Weierstraß-Institut
für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

Thermoviscoelasticity in Kelvin–Voigt rheology at large strains

In memory of Erwin Stein,
who advocated the importance of finite-strain elasticity in engineering practice

Alexander Mielke[1,2], Tomáš Roubíček[3,4]

submitted: April 8, 2019

[1] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: alexander.mielke@wias-berlin.de

[2] Institut für Mathematik
Humboldt-Universität zu Berlin
Rudower Chaussee 25
12489 Berlin-Adlershof
Germany

[3] Mathematical Institute
Charles University
Sokolovská 83
186 75 Praha 8
Czech Republic
E-Mail: tomas.roubicek@mff.cuni.cz

[4] Institute of Thermomechanics
Czech Academy of Sciences
Dolejškova 5
182 00 Praha 8
Czech Republic

No. 2584
Berlin 2019

2010 *Mathematics Subject Classification.* 35K55, 35Q74, 74A15, 74A30, 80A17.

Key words and phrases. Finite-strain elasticity, static frame indifference, time-dependent frame indifference, balance of total energy, energy estimates, hyperstress regularization, uniform positive of determinant, viscous heating, staggered time-incremental minimization scheme.

A.M. is grateful for the hospitality and support of Charles University and for partial support by Deutsche Forschungsgemeinschaft (DFG) via the SFB 1114 *Scaling Cascades in Complex Systems* (subproject B01 "Fault Networks and Scaling Properties of Deformation Accumulation"). T.R. is thankful for hospitality and support of the Weierstraß-Institut Berlin. Also the partial support of the Czech Science Foundation projects 17-04301S (as for the focus to dissipative evolutionary systems), 18-03834S (as for the application in modeling of shape-memory alloys), and 19-29646L (as for the focus to large strains), as well as through the institutional support RVO: 61388998 (ČR).

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Thermoviscoelasticity in Kelvin–Voigt rheology at large strains

In memory of Erwin Stein,
who advocated the importance of finite-strain elasticity in engineering practice

Alexander Mielke, Tomáš Roubíček

Abstract

The frame-indifferent thermodynamically-consistent model of thermoviscoelasticity at large strain is formulated in the reference configuration with using the concept of the second-grade nonsimple materials. We focus on physically correct viscous stresses that are frame indifferent under time-dependent rotations. Also elastic stresses are frame indifferent under rotations and respect positivity of the determinant of the deformation gradient. The heat transfer is governed by the Fourier law in the actual deformed configuration, which leads to a nontrivial description when pulled back into the reference configuration. Existence of weak solutions in the quasistatic setting, i.e. inertial forces are ignored, is shown by time discretization.

1 Introduction

For a long time, thermoviscoelasticity was considered as a quite difficult problem even at small strains, mainly because of the nonlinear coupling with the heat-transfer equation which has no obvious variational structure; hence special techniques had to be developed. It took about two decades after the pioneering work by C.M. Dafermos [Daf82] in one space dimension that first three-dimensional studies occurred (cf. e.g. [BlG00, BoB03, Rou09]). The basic new ingredient was the L^1-theory for the nonlinear heat equation developed in [BD*97, BoG89]. At large strains, in simple materials, the problem is still recognized to be very difficult even for the case of mere viscoelasticity without coupling with temperature, and only few results are available if the physically relevant frame-indifference is respected, as articulated by J.M. Ball [Bal77], see also [Bal02, Bal10]. In particular, local-in-time existence [LeM13] or existence of measure-valued solutions [Dem00, DST01] are known for simple materials. Further examples in this direction are [Tve08] for a general three-dimensional theory, but not respecting frame indifference and the determinant constraints, or [MOS13] for a one-dimensional theory using the variation structure. While the static theory for large-strain elasticity developed rapidly after [Bal77], there are still only few result for time-dependent processes respecting frame indifference as well as the determinant constraint. The first cases were restricted to rate-independent processes, such as elastoplasticity (cf. [MaM09, MiR16]) or crack growth (cf. [DaL10], see [MiR15, Sec. 4.2] for a survey. Recently the case of viscoplasticity was treated in [MRS18].

The main features of the model discussed in this work can be summarized in brief as follows: the thermo-visco-elastic continuum is formulated at large strains in a reference configuration, i.e. the Lagrangian approach. The concepts of 2nd-grade nonsimple material is used, which gives higher regularity of the deformation. The heat transfer is modeled by the Fourier law in the actual deformed configuration, but transformed (pulled back) into the reference configuration for the analysis. Our model respects both static frame-indifference of the free energy and dynamic frame indifference for the dissipation potential. Moreover, the local non-selfpenetration is realized by imposing a blowup of the free

DOI 10.20347/WIAS.PREPRINT.2584 Berlin 2019

energy if the determinant of the deformation gradient approaches 0 from above, however we do not enforce global non-selfpenetration. Also, we neglect inertial effect; cf. Remark 6.6 for more detailed discussion.

Let us highlight the important aspects of the presented model and their consequences:

(α) The temperature-dependence of the free energy creates adiabatic effects involving the rate of the deformation gradient. To handle this, the Kelvin-Voigt-type viscosity is used to control the rate of the deformation gradient. In addition, we separate the purely mechanical part, cf. (2.15) below, which allows us to decouple the singularities of large-strain elasticity from the heat equation.

(β) The heat transfer itself (and also the viscosity from (α)) is clearly rate dependent and the technique of rate-independent processes supported by variationally efficient energetic-solution concept cannot be used (which also prevents us from excluding possible global selfpenetration).

(γ) The equations for the solid continuum need to be formulated and analyzed in the fixed reference configuration but transport processes (here only the heat transfer) happen rather in the actual configuration and the pull-back procedure needs the determinant of the deformation gradient to be well away from 0. To achieve this, we exploit the concept of 2nd-grade nonsimple materials together with the results of T.J. Healey and S. Krömer [HeK09], which allow us to show that the determinant for the deformation gradient is bounded away from 0, see Section 3.1.

(δ) The transport coefficients depend on the deformation gradient because of the reasons in point (γ). For this, measurability in time is needed and thus the concept of global quasistatic minimization of deformation (as in rate-independent systems [MiR15] or in viscoplasticity in [MRS18]) would not be satisfactory; therefore we rather control the time derivative of the deformation, which can be done either by inertia (which is neglected in our work) or by the Kelvin-Voigt-type viscosity from (α).

(ϵ) The viscosity from (α) must satisfy time-dependent frame indifference as explained in [Ant98], thus it is dependent on the rate of the right Cauchy-Green tensor rather than on the rate of the deformation gradient itself. However, the adiabatic heat sources/sinks involve terms where the rate of the deformation gradient occurs directly. To control the latter by the former, we exploit results of P. Neff [Nef02] in the extension by W. Pompe [Pom03] for generalized Korn's inequalities, see Section 3.2. Here, again the mentioned concept of 2nd-grade nonsimple materials is used to control determinant of the deformation gradient, see (γ).

As mentioned above, our model heavily relies on the strain-gradient theories to describe materials, referred as nonsimple, or also multipolar or complex. This concept has been introduced long time ago, cf. [Tou62] or also e.g. [FrG06, MiE68, Pod02, Šil85, TrA86, BaC11] and in the thermodynamical concept also [Bat76]. In the simplest scenario, which is also used here, the stored-energy density depends only on the strain $F = \nabla y$ and on the first gradient ∇F of the strain. This case is called *2nd-grade nonsimple material*. Possible generalization using only certain parts of the 2nd in the spirit of [KPS19] still need to be explored.

The structure of the paper is as follows. In Section 2 we present the model in physical and mathematical terms. After the precise definition of our notion of solution, Theorem 2.2 presents the main existence result for global-in-time solutions for the large-strain thermoviscoelastic system, while Corollary 2.3 gives the corresponding existence result for viscoelasticity at large-strain and at constant temperature, which, to the knowledge of the authors, is also new. A related result for isothermal large-strain viscoelasticity is derived in [FrK18], but there the limit of small strains is treated.

In Section 4 we start the proof of the main result by introducing certain regularizations as well as a

time-incremental approach that is particularly constructed in such a split (sometimes called staggered) way that the deformation is first updated at fixed temperature and then the temperature is updated, where in some terms the old and in others the new deformation is used. Another important step in the analysis is the usage of an energy-like variable $w = \mathfrak{w}(\nabla y, \theta)$ instead of temperature θ, which enables us to exploit the balance-law structure of the heat equation; cf. [Mie13, MiM18] for arguments for the preference of energy in favor of temperature. As an intermediate result Proposition 5.1 provides the existence of solutions $(y_\varepsilon, \theta_\varepsilon)$ of the regularized problem.

In Section 6 we finally show that the limit $\varepsilon_k \to 0$ for $(y_{\varepsilon_k}, \theta_{\varepsilon_k}) \to (y, \theta)$ can be controlled in such a way that (y, θ) are the desired solutions. We conclude with a few remarks concerning potential generalizations and further applications of the methods.

2 Modeling of thermoviscoelastic materials in the reference configuration

We will use the Lagrangian approach and formulate the model in the reference (fixed) domain $\Omega \subset \mathbb{R}^d$ being bounded with smooth boundary Γ. We assume $d \geq 2$ although, of course, the rather trivial case $d = 1$ works too if $p \geq 2$ is assumed additionally to $p > d$ in (2.30) below. We will consider a fixed time horizon $T > 0$ and use the notation $I := [0, T]$, $Q := I \times \Omega$, and $\Sigma := I \times \Gamma$. For readers' convenience, Table 1 summarizes the main nomenclature used throughout the paper.

y deformation, $y(t,x) \in \mathbb{R}^d$,	ζ potential of dissipative forces,
θ absolute temperature,	ξ rate of dissipation (=heat production),
$(\cdot)^{\cdot}$ time derivative,	$\mathbb{K} = \mathbb{K}(\theta)$ material heat conductivity,
$\psi = \phi + \varphi$ free energy,	$\mathcal{K} = \mathcal{K}(F, \theta)$ pulled-back heat conductivity,
$\sigma_{\rm el} = \partial_F \psi$ elastic stress,	$C = F^\top F$ right Cauchy-Green tensor,
$\sigma_{\rm vi} = \partial_{\dot F} \zeta$ viscous stress,	κ heat-transfer coefficient on Γ,
$F = \nabla y$ deformation gradient,	$g : I \times \Omega \to \mathbb{R}^d$ a time-dependent dead force,
$\mathsf{G} = \nabla F = \nabla^2 y$ valued in $\mathbb{R}^{d\times d\times d}$,	$f : I \times \Gamma_{\rm N} \to \mathbb{R}^d$ a boundary traction,
w heat part of internal energy,	ℓ an external mechanical loading,
$\mathfrak{h}_{\rm el}$ elastic hyperstress,	Ω the reference domain,
$c_{\rm v} = c_{\rm v}(F, \theta)$ heat capacity,	Γ the boundary of Ω, $\Gamma = \Gamma_{\rm D} \cap \Gamma_{\rm N}$,
$\vec{q}$ heat flux,	$I := [0, T]$ the fixed time interval,
$\mathcal{M} = \Phi_{\rm el} + \mathcal{H}$ main mechanical energy,	$Q := I \times \Omega$,
$\mathcal{H}$ hyperstress energy,	$\Sigma := I \times \Gamma$,
$\Phi_{\rm cpl}$ coupling energy,	$\mathscr{H} = \mathscr{H}(\nabla F)$ the potential of $\mathfrak{h}_{\rm el}$,
$\Psi = \mathcal{M} + \Phi_{\rm cpl}$ free energy,	$\mathcal{Y}_0, \mathcal{Y}_{\rm id}$ sets of admissimble deformations,
$\mathcal{W}$ thermal energy,	$\mathrm{GL}^+(d) := \{A \in \mathbb{R}^{d\times d};\ \det A > 0\}$,
$\mathcal{E} = \mathcal{M} + \mathcal{W}$ total energy,	$\mathrm{SO}(d) := \{A \in \mathrm{GL}^+(d);\ A^\top A = \mathbb{I} = A A^\top\}$.

Table 1. *Summary of the basic notation used throughout the paper.*

To introduce our model in a broader context, we may define the *total free energy* and the *total dissipation potential*

$$\Psi(y,\theta) = \int_\Omega \big(\psi(\nabla y, \theta) + \mathscr{H}(\nabla^2 y)\big)\,\mathrm{d}x \quad \text{and} \quad \mathcal{R}(y, \dot y, \theta) = \int_\Omega \zeta(\nabla y, \nabla \dot y, \theta)\,\mathrm{d}x, \tag{2.1}$$

respectively. The mechanical evolution part can then be viewed as an abstract gradient flow

$$\mathrm{D}_{\dot y}\mathcal{R}(y,\dot y,\theta)+\mathrm{D}_y\Psi(y,\theta)=\ell(t)\ \text{ with }\langle\ell(t),y\rangle=\int_\Omega g(x,t)\cdot y(x)\,\mathrm{d}x+\int_{\Gamma_{\mathrm{N}}} f(x,t)\cdot y(x)\,\mathrm{d}S, \tag{2.2}$$

cf. also [Tve08, MOS13] for the isothermal case and [Mie11] for the general case. The sum of the conservative and the dissipative parts corresponds to the *Kelvin-Voigt rheological model* in the quasistatic variant (neglecting inertia). The notation "∂" is used for partial derivatives (here functional or later in Euclidean spaces), while $(\cdot)'$ will occasionally be used for functions of only one variable.

Writing (2.2) locally in the classical formulation, one arrives at the nonlinear parabolic 4th-order partial differential equation expressing quasistatic *momentum equilibrium*

$$\mathrm{div}\ \sigma+g=0\qquad\text{with}\quad\sigma=\sigma_{\mathrm{vi}}+\sigma_{\mathrm{el}}-\mathrm{div}\ \mathfrak{h}_{\mathrm{el}}, \tag{2.3}$$

where the viscous stress is $\sigma_{\mathrm{vi}}=\sigma_{\mathrm{vi}}(F,\dot F,\theta)$ and the elastic stress is $\sigma_{\mathrm{el}}=\sigma_{\mathrm{el}}(F,\theta)$, while $\mathfrak{h}_{\mathrm{el}}$ is a so-called hyperstress arising from the 2nd-grade nonsimple material concept, cf. e.g. [Pod02, Šil85, Tou62]. In view of the local potentials used in (2.2), we have

$$\sigma_{\mathrm{vi}}(F,\dot F,\theta)=\partial_{\dot F}\zeta(F,\dot F,\theta),\quad\sigma_{\mathrm{el}}(F,\theta)=\partial_F\psi(F,\theta),\quad\text{and}\quad\mathfrak{h}_{\mathrm{el}}(\mathsf{G})=\mathscr{H}'(\mathsf{G}), \tag{2.4}$$

where $\mathsf{G}\in\mathbb{R}^{d\times d\times d}$ is a placeholder for ∇F.

An important physical requirement is *static and dynamic frame indifference*. For the elastic stresses, static frame indifference means that

$$\sigma_{\mathrm{el}}(RF,\theta)=R\,\sigma_{\mathrm{el}}(F,\theta)\quad\text{and}\quad\mathfrak{h}_{\mathrm{el}}(R\mathsf{G})=R\mathfrak{h}_{\mathrm{el}}(\mathsf{G}) \tag{2.5a}$$

for all $R\in\mathrm{SO}(d)$, F and G. For the viscous stresses, dynamic frame indifference means that

$$\sigma_{\mathrm{vi}}(RF,\dot RF+R\dot F,\theta)=R\,\sigma_{\mathrm{vi}}(F,\dot F,\theta) \tag{2.5b}$$

for all smoothly time-varying $R:t\mapsto R(t)\in\mathrm{SO}(d)$, cf. [Ant98]. Note that R may depend on t but not on $x\in\Omega$, since frame-indifference relates to superimposing time-dependent *rigid-body motions*.

In terms of the thermodynamic potentials ζ, ψ, and $\mathscr{H}$, these frame indifferences read as

$$\psi(RF,\theta)=\psi(F,\theta),\quad\mathscr{H}(R\nabla F)=\mathscr{H}(\nabla F),\quad\text{and} \tag{2.6a}$$

$$\zeta(RF,\theta;(RF)^{\bullet})=\zeta(RF,\theta;\dot RF+R\dot F)=\zeta(F,\theta;\dot F) \tag{2.6b}$$

for R, F and ∇F as above. These frame indifferences imply the existence of reduced potentials $\hat\psi$, $\hat\zeta$, and $\hat{\mathscr{H}}$ such that

$$\zeta(F,\dot F,\theta)=\hat\zeta(C,\dot C,\theta),\quad\psi(F,\theta)=\hat\psi(C,\theta),\quad\text{and}\quad\mathscr{H}(\mathsf{G})=\hat{\mathscr{H}}(\mathsf{B}) \tag{2.7}$$

where $\mathsf{B}=\mathsf{G}^\top\cdot\mathsf{G}\in\mathbb{R}^{(d\times d)\times(d\times d)}$, and $C\in\mathbb{R}^{d\times d}_{\mathrm{sym}}$ is the right Cauchy-Green tensor $C=F^\top F$ with time derivative $\dot C=\dot F^\top F+F^\top\dot F$. More specifically, denoting $\mathsf{G}=[\mathsf{G}_{\alpha ij}]$ the placeholder for $\frac{\partial}{\partial x_j}F_{\alpha i}$ with $F_{\alpha i}$ the placeholder for $\frac{\partial}{\partial x_i}y_\alpha$, the exact meaning is $[\mathsf{G}^\top\cdot\mathsf{G}]_{ijkl}:=\sum_{\alpha=1}^d\mathsf{G}_{\alpha ij}\mathsf{G}_{\alpha kl}$ and $[F^\top F]_{ij}:=\sum_{\alpha=1}^d F_{\alpha i}F_{\alpha j}$. The ansatz (2.7) also means that

$$\sigma_{\mathrm{el}}(F,\theta):=\partial_F\psi(F;\theta)=2F\partial_C\hat\psi(F^\top F,\theta)=2F\partial_C\hat\psi(C,\theta), \tag{2.8a}$$

$$\mathfrak{h}_{\mathrm{el}}(\mathsf{G}):=\partial_{\mathsf{G}}\mathscr{H}(\mathsf{G})=2\mathsf{G}\partial_{\mathsf{B}}\hat{\mathscr{H}}(\mathsf{G}^\top\cdot\mathsf{G})=2\mathsf{G}\partial_{\mathsf{B}}\hat{\mathscr{H}}(\mathsf{B}), \tag{2.8b}$$

$$\sigma_{\mathrm{vi}}(F,\dot F,\theta):=\partial_{\dot F}\zeta(F,\dot F,\theta)=2F\partial_{\dot C}\hat\zeta(F^\top F,\dot F^\top F+F^\top\dot F,\theta)=2F\partial_{\dot C}\hat\zeta(C,\dot C,\theta). \tag{2.8c}$$

The simplest choice, which is adopted in this paper for avoiding unnecessary technicalities, is that the viscosity $\sigma_{\rm vi}$ is linear in $\dot C$. This is the relevant modeling choice for non-activated dissipative processes with rather moderate rates (in contrast to activated processes like plasticity having nonsmooth potentials that are homogeneous of degree 1 in a small-rate approximation). This linear viscosity leads to a potential which is quadratic in $\dot C$, viz.

$$\hat\zeta(C,\dot C,\theta) := \frac{1}{2}\,\dot C : \mathbb{D}(C,\theta)\dot C\,. \tag{2.9}$$

Although for this choice the material viscosity is linear, the geometrical nonlinearity arising from large strains is still a vital part of the problem due to the requirement of frame indifference. Note that $\sigma_{\rm vi}(F,\dot F,\theta)$ necessarily depends on F if we express $\dot C$ in terms of the velocity gradients $\dot F$, even if $\mathbb{D}$ is constant: $\sigma_{\rm vi}(F,\dot F,\theta) = 2F\mathbb{D}(C,\theta)(\dot F^\top F + F^\top \dot F)$. While we will be able to handle general dependence on F, it will be a crucial restriction that $\dot F \mapsto \sigma_{\rm vi}(F,\dot F,\theta)$ is linear.

Furthermore, the *specific dissipation rate* can be simply identified in terms of $\hat\zeta$ as

$$\begin{aligned}\xi(F,\dot F,\theta) &= \sigma_{\rm vi}(F,\dot F,\theta){:}\dot F = 2F\partial_{\dot C}\hat\zeta(F^\top F, \dot F^\top F + F^\top \dot F,\theta){:}\dot F\\ &= \partial_{\dot C}\hat\zeta(F^\top F, \dot F^\top F + F^\top \dot F,\theta){:}(\dot F^\top F + F^\top \dot F) = \partial_{\dot C}\hat\zeta(C,\dot C,\theta){:}\dot C.\end{aligned} \tag{2.10}$$

For our choice (2.9), we simply have $\xi(F,\dot F,\theta) = \mathbb{D}(C,\theta)\dot C{:}\dot C = 2\hat\zeta(C,\dot C,\theta) = 2\zeta(F,\dot F,\theta)$.

In brief, the standard thermodynamical arguments start from the free energy density ψ and the definition of *entropy* via $s = -\partial_\theta\psi$ (here $\mathscr{H}$ does play no role as it is chosen to be independent of θ) and the entropy equation

$$\theta\dot s = \xi - \operatorname{div}\,\vec q \tag{2.11}$$

with the dissipation rate ξ from (2.10) and the heat flux $\vec q$. We further use the formula $\dot s = -\partial^2_{\theta\theta}\psi\,\dot\theta - \partial^2_{F\theta}\psi{:}\dot F$ and the *Fourier law* formulated in the reference configuration

$$\vec q = -\mathcal{K}(F,\theta)\nabla\theta, \tag{2.12}$$

which will be specified later in (2.24). Altogether, we arrive at the coupled system

$$\operatorname{div}\big(\sigma_{\rm vi}(\nabla y,\nabla\dot y,\theta) + \sigma_{\rm el}(\nabla y,\theta) - \operatorname{div}\mathfrak{h}_{\rm el}(\nabla^2 y)\big) + g$$
$$\text{with}\ \ \sigma_{\rm vi}(F,\dot F,\theta) = \partial_{\dot F}\zeta(F,\dot F,\theta)\ \text{and}\ \ \sigma_{\rm el}(F,\theta) = \partial_F\psi(F,\theta)\,, \tag{2.13a}$$

$$c_{\rm v}(\nabla y,\theta)\dot\theta = \operatorname{div}\big(\mathcal{K}(\nabla y,\theta)\nabla\theta\big) + \xi(\nabla y,\nabla\dot y,\theta) + \theta\partial^2_{F\theta}\psi(\nabla y,\theta){:}\nabla\dot y$$
$$\text{with}\ \ c_{\rm v}(F,\theta) = -\theta\partial^2_{\theta\theta}\psi(F,\theta)\ \ \text{and}\ \ \xi\ \ \text{from (2.10)} \tag{2.13b}$$

on Q. We complete (2.13) by some boundary conditions. For simplicity, we only consider a mechanically fixed part $\Gamma_{\rm D}$ time independent undeformed (i.e. identity) while the whole boundary is thermally exposed with a phenomenological heat-transfer coefficient $\kappa \ge 0$:

$$\big(\sigma_{\rm vi}(\nabla y,\nabla\dot y,\theta) + \sigma_{\rm el}(\nabla y,\theta)\big)\vec n - \operatorname{div}_{\rm s}\big(\mathfrak{h}_{\rm el}(\nabla^2 y)\vec n\big) = f \qquad \text{on}\ \ \Gamma_{\rm N}, \tag{2.14a}$$
$$y(x) = x\ \ \text{(identity)} \qquad \text{on}\ \ \Gamma_{\rm D}, \tag{2.14b}$$
$$\mathfrak{h}_{\rm el}(\nabla^2 y){:}(\vec n\otimes\vec n) = 0 \qquad \text{on}\ \ \Gamma, \tag{2.14c}$$
$$\mathcal{K}(\nabla y,\theta)\nabla\theta\cdot\vec n + \kappa\theta = \kappa\theta_{\flat} \qquad \text{on}\ \ \Gamma, \tag{2.14d}$$

where $\vec{n}$ is the outward pointing normal vector, and $\theta_\flat$ is a given external temperature. Moreover, following [Bet86] the surface divergence "$\mathrm{div}_{\mathrm{s}}$" in (2.14a) is defined as $\mathrm{div}_{\mathrm{s}}(\cdot) = \mathrm{tr}\big(\nabla_{\mathrm{s}}(\cdot)\big)$, where $\mathrm{tr}(\cdot)$ denotes the trace and ∇_{s} denotes the surface gradient given by $\nabla_{\mathrm{s}} v = (\mathbb{I} - \vec{n}\otimes\vec{n})\nabla v = \nabla v - \frac{\partial v}{\partial \vec{n}}\vec{n}$. See (2.29) for a short mathematical derivation of the boundary conditions (2.14a) and (2.14c), and [Ste15, pp. 358-359] for the mechanical interpretation in second-order materials.

In order to facilitate the subsequent mathematical analysis, we assume a rather weak thermal coupling through the free energy (together with the coupling through the temperature-dependent viscous dissipation). To distinguish the particular coupling thermo-mechanical term from the purely mechanical one, we consider the explicit ansatz

$$\psi(F,\theta) = \varphi(F) + \phi(F,\theta) \quad \text{with} \quad \phi(F,0) = 0. \tag{2.15}$$

In applications, the internal energy e given by Gibbs' relation

$$e = e(F,\theta) = \psi(F,\theta) + \theta s = \psi(F,\theta) - \theta\partial_\theta\psi(F,\theta) = \psi(F,\theta) - \theta\partial_\theta\phi(F,\theta).$$

is often balanced. Here, we rather use the thermal part of the internal energy $w := e - \varphi(F)$. In view of the ansatz (2.15), we have

$$w = \mathfrak{w}(F,\theta) = \psi(F,\theta) - \theta\partial_\theta\phi(F,\theta) - \psi(F,0) = \phi(F,\theta) - \theta\partial_\theta\phi(F,\theta). \tag{2.16}$$

Note that $\mathfrak{w}(F,\cdot)$ is the primitive function of the specific heat $c_{\mathrm{v}}(F,\cdot)$ calibrated as $\mathfrak{w}(F,0) = 0$, so that also $e = \psi(F,0) + w$. The heat-transfer equation (2.13b) simplifies as

$$\dot{w} - \mathrm{div}\left(\mathcal{K}(\nabla y,\theta)\nabla\theta\right) = \xi(\nabla y, \nabla\dot{y},\theta) + \partial_F\phi(\nabla y,\theta){:}\nabla\dot{y} \ \text{ with } \ w = \mathfrak{w}(F,\theta)\,. \tag{2.17}$$

In particular, the purely mechanical stored energy φ does not occur in (2.16) and does not influence the heat production and transfer (2.17).

The energetics of the system (2.13)–(2.14) can be best described by introducing additional energy functionals as follows:

$$\mathcal{H}(y) := \int_\Omega \mathscr{H}(\nabla^2 y)\,\mathrm{d}x \quad \text{hyperstress energy,} \tag{2.18a}$$

$$\mathcal{M}(y) := \mathcal{H}(y) + \Phi_{\mathsf{el}}(y) \text{ with } \Phi_{\mathsf{el}}(y) := \int_\Omega \varphi(\nabla y)\,\mathrm{d}x \quad \text{main mech. energy,} \tag{2.18b}$$

$$\Phi_{\mathsf{cpl}}(y,\theta) := \int_\Omega \phi(\nabla y,\theta)\,\mathrm{d}x \quad \text{coupling energy,} \tag{2.18c}$$

$$\Psi(y,\theta) := \mathcal{M}(y) + \Phi_{\mathsf{cpl}}(y,\theta) \quad \text{free energy,} \tag{2.18d}$$

$$\mathcal{W}(y,\theta) := \int_\Omega \mathfrak{w}(\nabla y,\theta)\,\mathrm{d}x \quad \text{thermal energy,} \tag{2.18e}$$

$$\mathcal{E}(y,\theta) := \mathcal{M}(y) + \mathcal{W}(y,\theta) \quad \text{total energy.} \tag{2.18f}$$

An mechanical energy balance is revealed by testing (2.13a) by $\dot{y}$ and (2.13b) by 1, and using the boundary conditions after integration over Ω and using Green's formula twice together with another $(d-1)$-dimensional Green formula over Γ for (2.13a) and once again Green's formula for (2.13b). The last mentioned technique is related with the concept of nonsimple materials; for the details about how the boundary conditions are handled see e.g. [Rou13, Sect. 2.4.4]. This test of (2.13a) gives the mechanical energy balance:

$$\int_\Omega \underbrace{\xi(\nabla y,\nabla\dot{y},\theta)}_{\substack{\text{dissipation}\\ \text{rate}}} + \underbrace{\sigma_{\mathrm{el}}{:}\nabla\dot{y}}_{\substack{\text{mechanical}\\ \text{power}}}\,\mathrm{d}x + \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}(y) = \int_\Omega \underbrace{g\cdot\dot{y}}_{\substack{\text{power of the}\\ \text{bulk force}}}\mathrm{d}x + \int_{\Gamma_{\mathrm{N}}} \underbrace{f\cdot\dot{y}}_{\substack{\text{power of}\\ \text{the traction}}}\,\mathrm{d}S. \tag{2.19}$$

Using $\sigma_{\text{el}} = \partial_F\varphi + \partial_F\phi$ and integrating in time leads to the relation

$$\mathcal{M}(y(T)) + \int_0^t \int_\Omega \Big(\int_\Omega \xi(\nabla y, \nabla\dot{y}, \theta) + \partial_F\phi(\nabla y, \theta){:}\nabla\dot{y} \Big) \,\mathrm{d}x\,\mathrm{d}t = \mathcal{M}(y(0)) + \int_0^t \langle \ell, \dot{y}\rangle \,\mathrm{d}t. \tag{2.20}$$

that will be very useful for obtaining a priori estimates in the following sections.

Next, we test the heat equation in its simplified form (2.17) together with the boundary conditions (2.14d) by the constant function 1 (i.e. we merely integrated over Ω) and add the result to (2.20). After major cancellations we obtain the total energy balance:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(y,\theta) = \int_\Omega \underbrace{g\cdot\dot{y}}_{\substack{\text{power of mecha-}\\ \text{nical bulk load}}} \mathrm{d}x + \int_{\Gamma_{\text{N}}} \underbrace{f\cdot\dot{y}}_{\substack{\text{power of}\\ \text{the traction}}} \mathrm{d}S - \int_\Gamma \underbrace{\kappa(\theta-\theta_{\text{b}})}_{\substack{\text{power of the}\\ \text{external heating}}} \mathrm{d}S. \tag{2.21}$$

In particular, we see that the total energy is conserved up to the work induced by the external loadings or the flux of heat through the boundary.

From the entropy equation (2.11), we can read the total entropy balance (the *Clausius-Duhem inequality*):

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega s(t,x)\,\mathrm{d}x &= \int_\Omega \frac{\xi + \operatorname{div}(\mathcal{K}\nabla\theta)}{\theta}\,\mathrm{d}x = \int_\Omega \frac{\xi}{\theta} - \mathcal{K}\nabla\theta\cdot\nabla\frac{1}{\theta}\,\mathrm{d}x + \int_\Gamma \frac{\mathcal{K}\nabla\theta}{\theta}\cdot\vec{n}\,\mathrm{d}S \\ &= \int_\Omega \underbrace{\frac{\xi}{\theta} + \frac{\mathcal{K}\nabla\theta\cdot\nabla\theta}{\theta^2}}_{\substack{\text{entropy-production}\\ \text{rate}}}\,\mathrm{d}x \;+\; \int_\Gamma \frac{\mathcal{K}\nabla\theta}{\theta}\cdot\vec{n}\,\mathrm{d}S \geq \int_\Gamma \underbrace{\frac{-\vec{q}}{\theta}\cdot\vec{n}}_{\substack{\text{entropy flux}\\ \text{through boundary}}}\mathrm{d}S.\end{aligned} \tag{2.22}$$

This articulates, in particular, the second law of thermodynamics that the total entropy in the isolated systems (i.e. here $\vec{q} = 0$ on Γ) is nondecreasing with time provided $\mathcal{K} = \mathcal{K}(\nabla y, \theta)$ is positive semidefinite and the dissipation rate is non-negative.

It is certainly a very natural modeling choice that Fourier's law is formulated in the actual (also called the deformed) configuration in a simple form, namely the *actual heat flux* is given by

$$\vec{\mathsf{q}} = -\mathbb{K}(\theta)\nabla_z\theta, \quad \text{where } z = y(x) \text{ and } \theta(z) = \theta(y^{-1}(z)) \text{ for } z \in y(\Omega) \tag{2.23}$$

with the heat-conductivity tensor $\mathbb{K} = \mathbb{K}(x,\theta)$ considered as a material parameter possibly dependent on $x \in \Omega$. We transform (i.e. *pulled-back*) this *Fourier law* into the reference configuration via the heat flux $\vec{q}(x) = \mathcal{K}(x)\nabla\theta = \mathcal{K}(\nabla y(x))^\top\nabla_z\theta(y(x))$ and $\vec{q} = (\operatorname{Cof} F^\top)\vec{\mathsf{q}}$, because fluxes should be considered as $(d-1)$-forms. With (2.23) the usual transformation rule for 2nd-order contra-variant tensors yields the heat-conductivity tensor

$$\mathcal{K}(x,F,\theta) = (\operatorname{Cof} F^\top)\mathbb{K}(x,\theta)F^{-\top} = \frac{(\operatorname{Cof} F^\top)\mathbb{K}(x,\theta)\operatorname{Cof} F}{\det F} = (\det F)F^{-1}\mathbb{K}(x,\theta)F^{-\top} \tag{2.24}$$

if $\det F > 0$, whereas the case $\det F \leq 0$ is considered nonphysical, so $\mathcal{K}$ is then not defined. Here we used the standard shorthand notation $F^{-\top} = [F^{-1}]^\top = [F^\top]^{-1}$ and also the algebraic formula $F^{-1} = (\operatorname{Cof} F^\top)/\det F$. In what follows, we omit explicit x-dependence for notational simplicity. Let us emphasize that in our formulations $\nabla\theta$ is *not* treated as a vector, but a contravariant 1-form. Starting from $\theta(x) = \theta(y(x))$ the chain-rule gives $\nabla(x) = \nabla y(x)^\top\nabla_Y\boldsymbol{\theta}(y(x))$. It should be noted that (2.23) is rather formal argumentation, assuming injectivity of the deformation y and thus existence

of y^{-1}, which is however not guaranteed in our model; anyhow, handling only local non-selfpenetration while ignoring possible global selfpenetration is our modeling approach often accepted in engineering, too.

For the isotropic case $\mathbb{K}(\theta) = \varkappa(\theta)\mathbb{I}$, relation (2.24) can also be written by using the right Cauchy-Green tensor $C = F^\top F$ as $\mathcal{K} = \det(F)\varkappa(\theta)C^{-1}$, cf. e.g. [DSF10, Formula (67)] or [GoS93, Formula (3.19)] for the mass instead of the heat transport. In principle, $\mathbb{K}$ in (2.23) itself may also depend on $C = F^\top F$, which we omitted to emphasize that $\mathcal{K}$ in (2.24) will depend on F anyhow.

In what follows, we will use the (standard) notation for the Lebesgue L^p-spaces and $W^{k,p}$ for Sobolev spaces whose k-th distributional derivatives are in L^p-spaces and the abbreviation $H^k = W^{k,2}$. The notation $W^{1,p}_{\rm D}$ will indicate the closed subspace of $W^{1,p}$ with zero traces on $\Gamma_{\rm D}$. Moreover, we will use the standard notation $p' = p/(p-1)$. In the vectorial case, we will write $L^p(\Omega;\mathbb{R}^n) \cong L^p(\Omega)^n$ and $W^{1,p}(\Omega;\mathbb{R}^n) \cong W^{1,p}(\Omega)^n$. Thus, for example,

$$H^1_{\rm D}(\Omega;\mathbb{R}^d) := \big\{v \in L^2(\Omega;\mathbb{R}^d);\ \nabla v \in L^2(\Omega;\mathbb{R}^{d\times d}),\ \ v|_{\Gamma_{\rm D}} = 0\big\}. \tag{2.25}$$

For the fixed time interval $I = [0,T]$, we denote by $L^p(I;X)$ the standard Bochner space of Bochner-measurable mappings $I \to X$ with X a Banach space. Also, $W^{k,p}(I;X)$ denotes the Banach space of mappings from $L^p(I;X)$ whose k-th distributional derivative in time is also in $L^p(I;X)$. The dual space to X will be denoted by X^*. Moreover, $C_{\rm w}(I;X)$ denotes the Banach space of weakly continuous functions $I \to X$. The scalar product between vectors, matrices, or 3rd-order tensors will be denoted by "$\cdot$", "$:$", or "$\vdots$", respectively. Finally, in what follows, K denotes a positive, possibly large constant.

We consider an initial-value problem, imposing the initial conditions

$$y(0,\cdot) = y_0 \quad \text{and} \quad \theta(0,\cdot) = \theta_0 \quad \text{on } \ \Omega. \tag{2.26}$$

Having in mind the form (2.17) of the heat equation, we can now state the following definition for a weak solution:

Definition 2.1 (Weak solution). *A couple* $(y,\theta) : Q = [0,T]\times\Omega \to \mathbb{R}^d \times \mathbb{R}$ *is called a* weak solution *to the initial-boundary-value problem* (2.13) *&* (2.14) *&* (2.26) *if* $(y,\theta) \in C_{\rm w}(I;W^{2,p}(\Omega;\mathbb{R}^d)) \times L^1(I;W^{1,1}(\Omega))$ *with* $\nabla\dot y \in L^2(Q;\mathbb{R}^{d\times d})$, *if* $\min_Q \det\nabla y > 0$ *and* $y|_{\Sigma_{\rm D}} = \text{identity}$, *and if it satisfies the integral identity*

$$\int_0^T \bigg(\int_\Omega \Big(\big(\sigma_{\rm vi}(\nabla y, \nabla\dot y, \theta) + \sigma_{\rm el}(\nabla y,\theta)\big){:}\nabla z + \mathfrak{h}_{\rm el}(\nabla^2 y)\vdots\nabla^2 z \Big) \,{\rm d}x\,{\rm d}t \\ = \int_Q g{\cdot}z \,{\rm d}x\,{\rm d}t + \int_{\Sigma_{\rm N}} f{\cdot}z \,{\rm d}S\,{\rm d}t \tag{2.27a}$$

for all smooth $z : Q \to \mathbb{R}^d$ *with* $z = 0$ *on* $\Sigma_{\rm D}$ *together with* $y(0,\cdot) = y_0$, *and if*

$$\int_Q \mathcal{K}(\nabla y,\theta)\nabla\theta{\cdot}\nabla v - \big(\xi(\nabla y,\nabla\dot y,\theta) + \partial_F\phi(\nabla y,\theta){:}\nabla\dot y\big)v - \mathfrak{w}(\nabla y,\theta)\dot v \,{\rm d}x\,{\rm d}t \\ + \int_\Sigma \kappa\theta v \,{\rm d}S\,{\rm d}t = \int_\Sigma \kappa\theta_\flat v \,{\rm d}S\,{\rm d}t + \int_\Omega \mathfrak{w}(\nabla y_0,\theta_0)v(0)\,{\rm d}x \tag{2.27b}$$

for all smooth $v : Q \to \mathbb{R}$ *with* $v(T) = 0$, *where* $\mathfrak{w}$ *is defined in* (2.16).

At first sight, it seems that (2.27a) is not suited to apply the test function $z = \dot y$, which is the natural and necessary choice for deriving energy bounds. Obviously, we will not be able to obtain enough

control on $\nabla^2\dot y$. However, using the abstract chain rules provides in Section 3.3 this problem can be handled by extending $\mathcal{H}(y)=\int_\Omega \mathscr{H}(\nabla^2 y)\,\mathrm{d}x$ to a lower semicontinuous and convex functional on $H^1(\Omega;\mathbb{R}^d)$ by setting it ∞ outside $W^{2,p}(\Omega;\mathbb{R}^d)$, see the rigorous proof of (5.9) in Step 3 of the proof of Proposition 5.1.

It will be somewhat technical to see that the weak formulation (2.27a) is indeed selective enough, in the sense that for sufficiently smooth solutions one can indeed obtain the classical formulation (2.13) together with the boundary conditions (2.14), cf. also [Rou13, Sect. 2.4.4]. In particular, abbreviating $\sigma = \sigma_{\rm vi}(\nabla y,\nabla\dot y,\theta)+\sigma_{\rm el}(\nabla y,\theta)$, integrating by part once, and using the boundary conditions (2.14a,c) yields

$$\int_Q\Big(\big(\sigma-\operatorname{div}\mathfrak{h}_{\rm el}(\nabla^2 y)\big):\nabla z-g\cdot z\Big)\,\mathrm{d}x\,\mathrm{d}t=\int_{\Sigma_{\rm N}} f\cdot z\,\mathrm{d}S\,\mathrm{d}t-\int_\Sigma \mathfrak{h}_{\rm el}(\nabla^2 y)\vdots(\nabla z\otimes\vec n)\,\mathrm{d}S\,\mathrm{d}t. \tag{2.28}$$

We now want to show how the strong form (2.13a) and the associated boundary conditions (2.14a,c) follow from (2.28). For this goal, we apply Green's formula in the opposite direction to remove ∇ in front of the test function z. Using also the orthogonal decomposition of $\nabla z=\nabla_{\rm s}z+\frac{\partial}{\partial\vec n}z\otimes\vec n$ involving the surface gradient $\nabla_{\rm s}z$ and writing shortly $\mathfrak{h}$ for $\mathfrak{h}_{\rm el}(\nabla^2 y)\in\mathbb{R}^{d\times d\times d}$, relation (2.28) leads to the identity

$$\begin{aligned}&\int_Q\big(-\operatorname{div}\,\sigma+\operatorname{div}^2\mathfrak{h}-g\big)\cdot z\,\mathrm{d}x\,\mathrm{d}t\\&=\int_\Sigma\Big(\big(\sigma-\operatorname{div}\mathfrak{h}\big):(z\otimes\vec n)-\mathfrak{h}\vdots(\nabla z\otimes\vec n)\Big)\,\mathrm{d}x\,\mathrm{d}t+\int_{\Sigma_{\rm N}} f\cdot z\,\mathrm{d}S\,\mathrm{d}t\\&=\int_\Sigma\Big((\sigma-\operatorname{div}\mathfrak{h})\vec n\cdot z+\big(\mathfrak{h}:(\vec n\otimes\vec n)\big)\cdot\frac{\partial z}{\partial\vec n}+\mathfrak{h}\vec n:\nabla_{\rm s}z)\Big)\,\mathrm{d}S\,\mathrm{d}t-\int_{\Sigma_{\rm N}} f\cdot z\,\mathrm{d}S\,\mathrm{d}t\end{aligned}$$

Using the surface divergence $\operatorname{div}_{\rm s}$ and the projection $P_{\rm s}:A\mapsto A-A\vec n\otimes\vec n$ to the tangential part, we obtain the integration by parts formula (cf. [Bet86] or [Ste15, pp. 358-359])

$$\int_{\partial\Omega}A:\nabla_{\rm s}z\,\mathrm{d}S=\int_{\partial\Omega}(P_{\rm s}A):\nabla_{\rm s}z\,\mathrm{d}S=-\int_{\partial\Omega}\operatorname{div}_{\rm s}(P_{\rm s}A)\cdot z\,\mathrm{d}S,$$

where the surface Γ is now assumed to be sufficiently smooth. Using this with $A=\mathfrak{h}\vec n$ for the previous relation we find

$$\begin{aligned}&\int_Q\big(-\operatorname{div}\,\sigma+\operatorname{div}^2\mathfrak{h}-g\big)\cdot z\,\mathrm{d}x\,\mathrm{d}t\\&=\int_{\Sigma_{\rm N}}\Big((\sigma-\operatorname{div}\mathfrak{h})\vec n-\operatorname{div}_{\rm s}\big(P_{\rm s}(\mathfrak{h}\vec n)\big)-f\Big)\cdot z\,\mathrm{d}S\,\mathrm{d}t+\int_\Sigma\big(\mathfrak{h}:(\vec n\otimes\vec n)\big)\cdot\frac{\partial z}{\partial\vec n}\,\mathrm{d}S\,\mathrm{d}t,\end{aligned}\tag{2.29}$$

where we have used $z=0$ on $\Sigma_{\rm D}=\Sigma\setminus\Sigma_{\rm N}$. Now, taking z's with a compact support in Q, we obtain the equilibrium (2.13a) in the bulk. Next taking taking z's with zero traces on Σ but general $\frac{\partial z}{\partial\vec n}$, we obtain (2.14c). Note that the latter condition implies $P_{\rm s}(\mathfrak{h}\vec n)=\mathfrak{h}\vec n-\big(\mathfrak{h}:(\vec n\otimes\vec n)\big)\otimes\vec n=\mathfrak{h}\vec n$. Hence, taking finally general z's, we obtain (2.14a), as $P_{\rm s}$ can be dropped because of (2.14c).

Moreover, also note that, from the integral identity (2.27b), one can read $\mathfrak{w}(\nabla y(0),\theta(0))=\mathfrak{w}(\nabla y_0,\theta_0)$ from which $\theta(0)=\theta_0$ follows when taken the invertibility of $\mathfrak{w}(F,\cdot)$ and $y(0)=y_0$ into account.

Now we exploit the decomposition (2.15) of ψ into ϕ and φ, which allows us to impose coercivity assumptions for the purely elastic part ϕ that are independent of those for φ, namely

$$\exists\, p\in\,]d,\infty[\,\cap\,[2,\infty[,\ s>0,\ q\ge pd/(p{-}d)\ \ \exists\,\alpha,K,\hat\epsilon>0:$$

$$\varphi:\mathrm{GL}^+(d)\to\mathbb{R}^+ \text{ twice continuously differentiable},\ \forall\, F\in\mathrm{GL}^+(d):$$
$$\varphi(F)\ge\hat\epsilon|F|^s+\hat\epsilon/(\det F)^q, \tag{2.30a}$$

$$\phi:\mathrm{GL}^+(d){\times}\mathbb{R}^+\to\mathbb{R}^+ \text{ twice continuously differentiable},\ \forall\, F,\tilde F\in\mathrm{GL}^+(d),\ \theta\ge0:$$
$$\big|\phi(F,\theta){-}\phi(\tilde F,\theta)\big|\le K\big(1{+}|F|^{s/2}{+}|\tilde F|^{s/2}\big)|F{-}\tilde F|, \tag{2.30b}$$
$$\partial^2_{FF}\phi(F,\theta)\le K,\quad |\theta\partial^2_{F\theta}\phi(F,\theta)|\le K,\quad \hat\epsilon\le-\theta\partial^2_{\theta\theta}\phi(F,\theta)\le K, \tag{2.30c}$$

$$\mathscr{H}:\mathbb{R}^{d\times d\times d}\to\mathbb{R}^+ \text{ convex, continuously differentiable},\forall\, G\in\mathbb{R}^{d\times d\times d}:$$
$$\hat\epsilon|G|^p\le\mathscr{H}(G)\le K(1{+}|G|^p), \tag{2.30d}$$

$$\hat\zeta:\mathbb{R}^{d\times d}_{\rm sym}{\times}\mathbb{R}^{d\times d}_{\rm sym}{\times}\mathbb{R}\to\mathbb{R}^+ \text{ is continuous and } \forall\,(C,\dot C,\theta)\in\mathbb{R}^{d\times d}_{\rm sym}{\times}\mathbb{R}{\times}\mathbb{R}^{d\times d}_{\rm sym}:$$
$$\hat\zeta(C,\cdot,\theta):\mathbb{R}^{d\times d}_{\rm sym}\to\mathbb{R}^+ \text{ quadratic (cf. (2.9))},\ \ \alpha|\dot C|^2\le\hat\zeta(C,\dot C,\theta)\le K|\dot C|^2, \tag{2.30e}$$

$$\mathbb{K}:\mathbb{R}\to\mathbb{R}^{d\times d} \text{ is continuous, uniformly positive definite, and bounded,} \tag{2.30f}$$

$$g\in L^2(Q;\mathbb{R}^d),\quad f\in L^2(\Sigma_{\rm N};\mathbb{R}^d),\quad \kappa>0, \tag{2.30g}$$

$$y_0\in\mathcal{Y}_{\rm id}:=\{\,y\in W^{2,p}(\Omega;\mathbb{R}^d)\,;\ y|_{\Gamma_{\rm D}}=\text{identity}\,\},\quad \det(\nabla y_0)\ge\hat\epsilon, \tag{2.30h}$$

$$\theta_\flat\in L^1(\Sigma),\quad \theta_\flat\ge0,\quad \theta_0\in L^1(\Omega),\quad \theta_0\ge0,\quad \psi(\nabla y_0,\theta_0)\in L^1(\Omega), \tag{2.30i}$$

where $\mathrm{GL}^+(d)$ denotes the set of matrices in $\mathbb{R}^{d\times d}$ with positive determinant. The last assumption in (2.30c) means that $c_{\rm v}$ together with $c_{\rm v}^{-1}$ are bounded, which is a major restriction. However, it allows for a rather simple estimation in Lemma 6.3; for alternative, more general situations dealing with increasing $c_{\rm v}(\cdot)$ we refer to [KrR19, Sec. 8.3].

The function $w=\mathfrak{w}(F,\theta)$ defined in (2.16) satisfies $\mathfrak{w}(F,0)=0$ by (2.15). Moreover, we have $\partial_\theta\mathfrak{w}(F,\theta)=-\theta\partial^2_\theta\phi(F,\theta)$. Hence assumption (2.30c) implies, for all $F\in\mathrm{GL}^+(\mathbb{R}^d)$, the two-sided estimates

$$\begin{aligned}&\hat\epsilon\theta\le\mathfrak{w}(F,\theta)\le K\theta\quad\text{for all }\theta\ge0.\\ &\hat\epsilon|\theta_1{-}\theta_2|\le|\mathfrak{w}(F,\theta_1){-}\mathfrak{w}(F,\theta_2)|\le K|\theta_1{-}\theta_2|\quad\text{for all }\theta_1,\theta_2\ge0.\end{aligned} \tag{2.31}$$

The assumptions (2.30b,c) make the thermomechanical coupling through ϕ rather weak in order to allow for a simple handling of the mechanical part independently of the temperature. These restrictive assumptions are needed for our specific and simple way of approximation method rather than with the problem itself. E.g. the assumption in (2.30b) is used to facilitate the estimate (4.12), which allows us to control the difference between $\int_\Omega(\nabla y^k,\theta)\,\mathrm{d}x$ and $\int_\Omega(\nabla y^{k-1},\theta)\,\mathrm{d}x$ in terms of $\mathcal{M}(y^k)$, $\mathcal{M}(y^{k-1})$, and $\|\nabla y^k{-}\nabla y^{k-1}\|^2_{L^2}$. Moreover, after having derived uniform bounds on $|\nabla y^k|$ it will be exploited to show that the thermo-coupling stress $\partial_F\phi$ is bounded. Finally, (2.30d,h) makes the stored energy finite at time $t=0$.

It will be important that $\partial_F\phi(F,\theta)$ vanishes for $\theta=0$ (which follows from (2.15)), so that temperature stays non-negative if $\theta_0\ge0$ and $\theta_\flat\ge0$, as assumed.

We now state our main existence results, which will be proved in the following Sections 4 to 6. The method will be constructive, avoiding non-constructive Schauder fixed-point arguments, however some non-constructive attributes such as selections of converging subsequences will remain. More specifically, the proof is obtained by first making the a priori estimate for time-discretized solutions in, see Proposition 4.2, and then deriving an existence result for time-continuous solutions of an ε-regularized problem, see Proposition 5.1. Finally, Proposition 6.4 provides convergence for $\varepsilon\to0$.

Theorem 2.2 (Existence of energy-conserving weak solutions)**.** *Assume that the conditions* (2.30) *hold. The original initial-boundary-value problem* (2.13)–(2.14)–(2.26) *with* $\mathcal{K}$ *from* (2.24) *possesses at least one weak solution* (y,θ) *in the sense of Definition 2.1. In addition, these solutions satisfy* $\nabla\theta\in L^r(Q;\mathbb{R}^d)$ *for all* $1\le r<(d{+}2)/(d{+}1)$*, the mechanical energy balance* (2.19)*, and the total energy balance* (2.21).

As mentioned in the introduction, a lot of publications are devoted to the simpler isothermal viscoelasticity at largestrain, yet, in the multi-dimensional case, they do not satisfy all the necessary physical requirements. It is therefore worthwhile to present a version of our existence result by restricting it to this simpler case, for which a lot of assumptions are irrelevant or simplify. In particular, (2.15) simplifies as $\psi(F,\theta)=\varphi(F)$. Of course, our theory only works because we are using a non-degenerate second-grade material, where $\mathcal{H}(y):=\int_\Omega\mathscr{H}(\nabla^2y)\,\mathrm{d}x$ generates enough regularity to handle the geometric and physical nonlinearities. To the best of the authors knowledge, even the following result for isothermal viscoelasticity is new.

A similar regularization approach to isothermal large-strain viscoelasticity was considered in [FrK18], where the $\mathcal{H}(y)$ is multiplied with a small parameter that vanishes slower than the loading. Hence, the authors are able to show that their solutions are sufficiently close to the identity which allows them to exploit a simpler Korn's inequality obtained by a perturbation argument. Hence, to the best of the author's knowledge the following result is the first that allows for truly largestrains.

Corollary 2.3 (Viscoelasticity at constant temperature)**.** *Let* φ *satisfy* (2.30a)*, and let (2.30d-e,g-h) be satisfied with* $\hat\zeta=\hat\zeta(C,\dot C)$ *and with* $\psi=\varphi$*. Then, the initial-boundary-value problem* (2.13a)–(2.14a)–(2.26) *(with* θ *ignored) possesses at least one weak solution* y *in the sense that the integral identity* (2.27a) *holds. In addition, the mechanical energy balance* (2.20) *holds with* $\xi=\xi(F,\dot F)$ *and without the last term involving* $\partial_F\phi$*.*

Before going into the proof of our main result, we show that our conditions are general enough for a series of nontrivial applications:

Example 2.4 (Classical thermomechanical coupling)**.** The classical example of a free energy in thermomechanical coupling is given in the form

$$\psi(F,\theta)=\varphi(F)-a(\theta)\,\varphi_1(F)+c\theta(1-\log\theta),\tag{2.32}$$

i.e. $\phi(F,\theta)$ involves a term in the product form $-a(\theta)\varphi_1(F)$. For the purely mechanical part we may take the polyconvex energy $\varphi(F)=c_1|F|^s+c_2/(\det F)^q$ for $\det F>0$ and ∞ otherwise. For the thermomechanical coupling we obtain $c_{\mathrm v}(F,\theta)=-\theta\partial^2_{\theta\theta}\psi(F,\theta)=c+a''(\theta)\varphi_1(F)$, thus to have positivity of the heat capacity $c_{\mathrm v}$, we assume $a''(\theta)\ge0$ and $\varphi_1(F)\ge0$. Moreover, we have

$$w=\mathfrak{w}(F,\theta)=c\,\theta+\big(\theta a'(\theta){-}a(\theta)\big)\varphi_1(F)\quad\text{and}\quad\partial_F\phi(F,\theta)=a(\theta)\varphi_1'(F)..$$

Thus, we see that all assumptions in (2.30) can easily be satisfied, e.g. by choosing $a(\theta)=(1{+}\theta)^{-\alpha}$ with $\alpha>0$, which is smooth bounded and convex, and taking any $\phi_1\in C^2_{\mathrm c}(\mathbb{R}^{d\times d})$.

Example 2.5 (*Phase transformation in shape-memory alloys)***.** An interesting example of a free energy ψ occurs in modeling of austenite-martensite transformation in so-called shape-memory alloys:

$$\psi(F,\theta)=(1{-}a(\theta))\varphi_{\mathrm A}(F)+a(\theta)\varphi_{\mathrm M}(F)+\psi_0(\theta).$$

cf. e.g. [Rou04] and references therein. Here a denotes the volume fraction of the austenite versus martensite which is supposed to depend only on temperature. Of course, this is only a rather simplified

model. For, $\psi_0(\theta) = c\theta(1-\log\theta)$ it complies with the ansatz (2.32) with $\varphi(F) = \varphi_{\rm A}(F)$ and $\varphi_1(F) = \varphi_{\rm M}(F)-\varphi_{\rm A}(F)$. The heat capacity then reads as

$$c_{\rm v}(F,\theta) = \theta a''(\theta)[\varphi_{\rm A}-\varphi_{\rm M}](F) - \theta\psi_0''(\theta).$$

To ensure its positivity, ψ_0 is to be strictly concave in such a way that $\psi_0''(\theta) \le K/\theta$ and then $\inf_{(F,\theta)} \theta a''(\theta)[\varphi_{\rm A}-\varphi_{\rm M}](F) + K > 0$ is to (and can) be ensured by suitable modeling assumptions.

Example 2.6 (*Thermal expansion*)**.** Multiplicative decomposition $F = F_{\rm el}F_{\rm th}$ with the "thermal strain" $F_{\rm th} = \mathbb{I}/\mu(\theta)$ and the elastic strain $F_{\rm el}$ which enters the elastic part of the stored energy φ. This leads to

$$\psi(F,\theta) = \beta(\theta)\varphi(F_{\rm el}) + \phi(\theta) = \beta(\theta)\,\varphi\big(\mu(\theta)F\big) - \phi(\theta). \tag{2.33}$$

Unfortunately, (2.33) is inconsistent with the ansatz (2.15) because the contribution φ which has been important for our analysis due to uniform coercivity, cannot be identified in (2.33).

3 A few auxiliary results

In this subsection we provide a series of auxiliary results that are crucial to tackle the difficulties arising from large-strain theory. First we show how the theory developed by Healey and Krömer [HeK09] which allows us to show that a bound for the elastic energy $\mathcal{M}(y,\theta)$ provides lower bounds on the $\det\nabla y$. This can then be used to establish the validity of the Euler-Lagrange equations and useful λ-convexity result, which is needed for obtaining optimal energy estimates. Second we provide a version of Korn's inequality from Pompe [Pom03] that allows us to obtain dissipation estimates via $\mathcal{D}(y,\dot y,\theta) \ge c_0\|\dot y\|^2_{H^1(\Omega)}$. Finally, in Section 3.3 we provide abstract chain rules as derived in [MRS13, Sec. 2.2] that allows us to derive energy balances like (2.20) from the corresponding weak equations.

3.1 Local invertibility and Euler-Lagrange equations

A crucial point in large-strain theory is the blow-up of the energy density $\psi(F,\theta)$ for $\det F \searrow 0$. Thus, it is desirable to find a suitable positive lower bound for $\det\nabla y(t,x)$. The following theorem is an adaptation of the result in [HeK09, Thm. 3.1].

Theorem 3.1 (Positivity of determinant)**.** *Assume that the functional $\mathcal{M} : W^{2,p}(\Omega;\mathbb{R}^d) \to \mathbb{R}_\infty$ satisfies the assumption* (2.30a) *and* (2.30d)*. Then, for each $C_M > 0$ there exists a $C_{\rm HK} > 0$ such that all $y \in \mathcal{Y}_{\rm id}$ with $\mathcal{M}(y) \le C_M$ satisfy*

$$\|y\|_{W^{2,p}} \le C_{\rm HK},\ \ \|y\|_{C^{1,1-d/p}} \le C_{\rm HK},\ \ \det\nabla y(x) \ge \frac{1}{C_{\rm HK}},\ \ \|(\nabla y)^{-1}\|_{C^{1-d/p}} \le C_{\rm HK}. \tag{3.1}$$

Proof. We give the full proof, since our mixed boundary conditions are not covered in [HeK09]. From $\mathcal{M}(y) \le C_M$ and the coercivities of φ and $\mathscr{H}$ we obtain $\det\nabla y \ge 0$ a.e. in Ω and the a priori bounds

$$\|\nabla y\|_{L^s} + \|\big(\det(\nabla y)\big)^{-1}\|_{L^q} + \|\nabla^2 y\|_{L^p} \le C_M^{(1)}.$$

Together with the Dirichlet boundary conditions in $\mathcal{Y}_{\rm id}$ we obtain an a priori bound for y in $W^{2,p}(\Omega;\mathbb{R}^d)$ and hence also in $C^{1,\lambda}(\Omega;\mathbb{R}^d)$, where $\lambda = 1 - d/p > 0$. This proves the first two assertions.

In particular, the function $\delta : x \mapsto \det(\nabla y(x))$ is Hölder continuous as well with $\|\delta\|_{C^\alpha} \leq C_M^{(2)}$. Since Ω is a bounded Lipschitz domain, there exist a radius $r_* > 0$ and a constant $\alpha_* > 0$ such that for all $x \in \overline{\Omega}$ the sets $B_{r_*}(x) \cap \overline{\Omega}$ contains an interior cone $C_x = \big\{x+z \big| 0 < |z| < r_*,\ \frac{1}{|z|} z \in A(x)\big\}$ where the set $A(x) \subset \mathbb{S}^{d-1}$ of cone directions has a surface measure $\int_{A(x)} 1 \,\mathrm{d}S \geq \alpha_*$. Thus, using the Hölder continuity

$$\delta(y) \leq \delta(x) + C_M^{(2)} |x-y|^\lambda \quad \text{for all } x, y \in \overline{\Omega},$$

we can estimate as follows:

$$\begin{aligned}
\big(C_M^{(1)}\big)^q &\geq \int_\Omega \frac{1}{\delta(y)^q} \,\mathrm{d}y \geq \int_{\Omega \cap B_{r_*}(x)} \frac{1}{\big(\delta(x) + C_M^{(2)} |x-y|^\alpha\big)^q} \,\mathrm{d}y \\
&\geq \int_{\omega \in A(x)} \int_{r=0}^{r_*} \frac{r^{d-1} \,\mathrm{d}r}{\big(\delta(x) + C_M^{(2)} r^\alpha\big)^q} \,\mathrm{d}\omega \ \geq\ \frac{\alpha_*}{2^q} \int_{r=0}^{r_*} \frac{r^{d-1} \,\mathrm{d}r}{\max\{\delta(x)^q, (C_M^{(2)} r^\alpha)^q\}} \\
&\geq c_M^{(3)} \min\{\delta(x)^{-q}, \delta(x)^{-(q-d/\lambda)}\} \ =\ \frac{c_M^{(3)}}{\max\{\delta(x)^q, \delta(x)^{q-d/\lambda}\}},
\end{aligned}$$

where in the last estimate we crucially used the assumption $q > pd/(p-d)$ which implies $\lambda q > d$. Since in the last expression both exponents of $\delta(x)$ are positive, we obtain the explicit lower bound

$$\det \nabla y(x) = \delta(x) \geq \min\Big\{ \big(c_M^{(3)}\big)^{1/q} / C_M^{(1)}, \big(c_M^{(3)} / (C_M^{(1)})^q\big)^{\lambda/(\lambda q - d)} \Big\},$$

which gives the third assertion in (3.1).

The last assertion follows via the implicit function theorem. □

The most important part of the above result is that the determinant of ∇y is bounded away from 0. Hence, the function $f \mapsto \varphi(F)$, which is blows up for $\det F \searrow 0$, is evaluated only in a compact subset of $\mathrm{GL}^+(d) \subset \mathbb{R}^{d\times d}$ such that $\partial_F \phi$ and $\partial^2 \varphi$ exist. Again following [HeK09, Cor. 3.3] we obtain the Gâteaux differentiability of $\mathcal{M}$ and as well as a useful Λ-semiconvexity result.

Proposition 3.2 (Gâteaux derivative and Λ-semiconvexity)**.** *Assume that $\mathcal{M}$ satisfies* (2.30a) *and* (2.30d)*. Then, in each point $y \in \mathcal{Y}_{\mathrm{id}}$ with $\mathcal{M}(y) < \infty$ the Gâteaux derivative in all directions $h \in \mathcal{Y}_0 := \big\{ v \in W^{2,p}(\Omega) \,;\, v|_{\Gamma_{\mathrm{D}}} \big\}$ exists and has the form*

$$\mathrm{D}\mathcal{M}(y)[h] = \int_\Omega \Big(\mathrm{D}\mathscr{H}(\nabla^2 y) \vdots \nabla^2 h + \partial_F \varphi(\nabla y) : \nabla h \Big) \,\mathrm{d}x \tag{3.2}$$

Moreover, for each $C_M > 0$ there exists $\Lambda(C_M) > 0$ such that for all $y^{(1)}, y^{(2)} \in \mathcal{Y}_{\mathrm{id}}$ with $\mathcal{M}(y^{(j)}) \leq C_M$ and $\|\nabla y^{(1)} - \nabla y^{(2)}\|_{L^\infty} \leq 1/C_G$ we have C_G convexity

$$\mathcal{M}(y^{(2)}) \geq \mathcal{M}(y^{(1)}) + \mathrm{D}\mathcal{M}(y^{(1)})[y^{(2)} - y^{(1)}] - \Lambda(C_M) \|\nabla y^{(2)} - \nabla y^{(1)}\|_{L^2}^2. \tag{3.3}$$

Proof. We decompose $\mathcal{M} = \mathcal{H} + \Phi_{\mathrm{el}}$, see (2.18b). The differentiability of the convex functional $y \mapsto \mathcal{H}(y)$ on $W^{2,p}(\Omega;\mathbb{R}^d)$ is standard and follows from (2.30d). For treating Φ_{el} we use the embedding $W^{2,p}(\Omega) \subset C^{1,\lambda}(\Omega)$ and exploit the result $\det \nabla y(x) \geq 1/C_{\mathrm{HK}}$ from Theorem 3.1. For all $h \in W^{2,p}_{\Gamma_{\mathrm{D}}}(\Omega;\mathbb{R}^d)$ we find a $t_* > 0$ such that $\det\big(\nabla(y+th)(x)\big) > 1/(2C_{\mathrm{HK}})$ for all $t \in [-t_*, t_*]$ and all $x \in \Omega$. Hence,

$$\mathrm{D}\Phi_{\mathrm{el}}(y)[h] = \lim_{t\to 0} \frac{1}{t} \big(\Phi_{\mathrm{el}}(y+th) - \Phi_{\mathrm{el}}(y)\big) = \lim_{t\to 0} \int_\Omega \frac{1}{t} \big(\varphi(\nabla y + t\nabla h) - \varphi(\nabla y)\big) \,\mathrm{d}x,$$

and the limit passage is trivial as the convergence in the integrand is uniform.

To derive (3.3) that the convexity of $\mathscr{H}$ implies

$$\mathcal{H}(y^{(2)}) \geq \mathcal{H}(y^{(1)}) + \int_\Omega \mathrm{D}\mathscr{H}(\nabla^2 y^{(1)}) \vdots \big(\nabla^2 y^{(2)} - \nabla^2 y^{(1)}\big)\,\mathrm{d}x.$$

To treat the functional Φ_{el} we apply Theorem 3.1 to $y^{(1)}$ and $y^{(2)}$, which implies the pointwise bounds

$$|\nabla y^{(j)}(x)| \leq C_{\mathrm{HK}} \quad \text{and} \quad \det \nabla y^{(j)}(x) \geq 1/C_{\mathrm{HK}}.$$

Clearly there is a $\delta > 0$ such that all

$$\forall\, F_1, F_2 \in \mathbb{R}^{d\times d}\, \forall\, s \in [0,1]: \quad \left.\begin{array}{c} |F_1|, |F_2| \leq C_{\mathrm{HK}}, \\ \det F_1, \det F_2 \geq 1/C_{\mathrm{HK}} \end{array}\right\} \implies \det\big((1-s)F_1 + sF_2\big) \geq 1/(2C_{\mathrm{HK}}).$$

This we denote by $-\Lambda_*$ the minimum of smallest eigenvalue of of the matrices $\partial_F^2\varphi(F)$ where $F \in \mathbb{R}^{d\times d}$ runs through the compact set given by $|F| \leq C_{\mathrm{HK}}$ and $\det F \geq 1/(2C_{\mathrm{HK}})$. Hence, assuming $\|\nabla y^{(2)} - \nabla y^{(2)}\|_{L^\infty} \leq \delta$ we find

$$\begin{aligned}
&\Phi_{\mathrm{el}}(y^{(2)}) - \Phi_{\mathrm{el}}(y^{(1)}) - \mathrm{D}\Phi_{\mathrm{el}}(y^{(1)})[y^{(2)} - y^{(1)}] \\
&= \int_\Omega \Big(\varphi(\nabla y^{(2)}) - \varphi(\nabla y^{(1)}) - \partial_\varphi(y^{(1)}) : (\nabla y^{(2)} - \nabla y^{(1)})\Big)\,\mathrm{d}x \\
&= \int_\Omega \frac{1}{2} \int_{s=0}^1 \partial_F^2\varphi\big((1-s)\nabla y^{(1)} + s\nabla y^{(2)}\big)\big[\nabla y^{(2)} - \nabla y^{(1)}, \nabla y^{(2)} - \nabla y^{(1)}\big]\,\mathrm{d}s\,\mathrm{d}x \\
&\geq -\frac{\Lambda_*}{2} \int_\Omega |\nabla y^{(2)} - \nabla y^{(1)}|^2\,\mathrm{d}x.
\end{aligned}$$

This establishes the result with $\Lambda(C_M) := \max\{C_{\mathrm{CK}}, 1/\delta, \Lambda_*/2\}$. □

3.2 A generalized Korn's inequality

The following result will be crucial to show that the nonlinear viscosity depending on $F = \nabla y$ really controls the H^1 norm of of the rate $\dot{y}$. It relies on Neff's generalization [Nef02] of the Korn inequality, in the essential improvement obtained by Pompe [Pom03].

Theorem 3.3 (Generalized Korn's inequality)**.** *For a fixed $\lambda \in]0,1[$ and positive constants $K > 1$ define the set*

$$\mathsf{F}_K := \big\{\, F \in C^\lambda(\Omega;\mathbb{R}^{d\times d})\,;\, \|F\|_{C^\lambda} \leq K,\, \min_{x\in\Omega} \det F(x) \geq 1/K \,\big\}.$$

Then, for all $K > 1$ there exists a constant $c_K > 0$ such that for all $F \in \mathsf{F}_K$ we have

$$\forall\, v \in H^1(\Omega;\mathbb{R}^d) \text{ with } v|_{\Gamma_{\mathrm{D}}} = 0: \quad \int_\Omega \big|F^\top \nabla v + (\nabla v)^\top F\big|^2\,\mathrm{d}x \geq c_K \|v\|_{H^1}^2. \tag{3.4}$$

Proof. In [Pom03, Thm. 2.3] it is shown that (3.4) holds for any given $F \in \mathsf{F}_K$. Let us denote by $c(F) > 0$ the supremum of all possible such constants for the given F. By a perturbation argument it is easy to see that the mapping $F \mapsto c(F)$ is continuous with respect to the L^∞ norm in $C^0(\overline{\Omega};\mathbb{R}^{d\times d})$. Since F_K is a compact subset of $C^0(\overline{\Omega};\mathbb{R}^{d\times d})$ the infimum of c on F_K is attained at some $F_* \in \mathsf{F}_K$ by Weierstraß' extremum principle. Because of $c(F) \geq c(F_*)$, we conclude that (3.4) holds with $c_K = c(F_*)$. □

We emphasize that estimate (3.4) is not valid if F is not continuous, see [Pom03, Thm. 4.2]. This shows that without the in $W^{2,p}$ is crucial to control the rate of the strain $\nabla\dot{y}$, which is necessary to handle the thermomechanical coupling. The following corollary combines Theorems 3.1 and 3.3, by using the compact embedding $W^{2,p}(\Omega;\mathbb{R}^d) \subset C^{1,\lambda}(\Omega;\mathbb{R}^d)$.

Corollary 3.4 (Uniform generalized Korn's inequality on sublevels)**.** *Given any $C_M > 0$ there exists a $\overline{c}_K > 0$ such that for all $y \in \mathcal{Y}_{\rm id}$ with $\mathcal{M}(y) \le C_M$ we have the generalized Korn inequality*

$$\forall\, v \in H^1(\Omega;\mathbb{R}^d) \text{ with } v|_{\Gamma_{\rm D}} = 0 : \int_\Omega \left|(\nabla y)^\top \nabla v + (\nabla v)^\top \nabla y\right|^2 \mathrm{d}x \ge \overline{c}_K \|v\|_{H^1}^2. \tag{3.5}$$

3.3 Chain rules for energy functionals

Abstract chain rules for energy functionals $\mathcal{J} : X \to \mathbb{R}_\infty := \mathbb{R}\cup\{\infty\}$ on a Banach space concern the question under what conditions for an absolutely continuous curve $z : [0,T] \to X$ the composition $t \mapsto \mathcal{J}(z(t))$ is absolutely continuous and satisfies $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{J}(z(t)) = \langle \Xi(t), \dot{z}(t)\rangle$ for $\Xi \in \overline{\partial}\mathcal{J}(z(t))$, where $\overline{\partial}$ denotes a suitable subdifferential. In particular, this implies

$$\mathcal{J}(z(t_1)) = \mathcal{J}(z(t_0)) + \int_{t_0}^{t_1} \langle \Xi(t), \dot{z}(t)\rangle \,\mathrm{d}t \quad \text{for } 0 \le t_0 < t_1 \le T.$$

The case that X is a Hilbert space and $\mathcal{J}$ is convex and lower semicontinuous goes back to [Bré73, Lem. 3.3], see also [Bar10, Lemma 4.4]:

Proposition 3.5 (Chain rule for convex functionals in a Hilbert space)**.** *Let X be a Hilbert space and $\mathcal{J} : X \to \mathbb{R}_\infty := \mathbb{R}\cup\{\infty\}$ a lower semicontinuous and convex functionals. If the functions $z : [0,T] \to X$ and $\Xi : [0,T] \to X^*$ satisfy*

$$z \in H^1([0,T];X), \quad \Xi \in L^2([0,T];X^*), \text{ and } \quad \Xi(t) \in \partial\mathcal{J}(z(t)) \text{ a.e. in } [0,T],$$

where $\partial\mathcal{J}$ denotes the convex subdifferential, then

$$t \mapsto \mathcal{J}(z(t)) \text{ lies in } W^{1,1}(0,T) \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{J}(z(t)) = \langle \Xi(t), \dot{z}(t)\rangle \text{ a.e. in } [0,T].$$

A first generalization to Banach spaces X with separable dual X^* is given in [Vis96, Prop.XI.4.11]. We provide a slight generalization of the results in [MRS13, Sec. 2.2] that work for arbitrary reflexive Banach spaces and include also certain nonconvex functionals. The functional $\mathcal{J}$ is called *locally semiconvex*, if for all z with $\mathcal{J}(z) < \infty$ there exists a $\Lambda = \hat{\Lambda}(z) \ge 0$ and a balls $B_r(z) = \{\, \hat{z} \in X \,;\, \|\hat{z}-z\|_X \le r \,\}$ with $r = \hat{r}(z)$ the restriction $\mathcal{J}|_{B_r(z)}$ is Λ-semiconvex, viz.

$$\forall\, z_0, z_1 \in B_r(z)\, \forall\, s \in [0,1] : \ \mathcal{J}\big((1-s)z_0 + s z_1\big) \le (1-s)\mathcal{J}(z_0) + s\mathcal{J}(z_1) + \frac{\Lambda}{2}(s-s^2)\|z_1-z_0\|_X^2.$$

By $\overline{\partial}\mathcal{J}$ we denote the Fréchet subdifferential which is defined by

$$\overline{\partial}\mathcal{J}(z) = \big\{\, \Xi \in X^* \,;\, \mathcal{J}(\hat{z}) \ge \mathcal{J}(z) + \langle \Xi, \hat{z}-z\rangle - 2\hat{\Lambda}(z)\|\hat{z}-z\|_X^2 \text{ for } \hat{z} \in B_{\hat{r}(z)}(z) \,\big\}.$$

The next results follows by a simple adaptation of the proof of [MRS13, Prop. 2,4].

Proposition 3.6 (Chain rule for locally semiconvex functionals)**.** *Consider a separable reflexive Banach space, a $q \in]1,\infty[$ with $q' = q/(q-1)$, and $\mathcal{J} : X \to \mathbb{R}_\infty$ a lower semicontinuous and locally semiconvex functional. If the functions $z \in W^{1,q}([0,T];X)$ and $\Xi \in L^{q'}([0,T];X^*)$ satisfy*

$$\sup\big\{\,\mathcal{J}(z(t))\,;\,t\in[0,T]\,\big\}<\infty \quad \textit{and} \quad \Xi(t)\in\overline{\partial}\mathcal{J}(z(t)) \ \textit{a.e. in}\ [0,T],$$

then

$$t\mapsto\mathcal{J}(z(t)) \ \textit{lies in}\ W^{1,1}(0,T) \quad \textit{and} \quad \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{J}(z(t))=\langle\Xi(t),\dot z(t)\rangle \ \textit{a.e. in}\ [0,T].$$

Proof. The result follows by the fact that the image of z lies in $\mathrm{dom}\mathcal{J}=\{\,z\in X\,;\,\mathcal{J}(z)<\infty\,\}$ and is compact in Z. Hence there is one $\Lambda_*<\infty$ and one $r_*>0$ such that provides Λ_* semiconvexity on $B_{r_*}(z(t))$ for all $t\in[0,T]$. Hence, the results in the proof of [MRS13, Prop. 2,4] can be applied when choosing $\omega^R(\hat z,z)=\Lambda_*\|\hat z-z\|_X$ and using that fact that all needed arguments are local and use only information of $\mathcal{J}$ in a neighborhood of the image of z. □

4 Time discretization of a regularized problem

Before we construct solution by a suitable time-discretization, we introduce regularizations in two points. Firstly, we add a linear viscous damping which allows us to obtain simple a priori bounds for the strain rate $\nabla\dot y$, because in the first steps of the construction we are not yet in the position to exploiting the generalized Korn inequality of Theorem 3.3. Secondly, we modify the creation of heat through the viscous damping, which in the physically correct form leads to an L^1 source term which can only be handled in the first steps of the construction either.

Hence, introducing the regularization parameter $\varepsilon>0$ we consider the coupled system

$$\mathrm{div}\,\big(\sigma_{\rm vi}(\nabla y,\nabla\dot y,\theta)+\varepsilon\nabla\dot y+\sigma_{\rm el}(\nabla y,\theta)-\mathrm{div}\,\mathfrak{h}_{\rm el}(\nabla^2 y)\big)+g=0, \tag{4.1a}$$

$$\dot w-\mathrm{div}(\mathcal{K}(\nabla y,\theta)\nabla\theta)=\xi_\varepsilon^{\rm reg}(\nabla y,\nabla\dot y,\theta)+\partial_F\phi(\nabla y,\theta){:}\nabla\dot y \tag{4.1b}$$

$$w=\mathfrak{w}(\nabla y,\theta), \tag{4.1c}$$

$$\text{with } \xi_\varepsilon^{\rm reg}(F,\dot F,\theta):=\frac{\xi(F,\dot F,\theta)}{1+\varepsilon\,\xi(F,\dot F,\theta)},$$

where $\mathfrak{w}$ is from (2.16) and $\mathcal{K}$ from (2.24). This system is defined on Q and is complemented with regularized boundary and initial conditions

$$\big(\sigma_{\rm vi}(\nabla y,\nabla\dot y,\theta)+\varepsilon\nabla\dot y+\sigma_{\rm el}(\nabla y,\theta)\big)\vec n-\mathrm{div}_{\rm s}\big(\mathfrak{h}_{\rm el}(\nabla^2 y)\vec n\big)=f \qquad \text{on } \Sigma_{\rm N} \tag{4.2a}$$

$$y=\text{identity} \quad \text{on } \Sigma_{\rm D}, \qquad \mathfrak{h}_{\rm el}(\nabla^2 y):(\vec n\otimes\vec n)=0 \qquad \text{on } \Sigma, \tag{4.2b}$$

$$\mathcal{K}(\nabla y,\theta)\nabla\theta\cdot\vec n+\kappa\theta=\kappa\theta_{\flat,\varepsilon} \ \text{ with } \ \theta_{\flat,\varepsilon}:=\frac{\theta_\flat}{1+\varepsilon\theta_\flat}, \qquad \text{on } \Sigma, \tag{4.2c}$$

$$y(0,\cdot)=y_0 \quad \text{and} \quad \theta(0,\cdot)=\theta_{0,\varepsilon}:=\frac{\theta_0}{1+\varepsilon\theta_0} \qquad \text{on } \Omega. \tag{4.2d}$$

This system is solved by time discretization. For this we consider a constant time step $\tau>0$ such that T/τ is an integer, leading to an equidistant partition of the considered time interval $[0,T]$. (Let us emphasize, however, that a varying time-step and non-equidistant partitions can be easily implemented because we will always consider only first-order time differences and one-step formulas.)

For time discretization of the regularized system (4.1)–(4.2) we use the difference notation

$$\boldsymbol{\delta}_\tau f^k = \frac{1}{\tau}\big(f^k - f^{k-1}\big)$$

and define a staggered scheme, where first $y^{k-1}_{\varepsilon\tau}$ is updated to $y^k_{\varepsilon\tau}$ while keeping $\theta^{k-1}_{\varepsilon\tau}$ fixed, and then θ is updated implicitly by updating $w^{k-1}_{\varepsilon\tau}$ to $w^k_{\varepsilon\tau} = \mathfrak{w}(\nabla y^k_{\varepsilon\tau}, \theta^k_{\varepsilon\tau})$. More precisely, in the domain Ω we ask for

$$\begin{aligned}
&-\operatorname{div}\Big(\sigma_{\rm vi}\Big(\nabla y^{k-1}_{\varepsilon\tau}, \boldsymbol{\delta}_\tau \nabla y^k_{\varepsilon\tau}, \theta^{k-1}_{\varepsilon\tau}\Big) + \varepsilon\boldsymbol{\delta}_\tau\nabla y^k \\
&\qquad + \sigma_{\rm el}(\nabla y^k_{\varepsilon\tau}, \theta^{k-1}_{\varepsilon\tau}) - \operatorname{div}\mathfrak{h}_{\rm el}(\nabla^2 y^k_{\varepsilon\tau})\Big) = g^k_\tau := \frac{1}{\tau}\int_{(k-1)\tau}^{k\tau} g(t)\,\mathrm{d}t, \qquad (4.3a)\\
&\boldsymbol{\delta}_\tau w^k_{\varepsilon\tau} - \operatorname{div}(\mathcal{K}(\nabla y^{k-1}_{\varepsilon\tau}, \theta^{k-1}_{\varepsilon\tau})\nabla\theta^k_{\varepsilon\tau}) = \xi^{\rm reg}_\varepsilon(\nabla y^{k-1}_{\varepsilon\tau}, \nabla\delta_\tau y^k_{\varepsilon\tau}, \theta^{k-1}_{\varepsilon\tau}) \\
&\qquad\qquad + \partial_F\phi(\nabla y^k_{\varepsilon\tau}, \theta^k_{\varepsilon\tau}){:}\boldsymbol{\delta}_\tau\nabla y^k_{\varepsilon\tau} \qquad (4.3b)
\end{aligned}$$

together with the discrete variant of the boundary conditions (4.2) as

$$\begin{aligned}
&\Big(\sigma_{\rm vi}\Big(\nabla y^{k-1}_{\varepsilon\tau}, \boldsymbol{\delta}_\tau\nabla y^k_{\varepsilon\tau}, \theta^{k-1}_{\varepsilon\tau}\Big) + \varepsilon\boldsymbol{\delta}_\tau\nabla y^k_{\varepsilon\tau} + \sigma_{\rm el}(\nabla y^k_{\varepsilon\tau}, \theta^{k-1}_{\varepsilon\tau})\Big)\vec{n} \\
&\qquad - \operatorname{div}_{\rm s}\big(\mathfrak{h}_{\rm el}(\nabla^2 y^k_{\varepsilon\tau})\vec{n}\big) = f^k_\tau := \frac{1}{\tau}\int_{(k-1)\tau}^{k\tau} f(t)\,\mathrm{d}t \qquad \text{on } \Gamma_{\rm N}, \qquad (4.4a)\\
&y^k_{\varepsilon\tau} = \text{identity} \qquad \text{on } \Gamma_{\rm D}, \qquad\qquad \mathfrak{h}_{\rm el}(\nabla^2 y^k_{\varepsilon\tau}) : (\vec{n}\otimes\vec{n}) = 0 \qquad \text{on } \Gamma, \qquad (4.4b)\\
&\mathcal{K}(\nabla y^{k-1}_{\varepsilon\tau}, \theta^{k-1}_{\varepsilon\tau})\nabla\theta^k_{\varepsilon\tau}\cdot\vec{n} + \kappa\theta^k_{\varepsilon\tau} = \kappa\theta^k_{\flat,\varepsilon,\tau} := \frac{\kappa}{\tau}\int_{(k-1)\tau}^{k\tau}\theta_{\flat,\varepsilon}(t)\,\mathrm{d}t \qquad \text{on } \Gamma. \qquad (4.4c)
\end{aligned}$$

The main advantage is that the boundary-value problem (4.3a), (4.4a), and (4.4b) for $y^k_{\varepsilon\tau}$ are the Euler-Lagrange equation of a functional, so that solutions can be obtained by solving the global minimization problem

$$\begin{aligned}
y^k_{\varepsilon\tau} \in \mathsf{Arg\,Min}\Big\{ &\frac{1}{\tau}\mathcal{R}(y^{k-1}_{\varepsilon\tau}, y - y^{k-1}_{\varepsilon\tau}, \theta^{k-1}_{\varepsilon\tau}) + \frac{\varepsilon}{2\tau}\|\nabla y - \nabla y^{k-1}_{\varepsilon\tau}\|^2_{L^2} \\
&+ \Psi(y, \theta^{k-1}_{\varepsilon\tau}) - \langle \ell^k_\tau, y\rangle \;\Big|\; y\in\mathcal{Y}_{\rm id}\Big\}, \qquad (4.5)
\end{aligned}$$

where $\langle \ell^k_\tau, y\rangle = \int_\Omega g^k_\tau\cdot y\,\mathrm{d}x + \int_{\Gamma_{\rm N}} f^k_\tau\cdot y\,\mathrm{d}S$. Clearly, the Euler-Lagrange equation may have more solutions, however for deriving suitable a priori bounds, we will exploit the minimizing properties.

Similarly, the boundary value problem (4.3b) and (4.4c) for $\theta^k_{\varepsilon\tau}$, where $y^{k-1}_{\varepsilon\tau}$ and $y^k_{\varepsilon\tau}$ are given, has a variational structure. For this we define the functions $\phi_{\rm C}(F,\theta) := \int_0^\theta \phi(F,\hat\theta)\,\mathrm{d}\hat\theta$ and $W(F,\theta) = 2\phi_{\rm C}(F,\theta) - \theta\phi(F,\theta)$ to obtain the relation

$$\partial_\theta W(F,\theta) = \mathfrak{w}(F,\theta) = \phi(F,\theta) - \theta\partial_\theta\phi(F,\theta) \text{ and } \partial_\theta\partial_F\phi_{\rm C}(F,\theta) = \partial_F\phi(F,\theta). \qquad (4.6)$$

With $\partial^2_\theta W(F,\theta) = \partial_\theta\mathfrak{w}(F,\theta) = -\theta\partial^2_\theta\phi(F,\theta) \geq \hat\epsilon$ we see that $W(F,\cdot)$ is uniformly convex by assumption (2.30c). Thus, we can obtain solutions $\theta^k_{\varepsilon\tau}$ of (4.3b) and (4.4c) via the minimization problem

$$\begin{aligned}
\theta^k_{\varepsilon\tau} \in \mathsf{Arg\,Min}\Big\{ &\int_\Omega\Big(\frac{1}{\tau}\big(W(\nabla y^k_{\varepsilon\tau},\theta) - w^{k-1}_{\varepsilon\tau}\theta\big) + \frac{1}{2}\nabla\theta\cdot\mathcal{K}(\nabla y^{k-1}_{\varepsilon\tau}, \theta^{k-1}_{\varepsilon\tau})\nabla\theta\Big)\,\mathrm{d}x \\
&+ \int_\Omega\Big(-\xi^{\rm reg}_\varepsilon(\nabla y^{k-1}_{\varepsilon\tau}, \boldsymbol{\delta}_\tau y^k_\varepsilon, \theta^{k-1}_{\varepsilon\tau})\theta - \partial_F\phi_{\rm C}(\nabla y^k_{\varepsilon\tau},\theta) : \boldsymbol{\delta}_\tau\nabla y^k_{\varepsilon\tau}\Big)\,\mathrm{d}x \\
&+ \int_\Gamma \frac{\kappa}{2}\big(\theta - \theta^k_{\flat,\varepsilon,\tau}\big)^2\,\mathrm{d}S \;\Big|\; \theta\in H^1(\Omega),\ \theta\geq 0\Big\}. \qquad (4.7)
\end{aligned}$$

We emphasize that this staggered scheme is constructed in a very specific way by taking $\theta = \theta_{\varepsilon\tau}^{k-1}$ from the previous time step in the mechanics problem for $y_{\varepsilon\tau}^k$, see (4.5). For the construction of $\theta = \theta_{\varepsilon\tau}^k$ from the heat equation we have to use sometimes the explicit (backward) approximations $\theta_{\varepsilon\tau}^{k-1}$ and sometimes the implicit (forward) approximation $\theta_{\varepsilon\tau}^k$. Clearly, the former is simpler and it is used in the heat conduction tensor $\mathcal{K}(\nabla y_{\varepsilon\tau}^{k-1}, \theta_{\varepsilon\tau}^{k-1})$ and in the heat production $\xi_\varepsilon^{\mathsf{reg}}$. It is tempting to use the explicit choice $\theta_{\varepsilon\tau}^{k-1}$ also in the thermo-mechanical coupling term $\partial_F\phi(\nabla y_{\varepsilon\tau}^k, \theta)$:$\nabla\boldsymbol{\delta}_\tau y_\varepsilon^k$ (last term in (4.3b)) as it would simplify the energy balance, see Remark 6.1. However, as this term does not have a sign, we would not be able to guarantee positivity of $\theta_{\varepsilon\tau}^k$. Thus we are forced to use the more involved implicit term $\theta \mapsto \partial_F\phi_{\mathrm{C}}(\nabla y_\varepsilon^k, \theta)$:$\nabla\boldsymbol{\delta}_\tau y_\varepsilon^k$ in (4.7) instead of the simpler, linear choice $\theta \mapsto \theta\partial_F\phi(\nabla y_{\varepsilon\tau}^k, \theta_{\varepsilon\tau}^{k-1})$:$\nabla\boldsymbol{\delta}_\tau y_\varepsilon^k$. This choice may introduce a nonconvexity, so that $\theta_{\varepsilon\tau}^k$ may not be unique.

The following result states that we can obtain solutions $(y_{\varepsilon\tau}^k, \theta_{\varepsilon\tau}^k)$ of (4.3) and (4.4) by solving the minimization problems (4.5) and (4.7), alternatingly. For notational simplicity we have written the minimization problem (4.7) for θ with the constraint $\theta \geq 0$, however, for establishing the Euler-Lagrange (4.3b) and (4.4c) we need to show that non-negativity of θ comes even without imposing the constraint. This will be achieved by minimization over $\theta \in H^1(\Omega)$ after extending all functionals suitably for $\theta < 0$.

Proposition 4.1 (Time-discretized solutions via minimization)**.** *Let our assumptions* (2.30) *be satisfied. For $N \in \mathbb{N}$ set $\tau = T/N$ and $(y_{\varepsilon\tau}^0, \theta_{\varepsilon\tau}^0) = (y_0, \theta_{0,\varepsilon})$ as in* (4.2d)*. Then, for $k = 1, \dots, N$ we can iteratively find $(y_{\varepsilon\tau}^k, \theta_{\varepsilon\tau}^k) \in \mathcal{Y}_{\mathrm{id}} \times H^1_+(\Omega)$ by solving first the incremental global minimization problem* (4.5) *and then* (4.7)*. The global minimizers satisfy the time-discretized problem* (4.3)*.*

Proof. Mechanical step: We first show that the minimization problem in (4.5) has a solution for any $\theta_{\varepsilon\tau}^{k-1} \in H^1(\Omega)$ with $\theta_{\varepsilon\tau}^{k-1} \geq 0$. By assumption we have $\phi(F, \theta) \geq 0$ which implies $\Psi(y, \theta) \geq \mathcal{M}(y)$. Thus, the functional in the minimization problem is coercive on $\mathcal{Y}_{\mathrm{id}} \subset W^{2,p}(\Omega; \mathbb{R}^d)$. By lower semicontinuity in $W^{2,p}(\Omega; \mathbb{R}^d)$ we obtain the desired minimizer $y_{\varepsilon\tau}^k \in \mathcal{Y}_{\mathrm{id}}$ with $\mathcal{M}(y_{\varepsilon\tau}^k) < \infty$. Hence, Theorem 3.1 shows that the minimizer satisfies $\det \nabla y(x) \geq \delta > 0$. As in Proposition 3.2 we conclude that $y_{\varepsilon\tau}^k$ satisfies the Euler-Lagrange equation

$$\int_\Omega \Big(\partial_{\dot F}\zeta(\nabla y_{\varepsilon\tau}^{k-1}, \nabla\boldsymbol{\delta}_\tau y_\varepsilon^k, \theta_{\varepsilon\tau}^{k-1}) : \nabla z + \varepsilon\nabla\boldsymbol{\delta}_\tau y_\varepsilon^k : \nabla z + \partial_F\psi(\nabla y_\varepsilon^k, \theta_{\varepsilon\tau}^{k-1}) : \nabla z \Big)\,\mathrm{d}x \\ + \mathrm{D}\mathcal{H}(y_\varepsilon^k)[z] - \langle \ell_\tau^k, z\rangle \quad \text{for all } z \in \mathcal{Y}_0.$$

But this gives exactly (4.3a), (4.4a), and (4.4b).

Energy step: We now assume that $\theta_{\varepsilon\tau}^{k-1} \in H^1(\Omega)$ and $y_{\varepsilon\tau}^{k-1}, y_{\varepsilon\tau}^k \in \mathcal{Y}_{\mathrm{id}}$ are given with $\theta_{\varepsilon\tau}^{k-1} \geq 0$ and $\mathcal{M}(y_{\varepsilon\tau}^{k-1})$, $\mathcal{M}(y_{\varepsilon\tau}^k) < \infty$. With this, we show that a variant of the minimization problem (4.7) has a minimizer $\theta_{\varepsilon\tau}^k$. For this we extend the function ϕ, which satisfies $\phi(F, 0) = 0$ by assumption (2.15), continuously by $\phi(F, \theta) = 0$ whenever $\theta < 0$. As the functions $\mathfrak{w}$, ϕ_{C}, and W are defined through ϕ they all extend continuously differentiable for $\theta < 0$ to the constant value 0. Thus, the integrands in (4.7) are defined for all $\theta \in \mathbb{R}$ and we can minimize over $\theta \in H^1(\Omega)$, i.e. without the constraint $\theta \geq 0$.

Clearly, the extended functional is lower weakly semicontinuous on $H^1(\Omega)$ because of $\mathcal{K} \geq 0$. To show coercivity of the functional, we use that $\mathcal{M}(y_{\varepsilon\tau}^{k-1}) < \infty$ implies $\nabla y_{\varepsilon\tau}^{k-1} \in L^\infty$ and $\det \nabla y_{\varepsilon\tau}^{k-1}(x) \geq \delta > 0$. Hence, $\mathcal{K}$ given in (2.24) satisfies $\nabla\theta \cdot \mathcal{K}(\nabla y_{\varepsilon\tau}^{k-1}, \theta_{\varepsilon\tau}^{k-1})\nabla\theta \geq \alpha_*|\nabla\theta|^2$ for some $\alpha_* > 0$. Together with the boundary integral, where $\kappa > 0$ due to (2.30g), we have two terms that generate a lower bound $c_0\|\theta\|^2_{H^1(\Omega)} - C$.

For the remaining term we observe $W(F,\theta) \geq 0$ by construction, while $\frac{1}{\tau}w_{\varepsilon\tau}^{k-1}$ and $\xi_\varepsilon^{\mathrm{reg}}$ are given functions in $L^2(\Omega)$. Finally, the last bulk term involving $\partial_F\phi_{\mathrm{C}}$ we use (2.30b) giving $|\partial_F\phi(F,\theta)| \leq K(1+|F|^{s/2})$ and hence, because of $\nabla y_{\varepsilon\tau}^k \in L^\infty(\Omega;\mathbb{R}^{d\times d})$, we have

$$\big|\partial_F\phi_{\mathrm{C}}(\nabla y_{\varepsilon\tau}^k,\theta)\big| = \Big|\int_0^\theta \partial_F\phi(\nabla y_{\varepsilon\tau}^k,\hat\theta)\,\mathrm{d}\hat\theta\Big| \leq C_*|\theta|.$$

Together with $\boldsymbol{\delta}_\tau\nabla y_\varepsilon^k \in L^2(\Omega;\mathbb{R}^{d\times d})$ we have show that all remaining terms can be estimated from below by $-C\|\theta\|_{L^2(\Omega)}$.

In summary, we conclude that the extended functional in (4.7) is weakly lower semicontinuous and and coercive. Hence, a global minimizers θ_* exist and moreover these minimizers solve the associated Euler-Lagrange equation as $\partial_\theta W(F,\theta) = \mathfrak{w}(F,\theta)$ and $\partial_\theta\phi_{\mathrm{C}}(F,\theta) = \phi(F,\theta)$ depend continuously on θ.

To show that all global minimizers are non-negative we test the Euler-Lagrange equation by the negative part $\theta_*^- := \min\{\theta_*,0\}$ of θ_*, which is still an H^1 function:

$$\begin{aligned}
0 &= \int_\Omega \Big(\frac{1}{\tau}\mathfrak{w}(\nabla y_{\varepsilon\tau}^k,\theta_*)\theta_*^- - \frac{1}{\tau}w_{\varepsilon\tau}^{k-1}\theta_*^- + \nabla\theta_*\cdot\mathcal{K}(\nabla y_{\varepsilon\tau}^{k-1},\theta_{\varepsilon\tau}^{k-1})\nabla\theta_*^-\Big)\,\mathrm{d}x\\
&\quad + \int_\Omega \Big(-\xi_\varepsilon^{\mathrm{reg}}(\nabla y_{\varepsilon\tau}^{k-1},\boldsymbol{\delta}_\tau y_\varepsilon^k,\theta_{\varepsilon\tau}^{k-1})\theta_*^- - \theta_*^-\partial_F\phi(\nabla y_{\varepsilon\tau}^k,\theta_*) : \boldsymbol{\delta}_\tau\nabla y_{\varepsilon\tau}^k\Big)\,\mathrm{d}x\\
&\qquad\qquad + \int_\Gamma \big(\kappa\theta_*\theta_*^- - \theta_{\flat,\varepsilon,\tau}^k\theta_*^-\big)\,\mathrm{d}S\\
&\geq \int_\Omega \big(0 + p_2 + \alpha_*|\nabla\theta_*^-|^2 + p_4 + 0\big)\,\mathrm{d}x + \int_\Gamma \big(\kappa(\theta_*^-)^2 + p_7\big)\,\mathrm{d}S \geq\ c_0\|\theta_*^-\|_{H^1(\Omega)}^2.
\end{aligned}$$

In the first estimate we have used $w_{\varepsilon\tau}^{k-1} = \mathfrak{w}(\nabla y_{\varepsilon\tau}^{k-1},\theta_{\varepsilon\tau}^{k-1}) \geq \hat\epsilon\theta_{\varepsilon\tau}^{k-1} \geq 0$, $\xi_\varepsilon^{\mathrm{reg}} \geq 0$, and $\theta_{\flat,\varepsilon,\tau}^k \geq 0$ which gives the non-negativity of p_2, p_4, and p_7, while the first and fifth term vanish identically since for $\theta_* > 0$ we have $\theta_*^- = 0$ while for $\theta_* < 0$ we have $\mathfrak{w}(F,\theta_*) = 0$ and $\partial_F\phi(F,\theta_*) = 0$ (here we crucially use the implicit structure). Thus, we conclude $\theta_*^- = 0$ which is equivalent to $\theta_* \geq 0$.

Thus, choosing $\theta_{\varepsilon\tau}^k = \theta_*$ for any global minimizer of the extended functional we see that it is also a global minimizer of (4.7) and that the Euler-Lagrange equations hold. □

Considering discrete approximations $\big(y_{\varepsilon\tau}^k\big)_{k=0,\dots,T/\tau}$, we introduce a notation for the piecewise-constant and the piecewise affine interpolants defined respectively by

$$\left.\begin{aligned}
&\overline{y}_{\varepsilon\tau}(t) = y_{\varepsilon\tau}^k\,,\quad \underline{y}_{\varepsilon\tau}(t) = y_{\varepsilon\tau}^{k-1},\quad \text{and}\\
&y_{\varepsilon\tau}(t) = \frac{t-(k-1)\tau}{\tau}\,y_{\varepsilon\tau}^k + \frac{k\tau-t}{\tau}y_{\varepsilon\tau}^{k-1}
\end{aligned}\right\} \qquad \text{for } (k-1)\tau < t < k\tau,$$

$$\underline{y}_{\varepsilon\tau}(k\tau) = \overline{y}_{\varepsilon\tau}(k\tau) = y_{\varepsilon\tau}(k\tau) = y_{\varepsilon\tau}^k \qquad \text{for } k = 0,1,\dots,T/\tau. \tag{4.8}$$

The notations $\theta_{\varepsilon\tau}$, $\overline{\theta}_{\varepsilon\tau}$, and $\underline{\theta}_{\varepsilon\tau}$ or $w_{\varepsilon\tau}$ have analogous meanings. However, with $\overline{g}_\tau(t)$ we refer to the locally averaged loadings $\overline{g}_\tau(t) = g_\tau^k$ for $t \in\,]k\tau-\tau,k\tau]$ (cf. (4.3a)), and similarly for $\overline{f}_\tau$, $\overline{\ell}_\tau$ and $\overline{\theta}_{\flat,\varepsilon,\tau}$.

The following result provides the basic energy estimates where we will crucially use the carefully chosen semi-implicit scheme defined through the staggered minimization problems (4.5) and (4.7). Here also we will essentially rely regularizing viscous term $\varepsilon\Delta\dot y$, as $\mathcal{R}$ cannot be used because of the missing a priori bound for $y_{\varepsilon\tau}^k$ in $W^{2,p}(\Omega;\mathbb{R}^d)$. Moreover, we will exploit the fact that we have global minimizers in (4.5) rather than arbitrary solutions of the Euler-Lagrange equations (4.3a). This latter

argument works because we have neglected inertial terms in the momentum balance (2.27a) and hence in (4.3a). We refer to [KrR19] to cases where inertial effects are treated but in the isothermal case.

Proposition 4.2 (First a-priori estimates)**.** *Let* (2.30) *be satisfied, then for all* $\varepsilon > 0$ *there exists a* $K_\varepsilon > 0$ *such that the following holds. For* $\tau < 1/K_\varepsilon$ *the interpolants constructed from the discrete solutions* $(y^k_{\varepsilon\tau}, \theta^k_{\varepsilon\tau}) \in W^{2,p}(\Omega;\mathbb{R}^d) \times H^1(\Omega)$*,* $k = 1, ..., T/\tau$*, obtained in Proposition 4.1 satisfy the following estimates:*

$$\big\|y_{\varepsilon\tau}\big\|_{L^\infty(I;W^{2,p}(\Omega;\mathbb{R}^d))\cap H^1(I;H^1(\Omega;\mathbb{R}^d))} \le K_\varepsilon, \tag{4.9a}$$

$$\det\big(\nabla y_{\varepsilon\tau}(t,x)\big) \ge 1/K_\varepsilon \quad \text{a.e. on } Q, \tag{4.9b}$$

$$\big\|\overline{\theta}_{\varepsilon\tau}\big\|_{L^2(I;H^1(\Omega))\cap L^\infty(I;L^2(\Omega))} \le K_\varepsilon, \tag{4.9c}$$

$$\big\|\overline{w}_{\varepsilon\tau}\big\|_{L^2(I;H^1(\Omega))\cap L^\infty(I;L^2(\Omega)^*)} \le K_\varepsilon, \tag{4.9d}$$

$$\big\|w_{\varepsilon\tau}\big\|_{C(I;L^2(\Omega))\cap L^2([\tau,T],H^1(\Omega))\cap H^1(I;H^1(\Omega)^*)} \le K_\varepsilon, \tag{4.9e}$$

$$\big\|\theta_{\varepsilon\tau}\big\|_{C(I;L^2(\Omega))\cap L^2([\tau,T],H^1(\Omega))\cap H^1(I;H^1(\Omega)^*)} \le K_\varepsilon, \tag{4.9f}$$

We emphasize that we did not make any smoothness assumptions for θ_0, hence the regularized initial values $\theta^0_{\varepsilon\tau} := \theta_{0,\varepsilon}$ and $w^0_{\varepsilon\tau} := \mathfrak{w}(\nabla y_0, \theta_{0,\varepsilon})$ are not smooth. This explains, why we have to use the left-continuous interpolants in (4.9c) and (4.9d) and why in (4.9e) we have to exclude the interval $[0,\tau]$ in $L^2([\tau,T];H^1(\Omega))$.

Proof. As $y^k_{\varepsilon\tau}$ is a global minimizer, we can insert $y = y^{k-1}_{\varepsilon\tau}$ as testfunction in (4.5) to obtain the estimate (recall $\boldsymbol{\delta}_\tau y^k_\varepsilon = \frac{1}{\tau}(y^k_{\varepsilon\tau} - y^{k-1}_{\varepsilon\tau})$)

$$\Psi(y^k_{\varepsilon\tau}, \theta^{k-1}_{\varepsilon\tau}) - \Psi(y^{k-1}_{\varepsilon\tau}, \theta^{k-1}_{\varepsilon\tau}) + \tau\mathcal{R}(y^{k-1}_{\varepsilon\tau}, \boldsymbol{\delta}_\tau y^k_\varepsilon, \theta^{k-1}_{\varepsilon\tau}) + \frac{\varepsilon\tau}{2}\|\boldsymbol{\delta}_\tau y^k_\varepsilon\|^2_{L^2} \le \tau\langle \ell^k_\tau, \boldsymbol{\delta}_\tau y^k_\varepsilon\rangle. \tag{4.10}$$

The proof will be divided into three steps.

Step 1: Uniform energy bound. Using the decomposition $\Psi(y,\theta) = \mathcal{M}(y) + \Phi_{\rm cpl}(y,\theta)$, see (2.18b), we can write equivalently

$$\begin{aligned}&\mathcal{M}(y^k_{\varepsilon\tau}) - \mathcal{M}(y^{k-1}_{\varepsilon\tau}) + \tau\mathcal{R}(y^{k-1}_{\varepsilon\tau}, \boldsymbol{\delta}_\tau y^k_\varepsilon, \theta^{k-1}_{\varepsilon\tau}) + \frac{\varepsilon\tau}{2}\|\nabla\boldsymbol{\delta}_\tau y^k_\varepsilon\|^2_{L^2}\\ &\le \tau\langle \ell^k_\tau, \boldsymbol{\delta}_\tau y^k_\varepsilon\rangle + \int_\Omega \big(\phi(\nabla y^{k-1}_{\varepsilon\tau}, \theta^{k-1}_{\varepsilon\tau}) - \phi(\nabla y^k_{\varepsilon\tau}, \theta^{k-1}_{\varepsilon\tau})\big)\,{\rm d}x.\end{aligned} \tag{4.11}$$

To estimate the last term use the assumption (2.30b) on $|\partial_F\phi(F,\theta)|$ as follows

$$\begin{aligned}\phi(F_1,\theta) - \phi(F_2,\theta) &\le K(1{+}|F_1| + |F_2|)^{s/2}\,|F_1{-}F_2|\\ &\le \frac{K^2}{2\rho}(1{+}|F_1| + |F_2|)^s + \frac{\rho}{2}|F_1{-}F_2|^2,\end{aligned} \tag{4.12}$$

where $\rho > 0$ is arbitrary. Choosing $\rho = \varepsilon/(4\tau)$ and $F_j = \nabla y^{k+j-2}_{\varepsilon\tau}$ we can insert this into the estimate (4.11). Moreover we can use $\langle \ell^k_\tau, \boldsymbol{\delta}_\tau y^k_\varepsilon\rangle \le \|\ell^k_\tau\|_{H^{-1}}\|\boldsymbol{\delta}_\tau y^k_\varepsilon\|_{H^1} \le \|\ell^k_\tau\|_{H^{-1}}c_{\rm P}\|\nabla\boldsymbol{\delta}_\tau y^k_\varepsilon\|_{L^2}$ as $\boldsymbol{\delta}_\tau y^k_\varepsilon \in \mathcal{Y}_0$ and $\mathcal{R} \ge 0$. This leads to

$$\begin{aligned}&\mathcal{M}(y^k_{\varepsilon\tau}) - \mathcal{M}(y^{k-1}_{\varepsilon\tau}) + \frac{\varepsilon\tau}{2}\|\nabla\boldsymbol{\delta}_\tau y^k_\varepsilon\|^2_{L^2}\\ &\le \frac{2\tau c_P^2}{\varepsilon}\|\ell^k_\tau\|^2_{H^{-1}} + \frac{\varepsilon\tau}{8}\|\nabla\boldsymbol{\delta}_\tau y^k_\varepsilon\|^2_{L^2} + \frac{2\tau K^2}{\varepsilon}\int_\Omega \big(1{+}|\nabla y^k_{\varepsilon\tau}|{+}|\nabla y^{k-1}_{\varepsilon\tau}|\big)^s\,{\rm d}x + \frac{\varepsilon\tau}{8}\|\nabla\boldsymbol{\delta}_\tau y^k_\varepsilon\|^2_{L^2}.\end{aligned}$$

Using the coercivity assumption (2.30b) for ϕ the second-last term can be estimated by $\mathcal{M}$ again and setting $m_k := \mathcal{M}(y^k_{\varepsilon\tau})$ we obtain the recursive estimate

$$m_k - m_{k-1} + \frac{\varepsilon\tau}{4}\|\nabla\boldsymbol{\delta}_\tau y^k_\varepsilon\|^2_{L^2} \le \tau\overline{c}_\varepsilon\|\ell^k_\tau\|^2_{H^{-1}} + \tau C_\varepsilon(|\Omega|+m_k+m_{k-1}) \tag{4.13}$$

with $C_\varepsilon = 2{\cdot}3^sK^2/\varepsilon$ and $\overline{c}_\varepsilon = 2c_{\rm P}^2/\varepsilon$. In a first step we neglect the last term on the left-hand side and obtain

$$\big(1{-}\tau C_\varepsilon\big)m_k \le \big(1{+}\tau C_\varepsilon\big)m_{k-1} + \overline{c}_\varepsilon\tau\|\ell^k_\tau\|^2_{H^{-1}} + \tau C_\varepsilon|\Omega|.$$

We now restrict $\tau > 0$ via $\tau < 1/(2C_\varepsilon)$ by choosing $K_\varepsilon \ge 2C_\varepsilon$, so we can iterate the above estimate. With (2.30h) we have $m_0 := \Psi(y_0,\theta_0) < \infty$ and a simple induction yields the discrete Gronwall-type estimate (with $Q_\varepsilon = (1{+}\tau C_\varepsilon)/(1{-}\tau C_\varepsilon)$)

$$\begin{aligned} m_k &\le Q^k_\varepsilon m_0 + \frac{\tau}{1-\tau C_\varepsilon}\sum_{j=1}^k Q^{k-j}\big(\overline{c}_\varepsilon\|\ell^j_\tau\|^2_{H^{-1}}{+}C_\varepsilon|\Omega|\big)\\
&\le Q^k\Big(m_0 + 2\overline{c}_\varepsilon\big(\sum_{j=1}^k\tau\|\ell^j_\tau\|^2_{H^{-1}}\big) + k\tau\,2C_\varepsilon|\Omega|\Big)\\
&\le 4\mathrm{e}^{2C_\varepsilon T}\Big(\Psi(y_0,\theta_0) + 2\overline{c}_\varepsilon\int_0^T\|\ell(s)\|^2_{H^{-1}}\,\mathrm{d}s + 2TC_\varepsilon|\Omega|\Big) := \widetilde{K}_\varepsilon. \end{aligned} \tag{4.14}$$

Using Theorem 3.1 we obtain the desired uniform upper bound in (4.9a) for the interpolant $y_{\varepsilon\tau} : I = [0,T] \to \mathcal{Y}_{\rm id}$ in $L^\infty\big(I;W^{2,p}(\Omega;\mathbb{R}^d)\big)$ as well as the lower bound (4.9b) for the determinant.

Step 2: Dissipation bound. We return to (4.13) and add all estimates from $k=1$ to $N_\tau := T/\tau \in \mathbb{N}$ to obtain

$$\begin{aligned} \frac{\varepsilon}{4}\int_Q|\nabla\dot y_{\varepsilon\tau}|^2\,\mathrm{d}x\,\mathrm{d}t &= \frac{\varepsilon\tau}{4}\sum_{k=1}^{N_\tau}\|\nabla\boldsymbol{\delta}_\tau y^k_\varepsilon\|^2_{L^2}\\
&\le m_0 - m_{N_\tau} + \tau\sum_{k=1}^{N_\tau}\Big(\overline{c}_\varepsilon\|\ell^k_\tau\|^2_{H^{-1}} + C_\varepsilon(|\Omega|{+}m_{k-1}{+}m_k)\Big)\\
&\le \Psi(y_0,\theta_0) + \overline{c}_\varepsilon\|\ell\|^2_{L^2(I;H^{-1})} + C_\varepsilon T(|\Omega|{+}2\widetilde{K}_\varepsilon) =: \widehat{K}_\varepsilon. \end{aligned}$$

This provides the uniform bound for $y_{\varepsilon\tau}$ in $H^1(I;H^1(\Omega;\mathbb{R}^d))$, and (4.9a) is established.

Step 3: Temperature bounds. Testing the Euler-Lagrange equations (4.3b) and (4.4c) by $w^k_{\varepsilon\tau}$ yields the identity

$$\begin{aligned} \int_\Omega\Big(\frac{w^k_{\varepsilon\tau}-w^{k-1}_{\varepsilon\tau}}{\tau}\,w^k_{\varepsilon\tau} + \nabla w^k_{\varepsilon\tau}{\cdot}\mathcal{K}(\nabla y^{k-1}_{\varepsilon\tau},\theta^{k-1}_{\varepsilon\tau})\nabla\theta^k_{\varepsilon\tau}\Big)\,\mathrm{d}x + \int_\Gamma\kappa\theta^k_{\varepsilon\tau}w^k_{\varepsilon\tau}\,\mathrm{d}S\\
= \int_\Omega h^k_{\varepsilon\tau}w^k_{\varepsilon\tau}\,\mathrm{d}x + \int_\Gamma\kappa\theta^k_{\flat,\varepsilon,\tau}w^k_{\varepsilon\tau}\,\mathrm{d}S \end{aligned} \tag{4.15}$$

with $h^k_{\varepsilon\tau} := \xi^{\rm reg}_\varepsilon(\nabla y^{k-1}_{\varepsilon\tau},\nabla\boldsymbol{\delta}_\tau y^k_\varepsilon,\theta^{k-1}_{\varepsilon\tau}) + \partial_F\phi(\nabla y^k_{\varepsilon\tau},\theta^k_{\varepsilon\tau}){:}\nabla\boldsymbol{\delta}_\tau y^k_\varepsilon$.

Recalling $c_{\rm v}(F,\theta) = \partial_\theta\mathfrak{w}(F,\theta)$ we obtain the chain rule

$$\nabla w^k_{\varepsilon\tau} = \nabla\mathfrak{w}(\nabla y^k_{\varepsilon\tau},\theta^k_{\varepsilon\tau}) = \partial_F\mathfrak{w}(\nabla y^k_{\varepsilon\tau},\theta^k_{\varepsilon\tau}){:}\nabla^2 y^k_{\varepsilon\tau} + c_{\rm v}(\nabla y^k_{\varepsilon\tau},\theta^k_{\varepsilon\tau})\nabla\theta^k_{\varepsilon\tau}. \tag{4.16}$$

Moreover, we have the elementary estimate $\frac{1}{\tau}(w_{\varepsilon\tau}^k - w_{\varepsilon\tau}^{k-1})w_{\varepsilon\tau}^k \le \frac{1}{2\tau}\big((w_{\varepsilon\tau}^k)^2 - (w_{\varepsilon\tau}^{k-1})^2\big)$, and $\theta w = \theta\mathfrak{w}(F,\theta) \ge 0$ by the definition of $\mathfrak{w}$. Using additionally $c_{\rm v}(F,\theta) = -\theta\partial_\theta^2\phi(F,\theta) \ge \hat\epsilon$ (see (2.30c), the above identity (4.15) leads to

$$\begin{aligned}&\int_\Omega \Big(\frac{1}{2\tau}(w_{\varepsilon\tau}^k)^2 - \frac{1}{2\tau}(w_{\varepsilon\tau}^{k-1})^2 + \hat\epsilon\nabla\theta_{\varepsilon\tau}^k\cdot\mathcal{K}_{\varepsilon\tau}^k\nabla\theta_{\varepsilon\tau}^k\Big)\,{\rm d}x\\ &\quad\le \int_\Omega h_{\varepsilon\tau}^k w_{\varepsilon\tau}^k - \nabla\theta_{\varepsilon\tau}^k\cdot\mathcal{K}_{\varepsilon\tau}^k b_{\varepsilon\tau}^k\,{\rm d}x + \int_\Gamma \kappa\theta_{\flat,\varepsilon,\tau}^k w_{\varepsilon\tau}^k\,{\rm d}S. \qquad (4.17)\\ &\text{where } \mathcal{K}_{\varepsilon\tau}^k = \mathcal{K}(\nabla y_{\varepsilon\tau}^{k-1},\theta_{\varepsilon\tau}^{k-1}) \text{ and } b_{\varepsilon\tau}^k := \partial_F\mathfrak{w}(\nabla y_{\varepsilon\tau}^k,\theta_{\varepsilon\tau}^k){:}\nabla^2 y_{\varepsilon\tau}^k.\end{aligned}$$

Using uniform bounds for $\nabla y_{\varepsilon\tau}$ and $\det\nabla y_{\varepsilon\tau}$ from Step 1, the assumption (2.30f) on $\mathbb{K}$, as well as formula (2.24) we find a $\varkappa_\varepsilon$ such that

$$|\mathcal{K}_{\varepsilon\tau}^k| \le \varkappa_\varepsilon \quad\text{and}\quad a\cdot\mathcal{K}_{\varepsilon\tau}^k a \ge \frac{1}{\varkappa_\varepsilon}|a|^2 \text{ for all } a\in\mathbb{R}^d. \qquad (4.18)$$

Moreover, using $\partial_F\mathfrak{w} = \partial_F\phi - \theta\partial_{F\theta}^2\phi$ the assumptions (2.30b) and (2.30c) together with the uniform L^∞ bound for $\nabla y_{\varepsilon\tau}$ we find $\|\partial_F\mathfrak{w}(\nabla y_{\varepsilon\tau}^k,\theta_{\varepsilon\tau}^k)\|_{L^\infty}\le C_\varepsilon$. Realizing also that we have $\nabla^2 y_{\varepsilon\tau}^k$ already estimated in $L^p(\Omega;\mathbb{R}^{d\times d\times d})$ with $p\ge 2$ we obtain $\|b_{\varepsilon\tau}^k\|_{L^2}\le\overline{C}_\varepsilon$. For the right-hand side $h_{\varepsilon\tau}^k$ of (4.15) we have

$$\|h_{\varepsilon\tau}\|_{L^2} \le \|\xi_\varepsilon^{\rm reg}\|_{L^2} + \|\partial_F\phi(\nabla y_{\varepsilon\tau}^k,\theta_{\varepsilon\tau}^k)\|_{L^\infty}\|\nabla\boldsymbol{\delta}_\tau y_\varepsilon^k\|_{L^2} \le \overline{C}_\varepsilon\big(1+\|\boldsymbol{\delta}_\tau y_\varepsilon^k\|_{H^1}\big),$$

where we again used the L^∞ bounds for $\nabla y_{\varepsilon\tau}^k$. Finally, by definition we have $\theta_{\flat,\varepsilon}\in[0,1/\varepsilon]$, and (2.31) allows us to estimate w by θ, which yields the boundary estimate

$$\Big|\int_\Gamma \theta_{\flat,\varepsilon,\tau}^k w_{\varepsilon\tau}^k\,{\rm d}S\Big| \le \frac{1}{\varepsilon}\int_\Gamma K|\theta_{\varepsilon\tau}^k|\,{\rm d}S \le C_\varepsilon\|\theta_{\varepsilon\tau}^k\|_{H^1} \le \overline{C}_\varepsilon\big(\|w_{\varepsilon\tau}^k\|_{L^2}+\|\nabla\theta_{\varepsilon\tau}^k\|_{L^2}\big).$$

Based on the above estimates and introducing the abbreviations

$$\gamma_k := \|w_{\varepsilon\tau}^k\|_{L^2},\quad \Theta_k := \|\nabla\theta_{\varepsilon\tau}^k\|_{L^2},\ \text{and}\ \nu_k := \|\boldsymbol{\delta}_\tau y_\varepsilon^k\|_{H^1}$$

we can estimate the right-hand side in (4.17) via

$$\text{RHS} \le \overline{C}_\varepsilon(1{+}\nu_k)\gamma_k + C_\varepsilon\Theta_k + \overline{C}_\varepsilon(\gamma_k{+}\Theta_k) \le \overline{c}_\varepsilon\Big(\frac{1}{\alpha}+\nu_k^2+\gamma_k^2+\alpha\Theta^2\Big),$$

where $\alpha>0$ is arbitrary. Estimating the last term on the left-hand side in (4.17) from below by $\frac{\hat\epsilon}{\varkappa}\Theta_k^2$ we may choose $\alpha=\hat\epsilon/(2\varkappa\overline{c}_\varepsilon)$. After multiplying (4.17) by 2τ we obtain

$$\gamma_k^2 - \gamma_{k-1}^2 + \frac{\hat\epsilon}{2\varkappa}\Theta_k^2 \le \tau\hat{c}_\varepsilon\big(1+\nu_k^2+\gamma_k^2\big). \qquad (4.19)$$

Arguing as in Steps 1 and 2 for (4.13) and using $\gamma_0^2 = \int_\Omega w_{\varepsilon\tau}^0\,{\rm d}x \le K^2\int_\Omega\theta_{0,\varepsilon}^2\,{\rm d}x \le K^2|\Omega|/\varepsilon^2 < \infty$ (cf. (4.2d)) the left-continuous interpolants $\overline{\theta}_{\varepsilon\tau}$ and $\overline{w}_{\varepsilon\tau}$ satisfy the a priori estimates

$$\hat\epsilon\|\overline{\theta}_{\varepsilon\tau}\|_{L^\infty(I;L^2(\Omega))} \le \|\overline{w}_{\varepsilon\tau}\|_{L^\infty(I;L^2(\Omega))} = \sup_{k=0,\dots,N_\tau}\gamma_k \le K_\varepsilon \text{ and } \|\nabla\overline{\theta}_{\varepsilon\tau}\|_{L^2(Q)}^2 = \tau\sum_{k=1}^{N_\tau}\Theta_k^2 \le K_\varepsilon.$$

With $\theta\le\mathfrak{w}(F,\theta)/\hat\epsilon$ we immediately find (4.9c) for $\overline{\theta}_{\varepsilon\tau}$. The estimate (4.9d) follows by using (4.16) once again.

The uniform estimate the piecewise affine interpolant $w_{\varepsilon\tau}$ in the spaces $C(I;L^2(\Omega))\cap L^2([\tau,T],H^1(\Omega))$ follows from the previous estimates for $\overline{w}_{\varepsilon\tau}$. Finally, we note that the time derivative interpolant $w_{\varepsilon\tau}$ is equal to $\boldsymbol{\delta}_\tau w_\varepsilon^k$ on the intervals $](k-1)\tau,k\tau[$. We now use the Euler-Lagrange equations (4.3b) and (4.4c), which provides for $\boldsymbol{\delta}_\tau w_\varepsilon^k=\frac{1}{\tau}(w_\varepsilon^k-w_\varepsilon^{k-1})$ the estimate

$$\|\boldsymbol{\delta}_\tau w_\varepsilon^k\|_{(H^1)^*}\le C_\varepsilon^{\mathcal{K}}\|\nabla\theta_{\varepsilon\tau}^k\|_{L^2}+C_\varepsilon^{\xi}+C_\varepsilon^{\partial_F\phi}\|\boldsymbol{\delta}_\tau y_\varepsilon^k\|_{H^1}+C_\varepsilon^{\kappa}\big(\|\theta_{\varepsilon\tau}^k\|_{H^1}+|\Gamma|/\varepsilon\big).$$

Squaring and summation over $k=1,\dots,N_\tau$ gives the remaining uniform bound in (4.9e) for $\partial_t w_{\varepsilon\tau}$ in $L^2\big(I;H^1(\Omega)^*\big)$.

Using (2.31) once again, we bound the increments $\boldsymbol{\delta}_\tau\theta_\varepsilon^k$ via the pointwise estimate

$$\begin{aligned}\hat\epsilon|\boldsymbol{\delta}_\tau\theta_\varepsilon^k|&=\frac{\hat\epsilon}{\tau}|\theta_{\varepsilon\tau}^k-\theta_{\varepsilon\tau}^{k-1}|\le\frac{1}{\tau}|\mathfrak{w}(\nabla y_{\varepsilon\tau}^{k-1},\theta_{\varepsilon\tau}^k)-\mathfrak{w}(\nabla y_{\varepsilon\tau}^{k-1},\theta_{\varepsilon\tau}^{k-1})|\\&\le\frac{1}{\tau}|w_{\varepsilon\tau}^k-w_{\varepsilon\tau}^{k-1}|+\frac{1}{\tau}|\mathfrak{w}(\nabla y_{\varepsilon\tau}^k,\theta_{\varepsilon\tau}^k)-\mathfrak{w}(\nabla y_{\varepsilon\tau}^{k-1},\theta_{\varepsilon\tau}^k)|\le|\boldsymbol{\delta}_\tau w_\varepsilon^k|+C_\varepsilon|\nabla\boldsymbol{\delta}_\tau y_\varepsilon^k|.\end{aligned}$$

Taking the $H^1(\Omega)^*$ norm we obtain $\|\boldsymbol{\delta}_\tau\theta_\varepsilon^k\|_{H^1(\Omega)^*}\le K_\varepsilon\big(\|\boldsymbol{\delta}_\tau w_\varepsilon^k\|_{H^1(\Omega)^*}+\|\boldsymbol{\delta}_\tau y_\varepsilon^k\|_{H^1(\Omega)}\big)$, such that (4.9f) follows from (4.9e), (4.9a), and (4.9c).

This finishes the proof of Proposition 4.2. □

5 The limit $\tau\to0$ in the regularized problem

Using the above a priori estimates for the interpolants we will be able to extract convergent subsequences. First we will observe that the three different types of interpolants have to converge to the same limit. Next we want to pass to the limit in the discretized weak forms of the momentum balance and the heat equation. While most terms can be handled by compactness arguments or weak-convergence methods, there is one term that needs special attention namely the heat-source term $\xi_\varepsilon^{\rm reg}$ that is quadratic in $\nabla\dot y_\varepsilon$. Thus, it will be a crucial step to show strong convergence of $\dot y_{\varepsilon\tau}$ in $L^2(I;H^1(\Omega))$, which can be done by passing to the limit in a suitable discretized version of the mechanical energy balance (2.20). In this argument we will use the Λ-convexity derived in Proposition 3.2 to relate the mechanical energies $\mathcal{M}(y_{\varepsilon\tau}^{k-1})$ and $\mathcal{M}(y_{\varepsilon\tau}^k)$.

With the definition (4.8) for the three types of interpolants, we see that the following discretized version (5.1) of the momentum balance and heat equations (4.1) and (4.2) holds for the discrete solutions constructed in Proposition 4.1:

$$\begin{aligned}&-\operatorname{div}\big(\sigma_{\rm vi}(\nabla\underline{y}_{\varepsilon\tau},\nabla\dot{\underline{y}}_{\varepsilon\tau},\underline{\theta}_{\varepsilon\tau})+\varepsilon\nabla\dot{\underline{y}}_{\varepsilon\tau}+\sigma_{\rm el}(\nabla\overline{y}_{\varepsilon\tau},\underline{\theta}_{\varepsilon\tau})\\&\qquad\qquad-\operatorname{div}\mathfrak{h}_{\rm el}(\nabla^2\overline{y}_{\varepsilon\tau})\big)=\overline{g}_\tau\,,&\text{(5.1a)}\\&\dot w_{\varepsilon\tau}-\operatorname{div}\big(\mathcal{K}(\nabla\underline{y}_{\varepsilon\tau},\underline{\theta}_{\varepsilon\tau})\nabla\overline{\theta}_{\varepsilon\tau}\big)=\xi_\varepsilon^{\rm reg}(\nabla\underline{y}_{\varepsilon\tau},\nabla\dot{\underline{y}}_{\varepsilon\tau},\underline{\theta}_{\varepsilon\tau})+\partial_F\phi(\nabla\overline{y}_{\varepsilon\tau},\overline{\theta}_{\varepsilon\tau}):\nabla\dot{\underline{y}}_{\varepsilon\tau},&\text{(5.1b)}\\&\overline{w}_{\varepsilon\tau}=\mathfrak{w}(\nabla\overline{y}_{\varepsilon\tau},\overline{\theta}_{\varepsilon\tau}),&\text{(5.1c)}\end{aligned}$$

to hold on $Q=[0,T]\times\Omega$, while the regularized boundary conditions (4.4) read

$$\begin{aligned}&\big(\sigma_{\rm vi}\big(\nabla\underline{y}_{\varepsilon\tau},\nabla\dot{\underline{y}}_{\varepsilon\tau},\underline{\theta}_{\varepsilon\tau}\big)+\varepsilon\nabla\dot{\underline{y}}_{\varepsilon\tau}+\sigma_{\rm el}(\nabla\overline{y}_{\varepsilon\tau},\underline{\theta}_{\varepsilon\tau})\big)\vec n\\&\qquad\qquad-\operatorname{div}_{\rm s}\big(\mathfrak{h}_{\rm el}(\nabla^2\overline{y}_{\varepsilon\tau})\vec n\big)=\overline{f}_\tau&\text{on }\Sigma_{\rm N},&\text{(5.2a)}\\&\overline{y}_{\varepsilon\tau}=\text{ identity on }\Sigma_{\rm D},\qquad\qquad\mathfrak{h}_{\rm el}(\nabla^2\overline{y}_{\varepsilon\tau}):(\vec n\otimes\vec n)=0&\text{on }\Sigma,&\text{(5.2b)}\\&\mathcal{K}(\nabla\underline{y}_{\varepsilon\tau},\underline{\theta}_{\varepsilon\tau})\nabla\overline{\theta}_{\varepsilon\tau}\cdot\vec n+\kappa\overline{\theta}_{\varepsilon\tau}=\kappa\overline{\theta}_{\flat,\varepsilon,\tau}&\text{on }\Sigma.&\text{(5.2c)}\end{aligned}$$

Here it is essential that we have to use all three types of interpolants, e.g. $\overline{y}_{\varepsilon\tau}$, $\underline{y}_{\varepsilon\tau}$, and $y_{\varepsilon\tau}$. In particular, we emphasize that $t \mapsto w_{\varepsilon\tau}(t)$ is the piecewise affine interpolant of $\{w^k_{\varepsilon\tau}\}_{k=0,\dots,N_\tau}$, which does not coincide with $t \mapsto \mathfrak{w}(\nabla y_{\varepsilon\tau}(t), \theta_{\varepsilon\tau}(t))$ except at the nodal points $t = k\tau$.

Proposition 5.1 (Convergence for $\tau \to 0$)**.** *Let* (2.30) *hold, and let $\varepsilon > 0$ be fixed. Then, considering a sequence of time steps $\tau \to 0$, there is a subsequence (not relabeled) and limit functions $(y_\varepsilon, \theta_\varepsilon)$ such that*

$$y_{\varepsilon\tau} \to y_\varepsilon \qquad \text{weakly* in } L^\infty(I; W^{2,p}(\Omega;\mathbb{R}^d)) \cap H^1(I; H^1(\Omega;\mathbb{R}^d)), \tag{5.3a}$$

$$\theta_{\varepsilon\tau} \to \theta_\varepsilon \qquad \text{weakly in } L^2(I; H^1(\Omega)) \cap H^1(I; H^1(\Omega)^*). \tag{5.3b}$$

Moreover, any couple $(y_\varepsilon, \theta_\varepsilon)$ obtained by this way is a weak solution to the regularized initial-boundary-value problem (4.1)–(4.2)*.*

Proof. The proof consists of five steps.

Step 1: Extraction of convergent subsequences. As $\varepsilon > 0$ is still fixed, we can exploit the a priori estimates obtained in Proposition 4.2, namely (4.9a) and (4.9f). By Banach's selection principle, we choose a subsequence and some $(y_\varepsilon, \theta_\varepsilon)$ such that (5.3) holds. By the Aubin-Lions theorem combined with an interpolation, as $p > d$, we have also

$$\nabla y_{\varepsilon\tau} \to \nabla y_\varepsilon \qquad \text{uniformly in } L^\infty(Q;\mathbb{R}^{d\times d}), \tag{5.4a}$$

$$\theta_{\varepsilon\tau} \to \theta_\varepsilon \qquad \text{strongly in } L^s(Q) \quad \text{for all } s \in [1, \min\{4, 2+4/d\}[. \tag{5.4b}$$

Indeed, for the first result we use the continuous embedding $W^{1,p}(\Omega) \subset C^\alpha(\overline{\Omega})$ with $\alpha = 1 - d/p \in]0,1[$ and thus $\|\nabla y_{\varepsilon\tau}\|_{C^\alpha} \le K_0$. Moreover, (4.9a) yields the Hölder estimate

$$\big\|\nabla y_{\varepsilon\tau}(t_1) - \nabla y_{\varepsilon\tau}(t_2)\big\|_{L^2(\Omega;\mathbb{R}^d)} \le K_1 |t_1 - t_2|^{1/2} \quad \text{for all } t_1, t_2 \in I. \tag{5.5}$$

While the first part of (4.9a) yields just $\|\nabla y_{\varepsilon\tau}(t_1) - \nabla y_{\varepsilon\tau}(t_2)\|_{W^{1,p}(\Omega;\mathbb{R}^d)} \le K_0$. By interpolation, we find $\beta \in]0,\alpha[$ and $\lambda \in]0,1[$ such that we have the interpolation $\|\cdot\|_{C^\beta} \le C\|\cdot\|^{1-\lambda}_{C^\alpha}\|\cdot\|^\lambda_{L^2)}$ and conclude

$$\big\|\nabla y_{\varepsilon\tau}(t_1) - \nabla y_{\varepsilon\tau}(t_2)\big\|_{C^\beta(\bar\Omega;\mathbb{R}^d)} \le C K_0^{1-\lambda} K_1^\lambda |t_1 - t_2|^{\lambda/2}. \tag{5.6}$$

Thus, the sequence $\{\nabla y_{\varepsilon\tau}\}$ is uniformly bounded in $C^\gamma(\overline{Q})$ for $\gamma = \min\{\beta, \lambda/2\}$, and uniform convergence follows by the Arzelà-Ascoli theorem.

The convergence (5.4b) follows from (5.3b) by the Aubin-Lions theorem when interpolated with the estimate in $L^\infty(I; L^2(\Omega))$ which is contained implicitly in (5.3b).

Moreover, both convergences in (5.4) hold also for the piecewise constant interpolants because of the estimates $\|\nabla y_{\varepsilon\tau} - \nabla \underline{y}_{\varepsilon\tau}\|_{L^\infty(I;L^2(\Omega;\mathbb{R}^{d\times d}))} \le K\tau^{1/2}$ (and the same also for $\nabla\overline{y}_{\varepsilon\tau}$) and $\|\nabla\theta_{\varepsilon\tau} - \nabla\overline{\theta}_{\varepsilon\tau}\|_{L^\infty(I;H^1(\Omega;\mathbb{R}^d)^*)} \le K\tau^{1/2}$.

Similarly, using the a priory estimates (4.9d) and (4.9e) for $w_{\varepsilon\tau}$ and $\overline{w}_{\varepsilon\tau}$ yields

$$\begin{aligned} w_{\varepsilon\tau} &\rightharpoonup w_\varepsilon \quad \text{weakly in } L^2(I; H^1(\Omega)) \cap H^1(I; H^1(\Omega)^*) \\ \overline{w}_{\varepsilon\tau}, w_{\varepsilon\tau} &\to w_\varepsilon \quad \text{strongly in } L^s(Q) \text{ for all } s \in [1, \min\{4, 2+4/d\}[. \end{aligned} \tag{5.7}$$

Step 2: Convergence in the mechanical equation. Now the convergence in the discretized momentum balance (5.1a) can be done by the above weak convergences (5.3) because $\sigma_{\rm vi}$ is linear in terms of

$\dot F$ and by Minty's trick for the monotone operator induced by $\mathfrak{h}_{\rm el} = \mathscr{H}'$. For a reflexive Banach space X and a hemi-continuous, monotone operator $\mathbf{H} : X \to X^*$ Minty's trick means the implication

$$\left.\begin{aligned} &\mathbf{H}(u_\tau) = b_\tau, \quad u_\tau \rightharpoonup u \text{ in } X,\\ &b_\tau \rightharpoonup b \text{ in } X^*, \quad \langle b_\tau, u_\tau\rangle \to \langle b, u\rangle \end{aligned}\right\} \implies \mathbf{H}(u) = b. \tag{5.8}$$

We apply this for $\mathbf{H}$ defined by $\langle \mathbf{H}(y), z\rangle = \int_Q \mathfrak{h}_{\rm el}(\nabla^2 y(t,x)) \vdots \nabla^2 z(t,x)\,\mathrm{d}x\,\mathrm{d}t$, where $X = W^{2,p}(Q)$. Clearly, $\mathbf{H}$ is hemi-continuous and monotone. Choosing $u_\tau = \overline{y}_{\varepsilon\tau}$ the weak equations (5.1a) and (5.2) are interpreted as $\mathbf{H}(\overline{y}_{\varepsilon\tau}) = b_\tau$ with b_τ defined via

$$\langle b_\tau, z\rangle = -\int_\Omega \big(\sigma_{\rm vi}(\nabla \underline{y}_{\varepsilon\tau}, \nabla \dot{\overline{y}}_{\varepsilon\tau}, \underline{\theta}_{\varepsilon\tau}) + \varepsilon \nabla \dot{\overline{y}}_{\varepsilon\tau} + \sigma_{\rm el}(\nabla \overline{y}_{\varepsilon\tau}, \underline{\theta}_{\varepsilon\tau})\big) : \nabla z\,\mathrm{d}x\,\mathrm{d}t + \int_0^T \langle \overline{\ell}_\tau, z\rangle\,\mathrm{d}t.$$

We obtain $b_\tau \rightharpoonup b$ with b defined by

$$\langle b, z\rangle = -\int_\Omega \big(\sigma_{\rm vi}(\nabla y_\varepsilon, \nabla \dot y_\varepsilon, \theta_\varepsilon) + \varepsilon \nabla \dot y_\varepsilon + \sigma_{\rm el}(\nabla y_\varepsilon, \theta_\varepsilon)\big) : \nabla z\,\mathrm{d}x\,\mathrm{d}t + \int_0^T \langle \ell, z\rangle\,\mathrm{d}t,$$

because we can pass to the limit τ in all four terms separately. For the first term we applying the lower semicontinuity result [FoL07, Thm. 7.5] twice, namely for the integrands $f_\pm(x,(F,\theta),G) = \pm\sigma_{\rm vi}(F,G,\theta){:}\nabla z(x)$ which both are convex in G. The limit passage in the second term is simple weak convergence, and the fourth term converges because of $\overline{\ell}_\tau \to \ell$ in $L^2\big(I; H^1_{\rm D}(\Omega)^*\big)$. In the third term we exploit

$$\nabla \overline{y}_{\varepsilon\tau} \in \mathbb{F}(K_\varepsilon) := \big\{\, F \in \mathbb{R}^{d\times d};\, |F| \le K_\varepsilon,\ \det F \ge 1/K_\varepsilon \,\big\}$$

(see (4.9a) and (4.9b) from Proposition 4.2), such that using (2.30a) and (2.30b) the map $(F,\theta) \mapsto \sigma_{\rm el}(F,\theta) = \partial_F \varphi(F) + \partial_F \phi(F,\theta)$ is continuous and bounded on $\mathbb{F}(K_\varepsilon) \times \mathbb{R}^+$. Hence, with (5.4) and Lebesgue's dominated convergence theorem we obtain the desired convergence.

To use Minty's trick (5.8) we still need to check $\langle b_\tau, \overline{y}_{\varepsilon\tau}\rangle \to \langle b, y_\varepsilon\rangle$. However, as we have shown above b_τ is bounded (and hence weakly converging to b) in $L^2\big(I; H^1_{\rm D}(\Omega)^*\big)$ and $\overline{y}_{\varepsilon\tau} \to y_\varepsilon$ in $L^2\big(I; H^1_{\rm D}(\Omega)\big)$ strongly (by (5.4a), the result follows immediately. Hence, we conclude $\mathbf{H}(y_\varepsilon) = b$, which is nothing else than the regularized momentum balance (4.1a), (4.2a), and (4.2b).

Step 3: Balance of mechanical energy. For the limit passage in the heat equation we need strong L^2-convergence of $\nabla \dot y_{\varepsilon\tau}$ due to the viscous dissipation $\xi^{\rm reg}_\varepsilon(F,\dot F,\theta)$ that is nonlinear in $\dot F$. The strategy is to use the balance of mechanical energy as follows. Rewriting the regularized momentum balance (4.1a), (4.2a), and (4.2b) in the form

$$\mathrm{D}_{\dot y}\mathcal{R}(y_\varepsilon, \dot y_\varepsilon, \theta_\varepsilon) + \varepsilon \nabla \dot y_\varepsilon + \mathrm{D}\mathcal{M}(y_\varepsilon) + \mathrm{D}_y \Phi_{\rm cpl}(y_\varepsilon, \theta_\varepsilon) = \ell(t)$$

with $\mathcal{M}$ and $\Phi_{\rm cpl}$ defined in (2.18). We can now test with $\dot y_\varepsilon \in L^2(I; H^1_{\rm D}(\Omega))$ and use (after decomposing $\mathcal{M} = \mathcal{H} + \Phi_{\rm el}$, see (2.18)) the chain rule in Proposition 3.6 to obtain the balance of mechanical energy in the form

$$\begin{aligned} &\mathcal{M}(y_\varepsilon(T)) + \int_0^T \big(2\mathcal{R}(y_\varepsilon, \dot y_\varepsilon, \theta_\varepsilon) + \varepsilon \|\nabla \dot y_\varepsilon\|^2_{L^2}\big)\,\mathrm{d}t \\ &= \mathcal{M}(y_0) + \int_0^T \langle \ell, \dot y_\varepsilon\rangle\,\mathrm{d}t - \int_Q \partial_F \phi(\nabla y_\varepsilon, \theta_\varepsilon){:}\nabla \dot y_\varepsilon\,\mathrm{d}x\,\mathrm{d}t. \end{aligned} \tag{5.9}$$

Indeed, by Proposition 3.2 we know that $\mathcal{M}$ satisfies the assumptions of Proposition 3.6 with space $X = H^1_{\Gamma_{\rm D}}(\Omega;\mathbb{R}^d)$. Clearly, $y_\varepsilon \in H^1(I;X)$ and $\mathcal{M}(y_\varepsilon(t)) \le \widetilde{K}_\varepsilon$, see (4.14). Moreover, for

$$\Xi = \ell(t) - \mathrm{D}_{\dot y}\mathcal{R}(y_\varepsilon,\dot y_\varepsilon,\theta_\varepsilon) - \varepsilon\nabla\dot y_\varepsilon - \mathrm{D}_y\Phi_{\rm cpl}(y_\varepsilon,\theta_\varepsilon)$$

we have $\Xi(t) = \mathrm{D}\mathcal{M}(y(t))$ a.e. in $[0,T]$ and our a priori estimates provide $\Xi \in L^2([0,T];H^1_{\Gamma_{\rm D}}(\Omega)^*)$. Thus, (5.9) follows from Proposition 3.6.

Step 4: Strong convergence of strain rates. The next step is now to derive a similar mechanical energy balance for the time-discretized solutions, which is better than the previously used estimate (4.11). Passing to the limit $\tau\to0$ from the latter estimate we would arrive at an estimate like (5.9), but with $2\mathcal{R}$ and ε replaced by $\mathcal{R}$ and $\varepsilon/2$, respectively.

To improve the discrete bounds used in Proposition 4.2 we can exploit the a priori estimates $\mathcal{M}(y^k_{\varepsilon\tau}) \le K_\varepsilon$, which allow us to use the geodesic Λ-convexity result in Proposition 3.2. Instead of using the minimization property of $y^k_{\varepsilon\tau}$ in (4.5) we test the Euler-Lagrange equation (4.3a) with boundary conditions (4.4a) and (4.4b) by $y^k_{\varepsilon\tau}-y^{k-1}_{\varepsilon\tau}$ to obtain

$$\begin{aligned}\tau 2\mathcal{R}(y^{k-1}_{\varepsilon\tau},\boldsymbol{\delta}_\tau y^k_\varepsilon,\theta^{k-1}_{\varepsilon\tau}) + \tau\varepsilon\|\nabla\boldsymbol{\delta}_\tau y^k_\varepsilon\|^2_{L^2} + \mathrm{D}_y\mathcal{M}(y^k_{\varepsilon\tau})[y^k_{\varepsilon\tau}-y^{k-1}_{\varepsilon\tau}]\\ = \langle \ell^k_\tau, y^k_{\varepsilon\tau}-y^{k-1}_{\varepsilon\tau}\rangle - \mathrm{D}_y\Phi_{\rm cpl}(y^k_{\varepsilon\tau},\theta^{k-1}_{\varepsilon\tau})[y^k_{\varepsilon\tau}-y^{k-1}_{\varepsilon\tau}],\end{aligned}$$

where we have the correct factors $2\mathcal{R}$ and ε. To recover the energy values $\mathcal{M}(y^j_{\varepsilon\tau})$ we now eliminate the term involving $\mathrm{D}\mathcal{M}$ using the Λ-convexity estimate (3.3) with $y^{(1)} = y^k_{\varepsilon\tau}$ and $y^{(2)} = y^{k-1}_{\varepsilon\tau}$, which yields

$$\begin{aligned}\mathcal{M}(y^k_{\varepsilon\tau}) + \tau 2\mathcal{R}(y^{k-1}_{\varepsilon\tau},\boldsymbol{\delta}_\tau y^k_\varepsilon,\theta^{k-1}_{\varepsilon\tau}) + \big(\tau\varepsilon - \tau^2\Lambda(K_\varepsilon)\big)\|\nabla\boldsymbol{\delta}_\tau y^k_\varepsilon\|^2_{L^2}\\ \le \mathcal{M}(y^{k-1}_{\varepsilon\tau}) + \tau\langle \ell^k_\tau,\boldsymbol{\delta}_\tau y^k_\varepsilon\rangle - \mathrm{D}_y\Phi_{\rm cpl}(y^k_{\varepsilon\tau},\theta^{k-1}_{\varepsilon\tau})[\boldsymbol{\delta}_\tau y^k_\varepsilon].\end{aligned}$$

We now sum this inequality over $k = 1,,\dots,N_\tau$ and using the interpolants we obtain the integral estimate

$$\begin{aligned}\mathcal{M}(\overline{y}_{\varepsilon\tau}(T)) + \int_0^T 2\mathcal{R}(\underline{y}_{\varepsilon\tau},\dot y_{\varepsilon\tau},\underline{\theta}_{\varepsilon\tau})\,\mathrm{d}t + (\varepsilon-\tau\Lambda(K_\varepsilon))\int_Q|\nabla\dot y_{\varepsilon\tau}|^2\,\mathrm{d}x\,\mathrm{d}t\\ \le \mathcal{M}(y_0) + \int_0^T\left(\langle\overline{\ell}_\tau,\dot y_{\varepsilon\tau}\rangle - \int_\Omega \partial_F\phi(\nabla\overline{y}_{\varepsilon\tau},\underline{\theta}_{\varepsilon\tau})\,\mathrm{d}x\right)\mathrm{d}t.\end{aligned}\tag{5.10}$$

Using the the convergences (5.3) and (5.4) it is immediate to see that the all the terms on the right-hand side converge to the corresponding terms on the right-hand side in (5.9). Now denote the three terms on the left-hand side by $I^{(j)}_{\varepsilon\tau}$ and set $I^{(j)}_\varepsilon = \liminf_{\tau\to0^+} I^{(j)}_{\varepsilon\tau}$. Using lower semicontinuity arguments (use [FoL07, Thm. 7.5] once again for $I^{(2)}_{\varepsilon\tau}$) we find

$$\begin{aligned}\overline{y}_{\varepsilon\tau}(T) \rightharpoonup y_\varepsilon(T) \text{ in } W^{2,p}(\Omega;\mathbb{R}^d) \quad&\Longrightarrow\quad I^{(1)}_\varepsilon \ge \mathcal{M}(y_\varepsilon(T)),\\ \nabla\dot y_{\varepsilon\tau} \rightharpoonup \nabla\dot y_\varepsilon \text{ in } L^2(Q;\mathbb{R}^{d\times d}) \quad&\Longrightarrow\quad I^{(2)}_\varepsilon \ge \int_0^T 2\mathcal{R}(y_\varepsilon,\dot y_\varepsilon,\theta)\,\mathrm{d}t,\\ \nabla\dot y_{\varepsilon\tau} \rightharpoonup \nabla\dot y_\varepsilon \text{ in } L^2(Q;\mathbb{R}^{d\times d}) \quad&\Longrightarrow\quad I^{(3)}_\varepsilon \ge \varepsilon\|\nabla\dot y_\varepsilon\|^2_{L^2(Q)}.\end{aligned}\tag{5.11}$$

Thus, passing to the liminf on the left-hand side and to the limit on the right-hand side in (5.10) and comparing with (5.9) we obtain

$$I^{(1)}_\varepsilon + I^{(2)}_\varepsilon + I^{(3)}_\varepsilon \le \text{RHS} = \mathcal{M}(y_\varepsilon(T)) + \int_0^T\big(2\mathcal{R}(y_\varepsilon,\dot y_\varepsilon,\theta_\varepsilon) + \varepsilon\|\nabla\dot y_\varepsilon\|^2_{L^2}\big)\,\mathrm{d}t.$$

Together with (5.11) we conclude that we must have equality in all three cases after "$\Longrightarrow$". However, $\nabla\dot{y}_{\varepsilon\tau} \rightharpoonup \nabla\dot{y}_\varepsilon$ in $L^2(Q;\mathbb{R}^{d\times d})$ and

$$I_\varepsilon^{(3)} = \liminf_{\tau\to 0}(\varepsilon - \tau\Lambda(K_\varepsilon))\|\nabla\dot{y}_{\varepsilon\tau}\|^2_{L^2(Q)} = \varepsilon\|\nabla\dot{y}_\varepsilon\|^2_{L^2(Q)}$$

imply the desired strong convergence $\nabla\dot{y}_{\varepsilon\tau} \to \nabla\dot{y}_\varepsilon$ in $L^2(Q;\mathbb{R}^{d\times d})$.

Step 5: Limit in the heat equation. We first pass to the limit $\tau \to 0$ in the constitutive relation (5.1b), namely $\overline{w}_{\varepsilon\tau} = \mathfrak{w}(\nabla\overline{y}_{\varepsilon\tau}, \overline{\theta}_{\varepsilon\tau})$. The left-hand side converges to w_ε by (5.7), while the right-hand side converges to $\mathfrak{w}(\nabla y_\varepsilon, \theta_\varepsilon)$ by the continuity of $\mathfrak{w}$, the bound (2.31) and the convergences (5.4). Thus, $w_\varepsilon = \mathfrak{w}(\nabla y_\varepsilon, \theta_\varepsilon)$ is established, i.e. (4.1c) holds.

We write the heat equation (5.1b) with boundary conditions (5.2c) in the weak form

$$\begin{aligned}&\int_Q \Big(\dot{w}_{\varepsilon\tau} z + \nabla\overline{\theta}_{\varepsilon\tau}\cdot\mathcal{K}(\nabla\underline{y}_{\varepsilon\tau}, \underline{\theta}_{\varepsilon\tau})\nabla z\Big)\,\mathrm{d}x\,\mathrm{d}t + \int_\Sigma \kappa\big(\overline{\theta}_{\varepsilon\tau} - \overline{\theta}_{\flat,\varepsilon,\tau}\big) z\,\mathrm{d}S\,\mathrm{d}t\\ &= \int_Q \big(\xi_\varepsilon^{\mathrm{reg}}(\nabla\underline{y}_{\varepsilon\tau}, \nabla\dot{y}_{\varepsilon\tau}, \underline{\theta}_{\varepsilon\tau}) + \partial_F\phi(\nabla\overline{y}_{\varepsilon\tau}, \overline{\theta}_{\varepsilon\tau}) : \nabla\dot{y}_{\varepsilon\tau}\big) z\,\mathrm{d}x\,\mathrm{d}t\end{aligned} \tag{5.12}$$

for all $z \in L^\infty(I; H^1(\Omega))$. While we only have the weak convergences $\dot{w}_{\varepsilon\tau} \rightharpoonup \dot{w}_\varepsilon$ in $L^2\big(I; H^1(\Omega)^*\big)$ (see (5.7)) and $\nabla\overline{\theta}_{\varepsilon\tau} \rightharpoonup \nabla\theta_\varepsilon$ in $L^2(Q)$ (see (5.3b)), all other functions in (5.12) converge strongly. In particular, using the strong convergences $\nabla\dot{y}_{\varepsilon\tau} \to \nabla\dot{y}_\varepsilon$ in $L^2(Q;\mathbb{R}^{d\times d})$ and $0 \le \xi_\varepsilon^{\mathrm{reg}}(\nabla\underline{y}_{\varepsilon\tau}, \nabla\dot{y}_{\varepsilon\tau}, \underline{\theta}_{\varepsilon\tau}) \le K_\varepsilon$ we obtain

$$\xi_\varepsilon^{\mathrm{reg}}(\nabla\underline{y}_{\varepsilon\tau}, \nabla\dot{y}_{\varepsilon\tau}, \underline{\theta}_{\varepsilon\tau}) \to \xi_\varepsilon^{\mathrm{reg}}(\nabla y_\varepsilon, \nabla\dot{y}_\varepsilon, \theta_\varepsilon) \quad \text{strongly in } L^p(Q) \text{ for all } p \in [1,\infty[. \tag{5.13}$$

Thus, passing to the limit $\tau \to 0$ in (5.12) leads exactly to the weak form to the regularized heat equation (4.1b) with boundary condition (4.2c).

This conclude the proof of Proposition 5.1. □

6 Limit passage $\varepsilon \to 0$

In this final step of the proof of Theorem 2.2 we have to pass to the limit with the regularization parameter $\varepsilon \to 0$. As we are already in the time-continuous setting we are now able to make the formally derived total energy balance (2.21) for $\mathcal{E}$ rigorous for all $\varepsilon > 0$. From this we will be able to derive a priori bounds for $(y_\varepsilon, \theta_\varepsilon)$ that are independent of ε.

Remark 6.1 (Missing discrete estimate for the total energy)**.** *The derivation of the total energy balance is achieved by testing the momentum balance by $\dot{y}$ and the heat equation by the constant function 1. The corresponding step on the time-discrete level would be the test* (4.3a) *by $\boldsymbol{\delta}_\tau y^k$ and* (4.3b) *by 1. We would be able to use the desirable cancellation of the dissipation, namely $\xi_\varepsilon^{reg} - \xi \le 0$; however for the coupling terms*

$$\partial_F\phi(\nabla y_{\varepsilon\tau}^k, \theta_{\varepsilon\tau}^{k-1}) : \boldsymbol{\delta}_\tau\nabla y_\varepsilon^k \quad \textit{and} \quad \partial_F\phi(\nabla y_{\varepsilon\tau}^k, \theta_{\varepsilon\tau}^k) : \boldsymbol{\delta}_\tau\nabla y_\varepsilon^k,$$

which arise from (4.3a) *and* (4.3b) *respectively, we do not have any way to estimate the first against the second. Recall that we were forced to use the explicit/forward value $\theta_{\varepsilon\tau}^k$ to maintain positivity of the temperature.*

To exploit the balance of the total energy we have to strengthen the assumption on the leading $\ell(t)$, i.e. the functions g, and f, in (2.30g), namely

$$g \in W^{1,1}(I;L^2(\Omega;\mathbb{R}^d)), \qquad f \in W^{1,1}(I;L^2(\Gamma_{\mathsf{N}};\mathbb{R}^d)). \tag{6.1}$$

This implies that $t \mapsto \ell(t)$ lies in $W^{1,1}\big(I;H^1_{\Gamma_{\mathsf{D}}}(\Omega;\mathbb{R}^d)^*\big)$, which is what we will only need.

The new ε-independent estimates on $\nabla\dot y_\varepsilon$ in $L^2(Q)$ will be obtain by exploiting the Pompe's generalized Korn's inequality (cf. [Pom03]) as prepared in Theorem 3.3 above.

Lemma 6.2 (A-priori estimates for y_ε)**.** *Let the assumptions* (2.30) *and* (6.1) *hold. Then there exists a constant K such that for all $\varepsilon \in]0,1[$ and all weak solutions $(y_\varepsilon,\theta_\varepsilon)$ of the regularized problem* (4.1)-(4.2) *obtained in Proposition 5.1 we have the a priori estimates*

Then $\det(\nabla y_\varepsilon)>0$ on Q and the following estimates hold with K independent of $\varepsilon>0$:

$$\big\|y_\varepsilon\big\|_{L^\infty(I;W^{2,p}(\Omega;\mathbb{R}^d))} \le K, \tag{6.2a}$$

$$\det\big(\nabla y_\varepsilon(t,x)\big) \ge 1/K \ \text{ for all } (t,x)\in Q, \tag{6.2b}$$

$$\big\|\theta_\varepsilon\big\|_{L^\infty(I;L^1(\Omega))} \le K, \tag{6.2c}$$

$$\big\|\nabla\dot y_\varepsilon\big\|_{L^2(Q;\mathbb{R}^{d\times d})} \le K, \tag{6.2d}$$

$$\int_Q \xi(\nabla y_\varepsilon,\nabla\dot y_\varepsilon,\theta_\varepsilon)\,\mathrm{d}x\,\mathrm{d}t \le K, \tag{6.2e}$$

with q from (2.30a)*, where again* $\mathrm{sym}(\cdot)$ *denotes the symmetric part of a $(d\times d)$-matrix.*

Proof. We proceed in two steps that are close to estimates we have done in the time-discrete setting.

Step 1: Estimate for $\mathcal{E}(y_\varepsilon,\theta_\varepsilon)$. Using the derived regularity for the solution $(y_\varepsilon,\theta_\varepsilon)$ we see that a suitable variant of the total energy balance (2.21) holds. To be specific, we start from (5.9), which is also valid for arbitrary $t\in]0,T]$ in place of T, and add the time-integrated version of (4.1b) tested with the constant function $z\equiv 1$. Using $\mathcal{E}=\mathcal{M}+\mathcal{W}$ with $\mathcal{W}(y_\varepsilon,\theta_\varepsilon)=\int_\Omega w_\varepsilon\,\mathrm{d}x$ we find

$$\begin{aligned}&\mathcal{E}(y_\varepsilon(t),\theta_\varepsilon(t)) + \int_0^t\int_\Omega \Big(\xi(\nabla y_\varepsilon,\nabla\dot y_\varepsilon,\theta_\varepsilon)+\varepsilon|\nabla\dot y_\varepsilon|^2 - \xi_\varepsilon^{\mathsf{reg}}(\nabla y_\varepsilon,\nabla\dot y_\varepsilon,\theta_\varepsilon)\Big)\,\mathrm{d}x\,\mathrm{d}s\\ &= \mathcal{E}(y_\varepsilon(0),\theta_\varepsilon(0)) + \int_0^t\langle \ell(s),\dot y_\varepsilon(s)\rangle\,\mathrm{d}s + \int_0^t\int_\Gamma \kappa(\theta_\flat^\varepsilon-\theta_\varepsilon)\,\mathrm{d}S\,\mathrm{d}s.\end{aligned}$$

The importance is the cancellation of the term $\partial_F\phi:\nabla\dot y_\varepsilon$ and that the difference of the dissipation integrals has a sign.

Defining the auxiliary variable $E_\varepsilon(t):=\mathcal{E}(y_\varepsilon(t),\theta_\varepsilon(t))-\langle\ell(t),y_\varepsilon(t)\rangle$ and using $0\le\theta_\flat^\varepsilon\le\theta_\flat$ and $\theta_\varepsilon\ge 0$ gives

$$E_\varepsilon(t)\le E_\varepsilon(0)+\int_0^t\Big(\int_\Gamma \kappa\theta_\flat\,\mathrm{d}S-\langle\dot\ell(s),y_\varepsilon(s)\rangle\Big)\,\mathrm{d}s,$$

where we have integrated by parts the power of the external loadings, which was possible by the strengthened assumption (6.1).

With $\mathcal{E}\ge\mathcal{M}\ge\mathcal{H}$ and the coercivity of $\mathcal{H}$ we have $\|y\|_{H^1}\le c_1+c_2\mathcal{E}(y,\theta)$ and obtain

$$E_\varepsilon(t)\le E_\varepsilon(0)+\int_0^t\big(a(s)+b(s)E_\varepsilon(s)\big)\,\mathrm{d}s \ \text{ with } a,\ b\ge 0$$

and $a, b \in L^1(0,T)$, which follows from (6.1) for ℓ and (2.30i) for $\theta_\flat$. With $B(t) = \int_0^t b(s)\,\mathrm{d}s$ and $A(t) = \int_0^t a(s)\,\mathrm{d}s$ the Gronwall estimate yields the a priori estimate

$$\mathcal{E}_\varepsilon(t) \le \mathrm{e}^{B(t)}\big(E_\varepsilon(0) + A(t)\big) \le \mathrm{e}^{B(T)}\big(E^0 + A(T)\big) := M_1,$$

where we used $\mathcal{E}_\varepsilon(0) = \mathcal{E}(y_\varepsilon(0), \theta_\varepsilon(0)) \le \mathcal{E}(y_0, \theta_0) - \langle \ell(0), y_0\rangle =: E^0 < \infty$ by (2.30h), (2.30i), and (2.31). This immediately implies

$$\mathcal{M}(y_\varepsilon(t)) + \hat\epsilon\|\theta_\varepsilon(t)\|_{L^1(\Omega)} \le \mathcal{E}_\varepsilon(y_\varepsilon(t), \theta_\varepsilon(t)) \le M_2.$$

Hence, (6.2c) is established, whereas (6.2a) and (6.2b) follow by applying Theorem 3.1.

Step 2: Estimate for the strain rate $\nabla\dot y_\varepsilon$. We return to the mechanical energy balance (5.9) on the interval $I = [0,T]$. We recall that the dissipation function $\xi(F, \dot F, \theta)$ is assumed to control the symmetric part of $F^\top \dot F$ only, namely

$$\xi(F, \dot F, \theta) = 2\hat\zeta(F^\top F, F^T\dot F + \dot F^\top F, \theta) \ge \alpha|F^T\dot F + \dot F^\top F|^2.$$

Using our a priori bounds on $\mathcal{M}(y_\varepsilon(t))$, we can apply the generalized Korn's inequality a prepared in Corollary 3.4 with $y = y_\varepsilon(t,\cdot)$ and $v = \dot y_\varepsilon(t) \in H^1_{\Gamma_\mathrm{D}}(\Omega;\mathbb{R}^d)$ to obtain

$$\begin{aligned}
\alpha\bar c_K \int_0^T \|y_\varepsilon(t)\|_{H^1}^2\,\mathrm{d}t &\le \int_Q \alpha\big|\nabla y_\varepsilon^\top \nabla\dot y_\varepsilon + \nabla\dot y_\varepsilon^\top \nabla y_\varepsilon\big|^2\,\mathrm{d}x\,\mathrm{d}t \le \int_Q \xi(\nabla y_\varepsilon, \nabla\dot y_\varepsilon, \theta_\varepsilon)\,\mathrm{d}x\,\mathrm{d}t \\
&\le \mathcal{M}(y_0) - \mathcal{M}(y_\varepsilon(T)) + \int_0^T \big(\|\ell(t)\|_{(H^1)^*} + \|\partial_F\phi(\nabla y_\varepsilon, \theta_\varepsilon)\|_{L^\infty(Q)}\big)\|\dot y_\varepsilon(t)\|_{H^1}\,\mathrm{d}t,
\end{aligned}$$

where we used $|\partial_F\phi(F,\theta)|C(1+|F|)^s$ and $|\nabla y_\varepsilon(t,x)| \le K$, which follows from (6.2a). From this, (6.2d) and (6.2e) follow immediately. □

For the deformation y_ε we have all the estimates we need for passing to the limit. But we still need good a priori estimates for the temperature. Here the problem arises that the heating arising through the viscous dissipation $\xi(\nabla y_\varepsilon, \nabla\dot y_\varepsilon, \theta_\varepsilon)$ is only bounded in $L^1(Q)$. So, obtaining improved estimates we have to invoke special test functions developed by Boccardo and Gallouët [BoG89] for parabolic equations with measure-valued right-hand sides.

Proposition 6.3 (A priori estimates for θ_ε and w_ε)**.** *Under the conditions of Lemma 6.2, also the following estimates hold:*

$$\forall\, p \in \big[1, \tfrac{d+2}{d}\big[\ \exists\, C_p > 0\ \forall\,\varepsilon \in\,]0,1] : \quad \|\theta_\varepsilon\|_{L^p(Q)} + \|w_\varepsilon\|_{L^p(Q)} \le C_p, \tag{6.3a}$$

$$\forall\, r \in \big[1, \tfrac{d+2}{d+1}\big[\ \exists\, K_r > 0\ \forall\,\varepsilon \in\,]0,1] : \quad \|\nabla\theta_\varepsilon\|_{L^p(Q)} + \|\nabla w_\varepsilon\|_{L^p(Q)} \le K_r, \tag{6.3b}$$

$$\exists\, K > 0\ \forall\,\varepsilon \in\,]0,1[: \quad \big\|\dot w_\varepsilon\big\|_{L^1(I;H^{(d+3)/2}(\Omega)^*)} \le K. \tag{6.3c}$$

Proof. We follow the recipe in [BoG89] in the simplified variant of [FeM06], see also [MiN18]. For $\eta \in\,]0,1[$ we define the function $\chi_\eta : \mathbb{R}^+ \to \mathbb{R}^+$ via

$$\chi_\eta(0) = 0 \quad \text{and} \quad \chi_\eta'(w) := 1 - \frac{1}{(1+w)^\eta} \in [0,1].$$

Clearly, χ_η satisfies $\min\{0, w/2 - C_\eta\} \le \chi_\eta(w) \le w$ and $\chi_\eta''(w) = \dfrac{\eta}{(1+w)^{1+\eta}} > 0$.

Now testing (4.1b) with the test function $z = \chi_\eta' \circ w_\varepsilon$ amounts to applying the chain rule in Proposition 3.5 to the convex functional $\mathcal{J}(w) = \int_\Omega \chi_\eta(w(x))\,\mathrm{d}x$ on the space $X = H^1(\Omega)^*$. Indeed, from (5.3) and $w_\varepsilon = \mathfrak{w}(\nabla y_\varepsilon, \theta_\varepsilon)$ we have $w_\varepsilon \in L^2(I; H^1(\Omega)) \cap H^1(I; H^1(\Omega)^*))$, and the chain rule gives the first identity in the following calculation:

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \chi_\eta(w_\varepsilon)\,\mathrm{d}x = \int_\Omega \chi_\eta'(w_\varepsilon)\dot w_\varepsilon\,\mathrm{d}x\\
&= -\int_\Omega \chi_\eta''(w_\varepsilon)\nabla w_\varepsilon\cdot\mathcal{K}(\nabla y_\varepsilon,\theta_\varepsilon)\nabla\theta_\varepsilon\,\mathrm{d}x + \int_\Gamma \kappa(\theta_\flat^\varepsilon-\theta_\varepsilon)\,\mathrm{d}S\\
&\quad + \int_\Omega \chi_\eta'(w_\varepsilon)\Big(\xi_\varepsilon^{\mathrm{reg}}(\nabla y_\varepsilon,\nabla\dot y_\varepsilon,\theta_\varepsilon) + \partial_F\phi(\nabla y_\varepsilon,\theta_\varepsilon){:}\nabla\dot y_\varepsilon\Big)\,\mathrm{d}x.
\end{aligned}$$

Integration over $t\in I=[0,T]$ and using $\chi_\eta'(w)\in[0,1]$ and $\|\nabla y_\varepsilon\|_{L^\infty(Q)}\le K_\infty$ yield

$$\begin{aligned}
&\int_Q \chi_\eta''(w_\varepsilon)\nabla w_\varepsilon\cdot\mathcal{K}(\nabla y_\varepsilon,\theta_\varepsilon)\nabla\theta_\varepsilon\,\mathrm{d}x\,\mathrm{d}t\\
&\le \int_\Omega \chi_\eta(w_0)\,\mathrm{d}x + \int_\Sigma \kappa\theta_\flat\,\mathrm{d}S\,\mathrm{d}t + \int_Q\Big(\xi(\cdots) + C(1{+}K_\infty)^s|\nabla\dot y_\varepsilon|\Big)\,\mathrm{d}x\,\mathrm{d}t \le C,
\end{aligned}\tag{6.4}$$

where we used (2.30h), (2.30i), (6.2d), and (6.2e).

From this, we derive an a priori bound on ∇w_ε by setting $\widetilde{\mathcal{K}}_\varepsilon = \mathcal{K}(\nabla y_\varepsilon,\theta_\varepsilon)$ and estimate it as in (4.18) (see Step 3 of the proof of Proposition 4.2) by

$$|\widetilde{\mathcal{K}}_\varepsilon(t,x)|\le\varkappa \quad\text{and}\quad a\cdot\widetilde{\mathcal{K}}_\varepsilon(t,x)a\ge\frac{1}{\varkappa}|a|^2,$$

where $\varkappa$ is now independent of ε because of the ε-independent bound in (6.2a) and (6.2b). Moreover, ∇w_ε and $\nabla\theta_\varepsilon$ are related by

$$\nabla w_\varepsilon = \partial_\theta\mathfrak{w}(\nabla y_\varepsilon,\theta_\varepsilon)\nabla\theta_\varepsilon + \partial_F\mathfrak{w}(\nabla y_\varepsilon,\theta_\varepsilon):\nabla^2 y_\varepsilon.\tag{6.5}$$

With $\partial_\theta w(F,\theta) = -\theta\partial_\theta^2\phi(F,\theta)\le K$ we obtain

$$\begin{aligned}
\frac{1}{\varkappa}|\nabla w_\varepsilon|^2 &\le \nabla w_\varepsilon\cdot\widetilde{\mathcal{K}}_\varepsilon\nabla w_\varepsilon\\
&= \partial_\theta\mathfrak{w}(\nabla y_\varepsilon,\theta_\varepsilon)\,\nabla w_\varepsilon\cdot\widetilde{\mathcal{K}}_\varepsilon\nabla\theta_\varepsilon + \nabla w_\varepsilon\cdot\widetilde{\mathcal{K}}_\varepsilon\partial_F\mathfrak{w}(\nabla y_\varepsilon,\theta_\varepsilon):\nabla^2 y_\varepsilon\\
&\le K\nabla w_\varepsilon\cdot\widetilde{\mathcal{K}}_\varepsilon\nabla\theta_\varepsilon + \varkappa|\nabla w_\varepsilon|\,C(1{+}K_\infty)^s|\nabla^2 y_\varepsilon|\\
&\le K\nabla w_\varepsilon\cdot\widetilde{\mathcal{K}}_\varepsilon\nabla\theta_\varepsilon + \frac{1}{2\varkappa}|\nabla w_\varepsilon|^2 + C_*|\nabla^2 y_\varepsilon|^2.
\end{aligned}$$

Canceling $\frac{1}{2\varkappa}|\nabla w_\varepsilon|^2$, multiplying by $\chi''(w_\varepsilon)\in[0,1]$, and integrating over Q we employ (6.4) and arrive at

$$\frac{1}{\varkappa K}\int_Q \chi_\eta''(w_\varepsilon)|\nabla w_\varepsilon|^2\,\mathrm{d}x\,\mathrm{d}t \le \int_Q \chi''(w_\varepsilon)\Big(\nabla w_\varepsilon\cdot\widetilde{\mathcal{K}}_\varepsilon\nabla\theta_\varepsilon + C_*|\nabla^2 y_\varepsilon|^2\Big)\,\mathrm{d}x\,\mathrm{d}t\le C_3,$$

where the last integrand is bounded by (6.2a) and $p\ge 2$.

For $r\in[1,2[$ we set $p=2/(2{-}r)$, $p'=2/r$, and $q=(1{+}\eta)r/2$ and employ Hölder's estimate to obtain

$$\begin{aligned}
\|\nabla w_\varepsilon\|_{L^r(Q)}^r &= \int_Q (1{+}w_\varepsilon)^q\,\frac{|\nabla w_\varepsilon|^r}{(1{+}w_\varepsilon)^q}\,\mathrm{d}x\,\mathrm{d}t \;\le\; \|(1{+}w_\varepsilon)^q\|_{L^p(Q)}\Big\|\frac{|\nabla w_\varepsilon|^r}{(1{+}w_\varepsilon)^q}\Big\|_{L^{p'}(Q)}\\
&= \|1{+}w_\varepsilon\|_{L^{qp}(Q)}^q\Big(\int_Q\frac{|\nabla w_\varepsilon|^2}{(1{+}w_\varepsilon)^{1+\eta}}\,\mathrm{d}x\,\mathrm{d}t\Big)^{1/p'} \;\le\; \|1{+}w_\varepsilon\|_{L^{qp}(Q)}^q\big(\varkappa K C_3/\eta\big)^{1/p'},
\end{aligned}\tag{6.6}$$

where crucially relied on $p' = 2/r$, $\chi''(w) = \eta/(1+w)^{1+\eta}$, and the previous estimate. Using the a priori estimate $\|1+w_\varepsilon\|_{L^\infty(I;L^1(\Omega))} \le T|\Omega| + K =: K_1$ from (6.2c) we can now use the anisotropic Gagliardo-Nirenberg interpolation (see e.g. [MiN18, Lem. 4.2]) giving

$$\|1+w_\varepsilon\|_{L^{r/\lambda}(Q)} \le C\|1+w_\varepsilon\|^{1-\lambda}_{L^\infty(I;L^1(\Omega))}\big(\|1+w_\varepsilon\|_{L^\infty(I;L^1(\Omega))} + \|\nabla w_\varepsilon\|_{L^r(Q)}\big)^\lambda \text{ with } \lambda = \frac{d}{d+1}.$$

For inserting this into (6.6) we need $qp \le r/\lambda$ which gives the restriction $r \le 2 - (1+\eta)\lambda$.

Thus, for all $r \in [1, (d+2)/(d+1)[$ we find an $\eta = \eta_r \in]0, 1[$ such that the above estimates give

$$\|\nabla w_\varepsilon\|^r_{L^r(Q)} \le C_r\left(1 + \|\nabla w_\varepsilon\|^{q\lambda}_{L^r(Q)}\right),$$

and $q_r\lambda < q_r = (1+\eta_r)r/2 < r$ provide $\|\nabla w_\varepsilon\|_{L^r(Q)} \le K_r$. Using (6.5) and $\partial_\theta \mathfrak{w} \ge \hat\epsilon > 0$ we easily find $\|\nabla\theta_\varepsilon\|_{L^r(Q)} \le K_r$ and (6.3b) is established.

Applying the Gagliardo-Nirenberg interpolation once again gives assertion (6.3a).

Eventually, the a priori estimate (6.3c) is obtained estimating all other terms in (4.1b), when realizing that always $H^{(d+3)/2}(\Omega) \subset W^{1,\infty}(\Omega)$. □

We are now in the position to pass to the limit $\varepsilon \to 0$ in the regularized system (4.1)-(4.2), and thus provide the proof of our main existence result presented in Theorem 2.2. The approach is close to the convergence result presented in Proposition 5.1: first we extract converging subsequences and then pass to the limit in the mechanical momentum balance. This also provides the necessary strong convergence of the the strain rates that is needed to eventually pass to the limit in the heat heat equation.

Proposition 6.4 (Convergence for $\varepsilon \to 0$)**.** *Let again* (2.30) *and* (6.1) *hold. Then, considering the sequence of time steps $\varepsilon \to 0$, there is a subsequence $(y_\varepsilon, \theta_\varepsilon)$ of weak solutions to the regularized system* (4.1)-(4.2) *obtained in Proposition 5.1 such that, for some (y, θ), it holds*

$$y_\varepsilon \to y \qquad \text{weakly* in } L^\infty(I; W^{2,p}(\Omega;\mathbb{R}^d)) \cap H^1(I; L^2(\Omega;\mathbb{R}^d)) \text{ and} \tag{6.7a}$$

$$\theta_\varepsilon \to \theta \qquad \text{weakly in } L^r(I; W^{1,r}(\Omega)) \text{ for all } 1 \le r < (d+2)/(d+1). \tag{6.7b}$$

Moreover, every couple (y, θ) obtained in such a way is a weak solution, according Definition 2.1, of the boundary-value problem (2.13)–(2.14) *satisfying the initial values* (2.26)*.*

Proof. The proof follows the lines of the proof of Proposition 5.1, so we do not repeat all details of the arguments.

Step 1: Extraction of converging subsequences. Using the a priori estimates (6.2) and (6.3), Banach's selection principle allows us to choose a subsequence and some (y, θ) such that (6.7) holds. By the Aubin-Lions' theorem interpolated with the estimates (4.9a) and (4.9c), we have also

$$\nabla y_\varepsilon \to \nabla y \qquad \text{strongly in } L^\infty(Q;\mathbb{R}^{d\times d}) \text{ and} \tag{6.8a}$$

$$w_\varepsilon \to w \qquad \text{strongly in } L^p(Q) \text{ with any } 1 \le p < 1 + 2/d, \tag{6.8b}$$

$$\theta_\varepsilon \to \theta \qquad \text{strongly in } L^p(Q) \text{ with any } 1 \le p < 1 + 2/d. \tag{6.8c}$$

The proof of (6.8a) is similar to (5.4a). For (6.8b) we proceed as for (5.4b) by using the estimates on w_ε given in (6.3). Using the relation $w_\varepsilon = \mathfrak{w}(\nabla y_\varepsilon, \theta_\varepsilon)$ we also obtain the strong convergence (6.8c).

Step 2: Convergence in the mechanical equation. The limit passage in the momentum balance (4.1a)-(4.2) works as before, again using the Minty trick (5.8). Of course, the additional regularizing viscosity term $\varepsilon\nabla\dot{y}_\varepsilon$ vanishes because of our a priori bound (6.2d):

$$\left|\int_Q \varepsilon\nabla\dot{y}_\varepsilon{:}\nabla z\,\mathrm{d}x\,\mathrm{d}t\right| \le \varepsilon\big\|\nabla\dot{y}_\varepsilon\big\|_{L^2(Q;\mathbb{R}^{d\times d})}\big\|\nabla z\big\|_{L^2(Q;\mathbb{R}^{d\times d})} = C\varepsilon \to 0.$$

Step 3: Balance of mechanical energy. As in the proof of Proposition 5.1 we derive from the property that the limit couple (y,θ) solves the mechanical equation that the following mechanical energy relation holds:

$$\mathcal{M}(y(T)) + \int_0^T 2\mathcal{R}(y,\dot{y},\theta)\,\mathrm{d}t = \mathcal{M}(y_0) + \int_0^T \langle \ell,\dot{y}\rangle\,\mathrm{d}t - \int_Q \partial_F\phi(\nabla y,\theta){:}\nabla\dot{y}\,\mathrm{d}x\,\mathrm{d}t. \tag{6.9}$$

Step 4: Strong convergence of the symmetric strain rates. We can pass to the limit $\varepsilon\to 0$ in the mechanical energy relation (5.9). Comparing the result with (6.9) we obtain

$$\lim_{\varepsilon\to0}\int_Q \xi(\nabla y_\varepsilon,\nabla\dot{y}_\varepsilon,\theta_\varepsilon)\,\mathrm{d}x\,\mathrm{d}t = \int_Q \xi(\nabla y,\nabla\dot{y},\theta)\,\mathrm{d}x\,\mathrm{d}t. \tag{6.10}$$

To conclude strong convergence we use the special form (2.10), namely $\xi(F,\dot{F},\theta) = 2\hat{\zeta}(F^\top F, F^\top\dot{F}+\dot{F}^\top F,\theta)$. From the pointwise convergence $\theta_\varepsilon\to\theta$, the uniform convergence $F_\varepsilon := \nabla y_\varepsilon \to F = \nabla y$, and the weak convergence $\dot{F}_\varepsilon := \nabla\dot{y}_\varepsilon \rightharpoonup \dot{F}$ in $L^2(Q)$ we obtain

$$V_\varepsilon := F_\varepsilon^\top\dot{F}_\varepsilon+\dot{F}_\varepsilon^\top F_\varepsilon \rightharpoonup F^\top\dot{F}+\dot{F}^\top F =: V \ \text{ in } L^2(Q).$$

With the coercive and quadratic structure of $\hat{\zeta}$ assumed in (2.30e) we proceed as follows:

$$\begin{aligned}
2\alpha\|V_\varepsilon{-}V\|_{L^2(Q)}^2 &\le \int_Q 2\hat{\zeta}(C_\varepsilon,V_\varepsilon{-}V,\theta_\varepsilon)\,\mathrm{d}x\,\mathrm{d}t\\
&= \int_Q \Big(2\hat{\zeta}(C_\varepsilon,V_\varepsilon,\theta_\varepsilon) - 2V_\varepsilon : \mathbb{D}(C_\varepsilon,\theta_\varepsilon)V + 2\hat{\zeta}(C_\varepsilon,V,\theta_\varepsilon)\Big)\,\mathrm{d}x\,\mathrm{d}t\\
&= \int_Q \Big(\xi(F_\varepsilon,\dot{F}_\varepsilon,\theta_\varepsilon) - 2V_\varepsilon : \mathbb{D}(C_\varepsilon,\theta_\varepsilon)V + \xi(F,\dot{F},\theta)\Big)\,\mathrm{d}x\,\mathrm{d}t + \delta(\varepsilon),\\
&\qquad\qquad \text{with } \delta(\varepsilon) = \int_\Omega 2V : \big(\mathbb{D}(C_\varepsilon,\theta_\varepsilon){-}\mathbb{D}(C,\theta)\big)V\,\mathrm{d}x\,\mathrm{d}t.
\end{aligned}$$

We see that the first term converges by (6.10), while the second term converges by the weak convergence $V_\varepsilon\rightharpoonup V$ and the strong convergence $\mathbb{D}(C_\varepsilon,\theta_\varepsilon)V\to\mathbb{D}(C,\theta)V$ (as $\mathbb{D}$ is bounded and the arguments converge pointwise). Similarly, $\delta(\varepsilon)\to0$ by Lebesgue's dominated convergence theorem, and thus we conclude the strong convergence $\|V_\varepsilon{-}V\|_{L^2(Q)}\to0$.

Step 5: Limit passage in the heat equation. Testing the regularized heat equation (4.1b) with boundary conditions (4.2c) by smooth function v with $V(T,\cdot)\equiv0$ we find

$$\begin{aligned}
\int_Q \Big(\nabla\theta_\varepsilon\cdot\mathcal{K}(\nabla y_\varepsilon,\theta_\varepsilon)\nabla v - \big(\xi_\varepsilon^{\mathrm{reg}}(\nabla y_\varepsilon,\nabla\dot{y}_\varepsilon,\theta_\varepsilon)+\partial_F\phi(\nabla y_\varepsilon,\theta_\varepsilon){:}\nabla\dot{y}_\varepsilon\big)v - w_\varepsilon\dot{v}\Big)\,\mathrm{d}x\,\mathrm{d}t\\
+\int_\Sigma \kappa\theta_\varepsilon v\,\mathrm{d}S\,\mathrm{d}t = \int_\Sigma \kappa\theta_{\flat,\varepsilon}v\,\mathrm{d}S\,\mathrm{d}t + \int_\Omega \mathfrak{w}(\nabla y_0,\theta_{0,\varepsilon})v(0)\,\mathrm{d}x.
\end{aligned}\tag{6.11}$$

Here the first term passes to the limit by $\nabla\theta_\varepsilon \rightharpoonup \nabla\theta$ and $\mathcal{K}(\nabla y_\varepsilon,\theta_\varepsilon)\nabla v \to \mathcal{K}(\nabla y,\theta)$. In the second term we use

$$\xi_\varepsilon^{\rm reg}(\nabla y_\varepsilon,\nabla\dot y_\varepsilon,\theta_\varepsilon)=\frac{\xi(\nabla y_\varepsilon,\nabla\dot y_\varepsilon,\theta_\varepsilon)}{1+\varepsilon\xi(\nabla y_\varepsilon,\nabla\dot y_\varepsilon,\theta_\varepsilon)}=\frac{2\hat\zeta(C_\varepsilon,V_\varepsilon,\theta_\varepsilon)}{1+2\varepsilon\hat\zeta(C_\varepsilon,V_\varepsilon,\theta_\varepsilon)}\le 2K|V_\varepsilon|^2=:g_\varepsilon.$$

Because of Step 4, we know $V_\varepsilon\to V$ strongly in $L^2(Q;\mathbb{R}^{d\times d}_{\rm sym})$. Hence, we have $g_\varepsilon\to g:=K|V|^2$ in $L^1(Q)$ and may assume, after extracting another subsequence, $V_\varepsilon(t,x)\to V(t,x)$ a.e. in Q. By the uniform/pointwise convergence of C_ε and θ_ε for any $v\in C^0(\overline{Q})$ we obtain

$$g_\varepsilon\|v\|_{L^\infty(Q)}\ge\xi_\varepsilon^{\rm reg}(\nabla y_\varepsilon,\nabla\dot y_\varepsilon,\theta_\varepsilon)v\to\xi(\nabla y,\nabla\dot y,\theta)v\quad\text{a.e. in }Q.$$

As the majorants $g_\varepsilon\|v\|_{L^\infty(Q)}$ converge to $g\|v\|_{L^\infty(Q)}$ in $L^1(Q)$ the generalized dominated convergence theorem implies convergence of the second term in (6.11).

In the third term we have weak convergence of $\nabla\dot y_\varepsilon$ and strong convergence of $v\partial_F\phi(\nabla y_\varepsilon,\theta_\varepsilon)$. Similarly, the remaining four terms converge to the desired limits. Thus, we have shown that (y,θ) satisfy (2.27b), which finishes the proof of Proposition 6.4. □

Remark 6.5 (*Strong convergence of $y_{\varepsilon\tau}$ and y_ε*)**.** Strengthening monotonicity of $\mathfrak{h}_{\rm el}$, cf. (2.30d), for the strict monotonicity

$$\forall\, G_1,G_2\in\mathbb{R}^{d\times d\times d}:\qquad(\mathfrak{h}_{\rm el}(G_1)-\mathfrak{h}_{\rm el}(G_2))\vdots(G_1-G_2)\ge c_0|G_1-G_2|^p,$$

we use the argumentation after (5.11) to show $\overline{y}_{\varepsilon\tau}(t)\to y_\varepsilon(t)$ strongly in $W^{2,p}(\Omega;\mathbb{R}^d)$ for all $t\in[0,T]$. Similarly, in Proposition 6.4 one can show $y_\varepsilon(t)\to y(t)$ strongly in $W^{2,p}(\Omega;\mathbb{R}^d)$. Together with the L^∞-estimate (4.9a), we can also strengthen the weak* convergence (5.3a) in $L^\infty(I;W^{2,p}(\Omega;\mathbb{R}^d))$ to a strong convergence in $L^q(I;W^{2,p}(\Omega;\mathbb{R}^d))$ for all $q\in[1,\infty[$. The same applies to (6.7a).

Remark 6.6 (*Dynamical problems*)**.** Introducing the *kinetic energy* $\frac12\varrho|\dot y|^2$ with a mass density $\varrho=\varrho(x)>0$ leads to an *inertial force* $\varrho\ddot y$ in the momentum equation (2.13a), which would make the nonlinear problem hyperbolic. It is generally recognized as analytically very troublesome. Here, it would work for isothermal situation like in Corollary 2.3 if we would be able to work with weak convergence, i.e. $\mathscr{H}$ needs to be quadratic ($p=2$). Staying with $\mathcal{H}$ depending on the second gradient $\nabla^2 y$ we would be forced to give up the determinant constraint $\det\nabla y>0$, which is indeed possible if heat conduction is not considered. Alternatively, one may take $\mathcal{H}$ quadratic but coercive in Hilbert space norms $H^s(\Omega)$ with $s>1+d/2$, such that $H^s(\Omega)$ still embeds into $C^{1,\alpha}$ for some $\alpha>0$, cf. also [KrR19, Ch. 9.3]. In the anisothermal situation, it seems difficult to ensure that the acceleration $\ddot y\in L^2(I;H^{1+\kappa}(\Omega;\mathbb{R}^{d\times d})$ stays in duality with the velocity $\dot y$. The regularity seems difficult and the higher-order viscosity is inevitably very nonlinear to comply with frame-indifference while the corresponding generalization of Korn's inequality does not seem available.

Remark 6.7 (*Other transport processes: flow in porous media*)**.** Beside heat transport, one can also consider other transport processes in a similar way. The transport coefficients can be pulled back as in (2.24). For example, considering mass transport for a concentration c one has to make the free energy ψ also c-dependent and to augmenting it by a capillarity-like gradient term $\frac12\varkappa|\nabla c|^2$. The dissipation potential $\mathcal{R}$ will then be augmented by the nonlocal term $\frac12\|\mathcal{M}(\nabla y,c)^{1/2}\nabla\Delta^{-1}_{\mathcal{M}(\nabla y,c)}\dot c\|^2_{L^2(\Omega)}$ with $\Delta^{-1}_{\mathcal{M}}:f\mapsto\mu$ denoting the linear operator $H^1(\Omega)^*\to H^1(\Omega)$ defined by the weak solution μ to the equation $\operatorname{div}(\mathcal{M}\nabla\mu)=f$. Considering the mobility tensor $\mathbb{M}=\mathbb{M}(x,c)$, we can define the

pulled-back tensor $\mathcal{M}(x,F,c) := (\mathrm{Cof}\, F^\top)\mathbb{M}(x,c)\,\mathrm{Cof}\, F/\det F$ and augment the system for the diffusion equation of the Cahn-Hilliard type:

$$\mathrm{div}\,\big(\sigma_{\rm vi}(\nabla y,\nabla\dot y,\theta)+\partial_F\psi(F,c,\theta)-\mathrm{div}\,\mathfrak{h}_{\rm el}(\nabla^2 y)\big)+g=0\,, \tag{6.12a}$$

$$\dot c-\mathrm{div}\,\big(\mathcal{M}(\nabla y,c)\nabla\mu\big)=0\quad\text{with}\quad \mu=\partial_c\psi(\nabla y,c,\theta)-\varkappa\Delta c, \tag{6.12b}$$

$$\begin{aligned} c_{\rm v}(\nabla y,c,\theta)\dot\theta-\mathrm{div}\,\big(\mathcal{K}(\nabla y,\theta)\nabla\theta\big)&=\xi(\nabla y,\nabla\dot y,\theta)\\ &\quad+\theta\partial^2_{F\theta}\psi(\nabla y,c,\theta){:}\nabla\dot y+\nabla\mu\cdot\mathcal{M}(x,\nabla y,c)\nabla\mu \end{aligned} \tag{6.12c}$$

with $\sigma_{\rm vi}$ as in (2.13a), $c_{\rm v}(F,c,\theta)=-\theta\partial^2_{\theta\theta}\psi(F,c,\theta)$, and ξ from (2.10). In (6.12b), the variable μ is called a *chemical potential*. One can also augment the model by some inelastic (plastic or creep-type) strain like in [RoS18] where also the inertial forces have been involved and the viscosity ignored but the concept of small elastic strains imposed as a modeling assumption.

References

[Ant98] S. S. ANTMAN. Physically unacceptable viscous stresses. *Zeitschrift angew. Math. Physik*, 49, 980–988, 1998.

[BaC11] J. M. BALL and E. C. M. CROOKS. Local minimizers and planar interfaces in a phase-transition model with interfacial energy. *Calc. Var.*, 40, 501–538, 2011.

[Bal77] J. BALL. Constitutive inequalities and existence theorems in nonlinear elastostatics. In *Nonlinear Analysis and Mechanics: Heriot-Watt Symposium (Edinburgh, 1976), Vol. I*, pages 187–241. Res. Notes in Math., No. 17. Pitman, London, 1977.

[Bal02] J. M. BALL. Some open problems in elasticity. In P. Newton, P. Holmes, and A. Weinstein, editors, *Geometry, Mechanics, and Dynamics*, pages 3–59. Springer, New York, 2002.

[Bal10] J. M. BALL. Progress and puzzles in nonlinear elasticity. In J. Schröder and P. Neff, editors, *Poly-, Quasi- and Rank-One Convexity in Applied Mechanics*, volume 516 of *CISM Intl. Centre for Mech. Sci.*, pages 1–15. Springer, Wien, 2010.

[Bar10] V. BARBU. *Nonlinear Differential Equations of Monotone Types in Banach Spaces*. Springer, New York, 2010.

[Bat76] R. C. BATRA. Thermodynamics of non-simple elastic materials. *J. Elasticity*, 6, 451–456, 1976.

[BD*97] L. BOCCARDO, A. DALL'AGLIO, T. GALLOUËT, and L. ORSINA. Nonlinear parabolic equations with measure data. *J. Funct. Anal.*, 147, 237–258, 1997.

[Bet86] D. E. BETOUNES. Kinematics of submanifolds and the mean curvature normal. *Arch. Rational Mech. Anal.*, 96, 1–27, 1986.

[BlG00] D. BLANCHARD and O. GUIBÉ. Existence of a solution for a nonlinear system in thermoviscoelasticity. *Adv. Diff. Eq.*, 5, 1221–1252, 2000.

[BoB03] E. BONETTI and G. BONFANTI. Existence and uniqueness of the solution to a 3D thermoelastic system. *Electronic J. Diff. Eqs.*, 50, 1–15, 2003.

[BoG89] L. BOCCARDO and T. GALLOUËT. Non-linear elliptic and parabolic equations involving measure data. *J. Funct. Anal.*, 87, 149–169, 1989.

[Bré73] H. BRÉZIS. *Operateur Maximaux Monotones et Semi-groupes de Contractions dans les Espaces de Hilbert*. North-Holland, Amsterdam, 1973.

[Daf82] C. M. DAFERMOS. Global smooth solutions to the initial boundary value problem for the equations of one-dimensional thermoviscoelasticity. *SIAM J. Math. Anal.*, 13, 397–408, 1982.

[DaL10] G. DAL MASO and G. LAZZARONI. Quasistatic crack growth in finite elasticity with non-interpenetration. *Ann. Inst. H. Poinc. Anal. Non Lin.*, 27(1), 257–290, 2010.

[Dem00] S. DEMOULINI. Weak solutions for a class of nonlinear systems of viscoelasticity. *Arch. Rat. Mech. Anal.*, 155, 299–334, 2000.

[DSF10] F. P. DUDA, A. C. SOUZA, and E. FRIED. A theory for species migration in a finitely strained solid with application to polymer network swelling. *J. Mech. Phys. Solids*, 58, 515–529, 2010.

[DST01] S. DEMOULINI, D. STUART, and A. TZAVARAS. A variational approximation scheme for three dimensional elastodynamics with polyconvex energy. *Arch. Rat. Mech. Anal.*, 157, 325–344, 2001.

[FeM06] E. FEIREISL and J. MÁLEK. On the Navier-Stokes equations with temperature-dependent transport coefficients. *Diff. Equations Nonlin. Mech.*, pages 14pp.(electronic), Art.ID 90616, 2006.

[FoL07] I. FONSECA and G. LEONI. *Modern Methods in the Calculus of Variations: L^p spaces*. Springer, 2007.

[FrG06] E. FRIED and M. E. GURTIN. Tractions, balances, and boundary conditions for nonsimple materials with application to liquid flow at small-lenght scales. *Arch. Rational Mech. Anal.*, 182, 513–554, 2006.

[FrK18] M. FRIEDRICH and M. KRUŽÍK. On the passage from nonlinear to linearized viscoelasticity. *SIAM J. Math. Anal.*, 50, 4426–4456, 2018.

[GoS93] S. GOVINDJEE and J. C. SIMO. Coupled stress-diffusion: case II. *J. Mech. Phys. Solids*, 41, 863–887, 1993.

[HeK09] T. J. HEALEY and S. KRÖMER. Injective weak solutions in second-gradient nonlinear elasticity. *ESAIM: Control, Optim. & Cal. Var.*, 15, 863–871, 2009.

[KPS19] M. KRUŽÍK, P. PELECH, and A. SCHLÖMERKEMPER. Gradient polyconvexity in evolutionary models of shape-memory alloys. *J. Optim. Theo. Appl.*, 2019. Online first.

[KrR19] M. KRUŽÍK and T. ROUBÍČEK. *Mathematical Methods in Continuum Mechanics of Solids*. Springer, Switzerland, 2019.

[LeM13] M. LEWICKA and P. B. MUCHA. A local existence result for system of viscoelasticity with physical viscosity. *Evolution Equations & Control Theory*, 2, 337–353, 2013.

[MaM09] A. MAINIK and A. MIELKE. Global existence for rate-independent gradient plasticity at finite strain. *J. Nonlin. Science*, 19(3), 221–248, 2009.

[MiE68] R. MINDLIN and N. ESHEL. On first strain-gradient theories in linear elasticity. *Intl. J. Solid Structures*, 4, 109–124, 1968.

[Mie11] A. MIELKE. Formulation of thermoelastic dissipative material behavior using GENERIC. *Contin. Mech. Thermodyn.*, 23(3), 233–256, 2011.

[Mie13] A. MIELKE. Thermomechanical modeling of energy-reaction-diffusion systems, including bulk-interface interactions. *Discr. Contin. Dynam. Systems Ser. S*, 6(2), 479–499, 2013.

[MiM18] A. MIELKE and M. MITTNENZWEIG. Convergence to equilibrium in energy-reaction-diffusion systems using vector-valued functional inequalities. *J. Nonlin. Science*, 28(2), 765–806, 2018.

[MiN18] A. MIELKE and J. NAUMANN. On the existence of global-in-time weak solutions and scaling laws for Kolmogorov's two-equation model of turbulence. *Arch. Rational Mech. Anal.*, 2018. Submitted. WIAS preprint 2545, arXiv:1801.02039.

[MiR15] A. MIELKE and T. ROUBÍČEK. *Rate-Independent Systems – Theory and Application*. Springer, New York, 2015.

[MiR16] A. MIELKE and T. ROUBÍČEK. Rate-independent elastoplasticity at finite strains and its numerical approximation. *Math. Models Meth. Appl. Sci.*, 6, 2203–2236, 2016.

[MOS13] A. MIELKE, C. ORTNER, and Y. SENGÜL. An approach to nonlinear viscoelasticity via metric gradient flows. *SIAM J. Math. Anal.*, 46, 1317–1347, 2013.

[MRS13] A. MIELKE, R. ROSSI, and G. SAVARÉ. Nonsmooth analysis of doubly nonlinear evolution equations. *Calc. Var PDE*, 46(1-2), 253–310, 2013.

[MRS18] A. MIELKE, R. ROSSI, and G. SAVARÉ. Global existence results for viscoplasticity at finite strain. *Arch. Rational Mech. Anal.*, 227(1), 423–475, 2018.

[Nef02] P. NEFF. On Korn's first inequality with non-constant coefficients. *Proc. Royal Soc. Edinburgh*, 132A, 221–243, 2002.

[Pod02] P. PODIO-GUIDUGLI. Contact interactions, stress, and material symmetry, for nonsimple elastic materials. *Theor. Appl. Mech.*, 28-29, 261–276, 2002.

[Pom03] W. POMPE. Korn's First Inequality with variable coefficients and its generalization. *Comment. Math. Univ. Carolinae*, 44, 57–70, 2003.

[RoS18] T. ROUBÍČEK and U. STEFANELLI. Thermodynamics of elastoplastic porous rocks at large strains, earthquakes, and seismic waves. *SIAM J. Appl. Math.*, 78, 2597–2625, 2018.

[Rou04] T. ROUBÍČEK. Models of microstructure evolution in shape memory alloys. In P. P. Castaneda, J. Telega, and B. Gambin, editors, *Nonlin. Homogenization and its Appl. to Composites, Polycrystals and Smart Mater.*, pages 269–304. Kluwer, Dordrecht, 2004.

[Rou09] T. ROUBÍČEK. Thermo-visco-elasticity at small strains with L^1-data. *Quarterly Appl. Math.*, 67, 47–71, 2009.

[Rou13] T. ROUBÍČEK. *Nonlinear Partial Differential Equations with Applications*. Birkhäuser, Basel, 2nd edition, 2013.

[Šil85] M. ŠILHAVÝ. Phase transitions in non-simple bodies. *Arch. Rat. Mech. Anal.*, 88, 135–161, 1985.

[Ste15] P. STEINMANN. *Geometric Foundations of Continuum Mechanics*. Springer, 2015. LAMM Vol. 2.

[Tou62] R. TOUPIN. Elastic materials with couple stresses. *Arch. Rat. Mech. Anal.*, 11, 385–414, 1962.

[TrA86] N. TRIANTAFYLLIDIS and E. AIFANTIS. A gradient approach to localization of deformation. I. Hyperelastic materials. *J. Elast.*, 16, 225–237, 1986.

[Tve08] B. TVEDT. Quasilinear equations of viscoelasticity of strain-rate type. *Arch. Rational Mech. Anal.*, 189, 237–281, 2008.

[Vis96] A. VISINTIN. *Models of Phase Transitions*. Birkhäuser, Boston, 1996.