A robust model of pseudoelasticity in shape memory alloys*

S. Stupkiewicz†, H. Petryk

Institute of Fundamental Technological Research, Polish Academy of Sciences
Pawińskiego 5B, 02-106 Warsaw, Poland

Abstract

A model of pseudoelasticity in shape memory alloys is developed within the incremental energy minimization framework. Three constitutive functions are involved: the Helmholtz free energy and rate-independent dissipation that enter incrementally the minimized energy function, and the constraint function that defines the limit transformation strains. The proposed implementation is based on a unified augmented Lagrangian treatment of both the constitutive constraints and non-smooth dissipation function. A methodology for easy reformulation of the model from the small-strain to finite-deformation regime is presented. Finite-element computations demonstrate robustness of the finite-strain version of the model and illustrate the effects of tension–compression asymmetry and transversal isotropy of the surface of limit transformation strains.

Keywords: Shape memory alloys (SMA); Phase transformation; Energy methods; Finite element method; Augmented Lagrangian method

1 Introduction

Shape memory alloys (SMA) exhibit well-known spectacular effects (shape recovery, pseudoelasticity) associated with reversible martensitic phase transformation [1, 2]. In view of their wide applicability, modelling of SMA has been an active area of research in the last two decades, with the focus on a variety of aspects such as stress- and temperature-induced transformations, martensite variant reorientation, tension–compression asymmetry, cyclic effects, incomplete shape recovery, thermomechanical coupling, and others. The corresponding literature is very rich; a systematic review is not attempted here, see [3, 4] for an overview.

Inelastic transformation strains in SMA may reach 10 per cent, so that the assumption of infinitesimal strains may be inappropriate in general. The crystallographic theory of martensite [5, 6], which proved to successfully explain formation of martensitic microstructures, relies on the exact finite-strain kinematics. Subsequently, micromechanical approaches have been developed within the finite-strain framework, e.g., [7, 8, 9, 10, 11, 12].

*To appear in *International Journal for Numerical Methods in Engineering.*

†Corresponding author, email: sstupkie@ippt.gov.pl, tel: +48 22 8261281, fax: +48 22 8269815.

At the same time, the vast majority of existing phenomenological models of SMA polycrystals is restricted to the small-deformation framework. With exception of an early attempt in [13], finite-strain macroscopic models of SMA have attracted wider attention only recently [14, 15, 16, 17, 18, 19, 20, 21, 22]. Actually, reformulation of phenomenological models of SMA from the small-strain to finite-strain regime need not be a difficult task. In this paper it is shown that such a reformulation can be relatively easily done by employing the multiplicative decomposition of the deformation gradient jointly with exponential mapping of the tensorial internal variable that represents the logarithmic transformation strain.

Macroscopic models of SMA polycrystals inherently involve constitutive constraints imposed on the internal variables such as the volume fraction of the phases and the transformation strain. These constraints introduce additional complexity in the incremental schemes used in finite-element computations. Specifically, special treatments are needed in the situations corresponding to nucleation, saturation or completion of transformation. This requires careful consideration of the corresponding conditions and adequate branching of the algorithm, cf. [18, 21, 22]. Alternatively, regularization can be introduced, e.g., [15, 20], at the cost of some inaccuracy due to the approximation involved.

A different approach is proposed in this work. The constitutive constraints fit in a natural way the general framework of incremental energy minimization [23], thus leading to a constrained minimization problem to be solved at each load increment. The augmented Lagrangian method is subsequently used as a basis of an efficient and robust computational scheme in which the resulting unconstrained saddle-point problem is solved using the Newton method.

The adopted incremental energy minimization framework [23] has its origin in the criterion of path stability [24] and proved effective in non-convex minimization in plasticity using finite elements, e.g. [25, 26, 27]. In various versions, the approach has recently gained popularity in the mechanics of dissipative solids, e.g., [28, 29, 30, 31, 32, 33, 34], including modelling of SMA single crystals, e.g., [8, 35, 36]. Validity of the incremental energy minimization algorithm is limited to dissipation functions whose state-derivative exhibits a certain symmetry property that is needed to ensure consistency between the first- and second-order versions of the incremental energy minimization, see the detailed discussion in [23, 37].

SMA polycrystals are known to exhibit asymmetry of response in tension and compression [38]. This effect results from crystallography of martensitic transformation [39] and may be significantly affected by crystallographic texture [40, 7]. The asymmetry is usually included in the small-strain macroscopic models of SMA using the third invariant of stress [41] or transformation strain [42], depending on the adopted modelling framework. The tension–compression asymmetry is accounted for in the model of pseudoelasticity proposed in this paper, both in its small- and finite-strain versions. Additionally, the anisotropy of the material, e.g., due to crystallographic texture, is also included by adopting the concept of the surface of limit transformation strains [42].

The basic ingredients of the model developed here for describing pseudoelasticity in SMA can be summarized as follows:

(i) The material response is determined by *minimizing incrementally the total energy supply*, including both the free energy and *rate-independent dissipation of limited path-independence*.

(ii) Inequality constraints on the volume fraction of martensite and on the thermody-

namic driving force for phase transformation are enforced using a *single* Lagrange multiplier within the augmented Lagrangian method which leads to an *unconstrained* saddle-point problem.

(iii) Dependence of the maximum effective transformation strain on the stress direction is taken into account by introducing a *hypersurface in the transformation strain space* that represents the boundary of an admissible domain.

(iv) The methodology for easy reformulation of the model from the small-strain to *finite-deformation regime* is given.

The ingredient (ii) is new in modelling of shape memory alloys. The ingredient (iii) is essentially adopted from [42]. The general approach (i) has been developed earlier [23, 37], but its combination with (ii) and (iii) is new. The ingredient (iv) is new.

The focus of the paper is twofold: on a general formulation of a class of models with the four ingredients listed above, and on a robust and efficient finite-element implementation of a specific version of the model. The model includes the tension–compression asymmetry and anisotropy of limit transformation strains in the finite-deformation regime.

2 Illustrative one-dimensional model

Consider a one-dimensional model of pseudoelasticity governed by the following *Helmholtz free energy* function,

$$\phi(\varepsilon,\eta) = B(T-T_0)\eta + \frac{1}{2}E(\varepsilon-\eta\bar{\varepsilon}^{\mathrm{t}})^2, \qquad 0\le\eta\le 1, \tag{1}$$

for simplicity, formulated here for tension only, where $\varepsilon \ge 0$ is the strain, η the volume fraction of martensite, $\bar{\varepsilon}^{\mathrm{t}} > 0$ the given transformation strain, E the Young's modulus, and $B(T-T_0)$ the temperature-dependent chemical energy of transformation with $B > 0$ and $T > T_0$. This is the simplest and commonly accepted form of a free energy function in a one-dimensional model of a shape memory alloy. Throughout this paper, temperature T is regarded as given and therefore is excluded from the list of variables of ϕ.

Assume first that the response is governed by the energy minimization principle imposed on the free energy only. Accordingly, energy dissipation is disregarded at the moment. Under kinematic control, i.e., for given strain ε, the only unknown is the volume fraction η to be found as a solution of the following minimization problem

$$\min_{0\le\eta\le 1} \phi(\varepsilon,\eta) \qquad \text{(for given } \varepsilon\text{)}. \tag{2}$$

The constrained minimization problem (2) can be rewritten in the form

$$\min_{\eta} \tilde{\phi}(\varepsilon,\eta), \qquad \tilde{\phi}(\varepsilon,\eta) = \phi(\varepsilon,\eta) + I_{[0,1]}(\eta), \tag{3}$$

where $I_{[0,1]} : R \to \bar{R} = R \cup \{+\infty\}$ is the indicator function,

$$I_{[0,1]}(\eta) = \begin{cases} 0, & 0\le\eta\le 1, \\ +\infty, & \text{otherwise.} \end{cases} \tag{4}$$

A convenient way of solving the non-smooth minimization problem (3) is by applying the augmented Lagrangian method which transforms the minimization problem (3) to an equivalent saddle-point problem,

$$\min_{\eta}\max_{\lambda} L(\varepsilon,\eta,\lambda), \qquad L(\varepsilon,\eta,\lambda)=\phi(\varepsilon,\eta)+l_{\mathrm{b}}(\eta,\lambda), \tag{5}$$

where λ is a Lagrange multiplier, and $l_{\mathrm{b}}(\eta,\lambda)$ provides a suitable treatment of the indicator function enforcing the bound constraints $\eta^- \le \eta \le \eta^+$, cf. [43],

$$l_{\mathrm{b}}(\eta,\lambda)=\begin{cases} \lambda(\eta-\eta^-)+\dfrac{\varrho}{2}(\eta-\eta^-)^2 & \text{if } \hat\lambda<\varrho\eta^-,\\ -\dfrac{1}{2\varrho}\lambda^2 & \text{if } \varrho\eta^-\le\hat\lambda\le\varrho\eta^+,\\ \lambda(\eta-\eta^+)+\dfrac{\varrho}{2}(\eta-\eta^+)^2 & \text{if } \varrho\eta^+<\hat\lambda. \end{cases} \tag{6}$$

Here, $\varrho>0$ is a parameter, $\hat\lambda=\lambda+\varrho\eta$ is the augmented Lagrange multiplier, and in our case we have $\eta^-=0$, $\eta^+=1$.

The important known feature of the above augmented Lagrangian treatment is that the Lagrangian $L(\varepsilon,\eta,\lambda)$ is continuously differentiable and that the constraints are *enforced exactly* regardless of the value of parameter ϱ. To illustrate that, consider the following equations which express stationarity of L with respect to η and λ,

$$0=\frac{\partial L}{\partial\eta}=-f+\begin{cases} \lambda+\varrho(\eta-\eta^-) & \text{if } \hat\lambda<\varrho\eta^-,\\ 0 & \text{if } \varrho\eta^-\le\hat\lambda\le\varrho\eta^+,\\ \lambda+\varrho(\eta-\eta^+) & \text{if } \varrho\eta^+<\hat\lambda, \end{cases} \tag{7}$$

$$0=\frac{\partial L}{\partial\lambda}=\begin{cases} \eta-\eta^- & \text{if } \hat\lambda<\varrho\eta^-,\\ -\dfrac{1}{\varrho}\lambda & \text{if } \varrho\eta^-\le\hat\lambda\le\varrho\eta^+,\\ \eta-\eta^+ & \text{if } \varrho\eta^+<\hat\lambda, \end{cases} \tag{8}$$

where $f=-\partial\phi/\partial\eta$ is the thermodynamic driving force for transformation at the current strain ε. Note that $\rho>0$ implies convexity of L with respect to η. Equation (8) enforces a state-dependent constraint on η or λ, so that either $\lambda=0$ or $\eta=\eta^{\pm}$. Further, it follows from (7) and (8) that at the solution we have $\lambda=f$, and $\lambda=f=0$ whenever η varies in the range $0<\eta<1$. Figure 1 shows the indicator function $I_{[0,1]}$, the corresponding graph of the Lagrange multiplier λ, and the resulting stress-strain relationship, where $\sigma=E(\varepsilon-\eta\bar\varepsilon^{\mathrm{t}})$.

Introduce now the *rate-independent dissipation* in the following incremental form

$$\Delta D=f_{\mathrm{c}}|\Delta\eta|, \qquad f_{\mathrm{c}}>0, \quad \Delta\eta=\eta-\eta_n, \tag{9}$$

where η_n is the volume fraction at the end of the previous time increment. Here, the increments can be arbitrarily small, so that the above formulation covers both the time-discrete form and the time-continuous rate-form $\dot D=f_{\mathrm{c}}|\dot\eta|$ in the limit. ΔD represents energy dissipation related to jumps from one energy well to another, which corresponds to the transition from the parent to product phase. It is assumed that the increment, symbolized by Δ, is sufficiently small to treat all changes within the increment as monotonic. Note that the incremental dissipation does not vanish with reducing the rate of change of η with respect to a natural time to zero as it is independent of that rate.

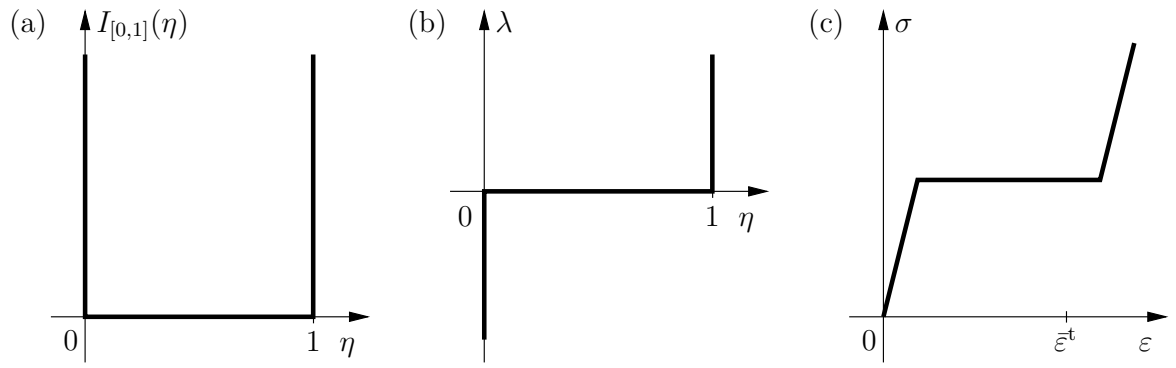


Figure 1: Bound constraints on η in the minimization problem (3): (a) indicator function $I_{[0,1]}(\eta)$, (b) graph of the Lagrange multiplier λ, (c) stress-strain relationship (no dissipation).

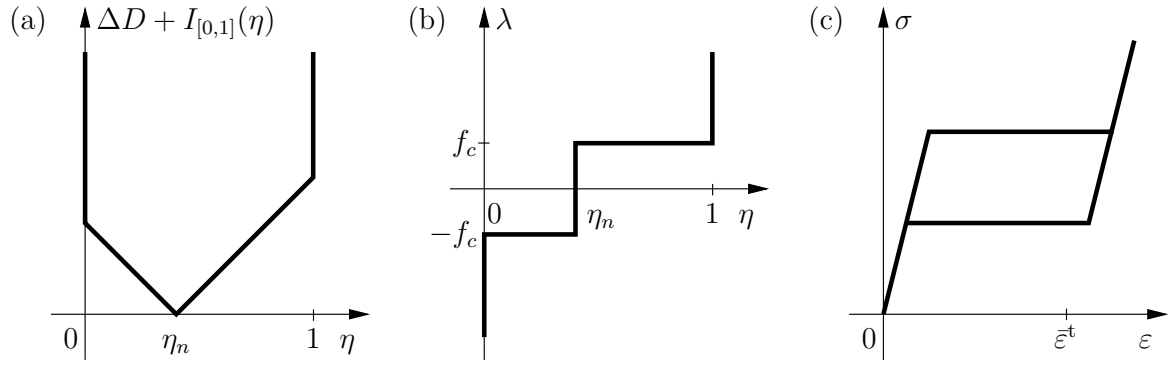


Figure 2: Minimization problem (11) involving bound constraints on η and rate-independent dissipation (9): (a) the nondifferentiable component, (b) graph of the Lagrange multiplier λ, (c) stress-strain relationship.

Using the rate-independence property of ΔD, the material response is postulated to be governed by the incremental energy minimization principle (to be discussed in more detail later):

$$\min_{0\le\eta\le 1} \Delta e, \qquad \Delta e = \Delta\phi + \Delta D \qquad \text{(for given } \varepsilon\text{)}, \tag{10}$$

with $\eta = \eta_n + \Delta\eta$. The minimization problem (10) can be equivalently expressed using the indicator function $I_{[0,1]}(\eta)$,

$$\min_{\eta} \Delta\tilde{e}, \qquad \Delta\tilde{e} = \Delta\phi + \Delta D + I_{[0,1]}(\eta). \tag{11}$$

The nondifferentiable component of $\Delta\tilde{e}$ in the minimization problem (11) includes now ΔD and is illustrated in Fig. 2(a).

Again, the augmented Lagrangian treatment of the nondifferentiable component in the minimization problem (11) is introduced, which leads to the following unconstrained saddle-point problem

$$\min_{\eta}\max_{\lambda} L(\varepsilon,\eta,\lambda), \qquad L(\varepsilon,\eta,\lambda) = \phi(\varepsilon,\eta) + l_{\mathrm{t}}(\Delta\eta,\lambda), \tag{12}$$

where

$$l_{\mathrm{t}}(\Delta\eta,\lambda)=\begin{cases}\lambda(\Delta\eta-\Delta\eta^{-})+\dfrac{\varrho}{2}(\Delta\eta-\Delta\eta^{-})^2-f_{\mathrm{c}}\Delta\eta^{-} & \text{if } \hat{\lambda}\le\varrho\Delta\eta^{-}-f_{\mathrm{c}},\\ -\dfrac{1}{2\varrho}(\lambda^2+2f_{\mathrm{c}}\hat{\lambda}+f_{\mathrm{c}}^2) & \text{if } \varrho\Delta\eta^{-}-f_{\mathrm{c}}<\hat{\lambda}<-f_{\mathrm{c}},\\ (\lambda+\dfrac{\varrho}{2}\Delta\eta)\Delta\eta & \text{if } -f_{\mathrm{c}}\le\hat{\lambda}\le f_{\mathrm{c}},\\ -\dfrac{1}{2\varrho}(\lambda^2-2f_{\mathrm{c}}\hat{\lambda}+f_{\mathrm{c}}^2) & \text{if } f_{\mathrm{c}}<\hat{\lambda}<\varrho\Delta\eta^{+}+f_{\mathrm{c}},\\ \lambda(\Delta\eta-\Delta\eta^{+})+\dfrac{\varrho}{2}(\Delta\eta-\Delta\eta^{+})^2+f_{\mathrm{c}}\Delta\eta^{+} & \text{if } \varrho\Delta\eta^{+}+f_{\mathrm{c}}\le\hat{\lambda}.\end{cases}\tag{13}$$

Here, $\hat{\lambda}=\lambda+\varrho\Delta\eta$ is the augmented Lagrange multiplier, while $\Delta\eta^{-}=-\eta_n\le 0$ and $\Delta\eta^{+}=1-\eta_n\ge 0$ are the bounds enforced on the volume fraction increment $\Delta\eta$. Function l_{t} in (13) is a generalization of the function corresponding to one of the cases of nondifferentiable polyhedral functions given in [43, pp. 168 and 173]. In analogy to the analysis in [43], the constrained minimization problem (10) can be shown to be equivalent to the unconstrained saddle-point problem (12) under the assumptions imposed on ϕ and ΔD.

The Lagrangian $L(\varepsilon,\eta,\lambda)$ in the saddle-point problem (12) with l_{t} defined by (13) is a continuously differentiable function of η and λ. Moreover, by construction, l_{t} is convex with respect to η and concave with respect to λ, and the same properties are shared by L if ϕ is convex with respect to η. Accordingly, solution to the saddle-point problem (12) can be found by solving directly the equations which express the stationarity conditions

$$0=\frac{\partial L}{\partial\eta}=-f+\begin{cases}\lambda+\varrho(\Delta\eta-\Delta\eta^{-}) & \text{if } \hat{\lambda}\le\varrho\Delta\eta^{-}-f_{\mathrm{c}},\\ -f_{\mathrm{c}} & \text{if } \varrho\Delta\eta^{-}-f_{\mathrm{c}}<\hat{\lambda}<-f_{\mathrm{c}},\\ \lambda+\varrho\Delta\eta & \text{if } -f_{\mathrm{c}}\le\hat{\lambda}\le f_{\mathrm{c}},\\ f_{\mathrm{c}} & \text{if } f_{\mathrm{c}}<\hat{\lambda}<\varrho\Delta\eta^{+}+f_{\mathrm{c}},\\ \lambda+\varrho(\Delta\eta-\Delta\eta^{+}) & \text{if } \varrho\Delta\eta^{+}+f_{\mathrm{c}}\le\hat{\lambda},\end{cases}\tag{14}$$

$$0=\frac{\partial L}{\partial\lambda}=\begin{cases}\Delta\eta-\Delta\eta^{-} & \text{if } \hat{\lambda}\le\varrho\Delta\eta^{-}-f_{\mathrm{c}},\\ -\dfrac{1}{\varrho}(\lambda+f_{\mathrm{c}}) & \text{if } \varrho\Delta\eta^{-}-f_{\mathrm{c}}<\hat{\lambda}<-f_{\mathrm{c}},\\ \Delta\eta & \text{if } -f_{\mathrm{c}}\le\hat{\lambda}\le f_{\mathrm{c}},\\ -\dfrac{1}{\varrho}(\lambda-f_{\mathrm{c}}) & \text{if } f_{\mathrm{c}}<\hat{\lambda}<\varrho\Delta\eta^{+}+f_{\mathrm{c}},\\ \Delta\eta-\Delta\eta^{+} & \text{if } \varrho\Delta\eta^{+}+f_{\mathrm{c}}\le\hat{\lambda},\end{cases}\tag{15}$$

where the right-hand expressions are continuous functions of η and λ. The solution can be shown to be unique if ϕ is strictly convex. Again, Eq. (15) enforces a state-dependent constraint on η or λ, and at the solution we have $\lambda=f$. The graph of λ is illustrated in Fig. 2(b), and the resulting stress-strain response corresponding to a complete forward-reverse transformation cycle is sketched in Fig. 2(c).

Remark 1. Equations (14)–(15) can be solved iteratively using a Newton-like method (second-order iterations) which is particularly attractive in the case of a nonlinear minimization problem, as discussed in the subsequent sections. For ϕ quadratic, as in (1), the functions in the stationarity conditions (14)–(15) are actually piecewise-linear continuous functions.

Remark 2. The key feature of the present new formulation (12)–(13) of the non-smooth minimization problem (11) is that both the bound constraint on η and the nondifferentiable dissipation (9) are treated by a *single* Lagrange multiplier. As shown above, the stationarity conditions (14)–(15) imply that the Lagrange multiplier λ coincides with the driving force $f = -\partial\phi/\partial\eta$ (at the end of the increment). Moreover, $|f| \leq f_c$ whenever $0 < \eta < 1$ and $f = \pm f_c$ whenever $\Delta\eta \neq 0$ and $0 < \eta < 1$. This corresponds to the following transformation criterion

$$|f| - f_c \leq 0, \quad (|f| - f_c)\Delta\eta = 0, \quad \Delta\eta|f| = f|\Delta\eta| \qquad (\text{for } 0 < \eta < 1), \tag{16}$$

cf. the graph of $\lambda = f$ in Fig. 2(b).

3 Three-dimensional model in the small-strain regime

3.1 Surface of limit transformation strains

The total strain $\boldsymbol{\varepsilon}$ in the geometrically linear theory is additively decomposed into elastic part $\boldsymbol{\varepsilon}^{\mathrm{e}}$ and transformation part $\boldsymbol{\varepsilon}^{\mathrm{t}}$,

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^{\mathrm{e}} + \boldsymbol{\varepsilon}^{\mathrm{t}}. \tag{17}$$

Define the set $\mathcal{P}$ of admissible effective transformation strains $\boldsymbol{\varepsilon}^{\mathrm{t}}$ as follows

$$\mathcal{P} = \{\boldsymbol{\varepsilon}^{\mathrm{t}}\colon\ \operatorname{tr}\boldsymbol{\varepsilon}^{\mathrm{t}} - \eta\bar{\varepsilon}^{\mathrm{t}}_{\mathrm{vol}} = 0,\ g(\boldsymbol{\varepsilon}^{\mathrm{t}}_{\mathrm{dev}}) \leq 0\}, \tag{18}$$

where η denotes the volume fraction of martensite, $\boldsymbol{\varepsilon}^{\mathrm{t}}_{\mathrm{dev}}$ denotes the deviatoric part of $\boldsymbol{\varepsilon}^{\mathrm{t}}$, $\boldsymbol{\varepsilon}^{\mathrm{t}} = \frac{1}{3}\eta\bar{\varepsilon}^{\mathrm{t}}_{\mathrm{vol}}\mathbf{1} + \boldsymbol{\varepsilon}^{\mathrm{t}}_{\mathrm{dev}}$, $\bar{\varepsilon}^{\mathrm{t}}_{\mathrm{vol}}$ is a material parameter.

The surface $g = 0$ of limit transformation strains is not a new concept, cf. e.g. [44]. There is much flexibility in defining function g here. We assume that $g(r\boldsymbol{\varepsilon}_{\mathrm{dev}})$ is a continuously increasing function of $r \geq 0$ for every given deviator $\boldsymbol{\varepsilon}_{\mathrm{dev}} \neq \mathbf{0}$, and the zero level set of g, $\bar{\mathcal{Z}} = \{\boldsymbol{\varepsilon}_{\mathrm{dev}} : g(\boldsymbol{\varepsilon}_{\mathrm{dev}}) = 0\}$, is a smooth, closed and strictly convex hypersurface in the deviatoric space (hence, zero strain lies inside $\bar{\mathcal{Z}}$ and $g(\mathbf{0}) < 0$). Otherwise, function g is arbitrary, which allows for modelling of directional dependence of transformation strains. In particular, tension–compression asymmetry and also anisotropy due to crystallographic texture can be included in the model, cf. [42].

In the definition (18), we admit a nonzero volumetric part of the transformation strain, $\bar{\varepsilon}^{\mathrm{t}}_{\mathrm{vol}} \neq 0$, although it is typically small in real shape memory alloys. In the vast majority of existing models of SMA, the volumetric transformation strain is assumed equal to zero, and for $\bar{\varepsilon}^{\mathrm{t}}_{\mathrm{vol}} = 0$ the definition (18) simplifies to that adopted in [42].

Pseudoelasticity is associated with martensitic transformation under stress that induces a preferable orientation of martensite variants. Accordingly, we assume here that the martensite is *fully oriented* so that the transformation strain can be expressed as

$$\boldsymbol{\varepsilon}^{\mathrm{t}} = \eta\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}, \qquad 0 \leq \eta \leq 1, \qquad \bar{\boldsymbol{\varepsilon}}^{\mathrm{t}} \in \bar{\mathcal{P}} = \{\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}\colon\ \eta(\operatorname{tr}\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}} - \bar{\varepsilon}^{\mathrm{t}}_{\mathrm{vol}}) = 0,\ \eta g(\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}_{\mathrm{dev}}) = 0\}, \tag{19}$$

where $\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}$ is the transformation strain of the fully oriented martensite phase. $\bar{\mathcal{P}}$ is the set of limit transformation strains which, as mentioned above, can be asymmetric and affected by anisotropy in general. When $\eta = 0$, the transformation strain $\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}$ is indeterminate, hence the equality constraints are multiplied by η in the definition (19) of $\bar{\mathcal{P}}$.

A suitable function g is to be specified; an example is given in Section 5.

Remark 3. The properties of function g, assumed in the paragraph that follows Eq. (18), ensure that both η and $\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}$ in Eq. $(19)_1$ are uniquely defined by any admissible $\boldsymbol{\varepsilon}^{\mathrm{t}} \neq \mathbf{0}$. Therefore, either a single variable $\boldsymbol{\varepsilon}^{\mathrm{t}}$ or a pair $(\eta, \bar{\boldsymbol{\varepsilon}}^{\mathrm{t}})$ can be equivalently used in the constitutive model. In the proposed computational treatment, the latter representation is employed.

3.2 Free energy

In the present work, the attention is restricted to isothermal pseudoelastic behaviour of SMA. The Helmholtz free energy function per unit volume of the material at given temperature T is thus adopted in the following form,

$$\phi(\boldsymbol{\varepsilon}, \bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}, \eta) = \phi_0 + \phi_{\mathrm{chem}}(\eta) + \phi_{\mathrm{el}}(\boldsymbol{\varepsilon}, \bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}, \eta) + \phi_{\mathrm{int}}(\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}, \eta) \tag{20}$$

where ϕ_0 is the (constant) free energy of austenite in a stress-free state, $\phi_{\mathrm{chem}} > 0$ is the chemical energy of austenite-to-martensite transformation in the pseudoelastic regime, ϕ_{el} the macroscopic elastic strain energy, and ϕ_{int} the internal interaction energy (both ϕ_0 and ϕ_{chem} depend on temperature, but this dependence is not indicated explicitly). The volume fraction η and the transformation strain $\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}$ of the oriented martensite are the internal variables subject to the following constraints,

$$0 \leq \eta \leq 1, \qquad \eta(\operatorname{tr} \bar{\boldsymbol{\varepsilon}}^{\mathrm{t}} - \bar{\varepsilon}^{\mathrm{t}}_{\mathrm{vol}}) = 0, \qquad \eta g(\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}_{\mathrm{dev}}) = 0, \tag{21}$$

which follow from (19). The free energy ϕ is assumed strictly convex with respect to η, but its actual form is to much extent arbitrary. In particular, different elastic properties of the austenite and martensite as well as initial and transformation-induced elastic anisotropy can be included in the elastic strain energy ϕ_{el}.

The three constituents of the free energy function (20) are to be specified; an example is given in Section 5. More advanced forms of ϕ_{int} can be deduced from micromechanical analysis, cf. e.g. [45].

3.3 Dissipation function

The model is completed by specifying the incremental dissipation function ΔD. It should include the dissipation associated with the forward and reverse austenite-to-martensite transformation. The simplest proposal is

$$\Delta D = f_{\mathrm{c}} |\Delta \eta|, \tag{22}$$

where $f_{\mathrm{c}} > 0$ and $\Delta\eta = \eta - \eta_n$, precisely as in (9). Dissipation due to reorientation of martensite, i.e., due to evolution of $\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}_{\mathrm{dev}}$, is disregarded in (22). This is certainly an oversimplification, adopted sometimes in earlier models, e.g. [46, 41], though it is not restrictive for nearly proportional loading–unloading paths. Moreover, f_{c} should in principle vary with η, which is disregarded here. On the other hand, the assumption of a constant f_{c} correctly predicts that the area of a hysteresis loop upon returning to a fully austenitic state does not depend on the stress direction, cf. [38]. Material parameter f_{c} is to be specified so that the width of the hysteresis loop is adequately represented.

The formula (22) is a special case of the following proposal

$$\Delta D = \hat{D}(\Delta\boldsymbol{\varepsilon}^{\mathrm{t}}, \boldsymbol{\varepsilon}^{\mathrm{t}}_n), \tag{23}$$

where $\hat{D}$ is a non-negative *function* of an admissible increment $\Delta\boldsymbol{\varepsilon}^{\mathrm{t}}$ of the effective transformation strain; the function depends on the transformation strain $\boldsymbol{\varepsilon}^{\mathrm{t}}_n$ at the end of the previous time step. The effect of $\Delta\eta$ in (22) is a special case of the effect of $\Delta\boldsymbol{\varepsilon}^{\mathrm{t}}$ in (23), cf. Remark 3. The form (23) of ΔD can include the dissipation due to reorientation of martensite. It satisfies also the condition of path-independence when limited to sufficiently small increments of monotonic transformations, which is essential for applicability of the incremental energy minimization to a discrete time-step, cf. [37]. For consistency with the assumption of rate-independent dissipation it is required that $\hat{D}$, in the limit, be a homogeneous function of degree one in *infinitesimal* $\Delta\boldsymbol{\varepsilon}^{\mathrm{t}}$. The general form of ΔD in (23) can be included in the incremental energy minimization framework introduced below, but it is not suitable for the specific algorithmic treatment based on the augmented Lagrangian method, cf. Section 3.5.

3.4 Incremental energy minimization

Evolution of the fields of strain, stress and internal variables η and $\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}$ within a material body $\mathcal{B}$ is determined by minimization of the increment of energy supplied to the system starting from a given state, cf. Petryk [23],

$$\Delta E = \Delta\Phi + \Delta\mathcal{D} + \Delta\Omega \;\to\; \min_{\tilde{\boldsymbol{u}},\tilde{\bar{\boldsymbol{\varepsilon}}}^{\mathrm{t}}} \quad \text{subject to kinematical and constitutive constraints.} \tag{24}$$

The minimization is performed with respect to unknown fields at the end of the increment. In the absence of interfacial terms,

$$\Phi = \int_{\mathcal{B}} \phi \,\mathrm{d}V, \qquad \Delta\mathcal{D} = \int_{\mathcal{B}} \Delta D \,\mathrm{d}V \tag{25}$$

are volume integrals of the local free energy and dissipation functions, respectively, and $\Delta\Omega$ is the increment of potential energy of the loads which are assumed conservative. In the case of purely kinematic control, $\Delta\Omega$ vanishes.

Minimization (24) performed with respect to a displacement field $\tilde{\boldsymbol{u}}$ over B for a fixed transformation strain field $\tilde{\boldsymbol{\varepsilon}}^{\mathrm{t}}$, viz.

$$\Delta E(\tilde{\boldsymbol{u}}) = \Delta\Phi + \Delta\Omega \;\to\; \min_{\tilde{\boldsymbol{u}}} \quad \text{subject to kinematical constraints for given } \tilde{\boldsymbol{\varepsilon}}^{\mathrm{t}}, \tag{26}$$

reduces in the usual way to finding an elastically stable configuration of the material body under the loads associated with $\Delta\Omega$. The resulting stress field satisfies all the equilibrium conditions at the end of the incremental step for a prescribed $\tilde{\boldsymbol{\varepsilon}}^{\mathrm{t}}$. The actual field of $\boldsymbol{\varepsilon}^{\mathrm{t}}$, or of η and $\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}$ taken jointly, cf. Remark 3, is to be determined from the incremental constitutive relationships.

The key point now is that the local incremental constitutive law need not be formulated separately, rather, it follows from the incremental energy minimization (24) with respect to a transformation strain field $\tilde{\boldsymbol{\varepsilon}}^{\mathrm{t}}$ for a fixed displacement field $\tilde{\boldsymbol{u}}$ over B. Since the constitutive law is assumed to be local, relating increments in η and $\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}$ to the increment of strain $\boldsymbol{\varepsilon}$ at a material point, and $\Delta\Omega$ does not depend on $\tilde{\boldsymbol{\varepsilon}}^{\mathrm{t}}$, the minimization can be applied pointwise to the local contribution $\Delta e = \Delta\phi + \Delta D$ to ΔE at a material point. The total strain $\boldsymbol{\varepsilon}$ is now treated as prescribed, the quantities $(\boldsymbol{\varepsilon}_n, \eta_n, \bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}_n)$ corresponding to the end of the previous time step $t = t_n$ are known, and with the use of functions (20) and (22) the local incremental minimization problem becomes

$$\min_{\eta,\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}} \Delta e(\boldsymbol{\varepsilon}, \bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}, \eta) \quad \text{subject to constraints (21)}, \tag{27}$$

where

$$\Delta e(\boldsymbol{\varepsilon}, \bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}, \eta) = \phi(\boldsymbol{\varepsilon}, \bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}, \eta) - \phi(\boldsymbol{\varepsilon}_n, \bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}_n, \eta_n) + f_{\mathrm{c}}|\eta - \eta_n|. \tag{28}$$

The algorithmic treatment of the above minimization problem is to be specified. A proposal motivated by the one-dimensional formulation given in Section 2 is presented below.

Remark 4. It is of primary importance that the value of the adopted dissipation function (22) depends merely on the initial and final values of η. Hence, not only the required symmetry condition for the state derivative of the dissipation function [23] is satisfied trivially, but also the dissipation is path-independent within *any finite* transformation step provided η would vary monotonically within that step also for arbitrarily fine time-discretization. It follows that as long as the above condition holds true then an *exact* solution can be determined by applying pretty large steps in external loading conditions. It means that no approximation is induced by time discretization since the solution does not depend on the step size. This is a fundamental property of the model developed in this paper.

3.5 Augmented Lagrangian formulation

The augmented Lagrangian treatment of the non-smooth constrained minimization problem (27) is analogous to that applied to the illustrative example in Section 2. The Lagrangian function L is introduced

$$L(\boldsymbol{\varepsilon}, \bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}, \eta, \lambda, \mu, \kappa) = \phi(\boldsymbol{\varepsilon}, \bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}, \eta) + l_{\mathrm{t}}(\Delta\eta, \lambda) + \eta\mu g(\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}_{\mathrm{dev}}) + \eta\kappa(\operatorname{tr}\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}} - \bar{\varepsilon}^{\mathrm{t}}_{\mathrm{vol}}), \tag{29}$$

where function l_{t} given by (13) introduces both the inequality constraints $(21)_1$ and the dissipation increment (22), while μ and κ are the usual Lagrange multipliers enforcing the equality constraints in (21). The minimization problem (27) is then transformed, in analogy to the one-dimensional example of Section 2, to the following saddle-point problem

$$\min_{\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}, \eta} \max_{\lambda, \mu, \kappa} L(\boldsymbol{\varepsilon}, \bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}, \eta, \lambda, \mu, \kappa). \tag{30}$$

In the above augmented Lagrangian formulation, the saddle-point problem (30) is smooth and unconstrained, hence the stationarity conditions

$$\frac{\partial L(\boldsymbol{\varepsilon}, \mathbf{h})}{\partial \mathbf{h}} = \mathbf{0}, \qquad \mathbf{h} = \{\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}, \eta, \lambda, \mu, \kappa\}, \tag{31}$$

can be solved directly, e.g., using the Newton method. Upon solving the stationarity conditions (31) for the unknown parameters $\mathbf{h}$, the stress $\boldsymbol{\sigma}$ is determined simply as

$$\boldsymbol{\sigma} = \frac{\partial \phi}{\partial \boldsymbol{\varepsilon}} = \boldsymbol{L}(\boldsymbol{\varepsilon} - \eta\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}). \tag{32}$$

Details of the algorithmic treatment, including a special treatment of the ill-posedness at $\eta \approx 0$, are provided in Section 6.

4 Extension to the finite-strain regime

Our goal is now to extend the approach presented above to the finite deformation setting. In analogy to the widely used multiplicative finite-strain plasticity framework, a multiplicative decomposition of the deformation gradient $\boldsymbol{F}$ into elastic part $\boldsymbol{F}^{\mathrm{e}}$ and inelastic (transformation) part $\boldsymbol{F}^{\mathrm{t}}$ is postulated,

$$\boldsymbol{F} = \boldsymbol{F}^{\mathrm{e}}\boldsymbol{F}^{\mathrm{t}}, \tag{33}$$

as in the majority of finite-strain models of SMA. The following well-known relationships for the Cauchy-Green tensors hold,

$$\boldsymbol{C}^{\mathrm{t}} = (\boldsymbol{F}^{\mathrm{t}})^{\mathrm{T}}\boldsymbol{F}^{\mathrm{t}}, \qquad \boldsymbol{b}^{\mathrm{e}} = \boldsymbol{F}^{\mathrm{e}}(\boldsymbol{F}^{\mathrm{e}})^{\mathrm{T}} = \boldsymbol{F}(\boldsymbol{C}^{\mathrm{t}})^{-1}\boldsymbol{F}^{\mathrm{T}}. \tag{34}$$

The configuration of stress-free austenite is adopted as a reference configuration so that $\boldsymbol{F}^{\mathrm{t}} = \boldsymbol{1}$ for $\eta = 0$, where η is the volume fraction of martensite in the reference configuration.

The main question now is how to define the set of admissible effective transformation strains. In the present model, we postulate that this set is defined in terms of the logarithmic strain $\boldsymbol{e}^{\mathrm{t}}$,

$$\boldsymbol{e}^{\mathrm{t}} = \log \boldsymbol{U}^{\mathrm{t}}, \qquad \boldsymbol{F}^{\mathrm{t}} = \boldsymbol{R}^{\mathrm{t}}\boldsymbol{U}^{\mathrm{t}}, \qquad (\boldsymbol{R}^{\mathrm{t}})^{-1} = (\boldsymbol{R}^{\mathrm{t}})^{\mathrm{T}}, \tag{35}$$

using the definition

$$\mathcal{P} = \{\boldsymbol{e}^{\mathrm{t}}\colon\ (\operatorname{tr}\boldsymbol{e}^{\mathrm{t}} - \eta\bar{\varepsilon}^{\mathrm{t}}_{\mathrm{vol}}) = 0,\ g(\boldsymbol{e}^{\mathrm{t}}_{\mathrm{dev}}) \leq 0\}. \tag{36}$$

On account of the basic properties of the logarithmic strain measure, the volumetric and shape strains can be separated in an additive manner, so that formula (36) is formally identical to definition (18). Furthermore, in analogy to the small-strain model, we can adopt the following assumption,

$$\boldsymbol{e}^{\mathrm{t}} = \eta\bar{\boldsymbol{e}}^{\mathrm{t}}, \qquad 0 \leq \eta \leq 1, \qquad \bar{\boldsymbol{e}}^{\mathrm{t}} \in \bar{\mathcal{P}} = \{\bar{\boldsymbol{e}}^{\mathrm{t}}\colon\ \eta(\operatorname{tr}\bar{\boldsymbol{e}}^{\mathrm{t}} - \bar{\varepsilon}^{\mathrm{t}}_{\mathrm{vol}}) = 0,\ \eta g(\bar{\boldsymbol{e}}^{\mathrm{t}}_{\mathrm{dev}}) = 0\}, \tag{37}$$

where $\bar{\boldsymbol{e}}^{\mathrm{t}}$ is the (logarithmic) transformation strain of the oriented martensite phase.

Inverting $(35)_1$ and using the properties of the tensor exponential we have

$$\boldsymbol{U}^{\mathrm{t}} = \exp(\boldsymbol{e}^{\mathrm{t}}), \qquad (\boldsymbol{C}^{\mathrm{t}})^{-1} = \exp(-2\boldsymbol{e}^{\mathrm{t}}). \tag{38}$$

Accordingly, the volume change due to transformation is equal to

$$J^{\mathrm{t}} = \det\boldsymbol{F}^{\mathrm{t}} = \det\boldsymbol{U}^{\mathrm{t}} = \exp(\eta\bar{\varepsilon}^{\mathrm{t}}_{\mathrm{vol}}). \tag{39}$$

In particular, for $\bar{\varepsilon}^{\mathrm{t}}_{\mathrm{vol}} = 0$, we have $J^{\mathrm{t}} = 1$ so that deformation due to phase transformation is guaranteed to be incompressible.

In analogy to (20), the Helmholtz free energy, now per unit volume in the reference configuration of stress-free austenite, is adopted in the following general form

$$\phi(\boldsymbol{F}, \bar{\boldsymbol{e}}^{\mathrm{t}}, \eta) = \phi_0 + \phi_{\mathrm{chem}}(\eta) + \phi_{\mathrm{el}}(\boldsymbol{F}, \bar{\boldsymbol{e}}^{\mathrm{t}}, \eta) + \phi_{\mathrm{int}}(\bar{\boldsymbol{e}}^{\mathrm{t}}, \eta). \tag{40}$$

The specific form of the elastic strain energy function ϕ_{el} is to much extent arbitrary, except that the resulting function ϕ should be strictly convex with respect to η. As in the

small-strain model, different elastic properties of the austenite and martensite as well as initial and transformation-induced elastic anisotropy can be included in the elastic strain energy $\phi_{\rm el}$.

If the material is taken to be elastically isotropic and of identical elastic properties of austenite and martensite then a suitable elastic strain energy function (e.g., of the form specified in Section 5) can be adopted as follows

$$\phi_{\rm el}(\boldsymbol{F}, \bar{\boldsymbol{e}}^{\rm t}, \eta) = J^{\rm t}\hat{\phi}_{\rm el}(\boldsymbol{b}^{\rm e}), \tag{41}$$

where $J^{\rm t}$, the transformed unit reference volume at zero stress, depends on η through (39), while $\boldsymbol{b}^{\rm e}$ is the elastic left Cauchy–Green tensor, directly expressed in terms of the total deformation gradient $\boldsymbol{F}$ and internal variables η and $\bar{\boldsymbol{e}}^{\rm t}$, viz.

$$\boldsymbol{b}^{\rm e} = \boldsymbol{F}\exp(-2\eta\bar{\boldsymbol{e}}^{\rm t})\boldsymbol{F}^{\rm T}. \tag{42}$$

The remaining ingredients of the model remain essentially unchanged with respect to the small-strain version, in particular, the simplest dissipation function (22) and the incremental energy minimization rule. Hence, the structure of the small-strain model is fully preserved here with $\boldsymbol{\varepsilon}$ and $\bar{\boldsymbol{\varepsilon}}^{\rm t}$ replaced by $\boldsymbol{F}$ and $\bar{\boldsymbol{e}}^{\rm t}$, respectively. In particular, the finite-strain counterpart to the minimization problem (27) reads

$$\min_{\bar{\boldsymbol{e}}^{\rm t},\eta} \Delta e(\boldsymbol{F}, \bar{\boldsymbol{e}}^{\rm t}, \eta) \quad \text{subject to} \;\; 0 \le \eta \le 1, \;\; \eta(\operatorname{tr}\bar{\boldsymbol{e}}^{\rm t} - \bar{\varepsilon}^{\rm t}_{\rm vol}) = 0, \;\; \eta g(\bar{\boldsymbol{e}}^{\rm t}_{\rm dev}) = 0, \tag{43}$$

where

$$\Delta e(\boldsymbol{F}, \bar{\boldsymbol{e}}^{\rm t}, \eta) = \phi(\boldsymbol{F}, \bar{\boldsymbol{e}}^{\rm t}, \eta) - \phi(\boldsymbol{F}_n, \bar{\boldsymbol{e}}^{\rm t}_n, \eta_n) + f_{\rm c}|\eta - \eta_n|, \tag{44}$$

and its augmented Lagrangian treatment yields the following saddle-point problem

$$\min_{\bar{\boldsymbol{e}}^{\rm t},\eta} \max_{\lambda,\mu,\kappa} L(\boldsymbol{F}, \bar{\boldsymbol{e}}^{\rm t}, \eta, \lambda, \mu, \kappa), \tag{45}$$

where

$$L(\boldsymbol{F}, \bar{\boldsymbol{e}}^{\rm t}, \eta, \lambda, \mu, \kappa) = \phi(\boldsymbol{F}, \bar{\boldsymbol{e}}^{\rm t}, \eta) + l_{\rm t}(\Delta\eta, \lambda) + \eta\mu g(\bar{\boldsymbol{e}}^{\rm t}_{\rm dev}) + \eta\kappa(\operatorname{tr}\bar{\boldsymbol{e}}^{\rm t} - \bar{\varepsilon}^{\rm t}_{\rm vol}). \tag{46}$$

Again, the stationarity conditions, such as those in (31) with $\mathbf{h} = \{\bar{\boldsymbol{e}}^{\rm t}, \eta, \lambda, \mu, \kappa\}$, can be solved directly using the Newton method, and the first Piola–Kirchhoff stress $\boldsymbol{P}$ is given by

$$\boldsymbol{P} = \frac{\partial\phi}{\partial\boldsymbol{F}}. \tag{47}$$

Details of the algorithmic treatment are provided in Section 6.

If elastic isotropy is assumed in the formulation presented above then rotation $\boldsymbol{R}^{\rm t}$ is arbitrary and does not appear in the kinematic relationships (38) and (42). In the case of elastic anisotropy, a free rotation in $\boldsymbol{F}^{\rm t}$ must be eliminated, for instance, by assuming that

$$\boldsymbol{R}^{\rm t} = \boldsymbol{1}, \qquad \boldsymbol{F}^{\rm t} = \boldsymbol{U}^{\rm t} = \exp(\eta\bar{\boldsymbol{e}}^{\rm t}), \tag{48}$$

so that the elastic right Cauchy-Green deformation tensor can be uniquely determined in terms of $\boldsymbol{F}$ and $\eta\bar{\boldsymbol{e}}^{\rm t}$ as follows:

$$\boldsymbol{C}^{\rm e} = (\boldsymbol{F}^{\rm e})^{\rm T}\boldsymbol{F}^{\rm e}, \qquad \boldsymbol{F}^{\rm e} = \boldsymbol{F}(\boldsymbol{F}^{\rm t})^{-1} = \boldsymbol{F}\exp(-\eta\bar{\boldsymbol{e}}^{\rm t}). \tag{49}$$

In this formulation, the elastic anisotropy induced by transformation (due to privileged orientation of stress-induced martensite) can be included in the model.

5 The basic specifications

The structure of the finite-strain pseudoelasticity model defined above involves the three constitutive functions: function g that defines the surface of limit transformation strains in (37), free energy function ϕ of a general structure (40), and incremental dissipation function ΔD. Obviously, these three functions must be specified for computational purposes. Then, the material response to a given loading program can be determined by applying the incremental energy minimization, which has been reduced to solving step-by-step the saddle-point problem (45). The simplest form of the dissipation function has already been given in (22). It remains to specify functions g and ϕ; our basic specifications are provided below in both the geometrically linear and nonlinear setting.

5.1 Surface of limit transformation strains

Function $g(\boldsymbol{\varepsilon}^{\mathrm{t}}_{\mathrm{dev}})$ is specified here for the case of a transversely isotropic material with the axis of rotational symmetry denoted by $\boldsymbol{m}$. Such anisotropy may be related to crystallographic texture due to prior plastic forming (transverse isotropy is typical for drawing; it may also result from rolling). In view of $I_1 = \operatorname{tr}\boldsymbol{\varepsilon}^{\mathrm{t}}_{\mathrm{dev}} = 0$ and $\|\boldsymbol{m}\| = 1$, an isotropic function of tensors $\boldsymbol{\varepsilon}^{\mathrm{t}}_{\mathrm{dev}}$ and $\boldsymbol{m}\otimes\boldsymbol{m}$ (jointly) is known to be a function of the following four invariants

$$I_2 = -\frac{1}{2}\mathrm{tr}(\boldsymbol{\varepsilon}^{\mathrm{t}}_{\mathrm{dev}})^2, \quad I_3 = \det\boldsymbol{\varepsilon}^{\mathrm{t}}_{\mathrm{dev}}, \quad I_4 = \boldsymbol{m}\cdot\boldsymbol{\varepsilon}^{\mathrm{t}}_{\mathrm{dev}}\boldsymbol{m}, \quad I_5 = \boldsymbol{m}\cdot(\boldsymbol{\varepsilon}^{\mathrm{t}}_{\mathrm{dev}})^2\boldsymbol{m}. \tag{50}$$

A suitable function g involving only one mixed invariant I_4 is adopted in the form

$$g(\boldsymbol{\varepsilon}^{\mathrm{t}}_{\mathrm{dev}}) = \left[(-I_2)^{3/2} - bI_3 - cI_4^3\right]^{1/3} - a, \tag{51}$$

where a, b and c are constitutive parameters. The function g in Eq. (51) is a simple modification of the function proposed in [42], $g = (-I_2)^{3/2} - bI_3 - cI_4^3 - a^*$, which has the same zero level set, but the former appears more suitable in computational practice.

The specific form of function g allows us to describe tension-compression asymmetry (using the third invariant I_3) and anisotropy in the form of transversal isotropy (using the mixed invariant I_4). For $c = 0$, function g describes an isotropic material exhibiting tension–compression asymmetry. For $b = c = 0$, an isotropic response with no tension–compression asymmetry is obtained.

In case of the finite-strain model, function g is specified by the expressions provided above with $\boldsymbol{\varepsilon}^{\mathrm{t}}_{\mathrm{dev}}$ replaced by $\boldsymbol{e}^{\mathrm{t}}_{\mathrm{dev}}$. For $b = c = 0$ and for isochoric transformation ($\bar{\varepsilon}^{\mathrm{t}}_{\mathrm{vol}} = 0$), the set $\mathcal{P}$ of admissible effective transformation strains reduces to the constraint $\|\boldsymbol{e}^{\mathrm{t}}\| \le \varepsilon_{\mathrm{L}}$ on the logarithmic transformation strain $\boldsymbol{e}^{\mathrm{t}}$ with $\operatorname{tr}\boldsymbol{e}^{\mathrm{t}} = 0$, as in [14, 22]. In other finite-strain models [15, 18, 19, 20, 21], an analogous constraint is imposed on the Green strain tensor, $\|\boldsymbol{E}^{\mathrm{t}}\| \le \varepsilon_{\mathrm{L}}$, while the incompressibility constraint must be enforced separately.

5.2 Free energy

The free energy comprises three essential parts that are specified below (ϕ_0 is a constant at fixed temperature and may be left unspecified). The chemical energy ϕ_{chem} is usually assumed to be a linear function of temperature T and is proportional to the volume fraction of martensite, viz.

$$\phi_{\mathrm{chem}}(\eta) = B(T - T_0)\eta, \qquad B > 0. \tag{52}$$

The present model is restricted to the pseudoelastic regime, hence we assume that $B(T - T_0) > f_c$ so that austenitic state can be recovered in stress-free conditions.

Assuming elastic isotropy and identical properties of austenite and martensite, the elastic strain energy $\phi_{\rm el}$ in the small-strain regime has a usual form,

$$\phi_{\rm el}(\boldsymbol{\varepsilon}, \bar{\boldsymbol{\varepsilon}}^{\rm t}, \eta) = \hat{\phi}_{\rm el}(\boldsymbol{\varepsilon}^{\rm e}) = \mu \,{\rm tr}(\boldsymbol{\varepsilon}^{\rm e}_{\rm dev})^2 + \frac{1}{2}\,\kappa ({\rm tr}\,\boldsymbol{\varepsilon}^{\rm e})^2, \qquad \boldsymbol{\varepsilon}^{\rm e} = \boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\rm t}, \tag{53}$$

where $\mu = E/2(1+\nu)$ and $\kappa = E/3(1-2\nu)$ are the shear modulus and the bulk modulus, respectively, E is the Young's modulus, and ν is the Poisson's ratio. The finite strain counterpart of (53) is adopted in the following form [47],

$$\hat{\phi}_{\rm el}(\boldsymbol{b}^{\rm e}) = \frac{1}{2}\,\mu({\rm tr}\,\bar{\boldsymbol{b}}^{\rm e} - 3) + \frac{1}{4}\,\kappa(\det \boldsymbol{b}^{\rm e} - 1 - \log(\det \boldsymbol{b}^{\rm e})), \qquad \bar{\boldsymbol{b}}^{\rm e} = (\det \boldsymbol{b}^{\rm e})^{-1/3}\,\boldsymbol{b}^{\rm e}. \tag{54}$$

The interaction energy is adopted in the following simple form

$$\phi_{\rm int}(\eta) = \frac{1}{2} H \eta^2, \tag{55}$$

which introduces hardening associated with increasing η. We do not include another term ($\frac{1}{2} H_{\rm grain} \bar{\boldsymbol{\varepsilon}}^{\rm t} \cdot \bar{\boldsymbol{\varepsilon}}^{\rm t} \eta^2$) adopted by Peultier et al. [45] since it seems to be not appropriate in the case of asymmetry of transformation strains in tension and compression addressed here. In fact, in a typical situation when the transformation strain is greater in tension than in compression, the slope of a stress-strain curve would be higher in tension than in compression, and that would contradict experimental observations, e.g., [38, 7].

6 Finite element implementation

The finite element implementation of the model, both in the small- and finite-strain regime, is based on the stationarity conditions that follow from the augmented Lagrangian treatment of the local incremental energy minimization problem, as introduced above. In the context of the finite element method, the *constitutive update problem* at a typical Gauss (integration) point is formulated as follows: given the strain $\boldsymbol{\varepsilon}$ (deformation gradient $\boldsymbol{F}$) at the current[1] time step $t = t_{n+1}$ and the internal variables $\mathbf{h}_n$ at the previous time step $t = t_n$, find the current internal variables $\mathbf{h}$, the current stress $\boldsymbol{\sigma}$ (first Piola-Kirchhoff stress $\boldsymbol{P}$), and the algorithmic constitutive tangent $\partial\boldsymbol{\sigma}/\partial\boldsymbol{\varepsilon}$ ($\partial\boldsymbol{P}/\partial\boldsymbol{F}$). The basic equations of the constitutive update problem are provided in Sections 6.1 and 6.2. Additional modifications that improve robustness of the algorithm are outlined in Appendix A.

The standard issues concerning the finite element formulation, including the virtual work principle, finite element discretization, numerical integration, assembly and solution of the global equilibrium equations, are not discussed here.

Finite element implementation has been carried out using the symbolic code generation system *AceGen/AceFEM* [48, 49]. This system extends the symbolic capabilities of *Mathematica* (www.wolfram.com) with the automatic differentiation (AD) technique, simultaneous optimization of expressions, and automatic generation of program code. Accordingly, in the presentation below, we indicate the differentiation whenever necessary in the formulation, but we do not provide explicit expressions resulting from differentiation as these are obtained automaticaly using the AD technique implemented in *AceGen*.

[1]The subscript $n+1$ denoting the quantities at the current time step $t = t_{n+1}$ is omitted for brevity.

Application of these automation tools leads to *exact* linearization of the nonlinear finite element equations (consistent tangent matrix) which guarantees quadratic convergence rate of the global Newton method and results in a robust and efficient finite-element implementation of the model, as illustrated by the numerical examples below.

6.1 Constitutive update problem: small-strain model

The principle of the algorithmic treatment of the present local incremental minimization problem is rather simple, although, as shown later, it cannot be applied directly as the problem is ill-posed for $\eta \approx 0$.

The constitutive update problem amounts to solving the stationarity conditions (31), which are written here in the form

$$\mathbf{Q}(\boldsymbol{\varepsilon}, \mathbf{h}) = \frac{\partial L(\boldsymbol{\varepsilon}, \mathbf{h})}{\partial \mathbf{h}} = \mathbf{0}, \tag{56}$$

with the Lagrangian function L given by (30) and $\mathbf{h} = \{\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}, \eta, \lambda, \mu, \kappa\}$ (the local residual $\mathbf{Q}$ depends algorithmically also on $\mathbf{h}_n$; however, this dependence is not indicated here). The nonlinear equations (56) are solved iteratively using the Newton method,

$$\mathbf{h}^{i+1} = \mathbf{h}^i + \Delta\mathbf{h}^i, \qquad \Delta\mathbf{h}^i = -\left(\frac{\partial \mathbf{Q}}{\partial \mathbf{h}}\right)^{-1} \mathbf{Q}(\boldsymbol{\varepsilon}, \mathbf{h}^i), \tag{57}$$

where $\partial\mathbf{Q}/\partial\mathbf{h}$ is the tangent matrix of the local problem (dependent tangent operator in the terminology of [50]). The converged solution $\mathbf{h}$ depends on the strain $\boldsymbol{\varepsilon}$ through the local problem $\mathbf{Q}(\boldsymbol{\varepsilon}, \mathbf{h}) = \mathbf{0}$. The derivative of this implicit dependence is obtained by taking the total derivative of $\mathbf{Q}$ with respect to $\boldsymbol{\varepsilon}$, viz.

$$\mathbf{Q}(\boldsymbol{\varepsilon}, \mathbf{h}(\boldsymbol{\varepsilon})) = \mathbf{0}, \qquad \frac{\partial \mathbf{Q}}{\partial \boldsymbol{\varepsilon}} + \frac{\partial \mathbf{Q}}{\partial \mathbf{h}} \frac{\partial \mathbf{h}}{\partial \boldsymbol{\varepsilon}} = \mathbf{0}, \tag{58}$$

and upon solving $(58)_2$ we have

$$\frac{\partial \mathbf{h}}{\partial \boldsymbol{\varepsilon}} = -\left(\frac{\partial \mathbf{Q}}{\partial \mathbf{h}}\right)^{-1} \frac{\partial \mathbf{Q}}{\partial \boldsymbol{\varepsilon}} = -\left(\frac{\partial^2 L}{\partial \mathbf{h} \partial \mathbf{h}}\right)^{-1} \frac{\partial^2 L}{\partial \mathbf{h} \partial \boldsymbol{\varepsilon}}. \tag{59}$$

As indicated in (32), the stress $\boldsymbol{\sigma}$ can be obtained as

$$\boldsymbol{\sigma} = \frac{\partial \phi}{\partial \boldsymbol{\varepsilon}} = \frac{\partial L(\boldsymbol{\varepsilon}, \mathbf{h})}{\partial \boldsymbol{\varepsilon}}. \tag{60}$$

The algorithmic constitutive tangent is then determined as the total derivative of $\boldsymbol{\sigma}$ with respect to $\boldsymbol{\varepsilon}$ taking into account the implicit dependence of $\mathbf{h}$ on $\boldsymbol{\varepsilon}$ expressed by (59),

$$\frac{\partial \boldsymbol{\sigma}}{\partial \boldsymbol{\varepsilon}} = \frac{\partial^2 L}{\partial \boldsymbol{\varepsilon} \partial \boldsymbol{\varepsilon}} + \frac{\partial^2 L}{\partial \boldsymbol{\varepsilon} \partial \mathbf{h}} \frac{\partial \mathbf{h}}{\partial \boldsymbol{\varepsilon}} = \frac{\partial^2 L}{\partial \boldsymbol{\varepsilon} \partial \boldsymbol{\varepsilon}} - \frac{\partial^2 L}{\partial \boldsymbol{\varepsilon} \partial \mathbf{h}} \left(\frac{\partial^2 L}{\partial \mathbf{h} \partial \mathbf{h}}\right)^{-1} \frac{\partial^2 L}{\partial \mathbf{h} \partial \boldsymbol{\varepsilon}}. \tag{61}$$

The constitutive tangent is symmetric which is apparent from the structure of the right-most expression in (61).

As already mentioned, the procedure presented above fails when η, the volume fraction of martensite, is close to zero. A special treatment is thus need to circumvent this problem. To this end, let us examine the stationarity condition corresponding to $\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}$,

$$\frac{\partial L}{\partial \bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}} = \eta \left(-\frac{\partial \phi}{\partial \boldsymbol{\varepsilon}} + \mu \frac{\partial g(\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}_{\mathrm{dev}})}{\partial \bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}} + \kappa \frac{\partial (\operatorname{tr} \bar{\boldsymbol{\varepsilon}}^{\mathrm{t}})}{\partial \bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}} \right) = \mathbf{0}. \tag{62}$$

We note that $\partial \phi / \partial \boldsymbol{\varepsilon} = \boldsymbol{\sigma}$. Accordingly, when $\eta > 0$, this equation ensures that the stress deviator is proportional to the normal to the surface $g(\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}_{\mathrm{dev}}) = 0$. This means that $\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}$ depends on the direction of the current deviatoric stress, as expected. However, this dependence becomes singular as $\eta \to 0$. In fact, for $\eta = 0$, the response is purely elastic, and the free energy does not depend on $\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}$, hence the (incremental) energy is minimized for arbitrary $\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}$.

In order to arrive at a formulation that would be valid for all η, we notice that, upon scaling by $1/\eta$, the stationarity condition (62) yields the desired relationship between $\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}$ and $\boldsymbol{\sigma}$. A similar discussion applies to the stationarity conditions corresponding to μ and κ. Accordingly, the internal variables are grouped in two vectors $\mathbf{h}_1 = \{\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}, \mu, \kappa\}$ and $\mathbf{h}_2 = \{\eta, \lambda\}$. A scaled residual $\hat{\mathbf{Q}}_1$ is then introduced,

$$\hat{\mathbf{Q}}_1 = \frac{1}{\eta} \frac{\partial L(\boldsymbol{\varepsilon}, \mathbf{h}_1, \mathbf{h}_2)}{\partial \mathbf{h}_1}, \tag{63}$$

and the stationarity condition $\mathbf{Q}_1 = \partial L / \partial \mathbf{h}_1 = \mathbf{0}$ is replaced by $\hat{\mathbf{Q}}_1 = \mathbf{0}$, so that a modified local residual $\hat{\mathbf{Q}}$ is actually solved,

$$\hat{\mathbf{Q}}(\boldsymbol{\varepsilon}, \mathbf{h}) = \{\hat{\mathbf{Q}}_1, \mathbf{Q}_2\} = \mathbf{0}, \tag{64}$$

where $\mathbf{Q}_2 = \partial L / \partial \mathbf{h}_2 = \mathbf{0}$ are the original stationarity conditions corresponding to η and λ. With this modification, the local constitutive problem is well defined both for $\eta = 0$ and for $\eta \neq 0$.

The solution procedure follows that presented above with $\mathbf{Q}$ replaced by $\hat{\mathbf{Q}}$. The local constitutive problem is solved using the Newton method as in (57). In analogy to (59), the derivative of $\mathbf{h}$ with respect to $\boldsymbol{\varepsilon}$ is given by

$$\frac{\partial \mathbf{h}}{\partial \boldsymbol{\varepsilon}} = - \left(\frac{\partial \hat{\mathbf{Q}}}{\partial \mathbf{h}} \right)^{-1} \frac{\partial \hat{\mathbf{Q}}}{\partial \boldsymbol{\varepsilon}}. \tag{65}$$

Finally, the stress and the constitutive tangent are equal to

$$\boldsymbol{\sigma} = \frac{\partial L}{\partial \boldsymbol{\varepsilon}}, \qquad \frac{\partial \boldsymbol{\sigma}}{\partial \boldsymbol{\varepsilon}} = \frac{\partial^2 L}{\partial \boldsymbol{\varepsilon} \partial \boldsymbol{\varepsilon}} - \frac{\partial^2 L}{\partial \boldsymbol{\varepsilon} \partial \mathbf{h}} \left(\frac{\partial \hat{\mathbf{Q}}}{\partial \mathbf{h}} \right)^{-1} \frac{\partial \hat{\mathbf{Q}}}{\partial \boldsymbol{\varepsilon}}. \tag{66}$$

The symmetry of the tangent operator is not apparent now. However, the constitutive response is not affected by the modification introduced above, so the symmetry is preserved. This has also been verified numerically.

Remark 5. A domain of elastic response (or a kind of a "yield" surface) is not used in the present formulation. Accordingly, the constitutive update problem is to be solved always, also in a fully elastic state. This is in contrast to the commonly used plasticity-based formulations which typically employ predictor-corrector schemes, and the constitutive update

Table 1: Material parameters for NiTi shape memory alloy.

E	ν	$\Delta\phi_0$	H	f_c	ϵ_T	α	β	$\bar{\varepsilon}^\mathrm{t}_\mathrm{vol}$
51.7 GPa	0.3	33 MJ/m^3	4.5 MJ/m^3	11 MJ/m^3	0.075	1.4	1.1	0.00023

problem does not involve solution of nonlinear equations, if the trial state is elastic. At the same time, the plasticity-based formulations often involve another predictor-corrector scheme associated with saturation of transformation, cf. [20, 21]. Specifically, if the solution of the constitutive update problem is not admissible (in the present notation, that would correspond to $\eta > 1$ or $g(\boldsymbol{\varepsilon}^\mathrm{t}_\mathrm{dev}) > 0$) then the constitutive update problem must be solved again in a modified form. This is avoided in the present formulation as the transformation saturation condition is naturally handled by the augmented Lagrangian method, and the nonlinear equations of the constitutive update problem are solved only once.

6.2 Constitutive update problem: finite-strain model

Algorithmic treatment of the finite-strain model follows *exactly* that of the small strain model, including the modification introduced to properly handle the case of $\eta \approx 0$. The modified local residual is defined as

$$\hat{\mathbf{Q}}(\boldsymbol{F}, \mathbf{h}) = \{\hat{\mathbf{Q}}_1, \mathbf{Q}_2\} = \mathbf{0}, \qquad \hat{\mathbf{Q}}_1 = \frac{1}{\eta}\frac{\partial L(\boldsymbol{F}, \mathbf{h}_1, \mathbf{h}_2)}{\partial \mathbf{h}_1}, \qquad \mathbf{Q}_2 = \frac{\partial L(\boldsymbol{F}, \mathbf{h}_1, \mathbf{h}_2)}{\partial \mathbf{h}_2}, \tag{67}$$

where the Lagrangian function L is given by (46), $\mathbf{h}_1 = \{\bar{\boldsymbol{e}}^\mathrm{t}, \mu, \kappa\}$ and $\mathbf{h}_2 = \{\eta, \lambda\}$, and the resulting nonlinear equations are solved using the Newton method, as indicated in (57). In analogy to (65) and (66), we have

$$\frac{\partial \mathbf{h}}{\partial \boldsymbol{F}} = -\left(\frac{\partial \hat{\mathbf{Q}}}{\partial \mathbf{h}}\right)^{-1} \frac{\partial \hat{\mathbf{Q}}}{\partial \boldsymbol{F}} \tag{68}$$

and

$$\boldsymbol{P} = \frac{\partial L}{\partial \boldsymbol{F}}, \qquad \frac{\partial \boldsymbol{P}}{\partial \boldsymbol{F}} = \frac{\partial^2 L}{\partial \boldsymbol{F} \partial \boldsymbol{F}} - \frac{\partial^2 L}{\partial \boldsymbol{F} \partial \mathbf{h}} \left(\frac{\partial \hat{\mathbf{Q}}}{\partial \mathbf{h}}\right)^{-1} \frac{\partial \hat{\mathbf{Q}}}{\partial \boldsymbol{F}}. \tag{69}$$

As in the case of (61), the tangent operator is symmetric.

7 Numerical examples

7.1 Material parameters

All the computations reported in this paper have been carried out for a NiTi shape memory alloy, and the adopted model parameters are provided in Table 1. Material parameters characterizing the free energy function (E, ν, $\Delta\phi_0$ and H), the dissipation function (f_c), and the maximum strain in uniaxial tension (ϵ_T) have been adopted such that the pseudoelastic response in uniaxial tension matches the response of the model of Arghavani et al. [21] (pseudoelastic NiTi at $T = 37°\mathrm{C}$).

The present model is additionally capable of describing transverse isotropy and tension–compression asymmetry. These two effects are introduced using parameters α and β which

Table 2: Parameters specifying anisotropy and tension compression asymmetry.

	α	β	a	b	c
transverse isotropy, tension–compression asymmetry	1.4	1.1	0.0527	0.462	0.187
isotropy, tension–compression asymmetry	1.4	1.4	0.0527	1.210	0
isotropy, no tension–compression asymmetry	1	1	0.0650	0	0

describe the asymmetry of maximum transformation strains along the axis of transverse isotropy and in the perpendicular direction. Denoting the maximum tensile and compressive strain along the symmetry axis by ϵ_{T} and ϵ_{C}, respectively, we have $\alpha = \epsilon_{\mathrm{T}}/\epsilon_{\mathrm{C}}$. Similarly, for uniaxial tension and compression in the direction perpendicular to the symmetry axis, we have $\beta = \epsilon^*_{\mathrm{T}}/\epsilon^*_{\mathrm{C}}$. Parameters a, b and c in function g in (51) are then expressed in terms of ϵ_{T}, α and β by the following formulae:

$$a = \epsilon_{\mathrm{T}} \left[\frac{3\sqrt{3}}{4(1+\alpha^3)} \right]^{1/3}, \quad b = \frac{\sqrt{3}}{6} \frac{9\alpha^3\beta^3 - 7\alpha^3 + 7\beta^3 - 9}{(1+\alpha^3)(1+\beta^3)}, \quad c = \frac{2\sqrt{3}}{3} \frac{\alpha^3 - \beta^3}{(1+\alpha^3)(1+\beta^3)}. \tag{70}$$

Finally, the volumetric transformation strain $\bar{\varepsilon}^{\mathrm{t}}_{\mathrm{vol}}$ has been determined using the lattice constants reported in [51]. The influence of volumetric transformation strain on the overall response is small, but, for completeness, this effect has also been included in the computations.

In order to illustrate the effect of transverse isotropy and tension–compression asymmetry, the computations have been carried out for three sets of parameters α and β provided in Table 2. An isotropic material with tension–compression asymmetry is obtained for $\alpha = \beta \neq 1$, and an isotropic material with no tension–compression asymmetry is obtained for $\alpha = \beta = 1$. Note that a fixed value of $\epsilon_{\mathrm{T}} = 0.075$ is used in all three cases so that the response in uniaxial tension is identical for the three materials (in case of the transversely isotropic material, this applies to tension along the symmetry axis). The corresponding values of parameters a, b and c in function g are also provided in Table 2.

7.2 Uniaxial response

Uniaxial response predicted by the present model is illustrated in Fig. 3 for uniaxial tension and compression along the axis of transverse isotropy and in the direction perpendicular to this axis. The anisotropy is clearly visible and so is the tension–compression asymmetry in both directions.

For comparison, the stress–strain response in compression corresponding to $\alpha = \beta = 1$ (isotropic material with no tension–compression asymmetry) is also included in Fig. 3 (dashed line). The apparent difference between this response and the response in tension along the symmetry axis is due to finite deformation effects (note that the nominal stress is used in Fig. 3).

For proportional loading, the model is not sensitive to the increment size, see Remark 4. Indeed, the stress-strain response obtained using large load increments (maximum elongation reached in 4 increments; indicated by dots) follows exactly the response obtained using small load increments (50 increments; indicated by solid lines).

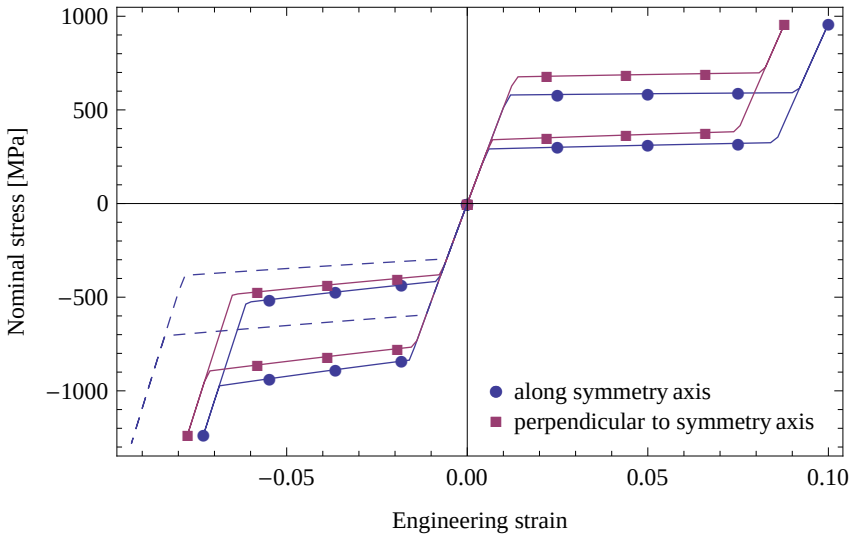


Figure 3: Uniaxial response of NiTi in tension and compression. The dashed line corresponds to $\alpha = \beta = 1$ (isotropy, no tension–compression asymmetry).

7.3 Proportional tension–torsion

Pseudoelastic response to proportional tension–torsion loading is shown in Fig. 4. In the initial, undeformed configuration, the tension axis has been aligned with the axis of transverse isotropy. This corresponds to a typical experimental arrangement in which a combined tension–torsion experiment is performed on a tubular specimen, and the material is transversely isotropic as a result of crystallographic texture, e.g., due to prior drawing or extrusion along the specimen axis.

Five proportional loading paths have been simulated as shown in Fig. 4(a), and the corresponding stress response is shown in Fig. 4(b). The hysteresis loops predicted in the stress response for combined tension and torsion are in a qualitative agreement with those observed experimentally by Helm and Haupt [52]. It has been checked that such hysteresis loops can only be predicted if both anisotropy and tension–compression asymmetry are adequately accounted for. The present model predicts also the hysteresis in the stress response for pure torsion which was not observed in [52]. This effect could possibly be avoided by adopting a more complex form of function g.

7.4 Helical spring

This example is adopted from Ref. [21]. A helical spring (two coils with wire diameter of 4 mm, external diameter of 24 mm, pitch size of 12 mm, and initial length of 28 mm) is clamped at one end and loaded by an axial force of 1525 N at the other end. Figure 5 presents the undeformed and the deformed finite element mesh (5400 eight-node F-bar elements [53], 22050 degrees of freedom). Finite deformations, and in particular finite rotations, are clearly visible in Fig. 5. Upon unloading, the spring recovers its initial shape.

Computations have been performed for the three sets of material parameters provided in Table 2. In case of the transversely isotropic material, the symmetry axis has been specified to follow the axis of the wire so that it is locally tangent to the helix. This corresponds to the drawing texture in the wire.

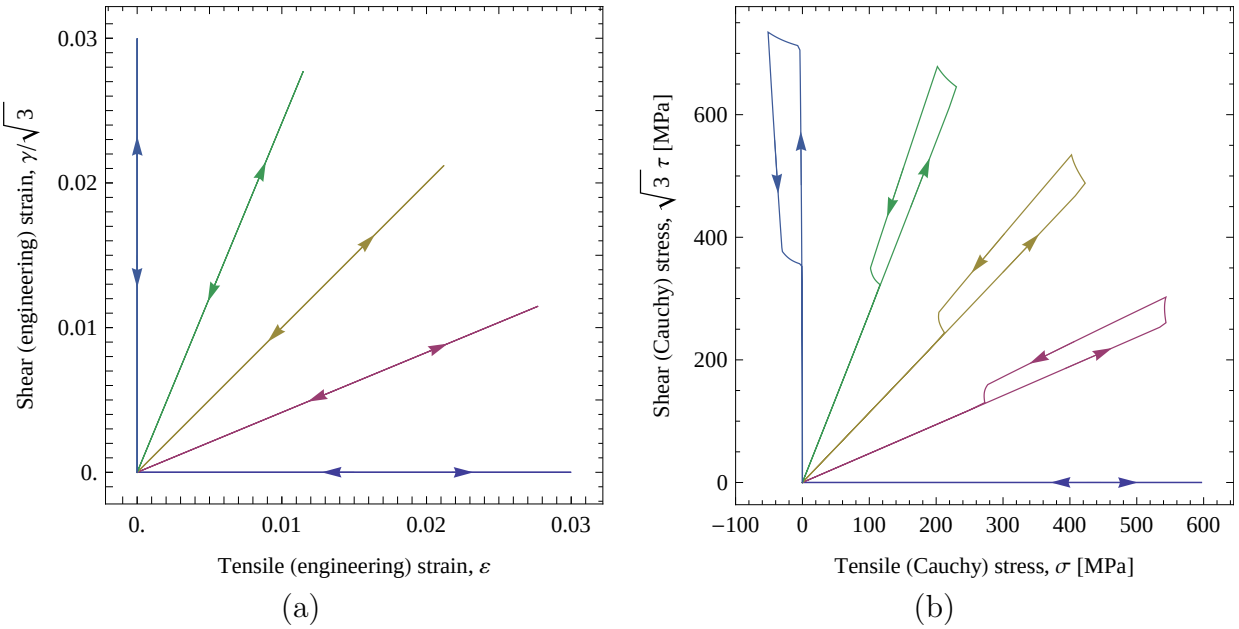


Figure 4: Proportional tension–torsion loading: (a) loading path in the strain space; (b) response in the stress space.

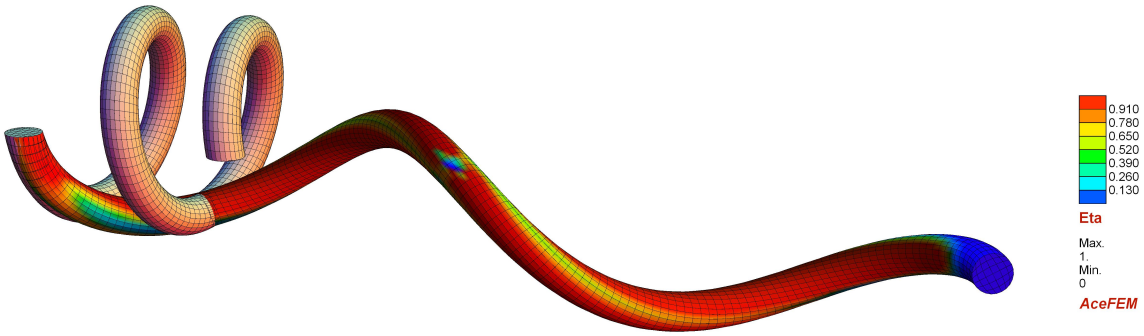


Figure 5: Helical spring: initial and deformed (maximum force) configurations. Color map indicates the distribution of the volume fraction η of martensite in the deformed configuration.

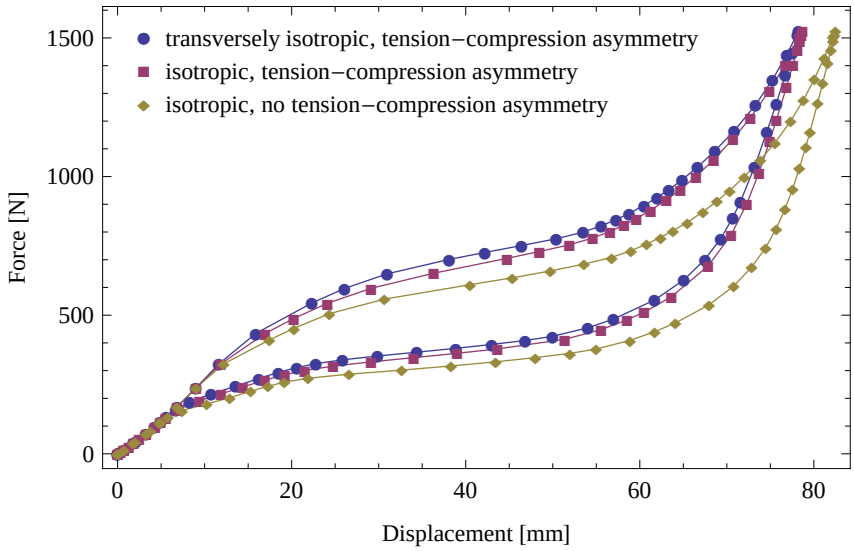


Figure 6: Helical spring: force–displacement curves obtained for the three sets of material parameters in Table 2.

Figure 6 shows the force–displacement curves obtained for the three material models. The effect of tension–compression asymmetry is significant; the effect of anisotropy is also visible though less pronounced. The markers in Fig. 6 indicate the progress of the incremental solution procedure with an automatic step size control.[2]

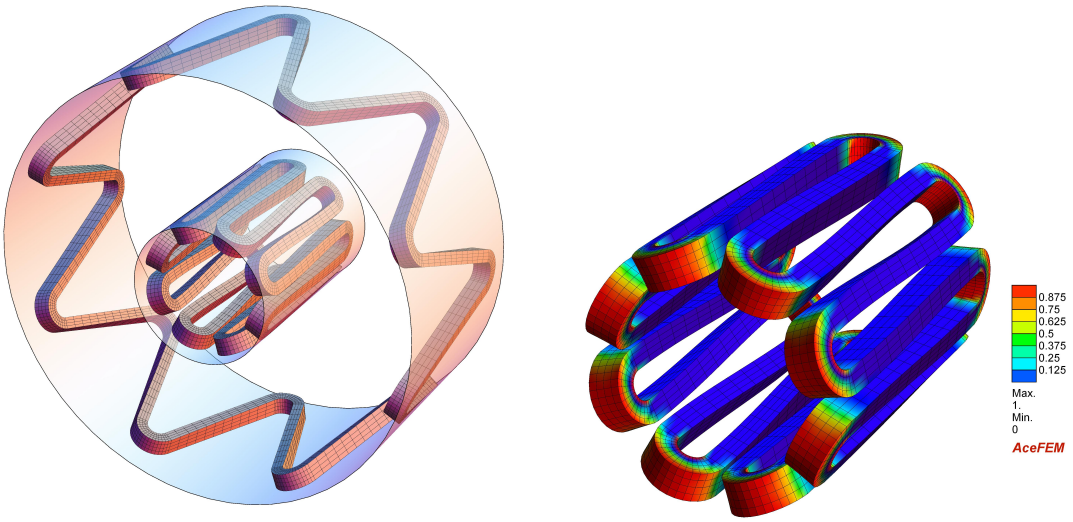
7.5 Crimping of a stent-like structure

Crimping of a stent-like structure is considered in this example, as shown in Fig. 7. The outer radius is 3.68 mm, the cross-section of the rod is 0.2×0.3 mm, so that the inner radius is 3.38 mm, the total length along the axis is 2.85 mm, the length of the straight portions of the rod is 2.4 mm, the included angle is 40 degree, and the mean radius of the curved portions is 0.3 mm. Crimping is imposed by frictionless contact interaction with a rigid cylinder of varying radius. The radius is reduced from $R_0 = 3.68$ mm to $0.32R_0 = 1.18$ mm, cf. Fig. 7(a), and subsequently increased back to R_0. The unilateral contact conditions are enforced using the augmented Lagrangian method [54, 55]. The finite element mesh consists of 7168 eight-node F-bar elements [53] with the total of 35838 degrees of freedom, of which 2240 are contact Lagrange multipliers.

Figure 10 shows the average confining pressure (in the current configuration, i.e., referred to the current radius and current length of the stent) as a function of the reduction of the radius. Qualitatively, the effect of anisotropy and tension–compression asymmetry is similar to that reported in Fig. 6 for the helical spring.

The markers in Fig. 8 indicate the solution obtained using large load increments (automatic step size control), while the solid lines correspond to a fixed, small load increment such that the complete loading cycle requires 136 increments. Due to a predominantly proportional deformation pattern, the former solution follows exactly the latter, as ex-

[2]All computations are performed using an automatic step size control procedure in which the desired number of Newton iterations is prescribed (here, equal to 8), and the current load increment is increased (decreased) when the number of iterations at the pervious increment is smaller (larger) then the desired number.



(a) (b)

Figure 7: Crimping of a stent-like structure: (a) initial and deformed configurations; (b) distribution of the volume fraction η of martensite at the maximum load.

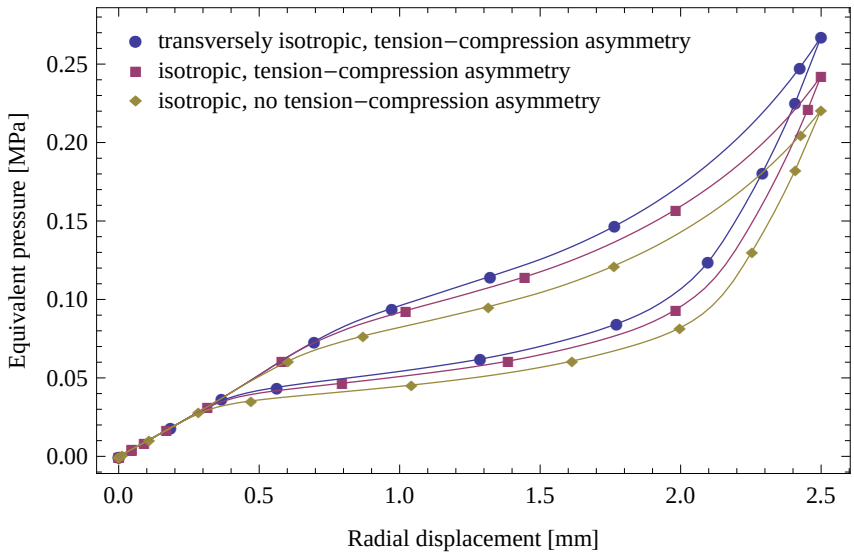
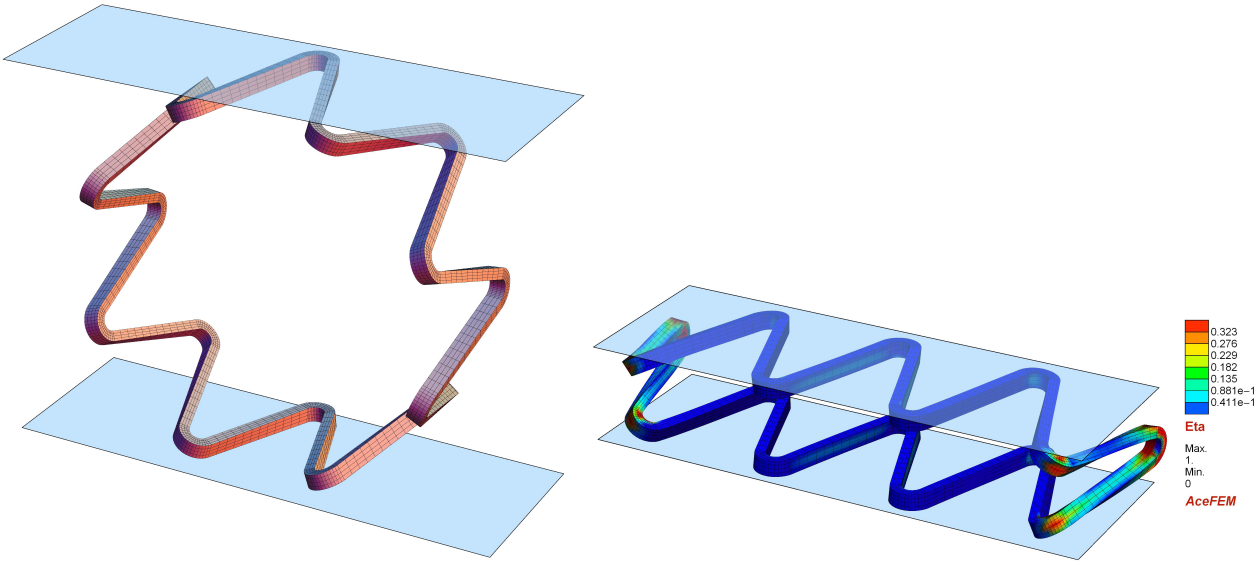


Figure 8: Crimping of a stent-like structure: average confining pressure as a function of radial displacement (markers—large load increments, solid lines—small increments).



(a) (b)

Figure 9: Compression of a stent-like structure: (a) initial configuration; (b) configuration corresponding to maximum height reduction (color map indicates distribution of the martensite volume fraction η).

pected, cf. Remark 4.

The results presented in Fig. 8 illustrate also the robustness of the present formulation and implementation. The incremental solution procedure can proceed with large load increments so that the complete loading–unloading cycle takes only 16 or 17 increments, depending on the material model specifications.

7.6 Compression of a stent-like structure

In this example, the stent-like structure of the previous subsection is compressed between two rigid planes (frictionless contact), cf. Fig. 9. The height is reduced from $2R_0$ to $0.5R_0$, and the initial shape is recovered upon unloading.

Figure 10 shows the compression force as a function of the height reduction. In addition to the three cases of pseudoelastic material behaviour, the response of a purely elastic material is also included in Fig. 10. Again, the solution obtained using large load increments (automatic step size control; indicated by markers) follows exactly the solution obtained using small load increments (indicated by solid lines). Depending on the material model, 49–53 increments are needed to simulate the complete loading–unloading cycle (33 increments in the elastic case). However, it is the initial elastic stage that requires small load increments due to contact zones evolving in a complex manner, while at the later stage the solution proceeds with relatively large increments.

In order to assess computational efficiency of the present model and its finite element implementation, in Table 3 we provide the number of steps, the number of global Newton iterations, the total computation time, and the average evaluation time of element tangent matrix and residual vector (K&R time; normalized by the evaluation time of the hyperelastic material). The number Newton iterations includes all iterations; the average number of iterations per step in the steps that successfully converged is between 6.5 and 7.9. For the three cases of pseudoelastic behaviour, the total computation time is

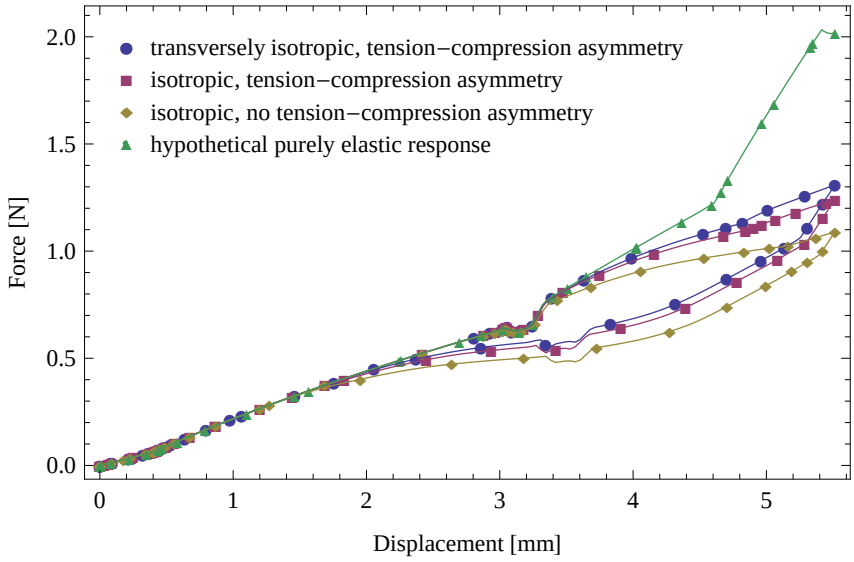


Figure 10: Compression of a stent-like structure: force-displacement curve.

Table 3: Compression of a stent-like structure: comparison of computational efficiency.

	Number of steps	Number of iterations	Total time [s]	Average K&R time
transverse isotropy, tension–compression asymmetry	50	558	766	4.59
isotropy, tension–compression asymmetry	53	550	769	4.69
isotropy, no tension–compression asymmetry	49	559	696	4.00
hypothetical purely elastic response	33	370	243	1.

about three times that of the hyperelastic material. The average K&R time directly characterizes the computational effort spent on iterative solution of the constitutive update problem. The average K&R time of the present SMA model is only 4–5 times that of the standard hyperelastic model (which does not involve any iterative loop at the element integration points).

8 Conclusions

A finite-strain three-dimensional model of pseudoelasticity in shape memory alloys has been developed within the incremental energy minimization framework. It has been shown that the constitutive constraints on the volume fraction of martensite and on the thermodynamic driving force for transformation can both be conveniently enforced using the augmented Lagrangian method with a *single* Lagrange multiplier. Directional stress-dependence of the maximum recoverable transformation strain has been incorporated by introducing a hypersurface in the transformation strain space. A methodology for an easy reformulation of the model from the small-strain to the finite-deformation regime has been proposed.

The focus of the work has been on the general formulation rather than on sophisticated constitutive modelling, still the model in its basic version is capable of modelling the tension–compression asymmetry of isotropic shape memory alloys. Furthermore, the effect of crystallographic texture leading to transversely isotropic material response is

also included in the model. The specific form of the model does not include dissipation due to reorientation of martensite which cannot be handled by the proposed algorithmic treatment in its present form.

The model has been implemented in the finite element code, and its robustness has been illustrated by numerical examples involving large deformations. In particular, it has been demonstrated that the incremental solution algorithm can proceed with large load increments, while the results are fully insensitive to the increment size within certain limits.

Appendix A. Refinements of the algorithm

Two additional minor modifications of the general approach are presented here which, however, significantly improve the robustness of the resulting finite-element formulation.

The incremental minimization problem (27) involves two equality constraints that result from the definition of the set $\bar{\mathcal{P}}$ of limit transformation strains in (19) (to fix the attention, the small-strain model is discussed here; the same argument applies to the finite-strain version). In the saddle-point problem (30), these constraints are enforced using the classical Lagrange multiplier technique. In the actual implementation, the augmented Lagrangian method is used instead, which simply amounts to the following replacement in the Lagrangian L in (29):

$$\eta\mu g(\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}_{\mathrm{dev}}) \quad \rightarrow \quad \eta\left(\mu g(\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}_{\mathrm{dev}}) + \frac{1}{2}\varrho^* g^2(\bar{\boldsymbol{\varepsilon}}^{\mathrm{t}}_{\mathrm{dev}})\right), \tag{71}$$

where $\varrho^* > 0$ is a parameter, and similarly for the second equality constraint. It has been observed that the convergence radius of the Newton method applied to solve the constitutive update problem (56)–(57) is significantly increased by selecting a sufficiently high value of ϱ^*, while, obviously, the solution itself is not affected (the value of 10^6 GPa has been used in the computations).

The convergence of the local constitutive problem has been further improved by modifying the local residual $\mathbf{Q}_2$. Specifically, the following residual is actually solved,

$$\mathbf{Q}_2^* = \frac{\partial}{\partial \mathbf{h}_2}[\phi(\boldsymbol{\varepsilon}, \mathbf{h}_1, \mathbf{h}_2) + l_{\mathrm{t}}(\mathbf{h}_2)] = \mathbf{0}, \tag{72}$$

instead of $\mathbf{Q}_2 = \partial L/\partial \mathbf{h}_2 = \mathbf{0}$. The difference between $\mathbf{Q}_2^*$ and $\mathbf{Q}_2$ is that the terms corresponding to the equality constraints enforced by Lagrange multipliers (as discussed above) are not included in $\mathbf{Q}_2^*$. However, we note that these terms vanish at the solution so that the solution is not affected, while the above modification appears highly beneficial for convergence of the Newton method.

Acknowledgement This work has been partially supported by the Ministry of Science and Higher Education in Poland under grant No. N N501 071935.

References

[1] K. Otsuka and C. M. Wayman, editors. *Shape Memory Materials*. Cambridge University Press, 1998.

[2] K. Bhattacharya. *Microstructure of martensite: why it forms and how it gives rise to the shape-memory effect*. Oxford University Press, Oxford, 2003.

[3] E. Patoor, D. C. Lagoudas, P. B. Entchev, L. C. Brinson, and X. Gao. Shape memory alloys, Part I: General properties and modeling of single crystals. *Mech. Mater.*, 38:391–429, 2006.

[4] D. C. Lagoudas, P. B. Entchev, P. Popov, E. Patoor, L. C. Brinson, and X. Gao. Shape memory alloys, Part II: Modeling of polycrystals. *Mech. Mater.*, 38:430–462, 2006.

[5] M. S. Wechsler, D. S. Lieberman, and T. A. Read. On the theory of the formation of martensite. *Trans. AIME J. Metals*, 197:1503–1515, 1953.

[6] J. M. Ball and R. D. James. Fine phase mixtures as minimizers of energy. *Arch. Ration. Mech. Anal.*, 100:13–50, 1987.

[7] P. Thamburaja and L. Anand. Polycrystalline shape-memory materials: effect of crystallographic texture. *J. Mech. Phys. Solids*, 49:709–737, 2001.

[8] M. Kružik, A. Mielke, and T. Roubiček. Modelling of microstructure and its evolution in shape-memory-alloy single-crystals, in particular in CuAlNi. *Meccanica*, 40:389–418, 2005.

[9] G. Maciejewski, S. Stupkiewicz, and H. Petryk. Elastic micro-strain energy at the austenite–twinned martensite interface. *Arch. Mech.*, 57:277–297, 2005.

[10] S. Stupkiewicz and H. Petryk. Finite-strain micromechanical model of stress-induced martensitic transformations in shape memory alloys. *Mater. Sci. Eng. A*, 438–440:126–130, 2006.

[11] E. Stein and G. Sagar. Theory and finite element computation of cyclic martensitic phase transformation at finite strain. *Int. J. Num. Meth. Engng.*, 74:1–31, 2008.

[12] A. Sengupta, P. Papadopoulos, and R. L. Taylor. Multiscale finite element modelling of superelasticity in Nitinol polycrystals. *Comp. Mech.*, 43:573–584, 2009.

[13] F. Auricchio and R. L. Taylor. Shape-memory alloys: modelling and numerical simulations of finite-strain superelastic behavior. *Comp. Meth. Appl. Mech. Engng.*, 143:175–194, 1997.

[14] Ch. Müller and O. T. Bruhns. A thermodynamic finite-strain model for pseudoelastic shape memory alloys. *Int. J. Plasticity*, 22:1658–1682, 2006.

[15] D. Helm. Thermomechanics of martensitic phase transitions in shape memory alloys. I. Constitutive theories for small and large deformations. *J. Mech. Mater. Struct.*, 2:87–112, 2007.

[16] H. Pan, P. Thamburaja, and F. S. Chau. An isotropic-plasticity-based constitutive model for martensitic reorientation and shape-memory effect in shape-memory alloys. *Int. J. Sol. Struct.*, 44:7688–7712, 2007.

[17] A. Ziółkowski. Three-dimensional phenomenological thermodynamic model of pseudoelasticity of shape memory alloys at finite strains. *Continuum Mech. Thermodyn.*, 19:379–398, 2007.

[18] S. Reese and D. Christ. Finite deformation pseudo-elasticity of shape memory alloys – Constitutive modelling and finite element implementation. *Int. J. Plasticity*, 24:455–482, 2008.

[19] D. Christ and S. Reese. A finite element model for shape memory alloys considering thermomechanical couplings at large strains. *Int. J. Sol. Struct.*, 46:3694–3709, 2009.

[20] V. Evangelista, S. Marfia, and E. Sacco. A 3D SMA constitutive model in the framework of finite strain. *Int. J. Num. Meth. Engng.*, 81:761–785, 2010.

[21] J. Arghavani, F. Auricchio, R. Naghdabadi, and A. Reali. On the robustness and efficiency of integration algorithms for a 3D finite strain phenomenological SMA constitutive model. *Int. J. Num. Meth. Engng.*, 85:107–134, 2011.

[22] J. Arghavani, F. Auricchio, and R. Naghdabadi. A finite strain kinematic hardening constitutive model based on Hencky strain: General framework, solution algorithm and application to shape memory alloys. *Int. J. Plasticity*, 27:940–961, 2011.

[23] H. Petryk. Incremental energy minimization in dissipative solids. *C. R. Mecanique*, 331:469–474, 2003.

[24] H. Petryk. On energy criteria of plastic instability. In *Plastic Instability, Proc. Considère Memorial*, pages 215–226. Ecole Nat. Ponts Chauss., Paris, 1985.

[25] H. Petryk and K. Thermann. On discretized plasticity problems with bifurcations. *Int. J. Sol. Struct.*, 29:745–765, 1992.

[26] H. Petryk and K. Thermann. Post-critical plastic deformation of biaxially stretched sheets. *Int. J. Sol. Struct.*, 33:689–705, 1996.

[27] H. Petryk and K. Thermann. Post-critical plastic deformation in incrementally nonlinear materials. *J. Mech. Phys. Solids*, 50:925–954, 2002.

[28] B. Fedelich and A. Ehrlaher. An analysis of stability of equilibrium and of quasi-static transformations on the basis of the dissipation function. *Eur. J. Mech. A/Solids*, 16:833–855, 1997.

[29] M. Ortiz and E. A. Repetto. Nonconvex energy minimization and dislocation structures in ductile single crystals. *J. Mech. Phys. Solids*, 47:397–462, 1999.

[30] A. Mielke, F. Theil, and V. I. Levitas. A variational formulation of rate-independent phase transformations using an extremum principle. *Arch. Ration. Mech. Anal.*, 162:137–177, 2002.

[31] C. Carstensen, K. Hackl, and A. Mielke. Non-convex potentials and microstructures in finite-strain plasticity. *Proc. R. Soc. Lond. A*, 458:299–317, 2002.

[32] C. Miehe, M. Lambrecht, and E. Gurses. Analysis of material instabilities in inelastic solids by incremental energy minimization and relaxation methods: evolving deformation microstructures in finite plasticity. *J. Mech. Phys. Solids*, 52:2725–2769, 2004.

[33] M. Schurig and A. Bertram. Relaxation in multi-mode plasticity with a rate-potential. *Comp. Mat. Sci.*, 32:524–531, 2005.

[34] J. Mosler. Variationally consistent modeling of finite strain plasticity theory with non-linear kinematic hardening. *Comp. Meth. Appl. Mech. Engng.*, 199:2753–2764, 2010.

[35] M. Arndt, M. Griebel, V. Novak, T. Roubiček, and P. Sittner. Martensitic transformation in NiMnGa single crystals: Numerical simulation and experiments. *Int. J. Plasticity*, 22:1943–1961, 2006.

[36] H. Petryk, S. Stupkiewicz, and G. Maciejewski. Interfacial energy and dissipation in martensitic phase transformations. Part II: Size effects in pseudoelasticity. *J. Mech. Phys. Solids*, 58:373–389, 2010.

[37] H. Petryk and S. Stupkiewicz. Interfacial energy and dissipation in martensitic phase transformations. Part I: Theory. *J. Mech. Phys. Solids*, 58:390–408, 2010.

[38] L. Orgeas and D. Favier. Stress-induced martensitic transformation of a NiTi alloy in isothermal shear, tension and compression. *Acta Mater.*, 46(15):5579–5591, 1998.

[39] P. Šittner and V. Novák. Anisotropy of martensitic transformations in modeling of shape memory alloy polycrystals. *Int. J. Plasticity*, 16:1243–1268, 2000.

[40] K. Gall and H. Sehitoglu. The role of texture in tension–compression asymmetry in polycrytalline NiTi. *Int. J. Plasticity*, 15:69–92, 1999.

[41] B. Raniecki and Ch. Lexcellent. Thermodynamics of isotropic pseudoelasticity in shape memory alloys. *Eur. J. Mech. A/Solids*, 17:185–205, 1998.

[42] A. Sadjadpour and K. Bhattacharya. A micromechanics inspired constitutive model for shape-memory alloys. *Smart Mater. Struct.*, 16:1751–1765, 2007.

[43] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Athena Scientific, Belmont, MA, 2nd edition, 1996.

[44] A. Bertram. Thermo-mechanical constitutive equations for the description of shape memory effects in alloys. *Nucl. Eng. Design*, 74:173–182, 1982.

[45] B. Peultier, T. Ben Zineb, and E. Patoor. Macroscopic constitutive law of shape memory alloy thermomechanical behaviour. Application to structure computation by FEM. *Mech. Mater.*, 38:510–524, 2006.

[46] B. Raniecki, C. Lexcellent, and K. Tanaka. Thermodynamic models of pseudoelastic behaviour of shape memory alloys. *Arch. Mech.*, 44:261–284, 1992.

[47] J. C. Simo and T. J. R. Hughes. *Computational Inelasticity*. Springer-Verlag, New York, 1998.

[48] J. Korelc. Multi-language and multi-environment generation of nonlinear finite element codes. *Engineering with Computers*, 18:312–327, 2002.

[49] J. Korelc. Automation of primal and sensitivity analysis of transient coupled problems. *Comp. Mech.*, 44:631–649, 2009.

[50] P. Michaleris, D. A. Tortorelli, and C. A. Vidal. Tangent operators and design sensitivity formulations for transient non-linear coupled problems with applications to elastoplasticity. *Int. J. Num. Meth. Engng.*, 37:2471–2499, 1994.

[51] Y. Kudoh, M. Tokonami, S. Miyazaki, and K. Otsuka. Crystal-structure of the martensite in Ti-49.2at%Ni alloy analyzed by the single-crystal X-ray-diffration method. *Acta Metall.*, 33:2049–2056, 1985.

[52] D. Helm and P. Haupt. Shape memory behaviour: modelling within continuum thermomechanics. *Int. J. Sol. Struct.*, 40:827–849, 2003.

[53] E. A. de Souza Neto, D. Perić, M. Dutko, and D. R. J. Owen. Design of simple low order finite elements for large strain analysis of nearly incompressible solids. *Int. J. Sol. Struct.*, 33:3277–3296, 1996.

[54] P. Alart and A. Curnier. A mixed formulation for frictional contact problems prone to Newton like solution methods. *Comp. Meth. Appl. Mech. Engng.*, 92:353–375, 1991.

[55] J. Lengiewicz, J. Korelc, and S. Stupkiewicz. Automation of finite element formulations for large deformation contact problems. *Int. J. Num. Meth. Engng.*, 85:1252–1279, 2011.