C-STATIONARITY FOR OPTIMAL CONTROL OF STATIC PLASTICITY WITH LINEAR KINEMATIC HARDENING

ROLAND HERZOG, CHRISTIAN MEYER, AND GERD WACHSMUTH

Abstract. An optimal control problem is considered for the variational inequality representing the stress-based (dual) formulation of static elastoplasticity. The linear kinematic hardening model and the von Mises yield condition are used. Existence and uniqueness of the plastic multiplier is rigorously proved, which allows for the re-formulation of the forward system using a complementarity condition.

In order to derive necessary optimality conditions, a family of regularized optimal control problems is analyzed, wherein the static plasticity problems are replaced by their viscoplastic approximations. By passing to the limit in the optimality conditions for the regularized problems, necessary optimality conditions of C-stationarity type are obtained.

1 Introduction

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

$$\begin{aligned} &a(\boldsymbol{\Sigma}, \boldsymbol{T} - \boldsymbol{\Sigma}) + b(\boldsymbol{T} - \boldsymbol{\Sigma}, \boldsymbol{u}) \geq 0 \quad \text{for all } \boldsymbol{T} \in \mathcal{K}, \\ &b(\boldsymbol{\Sigma}, \boldsymbol{v}) = \langle \ell, \boldsymbol{v} \rangle \quad \text{for all } \boldsymbol{v} \in V, \end{aligned} \tag{1.1}$$

where a and b are bilinear forms. The convex set $\mathcal{K}$ of admissible stresses is determined by the von Mises yield condition. The details are made precise below.

The optimization of elastoplastic systems is of significant importance for industrial deformation processes. The present paper can be viewed as a first step in this direction since the static system (1.1), despite of limited physical importance itself, appears as a time step of its quasi-static variants.

The optimal control of (1.1) leads to an infinite dimensional MPEC (mathematical program with equilibrium constraints). Due to the non-differentiability of the associated control-to-state map $\ell \mapsto (\boldsymbol{\Sigma}, \boldsymbol{u})$, the derivation of necessary optimality conditions is challenging. The same is true for the re-formulation of (1.1) as a complementarity system, which is the formulation used in this paper. It is well known that for the resulting MPCC (mathematical programs with complementarity constraints) classical constraint qualifications fail to hold.

To overcome these difficulties, several competing stationarity concepts have been developed, see for instance Scheel and Scholtes [2000] for an overview in the finite dimensional case. We follow classical arguments dating back to Barbu [1984], which lead to necessary optimality conditions of C-stationary type.

Let us briefly sketch these arguments, using the control of the obstacle problem as an example. This often used model problem is significantly simpler than (1.1) in

Date: February 24, 2012.

various aspects, and recalling the arguments will allow a comparison. The obstacle problem is to find, given $\ell \in H^{-1}(\Omega)$, the minimizer of $(1/2)(\nabla y, \nabla y)_\Omega - \langle \ell, y\rangle$, subject to $y \in \mathcal{K}_\psi = \{y \in H^1_0(\Omega) : y \le \psi\}$. Necessary and sufficient optimality conditions are given by the elliptic VI $(\nabla y, \nabla v - \nabla y)_\Omega \ge \langle \ell, v - y\rangle$ for all $v \in \mathcal{K}_\psi$.

The following classical arguments lead to a set of necessary optimality conditions of C-stationary type for the optimal control of the obstacle problem.

(1) Setting $\langle \xi, v\rangle := \langle \ell, v\rangle - (\nabla y, \nabla v)_\Omega$ for $v \in H^1_0(\Omega)$ defines an element $\xi \in H^{-1}(\Omega)$, which serves as a Lagrange multiplier associated with the constraint $y \in \mathcal{K}_\psi$, i.e., it belongs to the polar cone $\xi \in (\mathcal{K}_\psi - \psi)^\circ$ and the complementarity condition $\langle \xi, y - \psi\rangle = 0$ holds.
(2) The replacement of the constraint $y \in \mathcal{K}_\psi$ by a penalty term to the objective admits a differentiable control-to-state map $\ell \to y$. There is no difficulty in deriving necessary optimality conditions for the optimal control of this regularized problem.
(3) Since strict local solutions of the unregularized optimal control problem can be approximated by a sequence of regularized ones, we may pass to the limit in the latter. An optimality system of C-stationary type is obtained.

The following facts make the pursuit of this program significantly more difficult for the control of the VI (1.1), compared to the obstacle problem:

- The set of admissible stresses $\mathcal{K}$ is not a shifted cone like $\mathcal{K}_\psi$. Therefore the associated Lagrange multiplier λ, termed the plastic multiplier in the engineering literature, cannot be simply defined like ξ above. Its existence and uniqueness is a side result of the present paper.
- The set $\mathcal{K}$ is characterized by a pointwise nonlinear (indeed, quadratic) constraint function $\phi(\mathbf{\Sigma}) \le 0$. Since the stresses $\mathbf{\Sigma}$ involve the derivatives of the displacements $\boldsymbol{u}$, the problem at hand can be viewed as a VI with pointwise *nonlinear* constraints on the *gradient* of the state $\boldsymbol{u}$.
- The nonlinearity of $\phi(\mathbf{\Sigma})$ makes finding suitable, in particular differentiable regularizations a challenge, which play the role of the penalty terms in the regularized obstacle problem. The differentiability of this nonlinear Nemytzki operator is a nontrivial result and it requires recent regularity results for *quasi-linear* elasticity systems. The optimal control problems arising from this regularization represent a challenge in their own right.
- Finally, the passage to the limit requires more sophisticated arguments than the corresponding analysis for the obstacle problem. Due to the nonlinearity of the constraint $\phi(\mathbf{\Sigma}) \le 0$, the chain rule spawns additional nonlinear terms in the optimality conditions.

We mention that both, the obstacle problem and (1.1), can be interpreted as necessary and sufficient optimality conditions for a constrained optimization problem relating to the energy induced by the bilinear form $a(\cdot,\cdot)$. Therefore, these are also called the *lower-level* optimization problems, and the superimposed optimal control problem is referred to as the *upper-level* problem.

Let us put our work into perspective. There is an extensive list of contributions in the field of optimal control of VIs. In addition to the classical book of Barbu [1984], we refer to Mignot [1976], Mignot and Puel [1984], Friedman [1986], and Haslinger and Roubíček [1986] and the references therein. Nevertheless, the optimal control of VIs is still a very active field of research especially concerning their numerical treatment, see e.g. the recent publications Ito and Kunisch [2010], Hintermüller et al. [2009], Kunisch and Wachsmuth [2011], and Kunisch and Wachsmuth. The latter two contributions refine the classical penalty approach of Barbu [1984] and turn it into an efficient algorithm to solve optimal control problems of obstacle type. As mentioned before, we apply an analogous penalization technique in order to derive C-stationarity conditions which is much more delicate due to the differences of (1.1) to the obstacle problem described above. For the plastic torsion problem, which is structurally different from (1.1), first-order necessary optimality conditions are proved by Bermúdez and Saguez [1987]. The authors do not apply a penalization approach and obtain multipliers with considerably lower regularity in comparison with the multipliers derived here. Let us also mention the relaxation approaches considered in Bergounioux [1997] and **?** which could possibly also be applied to the problem under consideration, but would go beyond the scope of this paper.

The paper is organized as follows. The remainder of this section contains the presentation of the optimal control problem and the precise definition of notations and assumptions. Section 2 is devoted to the analysis of the lower-level problem. Existence and regularity of the plastic multiplier is rigorously proved in Section 2.1. We propose a regularization approach in Section 2.2 and show that it leads to a differentiable control-to-state map in Section 2.3. An estimate for the regularization error is proved in Section 2.4. Section 3 addresses the upper-level problem. Optimality systems for the regularized problem are obtained in Section 3.1. In Section 3.2 we discuss the approximation of optimal controls of the unregularized problem by regularized controls. The C-stationary optimality system is given in equations (3.3)–(3.6) in Section 3.3. Our main result is Theorem 3.16, which shows that all local minimizers of (**P**) are C-stationary.

1.1. Presentation of the Optimal Control Problem. In its strong form, the static problem of elastoplasticity with linear kinematic hardening reads

$$\left.\begin{aligned}
&& \mathbb{C}^{-1}\boldsymbol{\sigma} + \boldsymbol{\varepsilon}(\boldsymbol{u}) + \lambda\,(\boldsymbol{\sigma}^D + \boldsymbol{\chi}^D) &= \mathbf{0} && \text{in } \Omega,\\
&& \mathbb{H}^{-1}\boldsymbol{\chi} + \lambda\,(\boldsymbol{\sigma}^D + \boldsymbol{\chi}^D) &= \mathbf{0} && \text{in } \Omega,\\
&& \operatorname{div}\boldsymbol{\sigma} &= -\boldsymbol{f} && \text{in } \Omega,\\
&\text{with complem. conditions} & 0 \le \lambda \;\perp\; \phi(\boldsymbol{\Sigma}) &\le 0 && \text{in } \Omega,\\
&\text{and boundary conditions} & \boldsymbol{u} &= \mathbf{0} && \text{on } \Gamma_D,\\
&& \boldsymbol{\sigma}\cdot\boldsymbol{n} &= \boldsymbol{g} && \text{on } \Gamma_N = \Gamma\setminus\Gamma_D.
\end{aligned}\right\}\qquad(1.2)$$

The state variables consist of the stress and back stress $\boldsymbol{\Sigma} = (\boldsymbol{\sigma}, \boldsymbol{\chi})$, the displacement $\boldsymbol{u}$ and the plastic multiplier λ associated with the yield condition $\phi(\boldsymbol{\Sigma}) \le 0$ of von Mises type. The first two equations in (1.2), together with the complementarity conditions, represent the material law of static elastoplasticity. The tensors $\mathbb{C}^{-1}$ and $\mathbb{H}^{-1}$ are the inverse of the elasticity tensor (the compliance tensor) and of the hardening modulus, respectively, $\boldsymbol{\sigma}^D$ denotes the deviatoric part of $\boldsymbol{\sigma}$, while $\boldsymbol{\varepsilon}(\boldsymbol{u})$ is the linearized strain. The third equation in **??** The boundary conditions correspond to clamping on Γ_D and the prescription of boundary loads $\boldsymbol{g}$ on the remainder Γ_N.

The volume forces $\boldsymbol{f}$ and boundary loads $\boldsymbol{g}$ act as control variables. The optimal control, or upper-level problem under consideration reads

$$\left.\begin{aligned}
\text{Minimize}\quad & \frac{1}{2}\|\boldsymbol{u}-\boldsymbol{u}_d\|^2_{L^2(\Omega;\mathbb{R}^d)} + \frac{\nu_1}{2}\|\boldsymbol{f}\|^2_{L^2(\Omega;\mathbb{R}^d)} + \frac{\nu_2}{2}\|\boldsymbol{g}\|^2_{L^2(\Gamma_N;\mathbb{R}^d)}\\
\text{where}\quad & (\boldsymbol{\Sigma}, \boldsymbol{u}, \lambda) \text{ solves the static plasticity problem (1.2).}
\end{aligned}\right\}\qquad(\mathbf{P})$$

The objective expresses the goal of reaching as closely as possible a desired deformation $\boldsymbol{u}_d$. Objectives of this type are also relevant in future work for quasi-static variants of the problem, in order to approach a desired *final* deformation. In the

	state variable	test function	adjoint variable
generalized stresses	$\boldsymbol{\Sigma} = (\boldsymbol{\sigma}, \boldsymbol{\chi})$	$\boldsymbol{T} = (\boldsymbol{\tau}, \boldsymbol{\mu})$	$\boldsymbol{\Upsilon} = (\boldsymbol{\zeta}, \boldsymbol{\psi})$
displacement field	$\boldsymbol{u}$	$\boldsymbol{v}$	$\boldsymbol{w}$
	constraint		associated multiplier
plastic multiplier	$\lambda \geq 0$		μ
yield condition	$\phi(\boldsymbol{\Sigma}) \leq 0$, see (1.3)		θ
	control variable		
volume force	$\boldsymbol{f}$		
traction force	$\boldsymbol{g}$		
	constant		
yield stress	$\tilde{\sigma}_0$		

TABLE 1.1. Variables

interest of not further complicating the presentation, control constraints are not considered but they could be easily included with obvious modifications.

1.2. Notation and Assumptions.

Variables. Our notation follows Han and Reddy [1999] and Herzog and Meyer [2011] for the forward problem. Since the presentation of optimality conditions relies on adjoint variables and Lagrange multipliers associated with inequality constraints, additional variables are needed. For convenience, our notation is summarized in Table 1.1.

Function Spaces. Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary Γ in dimension $d \in \{2, 3\}$. We point out that the presented analysis is not restricted to the case $d \leq 3$, but for reasons of physical interpretation we focus on the two and three dimensional case. The boundary consists of two disjoint parts Γ_N and Γ_D. We denote by $\mathbb{S} := \mathbb{R}^{d\times d}_{\text{sym}}$ the space of symmetric d-by-d matrices, endowed with the inner product $\boldsymbol{A} : \boldsymbol{B} = \sum_{i,j=1}^{d} A_{ij} B_{ij}$, and we define

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

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

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

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

$$\phi(\boldsymbol{\Sigma}) = \big(|\boldsymbol{\sigma}^D + \boldsymbol{\chi}^D|^2 - \tilde{\sigma}_0^2\big)/2 \tag{1.3}$$

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

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

is the deviatoric part of $\boldsymbol{\sigma}$, and $\tilde{\sigma}_0$ is the yield stress. The yield function gives rise to the set of admissible generalized stresses

$$\mathcal{K} = \{\boldsymbol{\Sigma} \in S^2 : \phi(\boldsymbol{\Sigma}) \leq 0 \quad \text{a.e. in } \Omega\}. \tag{1.4}$$

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

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

for matrices $\boldsymbol{\Sigma} \in \mathbb{S}^2$ as well as for functions $\boldsymbol{\Sigma} \in S^2$. When considered as an operator in function space, $\mathcal{D}$ maps $S^2 \to S$. For later reference, we also remark that

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

holds.

Operators and Forms. We begin by defining the bilinear forms associated with (1.2). For $\boldsymbol{\Sigma} = (\boldsymbol{\sigma}, \boldsymbol{\chi}) \in S^2$ and $\boldsymbol{T} = (\boldsymbol{\tau}, \boldsymbol{\mu}) \in S^2$, let

$$a(\boldsymbol{\Sigma}, \boldsymbol{T}) = \int_\Omega \boldsymbol{\sigma} : \mathbb{C}^{-1}\boldsymbol{\tau}\,\mathrm{d}x + \int_\Omega \boldsymbol{\chi} : \mathbb{H}^{-1}\boldsymbol{\mu}\,\mathrm{d}x. \tag{1.5}$$

Here $\mathbb{C}^{-1}(x)$ and $\mathbb{H}^{-1}(x)$ are maps from $\mathbb{S}$ to $\mathbb{S}$ which may depend on the spatial variable x. For $\boldsymbol{\Sigma} = (\boldsymbol{\sigma}, \boldsymbol{\chi}) \in S^2$ and $\boldsymbol{v} \in V$, let

$$b(\boldsymbol{\Sigma}, \boldsymbol{v}) = -\int_\Omega \boldsymbol{\sigma} : \boldsymbol{\varepsilon}(\boldsymbol{v})\,\mathrm{d}x. \tag{1.6}$$

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

The bilinear forms induce operators

$$A : S^2 \to S^2, \quad \langle A\boldsymbol{\Sigma}, \boldsymbol{T}\rangle = a(\boldsymbol{\Sigma}, \boldsymbol{T}),$$
$$B : S^2 \to V', \quad \langle B\boldsymbol{\Sigma}, \boldsymbol{v}\rangle = b(\boldsymbol{\Sigma}, \boldsymbol{v}).$$

Here and throughout, $\langle\cdot, \cdot\rangle$ denotes the dual pairing between V and its dual V', or the scalar products in S or S^2, respectively. Moreover, $(\cdot, \cdot)_E$ refers to the scalar product of $L^2(E)$ where $E \subset \Omega$.

For convenience of the reader, all function spaces, operators and forms are summarized in Table 1.2.

Assumptions.

(1) The domain $\Omega \subset \mathbb{R}^d$, $d \geq 2$ is a bounded domain with Lipschitz boundary in the sense of [Grisvard, 1985, Chapter 1.2]. The boundary of Ω, denoted by Γ, consists of two disjoint measureable 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. Gröger [1989]. A characterization of regular domains for the case $d \in \{2, 3\}$ can be found in [Haller-Dintelmann et al., 2009, Section 5]. This class of domains covers a wide range of geometries.

We make these assumptions in order to apply the regularity results in Herzog et al. [2011] pertaining to systems of nonlinear elasticity. The latter appear in the 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.

(2) The yield stress $\tilde{\sigma}_0$ is assumed to be a positive constant. It equals $\sqrt{2/3}\,\sigma_0$, where σ_0 is the ~~initial~~ uni-axial yield stress.

space or set	definition	remark
$\mathbb{S}$	$\mathbb{R}^{d\times d}_{\text{sym}}$	symmetric d-by-d matrices
S	$L^2(\Omega;\mathbb{S})$	stress space
$\mathcal{K}$	$\{\boldsymbol{\Sigma} \in S^2 : \phi(\boldsymbol{\Sigma}) \le 0 \quad \text{a.e. in } \Omega\}$	admissible generalized stresses
V	$\{\boldsymbol{u} \in H^1(\Omega;\mathbb{R}^d) : \boldsymbol{u} = 0 \text{ on } \Gamma_D\}$	displacement space
U	$L^2(\Omega;\mathbb{R}^d) \times L^2(\Gamma_N;\mathbb{R}^d)$	control space
bilinear form	**definition**	**remark**
$a(\boldsymbol{\Sigma},\boldsymbol{T})$	$\int_\Omega \boldsymbol{\sigma} : \mathbb{C}^{-1}\boldsymbol{\tau}\,\mathrm{d}x + \int_\Omega \boldsymbol{\chi} : \mathbb{H}^{-1}\boldsymbol{\mu}\,\mathrm{d}x$	elasto-plastic energy
$b(\boldsymbol{\Sigma},\boldsymbol{v})$	$-\int_\Omega \boldsymbol{\sigma} : \boldsymbol{\varepsilon}(\boldsymbol{v})\,\mathrm{d}x$	weak form of div $\boldsymbol{\sigma}$
operator	**definition**	**remark**
$\mathbb{C}^{-1} : \mathbb{S} \to \mathbb{S}$		compliance tensor
$\mathbb{H}^{-1} : \mathbb{S} \to \mathbb{S}$		hardening tensor
$\boldsymbol{\varepsilon} : V \to S$	$\boldsymbol{\varepsilon}(\boldsymbol{v}) = (1/2)\big(\nabla \boldsymbol{v} + (\nabla \boldsymbol{v})^\top\big)$	(linearized) strain tensor
$\cdot^D : S \to S$	$\boldsymbol{\sigma}^D = \boldsymbol{\sigma} - (1/d)(\operatorname{trace}\boldsymbol{\sigma})\,\boldsymbol{I}$	deviatoric part
$\mathcal{D} : S^2 \to S$	$\mathcal{D}\boldsymbol{\Sigma} = \boldsymbol{\sigma}^D + \boldsymbol{\chi}^D$	sum of deviatoric parts
$\mathcal{D}^\star : S \to S^2$	$\mathcal{D}^\star\boldsymbol{\sigma} = (\boldsymbol{\sigma}^D,\, \boldsymbol{\sigma}^D)^\top$	adjoint of $\mathcal{D}$
$A : S^2 \to S^2$	$\langle A\boldsymbol{\Sigma},\, \boldsymbol{T}\rangle = a(\boldsymbol{\Sigma},\boldsymbol{T})$	elasto-plastic energy operator
$B : S^2 \to V'$	$\langle B\boldsymbol{\Sigma},\, \boldsymbol{v}\rangle = b(\boldsymbol{\Sigma},\boldsymbol{v})$	constraint operator

Table 1.2. Function spaces and operators

(3) $\mathbb{C}^{-1}$ and $\mathbb{H}^{-1}$ are elements of $L^\infty(\Omega;\mathcal{L}(\mathbb{S},\mathbb{S}))$, where $\mathcal{L}(\mathbb{S},\mathbb{S})$ denotes the space of linear operators $\mathbb{S} \to \mathbb{S}$. Both $\mathbb{C}^{-1}(x)$ and $\mathbb{H}^{-1}(x)$ are assumed to be uniformly coercive. Standard examples are isotropic and homogeneous materials, where

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

with Lamé constants μ and λ. (These constants appear only here and there is no risk of confusion with the plastic multiplier λ or the Lagrange multiplier μ.) In this case $\mathbb{C}^{-1}$ is coercive, provided that $\mu > 0$ and $d\,\lambda + 2\,\mu > 0$ hold. A common example for the hardening modulus is given by $\mathbb{H}^{-1}\boldsymbol{\chi} = \boldsymbol{\chi}/k_1$ with hardening constant $k_1 > 0$, see [Han and Reddy, 1999, Section 3.4].

(4) The desired displacement $\boldsymbol{u}_d$ is an element of $L^2(\Omega;\mathbb{R}^d)$. Moreover, ν_1 and ν_2 are positive constants.

Assumption (3) shows that $a(\boldsymbol{\Sigma},\boldsymbol{\Sigma}) \ge \underline{\alpha}\,\|\boldsymbol{\Sigma}\|^2_{S^2}$ for some $\underline{\alpha} > 0$.

2 Optimality Conditions and Regularization for the Lower-Level Problem

In this section, we address the lower-level problem.

$$\begin{aligned} &\text{Find} && (\boldsymbol{\Sigma},\boldsymbol{u}) \in S^2 \times V \\ &\text{satisfying} && a(\boldsymbol{\Sigma},\boldsymbol{T}-\boldsymbol{\Sigma}) + b(\boldsymbol{T}-\boldsymbol{\Sigma},\boldsymbol{u}) \ge 0 \quad \text{for all } \boldsymbol{T} \in \mathcal{K}, \\ & && b(\boldsymbol{\Sigma},\boldsymbol{v}) = \langle \ell,\, \boldsymbol{v}\rangle \quad \text{for all } \boldsymbol{v} \in V, \\ &\text{and} && \boldsymbol{\Sigma} \in \mathcal{K}. \end{aligned} \tag{L}$$

It is well known that given $\ell \in V'$, (L) has a unique solution $(\boldsymbol{\Sigma}, \boldsymbol{u})$, see, e.g., [Han and Reddy, 1999, Lemma 8.7] or [Herzog and Meyer, 2011, Proposition 3.1]. Moreover, (L) can be viewed as necessary and sufficient optimality conditions for the following energy minimization problem.

$$\left.\begin{aligned} \text{Minimize} \quad & \tfrac{1}{2} a(\boldsymbol{\Sigma}, \boldsymbol{\Sigma}) \\ \text{s.t.} \quad & b(\boldsymbol{\Sigma}, \boldsymbol{v}) = \langle \ell, \boldsymbol{v} \rangle \quad \text{for all } \boldsymbol{v} \in V, \\ \text{and} \quad & \boldsymbol{\Sigma} \in \mathcal{K}. \end{aligned}\right\} \tag{2.1}$$

The structure of this section is as follows. In Section 2.1 we give an equivalent reformulation of (2.1) in which the VI is replaced by an equivalent complementarity system involving the so-called plastic multiplier. This takes into account the particular yield function ϕ which characterizes the set of admissible stresses $\mathcal{K}$.

The derivation of optimality conditions for the upper-level problem ultimately requires the differentiability of the control-to-state map $(\boldsymbol{f}, \boldsymbol{g}) \mapsto (\boldsymbol{\Sigma}, \boldsymbol{u})$. Clearly, problem (2.1) does not enjoy this property. Therefore, the lower-level problem is regularized in Section 2.2 by penalizing the constraint $\boldsymbol{\Sigma} \in \mathcal{K}$. The desired differentiability is shown in Section 2.3.

In Section 2.4 we verify that the solutions of the regularized problems converge to those of the original problem (2.1).

2.1. Optimality Conditions Involving the Plastic Multiplier. We now give an equivalent characterization involving a Lagrange multiplier for the stress constraint. To this end, we recall from (1.4) the set of admissible generalized stresses. The gradient (w.r.t. the space S^2) of the yield function ϕ, defined in (1.3), is given by

$$\phi'(\boldsymbol{\Sigma}) = \begin{pmatrix} \boldsymbol{\sigma}^D + \boldsymbol{\chi}^D \\ \boldsymbol{\sigma}^D + \boldsymbol{\chi}^D \end{pmatrix} = \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}.$$

By formal Lagrangian calculus, we expect the following optimality conditions:

$$a(\boldsymbol{\Sigma}, \boldsymbol{T}) + b(\boldsymbol{T}, \boldsymbol{u}) + \big(\lambda, \mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T}\big)_\Omega = 0 \qquad \text{for all } \boldsymbol{T} = (\boldsymbol{\tau}, \boldsymbol{\mu}) \in S^2, \tag{2.2a}$$

$$b(\boldsymbol{\Sigma}, \boldsymbol{v}) = \langle \ell, \boldsymbol{v} \rangle \qquad \text{for all } \boldsymbol{v} \in V, \tag{2.2b}$$

$$0 \le \lambda \quad \perp \quad \phi(\boldsymbol{\Sigma}) \le 0 \qquad \text{a.e. in } \Omega, \tag{2.2c}$$

where $\lambda \perp \phi(\boldsymbol{\Sigma})$ represents the pointwise complementarity condition $\lambda\, \phi(\boldsymbol{\Sigma}) = 0$. A rigorous verification of (2.2a) is given in Theorem 2.2.

Remark 2.1. *In plasticity theory, the flow rule is often modeled using the so-called plastic multiplier λ. In the context of the* static *model of infinitesimal elastoplasticity, the flow rule reads $\lambda\, \phi'(\boldsymbol{\Sigma}) = \boldsymbol{P}$, where $\boldsymbol{P} = (\boldsymbol{p}, \boldsymbol{\xi})$ consists of the plastic strain $\boldsymbol{p}$ and the internal hardening variable $\boldsymbol{\xi}$. These satisfy the relations*

$$\boldsymbol{p} = \boldsymbol{\varepsilon}(\boldsymbol{u}) - \mathbb{C}^{-1}\boldsymbol{\sigma}, \tag{2.3a}$$

$$\boldsymbol{\xi} = \qquad - \mathbb{H}^{-1}\boldsymbol{\chi}. \tag{2.3b}$$

It is easy to check, using $\phi'(\boldsymbol{\Sigma}) = \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}$, that the plastic multiplier λ satisfies (2.2a). In addition, it also satisfies the complementarity condition (2.2c), see, e.g., [Han and Reddy, 1999, p. 60]. Therefore, the plastic multiplier can be interpreted as a Lagrange multiplier associated with the yield condition $\phi(\boldsymbol{\Sigma}) \le 0$.

Note that the expression $\big(\lambda, \mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T}\big)_\Omega$ in (2.2a) is well defined for $\lambda \in L^2(\Omega)$ and $\boldsymbol{\Sigma} \in \mathcal{K}$, which implies $\mathcal{D}\boldsymbol{\Sigma} \in L^\infty(\Omega; \mathbb{S})$. We now prove that (2.2) are indeed necessary and sufficient optimality conditions equivalent to (2.1).

Theorem 2.2. *Let $\ell \in V'$ be given.*

(a) *Suppose that* $(\boldsymbol{\Sigma}, \boldsymbol{u}) \in S^2 \times V$ *is the unique solution of* (2.1). *Then there exists a unique Lagrange multiplier* $\lambda \in L^2(\Omega)$ *such that* (2.2) *holds.*

(b) *If, on the other hand,* $(\boldsymbol{\Sigma}, \boldsymbol{u}, \lambda) \in S^2 \times V \times L^2(\Omega)$ *satisfies* (2.2)*, then* $(\boldsymbol{\Sigma}, \boldsymbol{u}) \in S^2 \times V$ *is the unique solution of* (2.1).

Before we prove Theorem 2.2, we need some auxiliary results. First we show a pointwise interpretation of the VI in (**L**). To this end, we define the pointwise bilinear forms

$$a_x(\boldsymbol{\Sigma}, \boldsymbol{T}) = [\boldsymbol{\sigma} : \mathbb{C}^{-1}\, \boldsymbol{\tau} + \boldsymbol{\chi} : \mathbb{H}^{-1}\, \boldsymbol{\mu}](x), \tag{2.4a}$$

$$b_x(\boldsymbol{\Sigma}, \boldsymbol{v}) = -[\boldsymbol{\sigma} : \boldsymbol{\varepsilon}(\boldsymbol{v})](x) \tag{2.4b}$$

for $\boldsymbol{\Sigma}, \boldsymbol{T} \in S^2$ and $\boldsymbol{v} \in V$.

For future reference, we define the active set, or plastic regime, for a given $\boldsymbol{\Sigma} \in \mathcal{K}$ as $\mathcal{A}(\boldsymbol{\Sigma}) = \{x \in \Omega : \phi(\boldsymbol{\Sigma})(x) = 0\}$. Its complement $\mathcal{I}(\boldsymbol{\Sigma}) = \{x \in \Omega : \phi(\boldsymbol{\Sigma})(x) < 0\}$ defines the inactive set, or elastic regime.

Lemma 2.3. *Let* $\boldsymbol{\Sigma} \in \mathcal{K}$ *satisfy* $a(\boldsymbol{\Sigma}, \boldsymbol{T} - \boldsymbol{\Sigma}) + b(\boldsymbol{T} - \boldsymbol{\Sigma}, \boldsymbol{u}) \geq 0$ *for all* $\boldsymbol{T} \in \mathcal{K}$. *Then this VI holds pointwise for almost all* $x \in \Omega$*, i.e.*

$$a_x(\boldsymbol{\Sigma}, \boldsymbol{T} - \boldsymbol{\Sigma}) + b_x(\boldsymbol{T} - \boldsymbol{\Sigma}, \boldsymbol{u}) \geq 0 \quad \text{for all } \boldsymbol{T} = (\boldsymbol{\tau}, \boldsymbol{\mu}) \in \mathcal{K}. \tag{2.5}$$

Moreover, for almost all $x \in \mathcal{I}(\boldsymbol{\Sigma})$*, we have*

$$a_x(\boldsymbol{\Sigma}, \boldsymbol{T}) + b_x(\boldsymbol{T}, \boldsymbol{u}) = 0 \quad \text{for all } \boldsymbol{T} = (\boldsymbol{\tau}, \boldsymbol{\mu}) \in S^2. \tag{2.6}$$

Proof. The pointwise interpretation of VIs w.r.t. the L^2-inner product is well established for the case of scalar functions, see for instance [Tröltzsch, 2010, Section 2.8]. The adaptation to the matrix valued situation can be done in a componentwise way. Equation (2.6) is a trivial consequence of (2.5). □

It is well known that the existence proof for Lagrange multipliers requires the verification of an appropriate constraint qualification. For infinite dimensional problems such as (2.1), standard constraint qualifications are those of Zowe and Kurcyusz [1979]. Concerning the inequality constraint, one has to verify that $\phi'(\boldsymbol{\Sigma})$ is surjective. In the case of (1.3), we have

$$\phi : S^2 \to L^1(\Omega), \qquad \phi'(\boldsymbol{\Sigma}) \in \mathcal{L}(S^2, L^1(\Omega)).$$

However, $\phi'(\boldsymbol{\Sigma})\,\boldsymbol{T} = \mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T} \in L^2(\Omega)$ for all $\boldsymbol{T} \in S^2$ because $\mathcal{D}\boldsymbol{\Sigma} \in L^\infty(\Omega; \mathbb{S})$ due to the structure of $\mathcal{K}$. This implies that $\phi'(\boldsymbol{\Sigma})$ cannot be surjective onto $L^1(\Omega)$.

To resolve this situation, given the solution $(\boldsymbol{\Sigma}, \boldsymbol{u}) \in S^2 \times V$ of (**L**), we define an auxiliary problem for $\boldsymbol{T} \in S^2$:

$$\left.\begin{aligned} \text{Minimize} \quad & a(\boldsymbol{\Sigma}, \boldsymbol{T}) + b(\boldsymbol{T}, \boldsymbol{u}) \\ \text{s.t.} \quad & \phi(\boldsymbol{T}) \leq 0 \quad \text{a.e. in } \mathcal{A}(\boldsymbol{\Sigma}). \end{aligned}\right\} \tag{$\mathbf{L}_{\text{aux}}$}$$

Problem ($\mathbf{L}_{\text{aux}}$) has a linear objective and the admissible set is convex but not bounded in S^2. Hence the existence of a solution is not a priori clear. However, we have

Lemma 2.4. $\boldsymbol{\Sigma}$ *is a solution of* ($\mathbf{L}_{\text{aux}}$).

Proof. We first observe that $\boldsymbol{\Sigma}$ is feasible for ($\mathbf{L}_{\text{aux}}$) since $\phi(\boldsymbol{\Sigma}) = 0$ holds on $\mathcal{A}(\boldsymbol{\Sigma})$ and $\phi(\boldsymbol{\Sigma}) < 0$ elsewhere. We need to show

$$a(\boldsymbol{\Sigma}, \boldsymbol{T} - \boldsymbol{\Sigma}) + b(\boldsymbol{T} - \boldsymbol{\Sigma}, \boldsymbol{u}) \geq 0 \tag{2.7}$$

for all $\boldsymbol{T} \in S^2$ satisfying $\phi(\boldsymbol{T}) \le 0$ a.e. in $\mathcal{A}(\boldsymbol{\Sigma})$. Let $\boldsymbol{T} \in S^2$ with $\phi(\boldsymbol{T}) \le 0$ a.e. in $\mathcal{A}(\boldsymbol{\Sigma})$ be given. We have $\boldsymbol{T} = \boldsymbol{T}|_{\mathcal{A}(\boldsymbol{\Sigma})} + \boldsymbol{T}|_{\mathcal{I}(\boldsymbol{\Sigma})} \in \mathcal{K} + S^2$. Testing (2.6) with $\boldsymbol{T} - \boldsymbol{\Sigma}$ and integrating over $\mathcal{I}(\boldsymbol{\Sigma})$ gives

$$a(\boldsymbol{\Sigma}, (\boldsymbol{T} - \boldsymbol{\Sigma})|_{\mathcal{I}(\boldsymbol{\Sigma})}) + b((\boldsymbol{T} - \boldsymbol{\Sigma})|_{\mathcal{I}(\boldsymbol{\Sigma})}, \boldsymbol{u}) = 0.$$

Using (2.5) and integrating over $\mathcal{A}(\boldsymbol{\Sigma})$ we obtain

$$a(\boldsymbol{\Sigma}, (\boldsymbol{T} - \boldsymbol{\Sigma})|_{\mathcal{A}(\boldsymbol{\Sigma})}) + b((\boldsymbol{T} - \boldsymbol{\Sigma})|_{\mathcal{A}(\boldsymbol{\Sigma})}, \boldsymbol{u}) \ge 0.$$

Adding these inequalities shows (2.7). □

In the following lemma we show how the constraint qualifications of Zowe and Kurcyusz [1979] can be applied to prove the existence of a Lagrange multiplier for $(\mathbf{L}_{\text{aux}})$.

Lemma 2.5. *Let $(\boldsymbol{\Sigma}, \boldsymbol{u}) \in S^2 \times V$ be the solution of $(\mathbf{L})$. Then there exists a Lagrange multiplier $\lambda \in L^2(\Omega)$, such that $(\boldsymbol{\Sigma}, \lambda)$ is a KKT point of $(\mathbf{L}_{\text{aux}})$, i.e.*

$$a(\boldsymbol{\Sigma}, \boldsymbol{T}) + b(\boldsymbol{T}, \boldsymbol{u}) + \big(\lambda, \mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T}\big)_{\mathcal{A}(\boldsymbol{\Sigma})} = 0 \quad \text{for all } \boldsymbol{T} = (\boldsymbol{\tau}, \boldsymbol{\mu}) \in S^2,$$
$$0 \le \lambda \quad \perp \quad \phi(\boldsymbol{\Sigma}) \le 0 \quad \text{a.e. in } \mathcal{A}(\boldsymbol{\Sigma})$$

is fulfilled. Moreover, λ is unique.

Proof. Step (1): We restrict problem $(\mathbf{L}_{\text{aux}})$ to the set

$$S_{\text{aux}} := \left\{ \boldsymbol{T} \in S^2 : \mathcal{D}\boldsymbol{T}|_{\mathcal{A}(\boldsymbol{\Sigma})} \in L^\infty(\mathcal{A}(\boldsymbol{\Sigma}); \mathbb{S}) \right\}.$$

Let us remark that all functions $\boldsymbol{T}$ which fulfill the condition $\phi(\boldsymbol{T}) \le 0$ on $\mathcal{A}(\boldsymbol{\Sigma})$ also belong to the set S_{aux}. Furthermore, S_{aux} is a Banach space when endowed with the norm $\|\boldsymbol{T}\|_{S^2} + \operatorname{ess\,sup}_{x \in \mathcal{A}(\boldsymbol{\Sigma})} |\mathcal{D}\boldsymbol{T}|$. We consider the constraint ϕ as a function $S_{\text{aux}} \to L^\infty(\mathcal{A}(\boldsymbol{\Sigma}))$. It is straightforward to show that ϕ is continuously differentiable with

$$\phi'(\boldsymbol{T})\widetilde{\boldsymbol{T}} = \mathcal{D}\boldsymbol{T} : \mathcal{D}\widetilde{\boldsymbol{T}}|_{\mathcal{A}(\boldsymbol{\Sigma})}.$$

Since $|\mathcal{D}\boldsymbol{\Sigma}| = \tilde{\sigma}_0$ on $\mathcal{A}(\boldsymbol{\Sigma})$, we have

$$\phi'(\boldsymbol{\Sigma})(\boldsymbol{\Sigma}/\tilde{\sigma}_0^2 \cdot f) = f$$

for every $f \in L^\infty(\mathcal{A}(\boldsymbol{\Sigma}))$ and therefore $\phi'(\boldsymbol{\Sigma})$ is surjective from S_{aux} onto $L^\infty(\mathcal{A}(\boldsymbol{\Sigma}))$.

Now we obtain by [Zowe and Kurcyusz, 1979, eq. (1.4)] the existence of a multiplier $\tilde{\lambda} \in L^\infty(\mathcal{A}(\boldsymbol{\Sigma}))'$. We extend it to an element $\lambda \in L^\infty(\Omega)'$ by $\langle \lambda, f \rangle := \langle \tilde{\lambda}, f|_{\mathcal{A}(\boldsymbol{\Sigma})} \rangle$. To summarize Step (1), we have shown that the following are necessary optimality conditions for $(\mathbf{L}_{\text{aux}})$:

$$a(\boldsymbol{\Sigma}, \boldsymbol{T}) + b(\boldsymbol{T}, \boldsymbol{u}) + \big(\lambda, \mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T}\big)_\Omega = 0 \quad \text{for all } \boldsymbol{T} = (\boldsymbol{\tau}, \boldsymbol{\mu}) \in S_{\text{aux}}, \tag{2.8a}$$
$$\langle \lambda, f \rangle \ge 0 \quad \text{for all } f \in L^\infty(\Omega),\ f \ge 0, \tag{2.8b}$$
$$\langle \lambda, \phi(\boldsymbol{\Sigma}) \rangle = 0. \tag{2.8c}$$

Step (2): We show the L^2 regularity of λ, following an idea used in Rösch and Tröltzsch [2007]. Thanks to Theorem 1.24 by Yosida and Hewitt [1952], we can uniquely decompose $\lambda \in L^\infty(\Omega)'$ as

$$\lambda = \lambda_c + \lambda_p,$$

where λ_c is a countably additive measure and λ_p is a purely finitely additive measure. Since λ is non-negative, λ_c and λ_p are also non-negative [Yosida and Hewitt, 1952, Theorem 1.23]. Now we can characterize λ_p by [Yosida and Hewitt, 1952, Theorem 1.22]: There exists a non-increasing sequence $\Omega \supset E_1 \supset E_2 \supset \ldots \supset E_n$, with Lebesgue measure $\mu(E_n) \to 0$ and $\lambda_p(E_n) = \lambda_p(\Omega)$. Now we test (2.8a) with $\boldsymbol{T}_n = \boldsymbol{\Sigma} \cdot \chi_{E_n}$:

$$a(\boldsymbol{\Sigma}, \boldsymbol{T}_n) + b(\boldsymbol{T}_n, \boldsymbol{u}) + \tilde{\sigma}_0^2\, \lambda_c(E_n) = -\tilde{\sigma}_0^2\, \lambda_p(E_n) = -\tilde{\sigma}_0^2\, \lambda_p(\Omega).$$

Since the left hand side tends to zero as $n \to \infty$, we have $\lambda_p(\Omega) = 0$. We can show $\lambda_c(N) = 0$ in an analogous way for arbitrary sets N of Lebesgue measure zero. Thus by the Radon-Nikodym Theorem we have $\lambda = \lambda_c \in L^1(\Omega)$. By (2.8c) we get $\lambda|_{\mathcal{I}(\boldsymbol{\Sigma})} = 0$. The sign condition on λ follows from (2.8b). Let us rewrite (2.8a) as

$$\big(\lambda, \mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T}\big)_\Omega = -a(\boldsymbol{\Sigma}, \boldsymbol{T}) - b(\boldsymbol{T}, \boldsymbol{u}).$$

The right hand side is continuous with respect to $\boldsymbol{T}$ and with respect to the S^2 norm. Since S_{aux} is dense in S^2, the mapping $J : \boldsymbol{T} \mapsto (\lambda, \mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T})_\Omega$ is a continuous mapping from S^2 to $\mathbb{R}$ and its Riesz representation is given by $J_R = \lambda\, \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma} \in S^2$. In view of (2.8c), computing the norm of J_R gives

$$\|J_R\|_{S^2}^2 = 2\,\tilde{\sigma}_0^2 \int_{\mathcal{A}(\boldsymbol{\Sigma})} \lambda^2 \,\mathrm{d}x = 2\,\tilde{\sigma}_0^2\, \|\lambda\|_{L^2(\Omega)}^2$$

and therefore $\lambda \in L^2(\Omega)$. Now we show that (2.8a) holds for $\boldsymbol{T} \in S^2$. Since $\lambda \in L^2(\Omega)$ and $\mathcal{D}\boldsymbol{\Sigma}|_{\mathcal{A}(\boldsymbol{\Sigma})} \in L^\infty(\mathcal{A}(\boldsymbol{\Sigma}))$, the left hand side of (2.8a) is continuous with respect to $\boldsymbol{T}$ and with respect to the S^2 norm. Since S_{aux} is dense in S^2, (2.8a) holds for all $\boldsymbol{T} \in S^2$.

Step (3): The uniqueness of λ follows directly from (2.8a) by testing with $\boldsymbol{T}_f = \frac{1}{2\tilde{\sigma}_0^2}\, f\, \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}$, where $f \in L^2(\mathcal{A}(\boldsymbol{\Sigma}))$ is arbitrary:

$$(\lambda, f)_\Omega = \big(\lambda, \mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T}_f\big)_\Omega = -a(\boldsymbol{\Sigma}, \boldsymbol{T}_f) - b(\boldsymbol{T}_f, \boldsymbol{u}).$$

Thus the right hand side is an alternative representation of λ and therefore λ is unique due to the uniqueness of $\boldsymbol{\Sigma}$ and $\boldsymbol{u}$. □

Now we can prove Theorem 2.2.

Proof of Theorem 2.2. Statement (a) follows from Lemma 2.5.

In order to prove (b), we test equation (2.2a) with $\boldsymbol{T} = \widetilde{\boldsymbol{T}} - \boldsymbol{\Sigma}$, $\widetilde{\boldsymbol{T}} \in \mathcal{K}$ and obtain

$$\begin{aligned} a(\boldsymbol{\Sigma}, \widetilde{\boldsymbol{T}} - \boldsymbol{\Sigma}) + b(\widetilde{\boldsymbol{T}} - \boldsymbol{\Sigma}, \boldsymbol{u}) &= -\big(\lambda, \mathcal{D}\boldsymbol{\Sigma} : (\mathcal{D}\widetilde{\boldsymbol{T}} - \mathcal{D}\boldsymbol{\Sigma})\big)_\Omega \\ &= \big(\lambda, \tilde{\sigma}_0^2 - \mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\widetilde{\boldsymbol{T}}\big)_{\mathcal{A}(\boldsymbol{\Sigma})} \ge 0 \end{aligned}$$

since both factors are pointwise non-negative. Therefore the VI in (**L**) is fulfilled for all $\widetilde{\boldsymbol{T}} \in \mathcal{K}$, which implies that $(\boldsymbol{\Sigma}, \boldsymbol{u})$ is the solution of (2.1). □

2.2. Regularization by Penalization and Smoothing. The goal of this section is to define a relaxed version of (2.1) where $\boldsymbol{\Sigma} \in \mathcal{K}$ is replaced by a penalty term in the objective. This leads to a different optimality system where the complementarity condition between λ and $\phi(\boldsymbol{\Sigma})$ is converted into a one-to-one relation. However, this relation is not differentiable, and therefore needs to be smoothed.

We begin by specifying the penalized problem. A natural approach is to use the Moreau-Yosida approximation, which leads to the additional term

$$\|\boldsymbol{\Sigma} - P_{\mathcal{K}}(\boldsymbol{\Sigma})\|_{S^2}^2 = \frac{1}{2} \int_\Omega \max\big\{0, |\mathcal{D}\boldsymbol{\Sigma}| - \tilde{\sigma}_0\big\}^2 \,\mathrm{d}x \tag{2.9}$$

in the objective, where $P_{\mathcal{K}}$ denotes the orthogonal projection onto $\mathcal{K}$ w.r.t. the scalar product of S^2, i.e.,

$$\boldsymbol{\Sigma} - P_{\mathcal{K}}(\boldsymbol{\Sigma}) = \frac{1}{2} \max\big\{0, |\mathcal{D}\boldsymbol{\Sigma}| - \tilde{\sigma}_0\big\}\, \frac{1}{|\mathcal{D}\boldsymbol{\Sigma}|} \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}. \tag{2.10}$$

parameters	definition	remark
γ	$\gamma > 0$	penalty parameter $\gamma \to \infty$
ε	$\varepsilon > 0$	regularization parameter $\varepsilon \to 0$
operator	**definition**	**remark**
$P_{\mathcal{K}} : S^2 \to S^2$	see (2.10)	orthogonal projection onto $\mathcal{K}$
$p_{\gamma,\varepsilon} : \mathbb{R} \to \mathbb{R}$	see (2.12)	
$J_{\gamma,\varepsilon} : S^2 \to S^2$	see (2.12)	smooth replacement of $\lambda\, \phi'(\boldsymbol{\Sigma})$
$G_{\gamma,\varepsilon} : V' \to S^2 \times V$	see (2.14)	solution map of (2.13)
$G^{\boldsymbol{\Sigma}}_{\gamma,\varepsilon}$, $G^{\boldsymbol{u}}_{\gamma,\varepsilon}$		1st and 2nd component of $G_{\gamma,\varepsilon}$
variable	**definition**	**remark**
$\boldsymbol{\Sigma}_{\gamma,\varepsilon} \in S^2$	$G^{\boldsymbol{\Sigma}}_{\gamma,\varepsilon}(\ell)$	regularization of $\boldsymbol{\Sigma}$
$\boldsymbol{u}_{\gamma,\varepsilon} \in S^2$	$G^{\boldsymbol{u}}_{\gamma,\varepsilon}(\ell)$	regularization of $\boldsymbol{u}$
$\lambda_{\gamma,\varepsilon} \in L^2(\Omega)$	$p_{\gamma,\varepsilon}(\lvert\mathcal{D}\boldsymbol{\Sigma}_{\gamma,\varepsilon}\rvert)$, see (2.20)	regularization of λ

TABLE 2.1. Operators and variables associated with penalization and smoothing

The penalized problem becomes

$$\left.\begin{aligned} \text{Minimize} \quad & \frac{1}{2}a(\boldsymbol{\Sigma},\boldsymbol{\Sigma}) + \frac{\gamma}{2}\int \max\big\{0, |\mathcal{D}\boldsymbol{\Sigma}| - \tilde{\sigma}_0\big\}^2 \,\mathrm{d}x \\ \text{s.t.} \quad & b(\boldsymbol{\Sigma},\boldsymbol{v}) = \langle \ell, \boldsymbol{v}\rangle \quad \text{for all } \boldsymbol{v} \in V. \end{aligned}\right\} \tag{$\mathbf{L}_\gamma$}$$

We remark that this problem coincides with the problem $(\mathbf{L}_\gamma)$ in Herzog and Meyer [2011] (with γ replaced by $\gamma/2$). The unique solvability of $(\mathbf{L}_\gamma)$ was proved in [Herzog and Meyer, 2011, Proposition 4.4]. Its optimal solution is denoted by $(\boldsymbol{\Sigma}_\gamma, \boldsymbol{u}_\gamma)$ and it is characterized by

$$\begin{aligned} & a(\boldsymbol{\Sigma}_\gamma, \boldsymbol{T}) + b(\boldsymbol{T}, \boldsymbol{u}_\gamma) + \gamma\Big(\max\big\{0, 1 - \frac{\tilde{\sigma}_0}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|}\big\},\, \mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{T}\Big)_\Omega = 0, \\ & b(\boldsymbol{\Sigma}_\gamma, \boldsymbol{v}) = \langle \ell, \boldsymbol{v}\rangle \end{aligned} \tag{2.11}$$

for all $\boldsymbol{T} \in S^2$ and $\boldsymbol{v} \in V$. In order to smooth this optimality system, we replace $\max\{0,\cdot\}$ by $\max_\varepsilon$. The non-differentiability in $x = 0$ is locally smoothed. We require the following conditions.

Assumption 2.6. *For all $\varepsilon > 0$, the function $\max_\varepsilon : \mathbb{R} \to \mathbb{R}$ is of class $C^{1,1}$ and satisfies*

(1) $\max_\varepsilon(x) \ge \max\{0, x\}$,
(2) $\max_\varepsilon$ *is monotone increasing and convex,*
(3) $\max_\varepsilon(x) = \max\{0, x\}$ *for* $|x| \ge \varepsilon$.

It is easy to see that there exists a class of functions satisfying these requirements, and we refrain from fixing a certain choice of $\max_\varepsilon$ here. This leaves a choice for numerical implementations.

It is convenient to define

$$J_{\gamma,\varepsilon}(\boldsymbol{\Sigma}) = p_{\gamma,\varepsilon}(|\mathcal{D}\boldsymbol{\Sigma}|)\, \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma} \quad \text{where} \quad p_{\gamma,\varepsilon}(x) = \max\nolimits_\varepsilon\big(\gamma\,(1 - \tilde{\sigma}_0/x)\big), \tag{2.12}$$

which acts pointwise on functions in S^2. We thus obtain the following smoothed version of the optimality condition (2.11):

$$a(\boldsymbol{\Sigma}_{\gamma,\varepsilon}, \boldsymbol{T}) + b(\boldsymbol{T}, \boldsymbol{u}_{\gamma,\varepsilon}) + \langle J_{\gamma,\varepsilon}(\boldsymbol{\Sigma}_{\gamma,\varepsilon}), \boldsymbol{T}\rangle = 0 \quad \text{for all } \boldsymbol{T} \in S^2, \tag{2.13a}$$

$$b(\boldsymbol{\Sigma}_{\gamma,\varepsilon}, \boldsymbol{v}) = \langle \ell, \boldsymbol{v}\rangle \quad \text{for all } \boldsymbol{v} \in V. \tag{2.13b}$$

Note that the expression $\langle J_{\gamma,\varepsilon}(\boldsymbol{\Sigma}_{\gamma,\varepsilon}), \boldsymbol{T}\rangle$ is well defined for $\boldsymbol{T} \in S^2$, since $J_{\gamma,\varepsilon}(\boldsymbol{\Sigma}_{\gamma,\varepsilon}) \in S^2$ due to $p_{\gamma,\varepsilon}(|\mathcal{D}\boldsymbol{\Sigma}_{\gamma,\varepsilon}|) \in L^\infty(\Omega)$. Note further that in general $J_{\gamma,\varepsilon}$ need not have a primitive. Therefore (2.13) cannot be viewed as optimality conditions for some regularized version of ($\mathbf{L}_\gamma$). Nevertheless, the existence and uniqueness of a solution can be shown by the theory of monotone operators. We begin by verifying the following properties of $J_{\gamma,\varepsilon}$.

Lemma 2.7. (1) $p_{\gamma,\varepsilon}(x) \le \max\{\gamma, \varepsilon\}$ *holds for* $x \in \mathbb{R}^+$.
(2) $J_{\gamma,\varepsilon} : S^2 \to S^2$ *is a monotone operator, and*
(3) $\|J_{\gamma,\varepsilon}(\boldsymbol{\Sigma})\|_{S^2} \ge 2\,\gamma\, \|\boldsymbol{\Sigma} - P_{\mathcal{K}}(\boldsymbol{\Sigma})\|_{S^2}$.

Proof. Property (1) is an immediate consequence of Assumption 2.6.

Let $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2 \in S^2$ and set $m = \min\{p_{\gamma,\varepsilon}(|\mathcal{D}\boldsymbol{\Sigma}_1|), p_{\gamma,\varepsilon}(|\mathcal{D}\boldsymbol{\Sigma}_2|)\} \in L^2(\Omega)$. Now consider

$$\begin{aligned}
&\langle J_{\gamma,\varepsilon}(\boldsymbol{\Sigma}_1) - J_{\gamma,\varepsilon}(\boldsymbol{\Sigma}_2),\, \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\rangle \\
&\quad = \langle p_{\gamma,\varepsilon}(|\mathcal{D}\boldsymbol{\Sigma}_1|)\, \mathcal{D}\boldsymbol{\Sigma}_1 - p_{\gamma,\varepsilon}(|\mathcal{D}\boldsymbol{\Sigma}_2|)\, \mathcal{D}\boldsymbol{\Sigma}_2,\, \mathcal{D}(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2)\rangle \\
&\quad = \langle m\, \mathcal{D}(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2),\, \mathcal{D}(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2)\rangle \\
&\qquad + \langle [p_{\gamma,\varepsilon}(|\mathcal{D}\boldsymbol{\Sigma}_1|) - m]\, \mathcal{D}\boldsymbol{\Sigma}_1,\, \mathcal{D}(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2)\rangle \\
&\qquad + \langle [-p_{\gamma,\varepsilon}(|\mathcal{D}\boldsymbol{\Sigma}_2|) + m]\, \mathcal{D}\boldsymbol{\Sigma}_2,\, \mathcal{D}(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2)\rangle.
\end{aligned}$$

The first term is non-negative since $m \ge 0$. For the second and third terms, a pointwise distinction of cases shows their pointwise non-negativity and we conclude (2). By (2.9) we have

$$\begin{aligned}
\|\boldsymbol{\Sigma} - P_{\mathcal{K}}(\boldsymbol{\Sigma})\|_{S^2}^2 &= \frac{1}{2}\int_\Omega \max\left\{0, |\mathcal{D}\boldsymbol{\Sigma}| - \tilde\sigma_0\right\}^2 \mathrm{d}x \\
&= \frac{1}{2}\int_\Omega \max\left\{0, 1 - \tilde\sigma_0/|\mathcal{D}\boldsymbol{\Sigma}|\right\}^2 |\mathcal{D}\boldsymbol{\Sigma}|^2 \,\mathrm{d}x.
\end{aligned}$$

Applying Assumption 2.6 (1), we obtain

$$\begin{aligned}
\gamma^2\, \|\boldsymbol{\Sigma} - P_{\mathcal{K}}(\boldsymbol{\Sigma})\|_{S^2}^2 &= \frac{1}{2}\int_\Omega \max\left\{0, \gamma\,(1 - \tilde\sigma_0/|\mathcal{D}\boldsymbol{\Sigma}|)\right\}^2 |\mathcal{D}\boldsymbol{\Sigma}|^2 \,\mathrm{d}x \\
&\le \frac{1}{2}\int_\Omega \max\nolimits_\varepsilon \left\{0, \gamma\,(1 - \tilde\sigma_0/|\mathcal{D}\boldsymbol{\Sigma}|)\right\}^2 |\mathcal{D}\boldsymbol{\Sigma}|^2 \,\mathrm{d}x = \frac{1}{4}\, \|J_{\gamma,\varepsilon}(\boldsymbol{\Sigma})\|_{S^2}^2,
\end{aligned}$$

which shows (3). □

With the monotonicity of $J_{\gamma,\varepsilon}$ established, we recognize (2.13) as a nonlinear saddle-point problem with a monotone (1,1) block. Existence and uniqueness of solutions follow from Lemma A.1 with the settings

$$X = S^2, \quad A\boldsymbol{\Sigma} = a(\boldsymbol{\Sigma}, \cdot), \quad B\boldsymbol{\Sigma} = b(\boldsymbol{\Sigma}, \cdot), \quad J = J_{\gamma,\varepsilon}.$$

Proposition 2.8. *For any* $\ell \in V'$, (2.13) *has a unique solution*

$$G_{\gamma,\varepsilon}(\ell) = (G^{\boldsymbol{\Sigma}}_{\gamma,\varepsilon}(\ell), G^{\boldsymbol{u}}_{\gamma,\varepsilon}(\ell)) = (\boldsymbol{\Sigma}_{\gamma,\varepsilon}, \boldsymbol{u}_{\gamma,\varepsilon}) \in S^2 \times V. \tag{2.14}$$

Moreover, $\boldsymbol{\Sigma}_{\gamma,\varepsilon}$ *and* $\boldsymbol{u}_{\gamma,\varepsilon}$ *depend Lipschitz continuously on* ℓ*, with a Lipschitz constant* L *independent of* γ *and* ε.

Proof. We only need to show the Lipschitz dependence of $\boldsymbol{u}_{\gamma,\varepsilon}$. This result does not hold in the general context of Lemma A.1, but we need to exploit the special structure of $J_{\gamma,\varepsilon}$. To this end, we consider arbitrary inhomogeneities $\ell, \ell' \in V'$ with associated solutions $(\boldsymbol{\Sigma}_{\gamma,\varepsilon}, \boldsymbol{u}_{\gamma,\varepsilon})$, $(\boldsymbol{\Sigma}'_{\gamma,\varepsilon}, \boldsymbol{u}'_{\gamma,\varepsilon}) \in S^2 \times V$. Moreover, we choose $\boldsymbol{T} = (\boldsymbol{\tau}, -\boldsymbol{\tau})$ with $\boldsymbol{\tau} \in S$ arbitrary as test function in (2.13a) for ℓ and ℓ', respectively. Due to $\mathcal{D}\boldsymbol{T} = 0$ and by (2.12), $\langle J_{\gamma,\varepsilon}(\boldsymbol{\Sigma}_{\gamma,\varepsilon}), \boldsymbol{T}\rangle = \langle J_{\gamma,\varepsilon}(\boldsymbol{\Sigma}'_{\gamma,\varepsilon}), \boldsymbol{T}\rangle = 0$ holds. Thus we arrive at

$$b(\boldsymbol{T}, \boldsymbol{u}_{\gamma,\varepsilon} - \boldsymbol{u}'_{\gamma,\varepsilon}) = -a(\boldsymbol{\Sigma}_{\gamma,\varepsilon} - \boldsymbol{\Sigma}'_{\gamma,\varepsilon}, \boldsymbol{T}) \le \|a\| \|\boldsymbol{\Sigma}_{\gamma,\varepsilon} - \boldsymbol{\Sigma}'_{\gamma,\varepsilon}\|_{S^2} \|\boldsymbol{T}\|_{S^2}.$$

We recall, e.g., from [Herzog and Meyer, 2011, (2.8)], that $b(\cdot,\cdot)$ satisfies an inf–sup condition due to Korn's inequality. This implies that we can find $\boldsymbol{\tau} \in \mathcal{S}$ such that $b((\boldsymbol{\tau}, \boldsymbol{0}), \boldsymbol{u}_{\gamma,\varepsilon} - \boldsymbol{u}'_{\gamma,\varepsilon}) \ge \underline{\beta}\, \|\boldsymbol{\tau}\|_S \|\boldsymbol{u}_{\gamma,\varepsilon} - \boldsymbol{u}'_{\gamma,\varepsilon}\|_V$ and thus

$$\begin{aligned} \underline{\beta}\, \|\boldsymbol{u}_{\gamma,\varepsilon} - \boldsymbol{u}'_{\gamma,\varepsilon}\|_V &\le \frac{b((\boldsymbol{\tau}, -\boldsymbol{\tau}), \boldsymbol{u}_{\gamma,\varepsilon} - \boldsymbol{u}'_{\gamma,\varepsilon})}{\|\boldsymbol{\tau}\|_S} \\ &= \sqrt{2}\, \frac{b(\boldsymbol{T}, \boldsymbol{u}_{\gamma,\varepsilon} - \boldsymbol{u}'_{\gamma,\varepsilon})}{\|\boldsymbol{T}\|_{S^2}} \le \sqrt{2}\, \|a\| \|\boldsymbol{\Sigma}_{\gamma,\varepsilon} - \boldsymbol{\Sigma}'_{\gamma,\varepsilon}\|_{S^2}. \end{aligned}$$

The Lipschitz continuity then follows from the one for $\boldsymbol{\Sigma}_{\gamma,\varepsilon}$. □

2.3. Differentiability of the Control-to-State Map. In this section we prove that the solution map $\ell \mapsto (\boldsymbol{\Sigma}_{\gamma,\varepsilon}, \boldsymbol{u}_{\gamma,\varepsilon})$ of (2.13) is Fréchet differentiable from V' to $S^2 \times V$, see Theorem 2.12. This is a nontrivial result since $J_{\gamma,\varepsilon}$ is a nonlinear operator which acts pointwise (Nemytzki operator). And hence $J_{\gamma,\varepsilon}$ itself is not differentiable from S^2 to S^2, see, e.g., [Tröltzsch, 2010, Section 4.3.2] or Krasnoselskii et al. [1976].

The proof relies on Lemma A.2, which will be applied with the setting

$$Y_\delta = L^{2+\delta}(\Omega; \mathbb{S}^2), \quad W' = U = L^2(\Omega; \mathbb{R}^d) \times L^2(\Gamma_N; \mathbb{R}^d)$$

with some $\delta > 0$ specified in the sequel. The embedding $W' \hookrightarrow V'$ is given by

$$\langle R(\boldsymbol{f}, \boldsymbol{g}), \boldsymbol{v}\rangle := -\int_\Omega \boldsymbol{f} \cdot \boldsymbol{v}\, \mathrm{d}x - \int_{\Gamma_N} \boldsymbol{g} \cdot \boldsymbol{v}\, \mathrm{d}s, \quad \boldsymbol{v} \in V \tag{2.15}$$

for $(\boldsymbol{f}, \boldsymbol{g}) \in W'$. In order to apply Lemma A.2 we need to verify a Lipschitz condition of $G^{\boldsymbol{\Sigma}}_{\gamma,\varepsilon} : W' \to Y$ and the differentiability of $J_{\gamma,\varepsilon} : Y \to X$, which will be established in the following two propositions.

The first proposition relies on a recent regularity result for nonlinear elasticity systems Herzog et al. [2011]. In order to apply it, we need to reduce (2.13) to the displacement component $\boldsymbol{u}$. This requires the invertibility of $A + J_{\gamma,\varepsilon}$, which is addressed in the following lemma.

Lemma 2.9. *Let $\gamma, \varepsilon > 0$ be given. Then, for all $\delta \ge 0$, the operator $A + J_{\gamma,\varepsilon}$ maps $Y_\delta \to Y_\delta$ and it is invertible with Lipschitz continuous inverse.*

Proof. We recall from (2.4) and (2.12) that $A + J_{\gamma,\varepsilon}$ acts pointwise on functions in Y_δ. Let us define the pointwise operators $A_x, J_x : \mathbb{S}^2 \to \mathbb{S}^2$

$$A_x \boldsymbol{\Sigma} = \begin{pmatrix} \mathbb{C}(x)^{-1} \boldsymbol{\sigma} \\ \mathbb{H}(x)^{-1} \boldsymbol{\chi} \end{pmatrix}, \quad J_x(\boldsymbol{\Sigma}) = p_{\gamma,\varepsilon}(|\mathcal{D}\boldsymbol{\Sigma}|)\, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}.$$

These satisfy $|A_x \boldsymbol{\Sigma}| \le c\, |\boldsymbol{\Sigma}|$ and $|J_x(\boldsymbol{\Sigma})| \le c\, |\boldsymbol{\Sigma}|$ with a constant independent of x. This follows from the general Assumption (3) in Section 1 and from Lemma 2.7 (1). Consequently, $A + J_{\gamma,\varepsilon}$ maps $Y_\delta \to Y_\delta$.

Due to the same assumption, A_x is coercive on $\mathbb{S}^2$ with constant $\underline{\alpha}$ independent of x, and J_x is monotone and continuous. Hence $A_x + J_x$ is strongly monotone and hemi-continuous, and the Browder-Minty theorem implies the existence of a Lipschitz continuous inverse, with Lipschitz constant $\underline{\alpha}^{-1}$, independent of x.

From here it is easy to see that $(A + J_{\gamma,\varepsilon})^{-1}$ defined by the pointwise inverse, maps $Y_\delta \to Y_\delta$. □

Proposition 2.10. *For any $\gamma, \varepsilon > 0$, there exists $\delta > 0$ such that for any $(\boldsymbol{f}, \boldsymbol{g}) \in W'$, the solution $\boldsymbol{\Sigma}_{\gamma,\varepsilon}$ belongs to Y_δ. Moreover, $G^{\boldsymbol{\Sigma}}_{\gamma,\varepsilon}$ is globally Lipschitz $W' \to Y_\delta$.*

Proof. The reduction of (2.13) to the displacement variable is given by

$$B\,(A + J_{\gamma,\varepsilon})^{-1}(-B^\star \boldsymbol{u}_{\gamma,\boldsymbol{\varepsilon}}) = (\boldsymbol{f}, \boldsymbol{g}).$$

This nonlinear elasticity equation fits into the setting of [Herzog et al., 2011, Theorem 1.1], which implies the $W^{1,2+\delta}$ regularity for $\boldsymbol{u}_{\gamma,\boldsymbol{\varepsilon}}$ and its Lipschitz dependence on the data with some $\delta > 0$. Due to the Lipschitz continuity of $(A + J_{\gamma,\varepsilon})^{-1}$ (Lemma 2.9) and since

$$\boldsymbol{\Sigma}_{\gamma,\varepsilon} = (A + J_{\gamma,\varepsilon})^{-1}(-B^\star \boldsymbol{u}_{\gamma,\boldsymbol{\varepsilon}}) = (A + J_{\gamma,\varepsilon})^{-1}\boldsymbol{\varepsilon}(\boldsymbol{u}_{\gamma,\boldsymbol{\varepsilon}}),$$

the assertion is proved. □

Proposition 2.11. *Let $\gamma, \varepsilon > 0$ be given. Then for any $\delta > 0$, $J_{\gamma,\varepsilon} : Y_\delta \to S^2$ is Fréchet differentiable. The derivative $J_{\gamma,\varepsilon}(\boldsymbol{\Sigma})$ is given by*

$$J'_{\gamma,\varepsilon}(\boldsymbol{\Sigma})\,\boldsymbol{T} = p'_{\gamma,\varepsilon}(|\mathcal{D}\boldsymbol{\Sigma}|)\,\frac{\mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{T}}{|\mathcal{D}\boldsymbol{\Sigma}|}\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma} + p_{\gamma,\varepsilon}(|\mathcal{D}\boldsymbol{\Sigma}|)\,\mathcal{D}^\star\mathcal{D}\boldsymbol{T} \tag{2.16}$$

with

$$p'_{\gamma,\varepsilon}(x) = \max\nolimits'_\varepsilon\big(\gamma\,(1 - \tilde\sigma_0\,x^{-1})\big)\,\gamma\,\tilde\sigma_0\,x^{-2}.$$

It maps $S^2 \to S^2$ and it is positive semidefinite.

Proof. The result follows from general differentiability results for nonlinear Nemytzki operators, e.g., [Goldberg et al., 1992, Theorem 7], [Tröltzsch, 2010, Section 4.3.3]. Let us verify the conditions for the setting

$$p = 2 + \delta, \quad q = 2, \quad r = \tfrac{p\,q}{p-q} = \tfrac{2\,(2+\delta)}{\delta} = 2 + \tfrac{4}{\delta}.$$

We have already seen in Section 2.2 that $J_{\gamma,\varepsilon}$ maps S^2 into S^2 and hence it maps $Y_\delta = L^p(\Omega; \mathbb{S}^2)$ into $S^2 = L^q(\Omega; \mathbb{S}^2)$.

We observe next that $\max'_\varepsilon : \mathbb{R} \to \mathbb{R}$ is bounded. This follows from the compactness of $[-\varepsilon, \varepsilon]$ and $\max_\varepsilon(x) = \max\{0, x\}$ for $|x| \geq \varepsilon$. Furthermore, $\max'_\varepsilon$ is globally Lipschitz by assumption. We recall from the proof of Lemma 2.9 the mapping $J_x : \mathbb{S}^2 \to \mathbb{S}^2$ defined by $J_x(\boldsymbol{\Sigma}) = p_{\gamma,\varepsilon}(|\mathcal{D}\boldsymbol{\Sigma}|)\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}$. Moreover, J_x is continuously differentiable from $\mathbb{S}^2$ into $\mathbb{S}^2$ and thus it satisfies the Carathéodory condition.

The boundedness of $\max'_\varepsilon$ implies $\|J'_x(\boldsymbol{\Sigma})\|_{\mathcal{L}(\mathbb{S}^2,\mathbb{S}^2)} \leq L$ for all $\boldsymbol{\Sigma} \in \mathbb{S}^2$. The constant L depends on γ and ε. We conclude that the Nemytzki operator generated by J'_x maps $L^p(\Omega; \mathbb{S}^2)$ into $L^\infty(\Omega; \mathcal{L}(\mathbb{S}^2, \mathbb{S}^2))$ and in particular into $L^r(\Omega; \mathcal{L}(\mathbb{S}^2, \mathbb{S}^2))$. All conditions are verified. □

The main result of this section now follows from Lemma A.2 as announced in the beginning of this section.

Theorem 2.12. *For any $\gamma, \varepsilon > 0$, the solution map $G_{\gamma,\varepsilon} : \ell \mapsto (\boldsymbol{\Sigma}_{\gamma,\varepsilon}, \boldsymbol{u}_{\gamma,\varepsilon})$ of (2.13) is Fréchet differentiable from U to $S^2 \times V$. The derivative at $(\boldsymbol{\Sigma}_{\gamma,\varepsilon}, \boldsymbol{u}_{\gamma,\varepsilon}) = G_{\gamma,\varepsilon}(\ell)$ in the direction $\delta\ell \in U$ is given by the unique solution $(\delta\boldsymbol{\Sigma}, \delta\boldsymbol{u})$ of*

$$(A + J'_{\gamma,\varepsilon}(\boldsymbol{\Sigma}_{\gamma,\varepsilon}))\,\delta\boldsymbol{\Sigma} + B^\star\delta\boldsymbol{u} = \boldsymbol{0}, \tag{2.17a}$$

$$B\,\delta\boldsymbol{\Sigma} = \delta\ell. \tag{2.17b}$$

2.4. Convergence. As the final result for the lower-level problem, we show that the regularization is consistent with the original problem (**L**), i.e., we show the convergence of $\mathbf{\Sigma}_{\gamma,\varepsilon}$, $\boldsymbol{u}_{\gamma,\varepsilon}$ and $\lambda_{\gamma,\varepsilon}$ as $\gamma \to \infty$ and $\varepsilon \to 0$.

We first prove two preliminary results. This first lemma shows that the solutions $\mathbf{\Sigma}_{\gamma,\varepsilon}$ are admissible for $\mathbf{\Sigma} \in \mathcal{K}$ in the limit.

Lemma 2.13. *Let $\ell \in V'$. Then the solution of (2.13) satisfies*

$$\|\mathbf{\Sigma}_{\gamma,\varepsilon} - P_{\mathcal{K}}(\mathbf{\Sigma}_{\gamma,\varepsilon})\|_{S^2} \leq \frac{C}{\gamma}\|\ell\|_{V'}, \tag{2.18}$$

where $C > 0$ is independent of γ, ε and ℓ.

Proof. By testing (2.13a) with $J_{\gamma,\varepsilon}(\mathbf{\Sigma}_{\gamma,\varepsilon})$ one obtains

$$\begin{aligned}\|J_{\gamma,\varepsilon}(\mathbf{\Sigma}_{\gamma,\varepsilon})\|_{S^2}^2 &\leq |a(\mathbf{\Sigma}_{\gamma,\varepsilon}, J_{\gamma,\varepsilon}(\mathbf{\Sigma}_{\gamma,\varepsilon}))| + |b(J_{\gamma,\varepsilon}(\mathbf{\Sigma}_{\gamma,\varepsilon}), \boldsymbol{u}_{\gamma,\varepsilon})| \\ &\leq \|a\|\|\mathbf{\Sigma}_{\gamma,\varepsilon}\|_{S^2}\|J_{\gamma,\varepsilon}(\mathbf{\Sigma}_{\gamma,\varepsilon})\|_{S^2} + \|b\|\|J_{\gamma,\varepsilon}(\mathbf{\Sigma}_{\gamma,\varepsilon})\|_{S^2}\|\boldsymbol{u}_{\gamma,\varepsilon}\|_V \\ &\leq \max\{\|a\|, \|b\|\}\, \|J_{\gamma,\varepsilon}(\mathbf{\Sigma}_{\gamma,\varepsilon})\|_{S^2}\big(\|\mathbf{\Sigma}_{\gamma,\varepsilon}\|_{S^2} + \|\boldsymbol{u}_{\gamma,\varepsilon}\|_V\big).\end{aligned}$$

Lemma 2.7 (3) and Proposition 2.8 conclude the proof. □

The second lemma estimates the variation of the penalty term in admissible directions. Here and throughout, $\mu(E)$ denotes the Lebesgue measure of a set E.

Lemma 2.14. *Let $\mathbf{\Sigma} \in S^2$ and $\mathbf{\Sigma}_2 \in \mathcal{K}$ be given. Then*

$$\langle J_{\gamma,\varepsilon}(\mathbf{\Sigma}),\, \mathbf{\Sigma}_2 - \mathbf{\Sigma}\rangle \leq 2\,\tilde{\sigma}_0^2\, p_{\gamma,\varepsilon}(\tilde{\sigma}_0)\, \mu(\Omega)$$

holds.

Proof. Since $\mathbf{\Sigma}_2 \in \mathcal{K}$ we have $|\mathcal{D}\mathbf{\Sigma}_2| \leq \tilde{\sigma}_0$ a.e. If $|\mathcal{D}\mathbf{\Sigma}(x)| \geq \tilde{\sigma}_0$ then $\mathcal{D}\mathbf{\Sigma}(x) : (\mathcal{D}\mathbf{\Sigma}_2(x) - \mathcal{D}\mathbf{\Sigma}(x)) \leq 0$. On the other hand, if $|\mathcal{D}\mathbf{\Sigma}(x)| < \tilde{\sigma}_0$ we have $\mathcal{D}\mathbf{\Sigma}(x) : (\mathcal{D}\mathbf{\Sigma}_2(x) - \mathcal{D}\mathbf{\Sigma}(x)) \leq 2\,\tilde{\sigma}_0^2$. Hence we can estimate

$$\langle J_{\gamma,\varepsilon}(\mathbf{\Sigma}), \mathbf{\Sigma}_2 - \mathbf{\Sigma}\rangle = \int_\Omega p_{\gamma,\varepsilon}(|\mathcal{D}\mathbf{\Sigma}|)\, \mathcal{D}\mathbf{\Sigma} : (\mathcal{D}\mathbf{\Sigma}_2 - \mathcal{D}\mathbf{\Sigma})\,\mathrm{d}x \leq 2\,\tilde{\sigma}_0^2\, p_{\gamma,\varepsilon}(\tilde{\sigma}_0)\,\mu(\Omega),$$

which yields the assertion. □

The following theorem shows an error estimate for the solution of the regularized lower-level problems.

Theorem 2.15. *Let us denote by $(\mathbf{\Sigma}, \boldsymbol{u})$ the solution of (2.1) with right hand side $\ell \in V'$ and by $\mathbf{\Sigma}_{\gamma,\varepsilon}$ the solutions of the regularized problems (2.13) with right hand side $\ell_{\gamma,\varepsilon}$ for $\gamma, \varepsilon > 0$. Then we obtain*

$$\begin{aligned}\|\mathbf{\Sigma} - \mathbf{\Sigma}_{\gamma,\varepsilon}\|_{S^2}^2 &\leq C\,\left(\|\ell - \ell_{\gamma,\varepsilon}\|_{V'}\|\boldsymbol{u} - \boldsymbol{u}_{\gamma,\varepsilon}\|_V + \gamma^{-1}\,\|\ell\|_{V'}\|\ell_{\gamma,\varepsilon}\|_{V'} + \varepsilon\right), \\ \|\boldsymbol{u} - \boldsymbol{u}_{\gamma,\varepsilon}\|_V &\leq C\,\big(\|\ell - \ell_{\gamma,\varepsilon}\|_{V'}\|\boldsymbol{u} - \boldsymbol{u}_{\gamma,\varepsilon}\|_V + \gamma^{-1}\,\|\ell\|_{V'}\|\ell_{\gamma,\varepsilon}\|_{V'} + \varepsilon \\ &\qquad + \|\mathbf{\Sigma} - \mathbf{\Sigma}_{\gamma,\varepsilon}\|_{S^2}\big)\end{aligned}$$

where C is independent of ℓ, $\ell_{\gamma,\varepsilon}$, γ and ε.

Proof. The proof is based on the proof of [Herzog and Meyer, 2011, Theorem 4.10], which shows the result for $\varepsilon = 0$ and $\ell = \ell_{\gamma,\varepsilon}$. By Assumption 2.6 (3) and Lemma 2.14 we have

$$\big(J_{\gamma,\varepsilon}(\mathbf{\Sigma}_{\gamma,\varepsilon}), \mathbf{\Sigma} - \mathbf{\Sigma}_{\gamma,\varepsilon}\big) \leq C\,\varepsilon, \tag{2.19}$$

with $C = 2\,\mu(\Omega)\,\tilde{\sigma}_0^2$ independent of γ, ε.

Let $\boldsymbol{\tau} \in S$ be arbitrary. We set $\widetilde{\boldsymbol{T}} = (\boldsymbol{\tau}, -\boldsymbol{\tau}) \in S^2$.

Testing the VI in (**L**) with $\boldsymbol{T} = P_{\mathcal{K}}(\boldsymbol{\Sigma}_{\gamma,\varepsilon}) - \widetilde{\boldsymbol{T}} \in \mathcal{K}$ yields

$$a(\boldsymbol{\Sigma}, P_{\mathcal{K}}(\boldsymbol{\Sigma}_{\gamma,\varepsilon}) - \widetilde{\boldsymbol{T}} - \boldsymbol{\Sigma}) \geq b(\boldsymbol{\Sigma} - P_{\mathcal{K}}(\boldsymbol{\Sigma}_{\gamma,\varepsilon}) + \widetilde{\boldsymbol{T}}, \boldsymbol{u}).$$

Testing (2.13a) with $\boldsymbol{T} = \boldsymbol{\Sigma} + \widetilde{\boldsymbol{T}} - \boldsymbol{\Sigma}_{\gamma,\varepsilon}$ leads to

$$a(\boldsymbol{\Sigma}_{\gamma,\varepsilon}, \boldsymbol{\Sigma} + \widetilde{\boldsymbol{T}} - \boldsymbol{\Sigma}_{\gamma,\varepsilon}) = b(\boldsymbol{\Sigma}_{\gamma,\varepsilon} - \boldsymbol{\Sigma} - \widetilde{\boldsymbol{T}}, \boldsymbol{u}_{\gamma,\varepsilon}) + \big(J_{\gamma,\varepsilon}(\boldsymbol{\Sigma}_{\gamma,\varepsilon}), \boldsymbol{\Sigma}_{\gamma,\varepsilon} - \boldsymbol{\Sigma} - \widetilde{\boldsymbol{T}}\big).$$

Adding this inequality and equality yields

$$\begin{aligned}
&a(\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{\gamma,\varepsilon}, \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{\gamma,\varepsilon}) + b(\widetilde{\boldsymbol{T}}, \boldsymbol{u} - \boldsymbol{u}_{\gamma,\varepsilon}) \\
&\leq a(\boldsymbol{\Sigma}_{\gamma,\varepsilon} - \boldsymbol{\Sigma}, \widetilde{\boldsymbol{T}}) + a(\boldsymbol{\Sigma}, P_{\mathcal{K}}(\boldsymbol{\Sigma}_{\gamma,\varepsilon}) - \boldsymbol{\Sigma}_{\gamma,\varepsilon}) \\
&\quad - b(\boldsymbol{\Sigma}_{\gamma,\varepsilon} - P_{\mathcal{K}}(\boldsymbol{\Sigma}_{\gamma,\varepsilon}), \boldsymbol{u}) - b(\boldsymbol{\Sigma}_{\gamma,\varepsilon} - \boldsymbol{\Sigma}, \boldsymbol{u}_{\gamma,\varepsilon} - \boldsymbol{u}) - \big(J_{\gamma,\varepsilon}(\boldsymbol{\Sigma}_{\gamma,\varepsilon}), \boldsymbol{\Sigma}_{\gamma,\varepsilon} - \boldsymbol{\Sigma} - \widetilde{\boldsymbol{T}}\big) \\
&\leq \|a\| \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{\gamma,\varepsilon}\|_{S^2} \, \|\widetilde{\boldsymbol{T}}\|_{S^2} + \|\ell - \ell_{\gamma,\varepsilon}\|_{V'} \|\boldsymbol{u} - \boldsymbol{u}_{\gamma,\varepsilon}\|_V + C\,\gamma^{-1} \, \|\ell\|_{V'} \|\ell_{\gamma,\varepsilon}\|_{V'} + C\,\varepsilon,
\end{aligned}$$

where we used that $b(\boldsymbol{\Sigma}_{\gamma,\varepsilon} - \boldsymbol{\Sigma}, \boldsymbol{u}_{\gamma,\varepsilon} - \boldsymbol{u}) = \langle \ell_{\gamma,\varepsilon} - \ell, \, \boldsymbol{u}_{\gamma,\varepsilon} - \boldsymbol{u}\rangle$ and $\mathcal{D}\widetilde{\boldsymbol{T}} = 0$. Moreover, we employed Proposition 2.8, Lemma 2.13 and (2.19) for the last estimate.

This result is used with two different choices of $\widetilde{\boldsymbol{T}}$ to obtain the rates for $\boldsymbol{\Sigma}_{\gamma,\varepsilon}$ and $\boldsymbol{u}_{\gamma,\varepsilon}$.

Rate of $\{\boldsymbol{\Sigma}_{\gamma,\varepsilon}\}$: Choosing $\boldsymbol{\tau} = 0$ and the coercivity of a yields

$$\underline{\alpha} \, \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{\gamma,\varepsilon}\|_{S^2}^2 \leq \|\ell - \ell_{\gamma,\varepsilon}\|_{V'} \|\boldsymbol{u} - \boldsymbol{u}_{\gamma,\varepsilon}\|_V + C\,\gamma^{-1} \, \|\ell\|_{V'} \|\ell_{\gamma,\varepsilon}\|_{V'} + C\,\varepsilon$$

Rate of $\{\boldsymbol{u}_k\}$: By the inf-sup condition of b we have

$$\begin{aligned}
\underline{\beta} \, \|\boldsymbol{u} - \boldsymbol{u}_{\gamma,\varepsilon}\|_V &\leq \sup_{\|\boldsymbol{\tau}\|=1} b((\boldsymbol{\tau}, \boldsymbol{0}), \boldsymbol{u} - \boldsymbol{u}_{\gamma,\varepsilon}) = \sup_{\|\boldsymbol{\tau}\|=1} b(\widetilde{\boldsymbol{T}}, \boldsymbol{u} - \boldsymbol{u}_{\gamma,\varepsilon}) \\
&\leq \|\ell - \ell_{\gamma,\varepsilon}\|_{V'} \|\boldsymbol{u} - \boldsymbol{u}_{\gamma,\varepsilon}\|_V + C\,\gamma^{-1} \, \|\ell\|_{V'} \|\ell_{\gamma,\varepsilon}\|_{V'} \\
&\quad + C\,\varepsilon + \sqrt{2} \, \|a\| \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{\gamma,\varepsilon}\|_{S^2}.
\end{aligned}$$

□

In the sequel, we will frequently discuss the situation where $\ell_{\gamma,\varepsilon} \to \ell$ as $\gamma \to \infty$ and $\varepsilon \to 0$. We simply refer to this as $\ell_{\gamma,\varepsilon} \to \ell$.

Corollary 2.16. *For $\ell_{\gamma,\varepsilon} \to \ell$, we obtain (strong) convergence of $(\boldsymbol{\Sigma}_{\gamma,\varepsilon}, \boldsymbol{u}_{\gamma,\varepsilon}) \to (\boldsymbol{\Sigma}, \boldsymbol{u})$.*

The comparison of (2.13a) and (2.2a) gives rise to the definition

$$\lambda_{\gamma,\varepsilon} := p_{\gamma,\varepsilon}(|\mathcal{D}\boldsymbol{\Sigma}_{\gamma,\varepsilon}|). \tag{2.20}$$

From the definition of $p_{\gamma,\varepsilon}$, we see that $0 \leq \lambda_{\gamma,\varepsilon} \leq \max\{\gamma, \varepsilon\}$ holds. The last result of this section concerns the convergence $\lambda_{\gamma,\varepsilon} \to \lambda$. This is done in three steps.

Lemma 2.17. *With the notation of Theorem 2.15, let $\ell_{\gamma,\varepsilon} \to \ell$. Then*

$$\lambda_{\gamma,\varepsilon} \mathcal{D}\boldsymbol{\Sigma}_{\gamma,\varepsilon} \to \lambda \, \mathcal{D}\boldsymbol{\Sigma} \quad \textit{in } S.$$

Proof. We test (2.13a) and (2.2a) with $\boldsymbol{T} = (\boldsymbol{0}, \boldsymbol{\tau})$ and subtract to obtain

$$\langle \lambda \, \mathcal{D}\boldsymbol{\Sigma} - \lambda_{\gamma,\varepsilon} \mathcal{D}\boldsymbol{\Sigma}_{\gamma,\varepsilon}, \, \mathcal{D}\boldsymbol{T} \rangle = -a(\boldsymbol{T}, \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_{\gamma,\varepsilon})$$

for arbitrary $\boldsymbol{\tau} \in S$. Choosing $\boldsymbol{\tau} = \lambda \, \mathcal{D}\boldsymbol{\Sigma} - \lambda_{\gamma,\varepsilon} \mathcal{D}\boldsymbol{\Sigma}_{\gamma,\varepsilon}$ and using Corollary 2.16 finishes the proof. □

Lemma 2.18. *With the notation of Theorem 2.15, let $\ell_{\gamma,\varepsilon} \to \ell$. Then there exists a weakly convergent subsequence*

$$\lambda_{\gamma',\varepsilon'} \rightharpoonup \bar{\lambda} \quad \textit{in } L^2(\Omega).$$

The weak limit satisfies $\bar{\lambda} = \lambda$ a.e. on the set $\{x \in \Omega : \mathcal{D}\boldsymbol{\Sigma}(x) \neq 0\}$.

Proof. Due to the convergence of $\lambda_{\gamma,\varepsilon}\mathcal{D}\boldsymbol{\Sigma}_{\gamma,\varepsilon}$ in S we have

$$\|\lambda_{\gamma,\varepsilon}\mathcal{D}\boldsymbol{\Sigma}_{\gamma,\varepsilon}\|_S \leq C.$$

We also obtain $\lambda_{\gamma,\varepsilon} = 0$ on $A^-_{\gamma,\varepsilon} := \{x \in \Omega : \gamma\bigl(1 - \frac{\tilde{\sigma}_0}{|\mathcal{D}\boldsymbol{\Sigma}_{\gamma,\varepsilon}|}\bigr) \leq -\varepsilon\}$, see Assumption 2.6 (3). Now we can estimate

$$\begin{aligned}
\|\lambda_{\gamma,\varepsilon}\mathcal{D}\boldsymbol{\Sigma}_{\gamma,\varepsilon}\|_S^2 &= \int_{\Omega\setminus A^-_{\gamma,\varepsilon}} |\lambda_{\gamma,\varepsilon}|^2\,|\mathcal{D}\boldsymbol{\Sigma}_{\gamma,\varepsilon}|^2\,\mathrm{d}x \\
&\geq \bigl(\frac{\tilde{\sigma}_0\,\gamma}{\gamma+\varepsilon}\bigr)^2 \int_{\Omega\setminus A^-_{\gamma,\varepsilon}} |\lambda_{\gamma,\varepsilon}|^2\,\mathrm{d}x \\
&= \bigl(\frac{\tilde{\sigma}_0\,\gamma}{\gamma+\varepsilon}\bigr)^2\,\|\lambda_{\gamma,\varepsilon}\|^2_{L^2(\Omega)}.
\end{aligned}$$

Thus, the norm of $\lambda_{\gamma,\varepsilon}$ is bounded and there exists a weakly convergent subsequence $\lambda_{\gamma',\varepsilon'} \rightharpoonup \bar{\lambda}$ in $L^2(\Omega)$. Due to the strong convergence of the stresses we obtain

$$\lambda_{\gamma',\varepsilon'}\,\mathcal{D}\boldsymbol{\Sigma}_{\gamma',\varepsilon'} \rightharpoonup \bar{\lambda}\,\mathcal{D}\boldsymbol{\Sigma} \quad \text{in } L^1(\Omega;\mathbb{S}).$$

Thus we conclude from Lemma 2.17 that $\bar{\lambda}\,\mathcal{D}\boldsymbol{\Sigma} = \lambda\,\mathcal{D}\boldsymbol{\Sigma}$ must hold. Consequently, $\bar{\lambda} = \lambda$ holds on the set $\{x \in \Omega : \mathcal{D}\boldsymbol{\Sigma} \neq 0\}$. □

Finally, the next theorem shows the terms $\lambda_{\gamma,\varepsilon}$ converge strongly to the unique multiplier λ of the unregularized lower-level problem.

Theorem 2.19. *With the notation of Theorem 2.15, let $\ell_{\gamma,\varepsilon} \to \ell$. Then $\lambda_{\gamma,\varepsilon} \to \lambda$ holds in $L^2(\Omega)$.*

Proof. Due to $\lambda_{\gamma',\varepsilon'} \rightharpoonup \bar{\lambda}$ in $L^2(\Omega)$ and the weak lower semicontinuity of the norm we find

$$\liminf \|\lambda_{\gamma',\varepsilon'}\|_{L^2(\Omega)} \geq \|\bar{\lambda}\|_{L^2(\Omega)}.$$

From the previous proof we get the estimate

$$\|\lambda_{\gamma,\varepsilon}\mathcal{D}\boldsymbol{\Sigma}_{\gamma,\varepsilon}\|_S \geq \bigl(\frac{\tilde{\sigma}_0\,\gamma}{\gamma+\varepsilon}\bigr)\,\|\lambda_{\gamma,\varepsilon}\|_{L^2(\Omega)}.$$

Thus

$$\begin{aligned}
\|\bar{\lambda}\|_{L^2(\Omega)} &\leq \liminf \|\lambda_{\gamma',\varepsilon'}\|_{L^2(\Omega)} \leq \limsup \|\lambda_{\gamma',\varepsilon'}\|_{L^2(\Omega)} \\
&\leq \limsup \frac{\gamma'+\varepsilon'}{\tilde{\sigma}_0\,\gamma'}\|\lambda_{\gamma',\varepsilon'}\,\mathcal{D}\boldsymbol{\Sigma}_{\gamma',\varepsilon'}\|_S = \frac{1}{\tilde{\sigma}_0}\|\lambda\,\mathcal{D}\boldsymbol{\Sigma}\|_S = \|\lambda\|_{L^2(\Omega)}
\end{aligned}$$

in view of the complementarity condition (2.2c). Since $\bar{\lambda} = \lambda$ on $\{x \in \Omega : \mathcal{D}\boldsymbol{\Sigma} \neq \mathbf{0}\}$ was already shown in the previous lemma and since $\lambda = 0$ holds on $\{x \in \Omega : \mathcal{D}\boldsymbol{\Sigma} = \mathbf{0}\}$ we deduce from $\|\bar{\lambda}\|_{L^2(\Omega)} \leq \|\lambda\|_{L^2(\Omega)}$ that $\bar{\lambda} = \lambda$ holds a.e. in Ω. The estimate above further implies $\|\lambda_{\gamma',\varepsilon'}\|_{L^2(\Omega)} \to \|\lambda\|_{L^2(\Omega)}$ and together with the weak convergence this yields the strong convergence $\lambda_{\gamma',\varepsilon'} \to \lambda$ in $L^2(\Omega)$.

This shows that the limit is independent of the particular subsequence $\lambda_{\gamma',\varepsilon'}$ of $\lambda_{\gamma,\varepsilon}$. Hence we can deduce that the whole sequence converges to λ. □

3 Optimality Conditions for the Upper-Level Problem

In this section we consider the upper-level problem

$$\left.\begin{aligned}
\text{Minimize}\quad & \frac{1}{2}\|\boldsymbol{u}-\boldsymbol{u}_d\|^2_{L^2(\Omega;\mathbb{R}^d)} + \frac{\nu_1}{2}\|\boldsymbol{f}\|^2_{L^2(\Omega;\mathbb{R}^d)} + \frac{\nu_2}{2}\|\boldsymbol{g}\|^2_{L^2(\Gamma_N;\mathbb{R}^d)} \\
\text{s.t.}\quad & ((\boldsymbol{f},\boldsymbol{g}),\boldsymbol{\Sigma},\boldsymbol{u}) \in U \times S^2 \times V \\
\text{and}\quad & (\boldsymbol{\Sigma},\boldsymbol{u}) \text{ solves the static plasticity problem } (\mathbf{L})
\end{aligned}\right\} \qquad (\mathbf{P})$$

variable	definition	convergence
$(\boldsymbol{f}_\gamma, \boldsymbol{g}_\gamma) \in U$	local solution to ($\mathbf{P}_\gamma$)	strongly in U
$\boldsymbol{\Sigma}_\gamma \in S^2$	local solution to ($\mathbf{P}_\gamma$)	strongly in S^2
$\boldsymbol{u}_\gamma \in V$	local solution to ($\mathbf{P}_\gamma$)	strongly in V
$\lambda_\gamma \in L^2(\Omega)$	$p_\gamma(\lvert\mathcal{D}\boldsymbol{\Sigma}_\gamma\rvert)$, see (2.20)	strongly in $L^2(\Omega)$
$\boldsymbol{\Upsilon}_\gamma \in S^2$	Theorem 3.1	weakly in S^2
$\boldsymbol{w}_\gamma \in V$	Theorem 3.1	weakly in V
$\mu_\gamma \in L^2(\Omega)$	(3.7a)	weakly in $L^2(\Omega)$
$\theta_\gamma \in L^2(\Omega)$	(3.7b)	weakly in $L^2(\Omega)$
$\boldsymbol{Q}_\gamma \in S^2$	(3.12)	weakly in S^2

Table 3.1. Variables associated with the regularized control problem ($\mathbf{P}_\gamma$) and their convergence as proved in Theorem 3.16. Note that the dependence on the smoothing ε is suppressed since $\varepsilon \to 0$ as $\gamma \to \infty$.

and the regularized upper-level problem

$$\left.\begin{aligned} \text{Minimize} \quad & \frac{1}{2}\|\boldsymbol{u}_\gamma - \boldsymbol{u}_d\|^2_{L^2(\Omega;\mathbb{R}^d)} + \frac{\nu_1}{2}\|\boldsymbol{f}_\gamma\|^2_{L^2(\Omega;\mathbb{R}^d)} + \frac{\nu_2}{2}\|\boldsymbol{g}_\gamma\|^2_{L^2(\Gamma_N;\mathbb{R}^d)} \\ \text{s.t.} \quad & ((\boldsymbol{f}_\gamma, \boldsymbol{g}_\gamma), \boldsymbol{\Sigma}_\gamma, \boldsymbol{u}_\gamma) \in U \times S^2 \times V \\ \text{and} \quad & (\boldsymbol{\Sigma}_\gamma, \boldsymbol{u}_\gamma) \text{ solves the regularized problem (2.13)} \end{aligned}\right\} \quad (\mathbf{P}_\gamma)$$

The goal of this section is to derive optimality conditions for ($\mathbf{P}$). Since the constraint in ($\mathbf{P}$) are not differentiable, we perform the following steps.

(1) We derive optimality conditions for ($\mathbf{P}_\gamma$) in Section 3.1.
(2) We discuss in Section 3.2 which (local or global) optimal controls $(\boldsymbol{f}, \boldsymbol{g})$ of ($\mathbf{P}$) can be approximated by stationary controls $(\boldsymbol{f}_\gamma, \boldsymbol{g}_\gamma)$ satisfying the optimality conditions of ($\mathbf{P}_\gamma$) in the sense that $(\boldsymbol{f}_\gamma, \boldsymbol{g}_\gamma) \to (\boldsymbol{f}, \boldsymbol{g})$.
(3) We pass to the limit in the optimality conditions in Section 3.3 to obtain optimality conditions for ($\mathbf{P}$).

Notice that from now on we drop the second regularization parameter ε since we can consider ε a function of γ as we pass to the limit. The only requirement for this coupling is that $\varepsilon \to 0$ as $\gamma \to \infty$. For convenience of the reader, the variables associated with ($\mathbf{P}_\gamma$) are summarized in Table 3.1.

3.1. Optimality Conditions for the Regularized Problem. In Section 2.3 the differentiability of the control-to-state map was proved. Thus we can apply standard arguments to derive optimality conditions for the problem ($\mathbf{P}_\gamma$).

Theorem 3.1. *Let $(\boldsymbol{f}_\gamma, \boldsymbol{g}_\gamma, \boldsymbol{\Sigma}_\gamma, \boldsymbol{u}_\gamma)$ be a local optimal solution to ($\mathbf{P}_\gamma$). Then there exist adjoint states $(\boldsymbol{\Upsilon}_\gamma, \boldsymbol{w}_\gamma) \in S^2 \times V$ such that*

$$(A + J_\gamma'(\boldsymbol{\Sigma}_\gamma))\,\boldsymbol{\Upsilon}_\gamma + B^\star \boldsymbol{w}_\gamma = \mathbf{0}, \tag{3.1a}$$

$$B\boldsymbol{\Upsilon}_\gamma = -(\boldsymbol{u}_\gamma - \boldsymbol{u}_d), \tag{3.1b}$$

$$(\nu_1 \boldsymbol{f}_\gamma, \nu_2\, \boldsymbol{g}_\gamma) - R^\star \boldsymbol{w}_\gamma = \mathbf{0} \tag{3.1c}$$

holds.

Proof. By means of the control-to-state map G_γ, we define the reduced problem

$$\left.\begin{aligned} \text{Minimize}\quad & \frac{1}{2}\|G_\gamma^{\boldsymbol{u}}(\boldsymbol{f},\boldsymbol{g}) - \boldsymbol{u}_d\|^2_{L^2(\Omega;\mathbb{R}^d)} + \frac{\nu_1}{2}\|\boldsymbol{f}\|^2_{L^2(\Omega;\mathbb{R}^d)} + \frac{\nu_2}{2}\|\boldsymbol{g}\|^2_{L^2(\Gamma_N;\mathbb{R}^d)} \\ \text{s.t.}\quad & (\boldsymbol{f},\boldsymbol{g}) \in U. \end{aligned}\right\} \tag{3.2}$$

Clearly, if $(\boldsymbol{f}_\gamma, \boldsymbol{g}_\gamma, \boldsymbol{\Sigma}_\gamma, \boldsymbol{u}_\gamma)$ is a local optimal solution to $(\mathbf{P}_\gamma)$, then $(\boldsymbol{f}_\gamma, \boldsymbol{g}_\gamma)$ is local optimal for (3.2). According to Theorem 2.12, $G_\gamma^{\boldsymbol{u}} : U \to V$ is Fréchet-differentiable. A straightforward calculation, using the self-adjointness of the differential operator in (2.17a) and the definition of R in (2.15), shows that the gradient of the objective at $(\boldsymbol{f}_\gamma, \boldsymbol{g}_\gamma)$ is given by $(\nu_1 \boldsymbol{f}_\gamma, \nu_2 \boldsymbol{g}_\gamma) - R^\star \boldsymbol{w}_\gamma$, where $(\boldsymbol{\Upsilon}_\gamma, \boldsymbol{w}_\gamma) \in S^2 \times V$ is the solution of the adjoint system (3.1a)–(3.1b) with $(\boldsymbol{\Sigma}_\gamma, \boldsymbol{u}_\gamma) = G_\gamma(\boldsymbol{f}_\gamma, \boldsymbol{g}_\gamma)$. The local optimality of $(\boldsymbol{f}_\gamma, \boldsymbol{g}_\gamma)$ implies $(\nu_1 \boldsymbol{f}_\gamma, \nu_2 \boldsymbol{g}_\gamma) - R^\star \boldsymbol{w}_\gamma = \mathbf{0}$ since (3.2) is an unconstrained problem. □

Remark 3.2. *We remark that it is not straightforward to obtain the result of Theorem 3.1 by formulating the Lagrangian associated with $(\mathbf{P}_\gamma)$, i.e., by adding*

$$a(\boldsymbol{\Sigma}_\gamma, \boldsymbol{\Upsilon}_\gamma) + b(\boldsymbol{\Upsilon}_\gamma, \boldsymbol{u}_\gamma) + \langle J_\gamma(\boldsymbol{\Sigma}_\gamma),\, \boldsymbol{\Upsilon}_\gamma\rangle + b(\boldsymbol{\Sigma}_\gamma, \boldsymbol{w}_\gamma) - \langle \ell, \boldsymbol{w}_\gamma\rangle$$

to the objective. The reason is that $J_\gamma(\boldsymbol{\Sigma}_\gamma)$ is not Fréchet differentiable from S^2 into S^2, but only from Y_δ into S^2. The verification of standard constraint qualifications, e.g., Zowe and Kurcyusz [1979], fails, because the linearization of (2.13a) is not surjective from $Y_\delta \times V$ onto S^2.

3.2. Approximation of Solutions. We now address the question whether optimal controls of $(\mathbf{P})$ can be approximated by optimal controls of $(\mathbf{P}_\gamma)$. Two results which give a partial answer in the case $\varepsilon = 0$ were proved in [Herzog and Meyer, 2011, Section 5].

The following result parallels [Herzog and Meyer, 2011, Theorem 5.1]. Since it relies mainly on the consistency of the regularization, it is also applicable to our problem.

Theorem 3.3. *Let $\{\gamma_k\}$ be a sequence tending to ∞ and let $(\boldsymbol{f}_k, \boldsymbol{g}_k)$ denote a global solution to $(\mathbf{P}_{\gamma_k})$.*

1. *There exists an accumulation point $(\boldsymbol{f}, \boldsymbol{g})$.*
2. *Every weak accumulation point of $\{(\boldsymbol{f}_k, \boldsymbol{g}_k)\}$ is a strong accumulation point and a global solution of $(\mathbf{P})$.*

As a consequence we find that [Herzog and Meyer, 2011, Theorem 5.4] also holds true:

Theorem 3.4. *Suppose that $(\boldsymbol{f}, \boldsymbol{g})$ is a strict local optimum of $(\mathbf{P})$ in the topology of U. Let γ_k be an arbitrary sequence tending to ∞. Then there exists a sequence $(\boldsymbol{f}_k, \boldsymbol{g}_k)$ of local optimal solutions of $(\mathbf{P}_{\gamma_k})$ such that $(\boldsymbol{f}_k, \boldsymbol{g}_k) \to (\boldsymbol{f}, \boldsymbol{g})$ strongly in U.*

Following Barbu [1984] and Mignot and Puel [1984], it is even possible to approximate *all* local optima of $(\mathbf{P})$ by solutions of slightly modified problems.

Corollary 3.5. *Suppose that $(\bar{\boldsymbol{f}}, \bar{\boldsymbol{g}})$ is a local optimum of $(\mathbf{P})$ and define the modified problems $(\mathbf{P}')$ and $(\mathbf{P}'_\gamma)$ by adding the additional term*

$$\frac{1}{2}\|(\boldsymbol{f},\boldsymbol{g}) - (\bar{\boldsymbol{f}},\bar{\boldsymbol{g}})\|^2_U$$

to the objective in $(\mathbf{P})$ and $(\mathbf{P}_\gamma)$, respectively. Then $(\bar{\boldsymbol{f}}, \bar{\boldsymbol{g}})$ is a strict local optimum of $(\mathbf{P}')$ and it can be approximated by a sequence of local optimal solutions of the regularized problems $(\mathbf{P}'_\gamma)$, i.e., $(\boldsymbol{f}_\gamma, \boldsymbol{g}_\gamma) \to (\boldsymbol{f}, \boldsymbol{g})$ strongly in U for $\gamma \to \infty$, where $(\boldsymbol{f}_\gamma, \boldsymbol{g}_\gamma)$ is a local solution of $(\mathbf{P}'_\gamma)$.

Remark 3.6. *While* ($\mathbf{P}'_\gamma$) *and the above corollary are of rather theoretical interest since the unknown (local) optimal control appears in the corresponding objective, problem* ($\mathbf{P}_\gamma$) *and the associated Theorem 3.4 are of numerical relevance and give rise to a penalization algorithm similar to the methods developed e.g. in Hintermüller [2008] or Kunisch and Wachsmuth for the obstacle problem. In the next section, we will use* ($\mathbf{P}'_\gamma$) *and Corollary 3.5 to verify C-stationarity conditions for* every *local minimum of* ($\mathbf{P}$)*, which is not possible by invoking Theorem 3.4 since this result only addresses* strict *local optima.*

3.3. Convergence and C-Stationarity. In this section we pass to the limit in the optimality systems (3.1) as $\gamma \to \infty$ and $\varepsilon \to 0$. It is typical that such an approach leads to an optimality system of C-stationary type, see, e.g., [Mignot and Puel, 1984, Theorem 3.2] or Hintermüller [2008]. Proceeding formally, we expect the following system of C-stationarity, compare Scheel and Scholtes [2000]:

$$A\boldsymbol{\Sigma} + \lambda\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma} + B^\star\boldsymbol{u} = \mathbf{0}, \tag{3.3a}$$

$$B\boldsymbol{\Sigma} = R(\boldsymbol{f},\boldsymbol{g}), \tag{3.3b}$$

$$0 \le \lambda \quad \perp \quad \phi(\boldsymbol{\Sigma}) \le 0, \tag{3.3c}$$

$$A\boldsymbol{\Upsilon} + \lambda\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon} + \theta\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma} + B^\star\boldsymbol{w} = \mathbf{0}, \tag{3.4a}$$

$$B\boldsymbol{\Upsilon} = -(\boldsymbol{u} - \boldsymbol{u}_d), \tag{3.4b}$$

$$(\nu_1\,\boldsymbol{f}, \nu_2\,\boldsymbol{g}) - R^\star\boldsymbol{w} = \mathbf{0}, \tag{3.5}$$

$$\mathcal{D}\boldsymbol{\Sigma} : \mathcal{D}\boldsymbol{\Upsilon} - \mu = 0, \tag{3.6a}$$

$$\mu\,\lambda = 0, \tag{3.6b}$$

$$\theta\,\phi(\boldsymbol{\Sigma}) = 0, \tag{3.6c}$$

$$\theta\,\mu \ge 0. \tag{3.6d}$$

We remark that (3.3) ensures the feasibility of the loads $(\boldsymbol{f},\boldsymbol{g})$, stresses and displacements $(\boldsymbol{\Sigma},\boldsymbol{u})$ and plastic multiplier λ for ($\mathbf{P}$). Equations (3.4) and (3.5) are the result of passing to the limit in the adjoint system (3.1).

Note that in (3.1) the adjoint states $\boldsymbol{\Upsilon}_\gamma$ and $\boldsymbol{w}_\gamma$ serve to represent part of the gradient of the reduced objective. Remark 3.2 shows that, strictly speaking, they cannot be interpreted as Lagrange multipliers for the regularized state equation (2.13). By contrast, in (3.4), they are Lagrange multipliers pertaining to (3.3a) and (3.3b).

Finally, (3.6) contains information about the Lagrange multipliers, where μ belongs to the constraint $\lambda \ge 0$ and θ belongs to $\phi(\boldsymbol{\Sigma}) \le 0$. Notice that there is no multiplier associated with $\lambda\,\phi(\boldsymbol{\Sigma}) = 0$, which is characteristic for optimality conditions in case of MPCCs. We observe two slackness conditions (3.6b) and (3.6c), while the positivity is only required for the product of the multipliers, cf. (3.6d). This is typical for optimality conditions of C-stationary type.

Theorem 3.16 contains the final result of this section. Its proof requires the following main steps.

(1) We begin by defining μ_γ and θ_γ as regularized counterparts of μ and θ in order that the optimality system for ($\mathbf{P}_\gamma$) resembles (3.3)–(3.5).
(2) We derive a number of priori bounds for L^2 norms of various quantities (Lemma 3.7 through Lemma 3.10). This will later enable us to extract a weakly convergent subsequence.
(3) We prove estimates for the left hand sides of (3.6b) and (3.6c) with the regularized quantities (Propositions 3.12 and 3.13).

(4) We prove an auxiliary result (Proposition 3.15) which enables us to transfer the pointwise inequalities $\theta_k \, \mu_k \geq 0$ holding for the regularized multipliers to the limit case.
(5) We consider a sequence $(\boldsymbol{f}_k, \boldsymbol{g}_k)$ of local solutions to $(\mathbf{P}_\gamma)$ which converges weakly to a local solution $(\boldsymbol{f}, \boldsymbol{g})$ of $(\mathbf{P})$, made possible by the results in Section 3.2. We prove in Theorem 3.16 that (3.3)–(3.6) is satisfied.

From now on, let $(\boldsymbol{f}_\gamma, \boldsymbol{g}_\gamma)$ with corresponding states $(\boldsymbol{\Sigma}_\gamma, \boldsymbol{u}_\gamma)$ and adjoint states $(\boldsymbol{\Upsilon}_\gamma, \boldsymbol{w}_\gamma)$ denote an arbitrary stationary point for $(\mathbf{P}_\gamma)$, i.e., (3.1) holds.

Step (1):

We define the regularized multipliers by

$$\mu_\gamma := \mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{\Upsilon}_\gamma, \tag{3.7a}$$

$$\theta_\gamma := \max\nolimits_\varepsilon' \left(\gamma \left(1 - \frac{\tilde{\sigma}_0}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|} \right) \right) \frac{\gamma \, \tilde{\sigma}_0}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|^3} \, \mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{\Upsilon}_\gamma. \tag{3.7b}$$

Note that the definition implies $\theta_\gamma \, \mu_\gamma \geq 0$ a.e. in Ω. Using these terms, we can re-state (3.1a) as

$$A\boldsymbol{\Upsilon}_\gamma + B^\star \boldsymbol{w}_\gamma + \theta_\gamma \, \mathcal{D}^\star \mathcal{D}\boldsymbol{\Sigma}_\gamma + \lambda_\gamma \, \mathcal{D}^\star \mathcal{D}\boldsymbol{\Upsilon}_\gamma = \mathbf{0}. \tag{3.1a'}$$

Note that $\theta_\gamma \, \mathcal{D}^\star \mathcal{D}\boldsymbol{\Sigma}_\gamma$, $\lambda_\gamma \, \mathcal{D}^\star \mathcal{D}\boldsymbol{\Upsilon}_\gamma \in S^2$ due to Proposition 2.11 and $\lambda_\gamma \in L^\infty(\Omega)$ because of (2.20).

Step (2):

The standard a priori estimate for saddle-point problems, cf. [Quarteroni and Valli, 1994, Theorem 7.4.1] or [Ern and Guermond, 2004, Theorem 2.34], involves the norm of the upper left block. In the situation at hand, see (3.8), this is the norm of the operator $A + J_\gamma'(\boldsymbol{\Sigma}_\gamma)$, which goes to infinity as $\gamma \to \infty$. Owing to the special structure of our problem, however, we can prove a refined a priori estimate independent of γ.

Lemma 3.7. *For fixed $\boldsymbol{\Sigma}, \boldsymbol{\Pi} \in S^2$ and $\ell \in V'$, the unique solution $(\boldsymbol{\Upsilon}, \boldsymbol{w}) \in S^2 \times V$ of the linear saddle-point problem*

$$\begin{aligned} (A + J_\gamma'(\boldsymbol{\Sigma}))\boldsymbol{\Upsilon} + B^\star \boldsymbol{w} &= \boldsymbol{\Pi}, \\ B\boldsymbol{\Upsilon} &= \ell \end{aligned} \tag{3.8}$$

satisfies

$$\|\boldsymbol{\Upsilon}\|_{S^2} + \|\boldsymbol{w}\|_V \leq C \left(\|\boldsymbol{\Pi}\|_{S^2} + \|\ell\|_{V'} \right),$$

where C is independent of γ, ε and the other terms on the right hand side.

Proof. Owing to the inf-sup condition, we can find a unique $\boldsymbol{T}_\ell \in (\ker(B))^\perp$ such that $B\boldsymbol{T}_\ell = \ell$, which depends linearly on ℓ and satisfies $\|\boldsymbol{T}_\ell\|_{S^2} \leq c_B \, \|\ell\|_{V'}$, see for instance [Girault and Raviart, 1986, Chapter I, Lemma 4.1], [Quarteroni and Valli, 1994, Proposition 7.4.1]. The structure of $b(\cdot,\cdot)$ implies $\boldsymbol{T}_\ell = (\boldsymbol{\tau}_\ell, \mathbf{0})$.

Testing the first equation with $\boldsymbol{\Upsilon} - \widetilde{\boldsymbol{T}}_\ell$, where $\widetilde{\boldsymbol{T}}_\ell = (\boldsymbol{\tau}_\ell, -\boldsymbol{\tau}_\ell)$, leads to

$$\langle A\boldsymbol{\Upsilon}, \, \boldsymbol{\Upsilon} \rangle + \langle J_\gamma'(\boldsymbol{\Sigma}) \, \boldsymbol{\Upsilon}, \, \boldsymbol{\Upsilon} \rangle + \langle \boldsymbol{w}, \, B\boldsymbol{\Upsilon} - B\widetilde{\boldsymbol{T}}_\ell \rangle = \langle \boldsymbol{\Pi}, \, \boldsymbol{\Upsilon} - \widetilde{\boldsymbol{T}}_\ell \rangle + \langle A\boldsymbol{\Upsilon}, \, \widetilde{\boldsymbol{T}}_\ell \rangle.$$

Here we used that $\langle J_\gamma'(\boldsymbol{\Sigma}) \, \boldsymbol{\Upsilon}, \, \widetilde{\boldsymbol{T}}_\ell \rangle = 0$ since $\mathcal{D}\widetilde{\boldsymbol{T}}_\ell = \mathbf{0}$. By construction $B\boldsymbol{\Upsilon} - B\widetilde{\boldsymbol{T}}_\ell = \ell - \ell = \mathbf{0}$ holds. The positive semidefiniteness (see Proposition 2.11) of $J_\gamma'(\boldsymbol{\Sigma})$ and the coercivity of a imply

$$\begin{aligned} \underline{\alpha} \, \|\boldsymbol{\Upsilon}\|_{S^2}^2 &\leq \|\boldsymbol{\Pi}\|_{S^2} \, (\|\boldsymbol{\Upsilon}\|_{S^2} + \|\widetilde{\boldsymbol{T}}_\ell\|_{S^2}) + \|a\| \, \|\boldsymbol{\Upsilon}\|_{S^2} \, \|\widetilde{\boldsymbol{T}}_\ell\|_{S^2} \\ &\leq \|\boldsymbol{\Pi}\|_{S^2} \, (\|\boldsymbol{\Upsilon}\|_{S^2} + \sqrt{2} \, c_B \, \|\ell\|_{V'}) + \sqrt{2} \, c_B \, \|a\| \, \|\boldsymbol{\Upsilon}\|_{S^2} \, \|\ell\|_{V'}. \end{aligned}$$

Young's inequality then gives

$$\|\mathbf{\Upsilon}\|_{S^2} \le C_1 \left(\|\mathbf{\Pi}\|_{S^2} + \|\ell\|_{V'}\right),$$

where C_1 only depends on $\underline{\alpha}$, $\|a\|$ and c_B. Testing the first equation in (3.8) with $\boldsymbol{T} = (\boldsymbol{\tau}, -\boldsymbol{\tau})$ with $\boldsymbol{\tau} \in S$ arbitrary, an argument based on the inf-sup property of b analogously to the proof of Proposition 2.8 shows

$$\|\boldsymbol{w}\|_V \le C_2 \left(\|\mathbf{\Pi}\|_{S^2} + \|\ell\|_{V'}\right),$$

where C_2 only depends on $\underline{\alpha}$, $\|a\|$ and c_B. □

Since $G_\gamma(\mathbf{0}) = \mathbf{0}$, we have the following corollary.

Corollary 3.8. *The previous lemma and Proposition 2.8 show*

$$\|\mathbf{\Upsilon}_\gamma\|_{S^2} + \|\boldsymbol{w}_\gamma\|_V \le C \left(\|(\boldsymbol{f}_\gamma, \boldsymbol{g}_\gamma)\|_U + 1\right), \tag{3.9}$$

where C is independent of γ, ε and the other quantities on the right hand side.

To prepare the following estimates, we define three sets according to the argument of $\max_\varepsilon$ in (2.12). These sets correspond to those parts of the domain where the argument of $\max_\varepsilon$ is smaller than $-\varepsilon$, greater than ε, or in between.

$$A_\gamma^- := \Big\{x \in \Omega : \gamma\,\big(1 - \frac{\tilde{\sigma}_0}{|\mathcal{D}\mathbf{\Sigma}_\gamma|}\big) \le -\varepsilon\Big\} = \Big\{x \in \Omega : |\mathcal{D}\mathbf{\Sigma}_\gamma| \le \frac{\tilde{\sigma}_0\,\gamma}{\gamma + \varepsilon}\Big\}, \tag{3.10a}$$

$$A_\gamma^+ := \Big\{x \in \Omega : \gamma\,\big(1 - \frac{\tilde{\sigma}_0}{|\mathcal{D}\mathbf{\Sigma}_\gamma|}\big) \ge \quad \varepsilon\Big\} = \Big\{x \in \Omega : |\mathcal{D}\mathbf{\Sigma}_\gamma| \ge \frac{\tilde{\sigma}_0\,\gamma}{\gamma - \varepsilon}\Big\}, \tag{3.10b}$$

$$A_\gamma^0 := \Omega \setminus (A_\gamma^- \cup A_\gamma^+). \tag{3.10c}$$

We remark that $\lambda_\gamma = \theta_\gamma = 0$ on A_γ^- and

$$\lambda_\gamma = \gamma \left(1 - \frac{\tilde{\sigma}_0}{|\mathcal{D}\mathbf{\Sigma}_\gamma|}\right), \qquad \theta_\gamma = \frac{\gamma\,\tilde{\sigma}_0}{|\mathcal{D}\mathbf{\Sigma}_\gamma|^3}\,\mathcal{D}\mathbf{\Sigma}_\gamma : \mathcal{D}\mathbf{\Upsilon}_\gamma \quad \text{on } A_\gamma^+. \tag{3.11}$$

For convenience, we also introduce

$$\boldsymbol{Q}_\gamma := -A\mathbf{\Upsilon}_\gamma - B^\star \boldsymbol{w}_\gamma, \tag{3.12}$$

which is the adjoint counterpart of the generalized plastic strain $\boldsymbol{P}_\gamma := -A\mathbf{\Sigma}_\gamma - B^\star \boldsymbol{u}_\gamma$.

We now address the bilinear terms in (3.1a').

Lemma 3.9. *The estimate*

$$\|\theta_\gamma\, \mathcal{D}^\star\mathcal{D}\mathbf{\Sigma}_\gamma\|_{S^2}^2 + \|\lambda_\gamma\, \mathcal{D}^\star\mathcal{D}\mathbf{\Upsilon}_\gamma\|_{S^2}^2 \le \|\boldsymbol{Q}_\gamma\|_{S^2}^2 \tag{3.13}$$

holds.

Proof. Sorting terms in (3.1a') and taking the S^2 norms squared of both sides we arrive at

$$\int_\Omega \left(\theta_\gamma^2\,|\mathcal{D}^\star\mathcal{D}\mathbf{\Sigma}_\gamma|^2 + 4\,\theta_\gamma\,\lambda_\gamma\,\mathcal{D}\mathbf{\Sigma}_\gamma : \mathcal{D}\mathbf{\Upsilon}_\gamma + \lambda_\gamma^2\,|\mathcal{D}^\star\mathcal{D}\mathbf{\Upsilon}_\gamma|^2\right) \mathrm{d}x = \int_\Omega |\boldsymbol{Q}_\gamma|^2\,\mathrm{d}x.$$

We use again $\theta_\gamma\,\mathcal{D}\mathbf{\Sigma}_\gamma : \mathcal{D}\mathbf{\Upsilon}_\gamma \ge 0$ and $\lambda_\gamma \ge 0$ on Ω, which shows the claim. □

Lemma 3.10. *The multiplier θ_γ verifies the estimate*

$$\|\theta_\gamma\|_{L^2(\Omega)} \le \frac{1}{\sqrt{2}}\,\frac{\gamma + \varepsilon}{\tilde{\sigma}_0\,\gamma}\,\|\boldsymbol{Q}_\gamma\|_{S^2}. \tag{3.14}$$

Proof. We recall $\theta_\gamma = 0$ on A_γ^-. On the complement $\Omega \setminus A_\gamma^- = A_\gamma^0 \cup A_\gamma^+$, we have

$$|\mathcal{D}\boldsymbol{\Sigma}_\gamma| \geq \tilde\sigma_0\,\gamma/(\gamma+\varepsilon) \quad \text{and hence} \quad |\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_\gamma| \geq \sqrt{2}\,\tilde\sigma_0\,\gamma/(\gamma+\varepsilon).$$

Together with (3.13) this shows

$$\sqrt{2}\,\tilde\sigma_0\,\gamma/(\gamma+\varepsilon)\,\|\theta_\gamma\|_{L^2(\Omega)} \leq \|\theta_\gamma\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_\gamma\|_{S^2} \leq \|\boldsymbol{Q}_\gamma\|_{S^2},$$

which concludes the proof. □

Step (3):

We address (3.6b) and (3.6c) for the regularized quantities. This requires some preliminary estimates.

Lemma 3.11. *The following estimates hold on the set A_γ^+.*

$$\big\||\mathcal{D}\boldsymbol{\Sigma}_\gamma| - \tilde\sigma_0\big\|_{L^2(A_\gamma^+)} = \|\boldsymbol{\Sigma}_\gamma - P_{\mathcal{K}}(\boldsymbol{\Sigma}_\gamma)\|_{L^2(A_\gamma^+;\mathbb{S}^2)} \leq C\,\gamma^{-1}\,\|(\boldsymbol{f}_\gamma,\boldsymbol{g}_\gamma)\|_U, \tag{3.15}$$

$$\left\|\frac{\mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{\Upsilon}_\gamma}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|}\right\|_{L^2(A_\gamma^+)} \leq \frac{\|\boldsymbol{Q}_\gamma\|_{S^2}}{\sqrt{2}}\,\gamma^{-1} \tag{3.16}$$

with C independent of γ, ε and $\|(\boldsymbol{f}_\gamma,\boldsymbol{g}_\gamma)\|_U$.

Proof. To show the first relation, we use

$$|\boldsymbol{\Sigma}_\gamma - P_{\mathcal{K}}(\boldsymbol{\Sigma}_\gamma)| = \max\{0, |\mathcal{D}\boldsymbol{\Sigma}_\gamma| - \tilde\sigma_0\} = |\mathcal{D}\boldsymbol{\Sigma}_\gamma| - \tilde\sigma_0 \quad \text{on } A_\gamma^+$$

and Lemma 2.13.

For the second relation, we test (3.1a') with $\boldsymbol{T} = |\mathcal{D}\boldsymbol{\Sigma}_\gamma|^{-2}\,(\mathcal{D}\boldsymbol{\Sigma}_\gamma{:}\mathcal{D}\boldsymbol{\Upsilon}_\gamma)\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_\gamma\,\chi_{A_\gamma^+}$. Note that this function belongs to S^2 since $\|\boldsymbol{T}\|_{S^2} \leq \sqrt{2}\,\|D\boldsymbol{\Upsilon}_\gamma\|_S$. A straightforward calculation with $\boldsymbol{Q}_\gamma$ from (3.12) and using θ_γ and λ_γ as in (3.11) shows

$$\begin{aligned}
&\int_{A_\gamma^+} \boldsymbol{Q}_\gamma : \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_\gamma\,\frac{\mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{\Upsilon}_\gamma}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|^2}\,\mathrm{d}x \\
&= \int_{A_\gamma^+} (\theta_\gamma \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_\gamma + \lambda_\gamma\,\mathcal{D}^\star\mathcal{D}\boldsymbol{\Upsilon}_\gamma) : \mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_\gamma\,\frac{\mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{\Upsilon}_\gamma}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|^2}\,\mathrm{d}x \\
&= 2\int_{A_\gamma^+} \big(\theta_\gamma |\mathcal{D}\boldsymbol{\Sigma}_\gamma|^2 + \lambda_\gamma\,(\mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{\Upsilon}_\gamma)\big)\,\frac{\mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{\Upsilon}_\gamma}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|^2}\,\mathrm{d}x \\
&= 2\int_{A_\gamma^+} \left(\frac{\gamma\,\tilde\sigma_0}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|^3}|\mathcal{D}\boldsymbol{\Sigma}_\gamma|^2 + \gamma\left(1 - \frac{\tilde\sigma_0}{|\mathcal{D}\boldsymbol{\Sigma}|}\right)\right)(\mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{\Upsilon}_\gamma)\,\frac{\mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{\Upsilon}_\gamma}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|^2}\,\mathrm{d}x \\
&= 2\,\gamma\int_{A_\gamma^+} \frac{(\mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{\Upsilon}_\gamma)^2}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|^2}\,\mathrm{d}x = 2\,\gamma\left\|\frac{\mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{\Upsilon}_\gamma}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|}\right\|^2_{L^2(A_\gamma^+)}.
\end{aligned}$$

We estimate the left hand side of this chain of equations by

$$\int_{A_\gamma^+} \boldsymbol{Q}_\gamma : \frac{\mathcal{D}^\star\mathcal{D}\boldsymbol{\Sigma}_\gamma}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|}\,\frac{\mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{\Upsilon}_\gamma}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|}\,\mathrm{d}x \leq \sqrt{2}\,\|\boldsymbol{Q}_\gamma\|_{S^2}\left\|\frac{\mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{\Upsilon}_\gamma}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|}\right\|_{L^2(A_\gamma^+)}$$

to conclude the proof. □

We may now deduce an estimate relevant for (3.6b).

Proposition 3.12. *The estimate*

$$\|\lambda_\gamma\,\mu_\gamma\|_{L^1(\Omega)} \leq C\,(\varepsilon + \gamma^{-1})\,\|(\boldsymbol{f}_\gamma,\boldsymbol{g}_\gamma)\|_U\,\big(\|\mathcal{D}\boldsymbol{\Upsilon}_\gamma\|_S + \|\boldsymbol{Q}_\gamma\|_{S^2}\big) \tag{3.17}$$

holds with C independent of γ, ε and the other terms on the right hand side.

Proof. We recall that $\lambda_\gamma = 0$ on A_γ^-. On A_γ^0 we have $\lambda_\gamma = \max_\varepsilon\{\gamma\,\big(1 - \frac{\tilde\sigma_0}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|}\big)\} \le \varepsilon$ and thus

$$\begin{aligned}\|\lambda_\gamma\,\mu_\gamma\|_{L^1(A_\gamma^0)} &= \|\lambda_\gamma\,\mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{\Upsilon}_\gamma\|_{L^1(A_\gamma^0)} \\ &\le \varepsilon\,\|\mathcal{D}\boldsymbol{\Upsilon}_\gamma\|_S\,\|\mathcal{D}\boldsymbol{\Sigma}_\gamma\|_S \le C\,\varepsilon\,\|\mathcal{D}\boldsymbol{\Upsilon}_\gamma\|_S\,\|(\boldsymbol{f}_\gamma,\boldsymbol{g}_\gamma)\|_U.\end{aligned}$$

On the set A_γ^+, by definition $\lambda_\gamma = \gamma\,(1 - \tilde\sigma_0\,|\mathcal{D}\boldsymbol{\Sigma}_\gamma|^{-1})$ holds. Thus

$$\|\lambda_\gamma\,|\mathcal{D}\boldsymbol{\Sigma}_\gamma|\,\|_{L^2(A_\gamma^+)} = \gamma\,\||\mathcal{D}\boldsymbol{\Sigma}_\gamma| - \tilde\sigma_0\|_{L^2(A_\gamma^+)}.$$

The definition of μ_γ, together with the Cauchy Schwarz inequality and (3.15), (3.16) yield

$$\begin{aligned}\|\lambda_\gamma\,\mu_\gamma\|_{L^1(A_\gamma^+)} = \|\lambda_\gamma\,\mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{\Upsilon}_\gamma\|_{L^1(A_\gamma^+)} &\le \|\lambda_\gamma\,|\mathcal{D}\boldsymbol{\Sigma}_\gamma|\,\|_{L^2(A_\gamma^+)} \left\|\frac{\mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{\Upsilon}_\gamma}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|}\right\|_{L^2(A_\gamma^+)} \\ &\le C\,\gamma^{-1}\,\|(\boldsymbol{f}_\gamma,\boldsymbol{g}_\gamma)\|_U\,\|\boldsymbol{Q}_\gamma\|_{S^2}.\end{aligned}$$

Adding both estimates on A_γ^0 and A_γ^+ yields (3.17). □

We now address an inequality related to (3.6c).

Proposition 3.13. *The estimate*

$$\|\theta_\gamma\,\phi(\boldsymbol{\Sigma}_\gamma)\|_{L^1(\Omega)} \le C\,\frac{\varepsilon^2}{\gamma^2}\|\theta_\gamma\|_{L^2(A_\gamma^0)} + C\,\gamma^{-1}\|(\boldsymbol{f}_\gamma,\boldsymbol{g}_\gamma)\|_U\,\|\boldsymbol{Q}_\gamma\|_{S^2} \tag{3.18}$$

holds, where C is independent of γ, ε and the other terms on the right hand side.

Proof. Note that $\theta_\gamma = 0$ on A_γ^-, so we only need to consider A_γ^0 and A_γ^+. On A_γ^0 we use the simple estimate

$$\|\theta_\gamma\,\phi(\boldsymbol{\Sigma}_\gamma)\|_{L^1(A_\gamma^0)} \le \|\theta_\gamma\|_{L^2(A_\gamma^0)}\,\|\phi(\boldsymbol{\Sigma}_\gamma)\|_{L^2(A_\gamma^0)}.$$

It remains to consider the norm of $\phi(\boldsymbol{\Sigma}_\gamma)$ on A_γ^0:

$$\begin{aligned}\|\phi(\boldsymbol{\Sigma}_\gamma)\|^2_{L^2(A_\gamma^0)} &= \int_{A_\gamma^0}\left(\frac{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|^2 - \tilde\sigma_0^2}{2}\right)^2 \mathrm{d}x = \frac{1}{4}\int_{A_\gamma^0}(|\mathcal{D}\boldsymbol{\Sigma}_\gamma| + \tilde\sigma_0)^2\,(|\mathcal{D}\boldsymbol{\Sigma}_\gamma| - \tilde\sigma_0)^2\,\mathrm{d}x \\ &\le \frac{\mu(A_\gamma^0)}{4}\left(\frac{2\,\gamma - \varepsilon}{\gamma - \varepsilon}\right)^2\left(\frac{\varepsilon}{\gamma - \varepsilon}\right)^2\tilde\sigma_0^4 \le C\,\tilde\sigma_0^4\,\frac{\varepsilon^2}{\gamma^2}.\end{aligned}$$

Strictly speaking, we need $2\,\varepsilon < \gamma$ here, which is of no concern since later $\varepsilon \to 0$ as $\gamma \to \infty$. Using (3.11) we have

$$\begin{aligned}\int_{A_\gamma^+}|\theta_\gamma\,\phi(\boldsymbol{\Sigma}_\gamma)|\,\mathrm{d}x &= \int_{A_\gamma^+}\frac{\gamma\,\tilde\sigma_0}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|^3}|\mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{\Upsilon}_\gamma|\,\frac{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|^2 - \tilde\sigma_0^2}{2}\,\mathrm{d}x \\ &\le \frac{\gamma\,\tilde\sigma_0}{2}\left\|\frac{|\mathcal{D}\boldsymbol{\Sigma}_\gamma| + \tilde\sigma_0}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|^2}\right\|_{L^\infty(A_\gamma^+)}\left\|\frac{\mathcal{D}\boldsymbol{\Sigma}_\gamma : \mathcal{D}\boldsymbol{\Upsilon}_\gamma}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|}\right\|_{L^2(A_\gamma^+)}\||\mathcal{D}\boldsymbol{\Sigma}_\gamma| - \tilde\sigma_0\|_{L^2(A_\gamma^+)} \\ &\le C\,\gamma^{-1}\,\|(\boldsymbol{f}_\gamma,\boldsymbol{g}_\gamma)\|_U\,\|\boldsymbol{Q}_\gamma\|_{S^2},\end{aligned}$$

where we used (3.15) and (3.16) and

$$\left\|\frac{|\mathcal{D}\boldsymbol{\Sigma}_\gamma| + \tilde\sigma_0}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|^2}\right\|_{L^\infty(A_\gamma^+)} = \left\|\Big(1 + \frac{\tilde\sigma_0}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|}\Big)\frac{1}{|\mathcal{D}\boldsymbol{\Sigma}_\gamma|}\right\|_{L^\infty(A_\gamma^+)} \le \Big(1 + \frac{\gamma - \varepsilon}{\gamma}\Big)\Big(\frac{\gamma - \varepsilon}{\gamma\,\tilde\sigma_0}\Big) \le \frac{2}{\tilde\sigma_0},$$

refer to the definition of A_γ^+. Adding both estimates on A_γ^0 and A_γ^+ yields (3.18). □

Step (4):

Lemma 3.14. *Let $\varphi \in C_0^\infty(\Omega)$ and consider sequences $\boldsymbol{T}_k \in S^2$, $\boldsymbol{v}_k \in V$ satisfying*

$$\begin{aligned} \boldsymbol{T}_k &\rightharpoonup \boldsymbol{T} && \text{in } S^2, && \boldsymbol{v}_k \rightharpoonup \boldsymbol{v} && \text{in } V, \\ B\boldsymbol{T}_k &\to B\boldsymbol{T} && \text{in } V'. \end{aligned}$$

Then $b(\varphi\,\boldsymbol{T}_k, \boldsymbol{v}_k) \to b(\varphi\,\boldsymbol{T}, \boldsymbol{v})$.

Proof. The product rule (see for instance [Evans, 1998, Section 5.2.3]) yields

$$\boldsymbol{\varepsilon}(\varphi\,\boldsymbol{z}) = \varphi\,\boldsymbol{\varepsilon}(\boldsymbol{z}) + \frac{1}{2}\,(\nabla\varphi\,\boldsymbol{z}^\top + \boldsymbol{z}\,\nabla\varphi^\top) \quad \text{for all } \boldsymbol{z} \in V.$$

Testing with an arbitrary $\boldsymbol{R} = (\boldsymbol{\rho}, \boldsymbol{\pi}) \in S^2$ and integrating over Ω yields

$$-\,b(\boldsymbol{R}, \varphi\,\boldsymbol{z}) = -b(\varphi\,\boldsymbol{R}, \boldsymbol{z}) + \int_\Omega \boldsymbol{\rho} : (\boldsymbol{z}\,\nabla\varphi^\top)\,\mathrm{d}x. \tag{$*$}$$

Using $\boldsymbol{R} = \boldsymbol{T}_k = (\boldsymbol{\tau}_k, \boldsymbol{\mu}_k)$ and $\boldsymbol{z} = \boldsymbol{v}_k$ implies

$$-b(\boldsymbol{T}_k, \varphi\,\boldsymbol{v}_k) = -b(\varphi\,\boldsymbol{T}_k, \boldsymbol{v}_k) + \int_\Omega \boldsymbol{\tau}_k : (\boldsymbol{v}_k\,\nabla\varphi^\top)\,\mathrm{d}x.$$

We show that the first and last terms convergence, hence the middle term will converge as well. The term on the left hand side satisfies

$$b(\boldsymbol{T}_k, \varphi\,\boldsymbol{v}_k) = \langle B\boldsymbol{T}_k,\, \varphi\,\boldsymbol{v}_k\rangle_{V',V} \to \langle B\boldsymbol{T},\, \varphi\,\boldsymbol{v}\rangle_{V',V} = b(\boldsymbol{T}, \varphi\,\boldsymbol{v}),$$

since $B\boldsymbol{T}_k \to B\boldsymbol{T}$ in V' and $\boldsymbol{v}_k \rightharpoonup \boldsymbol{v}$ in V. Furthermore, the term on the right hand side satisfies

$$\int_\Omega \boldsymbol{\tau}_k : (\boldsymbol{v}_k\,\nabla\varphi^\top)\,\mathrm{d}x \to \int_\Omega \boldsymbol{\tau} : (\boldsymbol{v}\,\nabla\varphi^\top)\,\mathrm{d}x$$

since $\boldsymbol{T}_k \rightharpoonup \boldsymbol{T}$ in S^2, and $\boldsymbol{v}_k \rightharpoonup \boldsymbol{v}$ in V implies $\boldsymbol{v}_k \to \boldsymbol{v}$ in $L^2(\Omega; \mathbb{R}^d)$. Thus,

$$\begin{aligned} b(\varphi\,\boldsymbol{T}_k, \boldsymbol{v}_k) &= b(\boldsymbol{T}_k, \varphi\,\boldsymbol{v}_k) + \int_\Omega \boldsymbol{\tau}_k : (\boldsymbol{v}_k\,\nabla\varphi^\top)\,\mathrm{d}x \\ &\to b(\boldsymbol{T}\ \ , \varphi\,\boldsymbol{v}\ \ ) + \int_\Omega \boldsymbol{\tau}\ \ : (\boldsymbol{v}\,\nabla\varphi^\top)\,\mathrm{d}x = b(\varphi\,\boldsymbol{T}, \boldsymbol{v}) \end{aligned}$$

follows by ($*$) with $\boldsymbol{R} = \boldsymbol{T}$ and $\boldsymbol{z} = \boldsymbol{v}$. ☐

Proposition 3.15. *Let $\varphi \in C_0^\infty(\Omega)$ such that $\varphi \geq 0$ and consider sequences $\boldsymbol{\Upsilon}_k \in S^2$, $\boldsymbol{w}_k \in V$, $\lambda_k \in L^2(\Omega)$ with*

$$\begin{aligned} \boldsymbol{\Upsilon}_k &\rightharpoonup \boldsymbol{\Upsilon} && \text{in } S^2, && \boldsymbol{w}_k \rightharpoonup \boldsymbol{w} && \text{in } V, \\ B\boldsymbol{\Upsilon}_k &\to B\boldsymbol{\Upsilon} && \text{in } V', && \lambda_k \to \lambda && \text{in } L^2(\Omega). \end{aligned}$$

Assume further that $\lambda\,\mathcal{D}\boldsymbol{\Upsilon} \in S$ and $\lambda_k, \lambda \geq 0$ hold. Then

$$a(\boldsymbol{\Upsilon}_k, \varphi\,\boldsymbol{\Upsilon}_k) + b(\varphi\,\boldsymbol{\Upsilon}_k, \boldsymbol{w}_k) + \int_\Omega \varphi\,\lambda_k\,\mathcal{D}\boldsymbol{\Upsilon}_k : \mathcal{D}\boldsymbol{\Upsilon}_k\,\mathrm{d}x \leq 0$$

for all $k \in \mathbb{N}$ implies

$$a(\boldsymbol{\Upsilon}, \varphi\,\boldsymbol{\Upsilon}) + b(\varphi\,\boldsymbol{\Upsilon}, \boldsymbol{w}) + \int_\Omega \varphi\,\lambda\,\mathcal{D}\boldsymbol{\Upsilon} : \mathcal{D}\boldsymbol{\Upsilon}\,\mathrm{d}x \leq 0.$$

Proof. First we remark that $\boldsymbol{\Upsilon} \mapsto a(\boldsymbol{\Upsilon}, \varphi\,\boldsymbol{\Upsilon})$ is weakly lower semicontinuous. The term involving b converges by Lemma 3.14. Hence we need to consider only the last term.

In order to apply Egorov's Theorem, we extract a subsequence, still denoted by λ_k, such that $\lambda_k \to \lambda$ a.e. in Ω. For all $\varepsilon > 0$, there exists $\Omega_\varepsilon \subset \Omega$ such that $\mu(\Omega \setminus \Omega_\varepsilon) \le \varepsilon$ and $\lambda_k \to \lambda$ in $L^\infty(\Omega_\varepsilon)$. Now we estimate

$$\begin{aligned}
&\liminf_{k\to\infty} \int_{\Omega_\varepsilon} \varphi\, \lambda_k\, \mathcal{D}\boldsymbol{\Upsilon}_k : \mathcal{D}\boldsymbol{\Upsilon}_k \,\mathrm{d}x \\
&\quad = \lim_{k\to\infty} \int_{\Omega_\varepsilon} \varphi\, (\lambda_k - \lambda)\, \mathcal{D}\boldsymbol{\Upsilon}_k : \mathcal{D}\boldsymbol{\Upsilon}_k \,\mathrm{d}x + \liminf_{k\to\infty} \int_{\Omega_\varepsilon} \varphi\, \lambda\, \mathcal{D}\boldsymbol{\Upsilon}_k : \mathcal{D}\boldsymbol{\Upsilon}_k \,\mathrm{d}x \\
&\quad \ge 0 + \int_{\Omega_\varepsilon} \varphi\, \lambda\, \mathcal{D}\boldsymbol{\Upsilon} : \mathcal{D}\boldsymbol{\Upsilon} \,\mathrm{d}x,
\end{aligned}$$

since the first addend converges to 0 and for the second we can use the weak lower semicontinuity. By the assumption and $\Omega_\varepsilon \subset \Omega$ we have

$$a(\boldsymbol{\Upsilon}_k, \varphi\, \boldsymbol{\Upsilon}_k) + b(\varphi\, \boldsymbol{\Upsilon}_k, \boldsymbol{w}_k) + \int_{\Omega_\varepsilon} \varphi\, \lambda_k\, \mathcal{D}\boldsymbol{\Upsilon}_k : \mathcal{D}\boldsymbol{\Upsilon}_k \,\mathrm{d}x \le 0.$$

Since we considered all terms on the left hand side previously, we can pass to the limit and get

$$a(\boldsymbol{\Upsilon}, \varphi\, \boldsymbol{\Upsilon}) + b(\varphi\, \boldsymbol{\Upsilon}, \boldsymbol{w}) + \int_{\Omega_\varepsilon} \varphi\, \lambda\, \mathcal{D}\boldsymbol{\Upsilon} : \mathcal{D}\boldsymbol{\Upsilon} \,\mathrm{d}x \le 0.$$

Letting $\varepsilon \to 0$ and using the assumption $\lambda\, \mathcal{D}\boldsymbol{\Upsilon} : \mathcal{D}\boldsymbol{\Upsilon} \in L^1(\Omega)$ yields

$$a(\boldsymbol{\Upsilon}, \varphi\, \boldsymbol{\Upsilon}) + b(\varphi\, \boldsymbol{\Upsilon}, \boldsymbol{w}) + \int_{\Omega} \varphi\, \lambda\, \mathcal{D}\boldsymbol{\Upsilon} : \mathcal{D}\boldsymbol{\Upsilon} \,\mathrm{d}x \le 0.$$

□

Step (5):

Now we are ready to prove our main result.

Theorem 3.16. *Let $(\boldsymbol{f}, \boldsymbol{g})$ be a local optimal solution of* (**P**). *Let $(\boldsymbol{\Sigma}, \boldsymbol{u})$ and λ denote the associated stresses, displacements, and plastic multiplier. Then there exist adjoint stresses and displacements $(\boldsymbol{\Upsilon}, \boldsymbol{w}) \in S^2 \times V$ and Lagrange multipliers $\theta, \mu \in L^2(\Omega)$ such that the C-stationarity system* (3.3)–(3.6) *is satisfied.*

Proof. By Corollary 3.5, there exists a sequence $(\boldsymbol{f}_k, \boldsymbol{g}_k)$ of local optimal solutions of $(\mathbf{P}'_\gamma)$ for regularization parameters γ_k, ε_k, which converges strongly in U to $(\boldsymbol{f}, \boldsymbol{g})$. We need to prove only (3.4)–(3.6) since (3.3) is a consequence of $(\boldsymbol{f}, \boldsymbol{g})$ together with $(\boldsymbol{\Sigma}, \boldsymbol{u})$ and λ being admissible.

In view of Corollary 2.16 we have strong convergence of $\boldsymbol{\Sigma}_k \to \boldsymbol{\Sigma}$ in S^2 and $\boldsymbol{u}_k \to \boldsymbol{u}$ in V. By Corollary 3.8, the sequence of adjoint states $(\boldsymbol{\Upsilon}_k, \boldsymbol{w}_k)$ is bounded in $S^2 \times V$. Thus there exists a weakly convergent subsequence, still denoted by the index k, such that $(\boldsymbol{\Upsilon}_k, \boldsymbol{w}_k) \rightharpoonup (\boldsymbol{\Upsilon}, \boldsymbol{w})$ in $S^2 \times V$. Hence (3.1b) yields (3.4b).

In view of the compactness of $R^\star$, taking the limit in the counterpart of (3.1c) for $(\mathbf{P}'_\gamma)$, given by

$$(\nu_1 \boldsymbol{f}_k, \nu_2\, \boldsymbol{g}_k) - R^\star \boldsymbol{w}_k + (\boldsymbol{f}_k - \boldsymbol{f}, \boldsymbol{g}_k - \boldsymbol{g}) = \boldsymbol{0}, \tag{3.1c'}$$

implies (3.5).

Now we address the weak convergence of the multipliers. Theorem 2.19 implies that $\lambda_k \to \lambda$ in $L^2(\Omega)$. Together with the weak convergence $\boldsymbol{\Upsilon}_k \rightharpoonup \boldsymbol{\Upsilon}$ in S^2 this implies $\lambda_k\, \mathcal{D}^\star \mathcal{D}\boldsymbol{\Upsilon}_k \rightharpoonup \lambda\, \mathcal{D}^\star \mathcal{D}\boldsymbol{\Upsilon}$ in $L^1(\Omega; \mathbb{S}^2)$. Moreover, due to the weak convergence of $\boldsymbol{\Upsilon}_k$ and $\boldsymbol{w}_k$, the sequence $\boldsymbol{Q}_k$ is bounded in S^2 by definition. Together with Lemma 3.9 this implies that the weak limit $\lambda\, \mathcal{D}^\star \mathcal{D}\boldsymbol{\Upsilon}$ belongs to S^2, and therefore we obtain the weak convergence of $\lambda_k\, \mathcal{D}^\star \mathcal{D}\boldsymbol{\Upsilon}_k$ in S^2.

By Lemma 3.10, θ_k is bounded in $L^2(\Omega)$ and a subsequence converges weakly to some $\theta \in L^2(\Omega)$. This implies $\theta_k \, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}_k \rightharpoonup \theta \, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}$ in $L^1(\Omega; \mathbb{S}^2)$. Invoking again Lemma 3.9 shows $\theta \, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma} \in S^2$ and the weak convergence in S^2 of $\theta_k \, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}_k$.

This implies that for a subsequence,

$$A \boldsymbol{\Upsilon}_k + \theta_k \, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma}_k + \lambda_k \, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Upsilon}_k + B^\star \boldsymbol{w}_k \rightharpoonup A \boldsymbol{\Upsilon} + \theta \, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Sigma} + \lambda \, \mathcal{D}^\star \mathcal{D} \boldsymbol{\Upsilon} + B^\star \boldsymbol{w} \quad \text{in } S^2$$

and (3.1a') shows (3.4a).

Passing to the limit in the definition (3.7) of μ_k shows (3.6a). Moreover, due to $\mathcal{D} \boldsymbol{\Sigma} \in L^\infty(\Omega; \mathbb{S})$ and $\mathcal{D} \boldsymbol{\Upsilon} \in S = L^2(\Omega; \mathbb{S})$ we have $\mu \in L^2(\Omega)$.

Now we address (3.6b) and (3.6c). Due to Propositions 3.12 and 3.13 and the weak lower semicontinuity of the $L^1(\Omega)$-norm it is sufficient to prove the weak convergence of $\lambda_k \, \mu_k$ and $\theta_k \, \phi(\boldsymbol{\Sigma}_k)$ in $L^1(\Omega)$. We remark that $\mu_k \rightharpoonup \mu$ and $\phi(\boldsymbol{\Sigma}_k) \to \phi(\boldsymbol{\Sigma})$ only in $L^1(\Omega)$. However, as already shown, $\lambda_k \, \mathcal{D} \boldsymbol{\Upsilon}_k \rightharpoonup \lambda \, \mathcal{D} \boldsymbol{\Upsilon}$ in S. Together with $\boldsymbol{\Sigma}_k \to \boldsymbol{\Sigma}$ in S^2, we obtain $\lambda_k \, \mu_k = \lambda_k \, \mathcal{D} \boldsymbol{\Upsilon}_k : \mathcal{D} \boldsymbol{\Sigma}_k \rightharpoonup \lambda \, \mu$ in $L^1(\Omega)$. Using the weak convergence $\theta_k \, \mathcal{D} \boldsymbol{\Sigma}_k \rightharpoonup \theta \, \mathcal{D} \boldsymbol{\Sigma}$ in S and $\boldsymbol{\Sigma}_k \to \boldsymbol{\Sigma}$ in S^2, we obtain $\theta_k \, \phi(\boldsymbol{\Sigma}_k) = \theta_k \, \frac{1}{2} (\mathcal{D} \boldsymbol{\Sigma}_k : \mathcal{D} \boldsymbol{\Sigma}_k - \tilde{\sigma}_0^2) \rightharpoonup \theta \, \phi(\boldsymbol{\Sigma})$ in $L^1(\Omega)$, where we used the definition of ϕ (1.3). This yields (3.6b) and (3.6c).

Finally, we address (3.6d). By definition, $\theta_k \, \mu_k \geq 0$ holds a.e. in Ω. We test (3.1a') with $\varphi \, \boldsymbol{\Upsilon}_k$, where $\varphi \in C_0^\infty(\Omega)$ is ≥ 0 but otherwise arbitrary, and obtain

$$a(\boldsymbol{\Upsilon}_k, \varphi \, \boldsymbol{\Upsilon}_k) + b(\varphi \, \boldsymbol{\Upsilon}_k, \boldsymbol{w}_k) + \int_\Omega \varphi \, \lambda_k \, \mathcal{D} \boldsymbol{\Upsilon}_k : \mathcal{D} \boldsymbol{\Upsilon}_k \, \mathrm{d}x \leq 0,$$

where we used that $\theta_k \, \mathcal{D} \boldsymbol{\Sigma}_k : \mathcal{D} \boldsymbol{\Upsilon}_k = \theta_k \, \mu_k \geq 0$. Applying Proposition 3.15 and observing that $\lambda \, \mathcal{D} \boldsymbol{\Upsilon} \in S$ and $\lambda \geq 0$ yields

$$a(\boldsymbol{\Upsilon}, \varphi \, \boldsymbol{\Upsilon}) + b(\varphi \, \boldsymbol{\Upsilon}, \boldsymbol{w}) + \int_\Omega \varphi \, \lambda \, \mathcal{D} \boldsymbol{\Upsilon} : \mathcal{D} \boldsymbol{\Upsilon} \, \mathrm{d}x \leq 0$$

for all $\varphi \in C_0^\infty(\Omega)$ satisfying $\varphi \geq 0$. Testing (3.4a) with $\varphi \, \boldsymbol{\Upsilon}$ yields

$$a(\boldsymbol{\Upsilon}, \varphi \, \boldsymbol{\Upsilon}) + b(\varphi \, \boldsymbol{\Upsilon}, \boldsymbol{w}) + \int_\Omega \varphi \, \big\{ \lambda \, \mathcal{D} \boldsymbol{\Upsilon} : \mathcal{D} \boldsymbol{\Upsilon} + \theta \, \mathcal{D} \boldsymbol{\Sigma} : \mathcal{D} \boldsymbol{\Upsilon} \big\} \, \mathrm{d}x = 0.$$

This implies

$$\int_\Omega \varphi \, \theta \, \mathcal{D} \boldsymbol{\Sigma} : \mathcal{D} \boldsymbol{\Upsilon} \, \mathrm{d}x \geq 0 \quad \text{for all } \varphi \in C_0^\infty(\Omega) \text{ with } \varphi \geq 0$$

and thus

$$\theta \, \mu = \theta \, \mathcal{D} \boldsymbol{\Sigma} : \mathcal{D} \boldsymbol{\Upsilon} \geq 0 \quad \text{a.e. in } \Omega.$$

□

Remark 3.17. *We remark that the multipliers μ and θ possess comparatively high regularity since they are elements of $L^2(\Omega)$. Moreover, relation (3.6d) involving μ and θ pertaining to the inequalities (3.3c) holds in a pointwise a.e. sense. This is structurally different from the corresponding relation for optimal control problems for the obstacle problem, see for instance Mignot and Puel [1984] or [Hintermüller and Kopacka, 2009, eq. (4.1e)]. This is due to the fact the multiplier belonging to $y \leq \psi$ in the upper level problem belongs only to $H^{-1}(\Omega)$, and thus a pointwise interpretation is impossible.*

A Results for Saddle-Point Problems

Lemma A.1. *Let X be a Hilbert space and let and V be a reflexive Banach space. Let $A : X \to X$, $B : X \to V'$ be bounded linear operators. Furthermore, let A be coercive and let B fulfill the inf-sup condition. Suppose that the operator $J : X \to X$ is monotone and continuous. Then, for every $\ell \in V'$, the nonlinear saddle-point problem*

$$A\boldsymbol{\Sigma} + J(\boldsymbol{\Sigma}) + B^\star \boldsymbol{u} = \boldsymbol{0}, \tag{A.1a}$$

$$B\boldsymbol{\Sigma} = \ell \tag{A.1b}$$

has a unique solution $G(\ell) = (G^{\boldsymbol{\Sigma}}(\ell), G^{\boldsymbol{u}}(\ell)) = (\boldsymbol{\Sigma}_\ell, \boldsymbol{u}_\ell) \in X \times V$. $\boldsymbol{\Sigma}_\ell$ depends Lipschitz continuously on ℓ, with a Lipschitz constant independent of J.

Proof. Step (1): Existence and uniqueness of $\boldsymbol{\Sigma}_\ell$. We follow the null space approach for saddle-point problems and define

$$X_\ell := \{\boldsymbol{\Sigma} \in X : B\boldsymbol{\Sigma} = \ell\}.$$

Owing to the inf-sup condition, we can find a unique $\boldsymbol{T}_\ell \in (\ker(B))^\perp \cap X_\ell$ which depends linearly on ℓ and satisfies $\|\boldsymbol{T}_\ell\|_X \le c_B \, \|\ell\|_{V'}$, see for instance [Girault and Raviart, 1986, Chapter I, Lemma 4.1], [Quarteroni and Valli, 1994, Proposition 7.4.1]. Using an arbitrary $\boldsymbol{T} \in X_0$ as a test function in (A.1a) and decomposing $\boldsymbol{\Sigma}_\ell = \boldsymbol{T}_0 + \boldsymbol{T}_\ell \in X_0 + X_\ell$ leads us to the following reduced problem:

$$\begin{aligned}&\text{Find } \boldsymbol{T}_0 \in X_0\\ &\text{satisfying } a(\boldsymbol{T}_0, \boldsymbol{T}) + \langle J(\boldsymbol{T}_0 + \boldsymbol{T}_\ell), \boldsymbol{T}\rangle = -a(\boldsymbol{T}_\ell, \boldsymbol{T}) \quad \text{for all } \boldsymbol{T} \in X_0.\end{aligned}$$

We define a nonlinear operator $C : X_0 \to (X_0)'$ such that the left hand side becomes $\langle C(\boldsymbol{T}_0), \boldsymbol{T}\rangle$.

In order to apply the Browder-Minty theorem (see, e.g., [Zeidler, 1990, Theorem 25.1]), we verify the following properties.

- C is strongly monotone and coercive. This follows from the X-ellipticity of $a(\cdot,\cdot)$ with a constant $\underline{\alpha}$ independent of J, and from the monotonicity of J.
- C is continuous and thus hemi-continuous. This is an immediate consequence of the boundedness of $a(\cdot,\cdot)$ and the continuity of J.

Consequently, there exists a unique solution $\boldsymbol{T}_0$ to the reduced problem, which depends Lipschitz continuously on $\boldsymbol{T}_\ell$ and thus on ℓ, with a Lipschitz constant $\underline{\alpha}^{-1}$. This implies that $\boldsymbol{\Sigma}_\ell = \boldsymbol{T}_0 + \boldsymbol{T}_\ell$ depends Lipschitz continuously on ℓ with Lipschitz constant $L_{\boldsymbol{\Sigma}} = (1 + \underline{\alpha}^{-1})\, c_B$ independent of J.

Step (2): Existence and uniqueness for $\boldsymbol{u}_\ell$. It is a standard result from the theory of saddle-point problems, see, e.g., [Girault and Raviart, 1986, Chapter I, Lemma 4.1] or [Quarteroni and Valli, 1994, Proposition 7.4.1], that $BB^\star$ is boundedly invertible, and $\boldsymbol{u}_\ell$ satisfies

$$BB^\star \boldsymbol{u}_\ell = -B\,(A\boldsymbol{\Sigma}_\ell + J(\boldsymbol{\Sigma}_\ell)).$$ □

Lemma A.2. *Let the conditions of Lemma A.1 hold. Assume in addition that Y and W are normed linear spaces with continuous embeddings $Y \hookrightarrow X$ and $W' \hookrightarrow V'$. Suppose that the partial solution map $G^{\boldsymbol{\Sigma}}$ of (A.1) is locally Lipschitz as a function $W' \to Y$. Suppose moreover that J is Fréchet differentiable as a mapping $Y \to X$. At any $\boldsymbol{\Sigma} \in Y$, the derivative $J'(\boldsymbol{\Sigma})$ needs to possess a positive semidefinite extension which maps $X \to X$, i.e., $\langle J'(\boldsymbol{\Sigma})\, \delta\boldsymbol{\Sigma}, \delta\boldsymbol{\Sigma}\rangle \ge 0$ for all $\delta\boldsymbol{\Sigma} \in X$.*

Then G is Fréchet differentiable as a function $W' \to X \times V$. The derivative $(\delta\boldsymbol{\Sigma}, \delta\boldsymbol{u})$ at ℓ in the direction $\delta\ell$ is given by the unique solution of

$$(A + J'(\boldsymbol{\Sigma}))\, \delta\boldsymbol{\Sigma} + B^\star \delta\boldsymbol{u} = \mathbf{0}, \tag{A.2a}$$

$$B\, \delta\boldsymbol{\Sigma} = \delta\ell. \tag{A.2b}$$

Proof. The unique solvability of (A.2a) follows from standard arguments for linear saddle-point problems. We need to verify an estimate for the remainder. To this end, let $\ell, \delta\ell \in W'$ be given and set $\ell' = \ell + \delta\ell$ as well as the remainder terms $\boldsymbol{\Sigma}_r = \boldsymbol{\Sigma}_{\ell'} - \boldsymbol{\Sigma}_\ell - \delta\boldsymbol{\Sigma}$ and $\boldsymbol{u}_r = \boldsymbol{u}_{\ell'} - \boldsymbol{u}_\ell - \delta\boldsymbol{u}$. These satisfy

$$\begin{aligned} (A + J'(\boldsymbol{\Sigma}))\, \boldsymbol{\Sigma}_r + B^\star \boldsymbol{u}_r &= -\big(J(\boldsymbol{\Sigma}_{\ell'}) - J(\boldsymbol{\Sigma}_\ell) - J'(\boldsymbol{\Sigma}_\ell)(\boldsymbol{\Sigma}_{\ell'} - \boldsymbol{\Sigma}_\ell)\big) \\ B\boldsymbol{\Sigma}_r &= \mathbf{0}. \end{aligned}$$

The standard a-priori estimate for this saddle-point problem yields

$$\|\boldsymbol{\Sigma}_r\|_X + \|\boldsymbol{u}_r\|_V \le C\, \|J(\boldsymbol{\Sigma}_{\ell'}) - J(\boldsymbol{\Sigma}_\ell) - J'(\boldsymbol{\Sigma}_\ell)(\boldsymbol{\Sigma}_{\ell'} - \boldsymbol{\Sigma}_\ell)\|_X.$$

Since $J : Y \to X$ is Fréchet differentiable and $\boldsymbol{\Sigma}_\ell, \boldsymbol{\Sigma}_{\ell'} \in Y$ we have

$$\|J(\boldsymbol{\Sigma}_{\ell'}) - J(\boldsymbol{\Sigma}_\ell) - J'(\boldsymbol{\Sigma}_\ell)(\boldsymbol{\Sigma}_{\ell'} - \boldsymbol{\Sigma}_\ell)\|_X = o\,(\|\boldsymbol{\Sigma}_{\ell'} - \boldsymbol{\Sigma}_\ell\|_Y).$$

Due to the local Lipschitz continuity of $G^{\boldsymbol{\Sigma}} : W' \to Y$, the term on the right hand side is of order $o\,(\|\delta\ell\|_{W'})$, and the combination of all estimates leads to

$$\|\boldsymbol{\Sigma}_r\|_X + \|\boldsymbol{u}_r\|_V = o\,(\|\delta\ell\|_{W'}),$$

which concludes the proof. □

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

References

V. Barbu. *Optimal Control of Variational Inequalities*, volume 100 of *Research Notes in Mathematics*. Pitman, Boston, 1984.

M. Bergounioux. Optimal control of an obstacle problem. *Applied Mathematics and Optimization*, 36:147–172, 1997.

A. Bermúdez and C. Saguez. Optimality conditions for optimal control problems of variational inequalities. In *Control problems for systems described by partial differential equations and applications (Gainesville, Fla., 1986)*, volume 97 of *Lecture Notes in Control and Information Science*, pages 143–153. Springer, Berlin, 1987.

Alexandre Ern and Jean-Luc Guermond. *Theory and Practice of Finite Elements*. Springer, Berlin, 2004.

L. C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, Rhode Island, 1998.

A. Friedman. Optimal control for variational inequalities. *SIAM Journal on Control and Optimization*, 24(3):439–451, 1986.

V. Girault and P.-A. Raviart. *Finite Element Methods for Navier-Stokes Equations*. Springer, 1986.

H. Goldberg, W. Kampowsky, and F. Tröltzsch. On Nemytskij operators in L_p-spaces of abstract functions. *Mathematische Nachrichten*, 155:127–140, 1992. ISSN 0025-584X. doi: 10.1002/mana.19921550110.

P. Grisvard. *Elliptic Problems in Nonsmooth Domains*. Pitman, Boston, 1985.

K. Gröger. A $W^{1,p}$-estimate for solutions to mixed boundary value problems for second order elliptic differential equations. *Mathematische Annalen*, 283:679–687, 1989. doi: 10.1007/BF01442860.

R. Haller-Dintelmann, C. Meyer, J. Rehberg, and A. Schiela. Hölder continuity and optimal control for nonsmooth elliptic problems. *Applied Mathematics and Optimization*, 60(3):397–428, 2009. doi: 10.1007/s00245-009-9077-x.

W. Han and B. D. Reddy. *Plasticity*. Springer, New York, 1999.

J. Haslinger and T. Roubíček. Optimal control of variational inequalities. Approximation theory and numerical realization. *Applied Mathematics and Optimization*, 14:187–201, 1986.

R. Herzog and C. Meyer. Optimal control of static plasticity with linear kinematic hardening. *Journal of Applied Mathematics and Mechanics*, 91(10):777–794, 2011. doi: 10.1002/zamm.200900378.

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

M. Hintermüller. An active-set equality constrained Newton solver with feasibility restoration for inverse coefficient problems in elliptic variational inequalities. *Inverse Problems*, 24(3):034017, 23, 2008.

M. Hintermüller and I. Kopacka. Mathematical programs with complementarity constraints in function space: C- and strong stationarity and a path-following algorithm. *SIAM Journal on Optimization*, 20(2):868–902, 2009. ISSN 1052-6234. doi: 10.1137/080720681.

M. Hintermüller, I. Kopacka, and M.H. Tber. Recent advances in the numerical solution of MPECs in function space. In *Numerical Techniques for Optimization Problems with PDE Constraints*, volume 6 of *Oberwolfach Report No. 4/2009*, pages 36–40, Zurich, 2009. European Mathematical Society Publishing House.

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

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

K. Kunisch and D. Wachsmuth. Path-following for optimal control of stationary variational inequalities. *Computational Optimization and Applications*, pages 1–29. doi: 10.1007/s10589-011-9400-8.

K. Kunisch and D. Wachsmuth. Sufficient optimality conditions and semi-smooth newton methods for optimal control of stationary variational inequalities. *ESAIM: Control, Optimisation and Calculus of Variations*, 2011. doi: 10.1051/cocv/2011105.

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

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

A. Quarteroni and A. Valli. *Numerical Approximation of Partial Differential Equations*. Springer, Berlin, 1994.

A. Rösch and F. Tröltzsch. On regularity of solutions and Lagrange multipliers of optimal control problems for semilinear equations with mixed pointwise control-state constraints. *SIAM Journal on Control and Optimization*, 46(3):1098–1115, 2007.

H. Scheel and S. Scholtes. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Mathematics of Operations*

Research, 25(1):1–22, 2000.

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

K. Yosida and E. Hewitt. Finitely additive measures. *Transactions of the American Mathematical Society*, 72:46–66, 1952.

E. Zeidler. *Nonlinear Functional Analysis and its Applications*, volume II/B. Springer, New York, 1990.

J. Zowe and S. Kurcyusz. Regularity and stability for the mathematical programming problem in Banach spaces. *Applied Mathematics and Optimization*, 5(1): 49–62, 1979.

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

E-mail address: roland.herzog@mathematik.tu-chemnitz.de

URL: http://www.tu-chemnitz.de/herzog

TU Dortmund, Faculty of Mathematics, Vogelpothsweg 87, 44227 Dortmund, Germany

E-mail address: christian.meyer@math.tu-dortmund.de

URL: http://www.mathematik.uni-dortmund.de/de/personen/person/Christian Meyer.html

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

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

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