. Optim.* 20 (2010)(5), 2730 – 2753.

[21] Mignot, F. and Puel, J.-P., Contrôle optimal d'un système gouverné par une inéquation variationelle parabolique (in French). *C. R. Acad. Sci. Paris* 298 (1984)(12), 277 – 280.

[22] Rindler, F., Optimal control for nonconvex rate-independent evolution processes. *SIAM J. Control Optim.* 47 (2008)(6), 2773 – 2794.

[23] Rindler, F., Approximation of rate-independent optimal control problems. *SIAM J. Numer. Anal.* 47 (2009)(5), 3884 – 3909.

[24] Růžička, M., *Nichtlineare Funktionalanalysis* (in German). Berlin: Springer 2004.

[25] Scheel, H. and Scholtes, S., Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Math. Oper. Res.* 25 (2000)(1), 1 – 22.

[26] Tartar, L., *An Introduction to Sobolev Spaces and Interpolation Spaces.* Lect. Notes Unione Mat. Ital. 3. Berlin: Springer 2007.

[27] Wachsmuth, G., *Optimal control of quasistatic plasticity – An MPCC in function space.* PhD thesis, Chemnitz Univ. of Technology, Germany (2011).

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

[29] Wachsmuth, G., Optimal control of quasistatic plasticity with linear kinematic hardening II: Regularization and differentiability. *Z. Anal. Anwend.* 34 (2015)(4), 391 – 418.

[30] Yosida, K., *Functional Analysis.* Berlin: Springer 1965.

Received April 7, 2014; revised July 21, 2015