IMA Journal of Applied Mathematics (2011) **76**, 193–216
doi:10.1093/imamat/hxq058
Advance Access publication on December 16, 2010

A model of shape memory alloys taking into account plasticity

MARTIN KRUŽÍK
Institute of Information Theory and Automation of the Czech Academy of Sciences (ASCR), Pod vodárenskou věží 4, 182 08 Prague, Czech Republic and Faculty of Civil Engineering, Czech Technical University, Thákurova 7, 166 29 Prague, Czech Republic

AND

JOHANNES ZIMMER*
Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK
*Corresponding author: zimmer@maths.bath.ac.uk

[Received on 24 July 2009; revised on 11 May 2010; accepted on 16 November 2010]

We propose a phenomenological model for the evolutionary behaviour of shape memory alloys, where the possibility of plastic deformation is taken in to account. Two dissipative mechanisms are considered, namely the dissipation associated with solid–solid (martensitic) phase transformations and plastic dissipation. The plastic contribution may lead to an irreversibility of the evolution. The existence of a so-called energetic solution is established for a suitable relaxation of the problem in the space of Young measures.

Keywords: elasto-plasticity; energetic solution; plastic strain gradients.

1. The mathematical model

'Shape memory alloys' (SMAs) have been the focus of many investigations in the last decade. This interest can partially be attributed to the shape memory effect itself (see Section 1.1) but even more the non-convexity of the Helmholtz energy density due to the co-existence of several variants, which poses a significant mathematical challenge. An established setting for models for SMAs is that of non-linear elasticity. We deviate here from this context by including elasto-plastic effects. For SMAs, this seems to be a relatively new line of research. While SMAs can undergo many cycles of loading and unloading, plastic effects can have a significant influence on material properties. For example, cyclic plasticity may occur, which can negatively affect the performance of the material.

Microscopically, martensitic shape memory materials exhibit dislocations and thus plastic effects. For a superelastic NiTi wire, a transmission electron microscopy of the microstructure shows the presence of dislocations on the $\{110\}$ slip system (see Fig. 1, left panel). It is remarkable that the dislocation appear not after a large numbers of cycles; the photograph of Fig. 1 was taken after 10 tensile cycles. The corresponding hysteretic stress–strain relationship evolves during these 10 cycles due to plastic effects as shown in Fig. 1 (right panel). We refer the reader to Novák *et al*. (2009) and Delville *et al*. (2011) for transmission electron analysis of dislocations during cycling of martensitic materials.

Also, molecular dynamics (MD) simulations of martensitic materials show the creation of dislocations. Kastner & Ackland (2009) investigate a 2D model of martensitic materials, using binary Lennard–Jones potentials in a classic MD setting (a thermostatted version of the Verlet algorithm). Dislocation lines can be seen to originate in a setting with crystalline initial data.

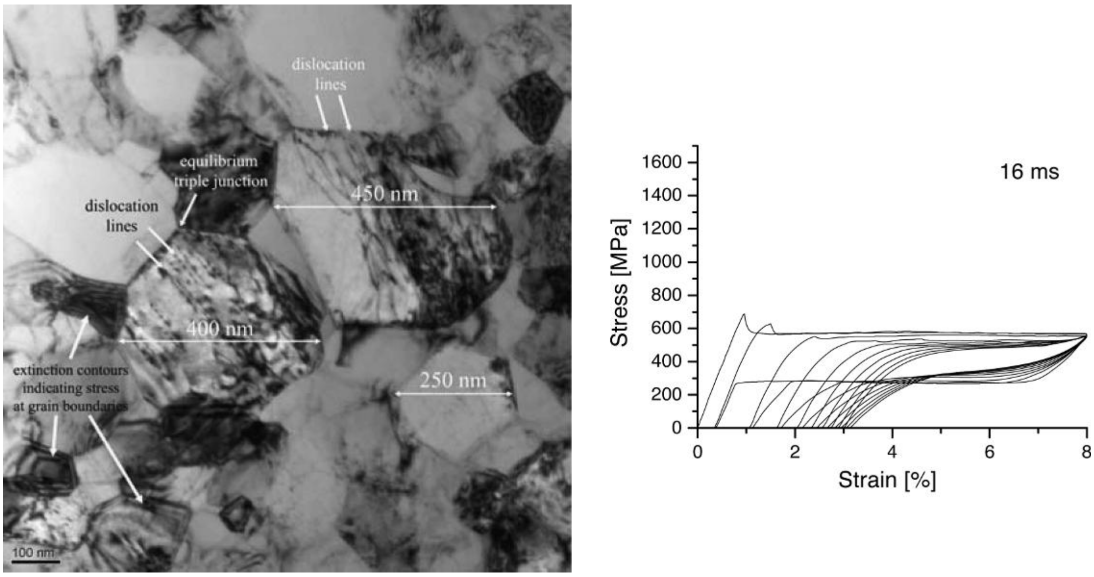

FIG. 1. Left: transmission electron microscopy of the microstructure in a of a NiTi wire shows the presence of dislocations on the {110} slip system after 10 tensile cycles. Right: the corresponding evolution of the hysteretic stress–strain loop over tensile cycling. Both figures courtesy of Šittner (see also Novák *et al.*, 2009).

Kundin *et al.* (2010) investigate plastic-elastic effects in martensitic materials with a phase field model. There, phase field variables are introduced for dislocations of different orientational variants of martensite and austenite; the dislocations move with the interface between adjacent phases. Again, the presence of dislocations is significant; for example, the martensitic volume fractions depend on the interdislocation distance (Kundin *et al.*, 2010). A full phase field model, with dislocations being able to move independently from interfaces, is however computationally much more costly and not available at present.

The above experiments and simulations motivate the mathematical analysis of models of shape memory materials accounting for plastic effects. Continuum mechanics is a framework in which existence can often be shown with relative ease for models of materials. Remarkably, the inclusion of plastic effects in macroscopic models of materials seems to be a new line of research. Auricchio *et al.* (2007) have recently investigated a 3D model for SMAs with inelastic effects. We propose a different model, based on the theory of gradient plasticity advocated by, e.g. Dillon & Kratochvíl (1970) and Gurtin (2000). To our knowledge, there is no other model of SMAs with plastic effects in the setting of multiplicative plasticity; we work in this setting since it then becomes possible to study large deformation phenomena.

The temporal evolution of the model presented here is rate independent. The framework of energetic solutions is a suitable description of such an evolution; it is sketched in Section 1.3.

One feature of the model is that it includes non-local terms, both for the plastic variables and for the volume fractions of the different phases and variants. The non-local terms we have chosen are phenomenological and among the simplest possible non-local expressions, being gradient terms. For the plastic variable, gradient plasticity is is an established model (Dillon & Kratochvíl, 1970; Gurtin, 2000). There are attempts to derive macroscopic non-local expressions, rather than assuming them as we do here, based on limit passages for statistical mechanics (Groma *et al.*, 2003; Kratochvíl & Sedláček, 2008). These attempts, while very valuable, are fraught with technical difficulties and thus have to resort to strong simplifying assumptions. Future research will have to address this limit passage further and derive non-local terms which can be used for the macroscopic description in place of the simple gradient expressions we use here.

The physical applicability of the model and the existence result obtained in this paper is demonstrated by numerical simulations in Section 4. For simplicity, we consider there the small strain version. The simulations show a pseudoelastic-plastic hysteresis loop (Fig. 3) of the same characteristics as the experimental one (Fig. 1). Both simulation and experiment show a significant increase of the permanent deformation of the austenite over an increasing number of cycles. Note that the number of cycles is small in both cases (6 for the simulation and 10 for the experiment). While the experiment concerns NiTi, which has 12 martensitic variants, we restrict the numerical simulation to a double-well setting. Yet, the increase of permanent deformation after 6 cycles is comparable ($<2.5\%$ in the experiment and about 1% in the simulation).

In the remainder of this section, we introduce the energetic approach to an elasto-plasticity problem of SMAs. After a brief overview over SMAs, microstructures and the associated non-quasiconvexity of the energy in Section 1.1, we describe in Section 1.2 the plasticity and the dissipation mechanisms employed in this paper. The mathematical formulation of the model is introduced in Section 2. The existence of a solution is proved via time discretization (Section 3).

A few remarks are in order about the context of the model. We consider energetic solutions to a rate-independent process. Energetic solutions have been proved to be very powerful for a number of applications, including non-convex problems (see, e.g. Francfort & Mielke, 2006 and Mielke & Roubíček, 2003). There and in the present situation, the proof of the existence of a solution serves as a first indication that the model is meaningful (the proof is constructive and shows that a time-incremental problem is well posed). Here, we present in Section 4 some 2D numerical simulations to further validate the model (in fact, a slight simplification of the model). The result shows the expected behaviour: Shape memory effects can be temperature induced or stress induced. We here work in the stress-induced setting. SMAs exhibit there a strongly non-linear (even hysteretic) stress–strain relationship, which is known as pseudoelasticity. The stress–strain relationship in the simulation exhibits a non-linearity as for the typical pseudoelastic regime but with additional plastic effects (see Fig. 3). Similarly, the plastic strain increases in the load experiment of Section 4 as shown in Fig. 3. The simulations thus show that plastic effects are captured by a small strain version of the model analysed here as are the pseudoelastic effects associated with shape memory effects. There is a wide range of other isothermal models for SMAs (see, e.g. Bhattacharya, 2003, Roubíček, 2004 for a survey of various models, both isothermal and thermodynamic, of SMAs), but the inclusion of plastic effects seems to be a relatively new area of research. The coupling of elastic and plastic effects for SMAs is a first step beyond purely elastic models (such as the limit problem of infinite plastic dissipation); it is desirable to include further effects, be it thermal or magnetic.

1.1 *Shape memory alloys*

SMAs are active materials and have been the subject of intensive theoretical and experimental research during the past decades. Existing or potential applications can be found, e.g. in medicine and mechanical or aerospace engineering. SMAs are crystalline materials that exhibit specific a 'hysteretic' stress/strain/temperature response; they have the ability to recover a trained shape after deformation and subsequent reheating. This is called the 'shape memory effect'. It is based on the ability of the SMA to rearrange atoms in different crystallographic configurations (in particular, with different symmetry groups). The stability depends on the temperature. Normally, at higher temperatures, a high-symmetry (e.g. cubic) lattice is stable, which is referred to as the 'austenite' phase. At lower temperatures, a lattice of lower symmetry (e.g. tetragonal, orthorhombic, monoclinic or triclinic) becomes stable, called the 'martensite' phase. Due to the loss of symmetry, this phase may occur in different 'variants'. The number of variants M, say, is the quotient of the order of the high-symmetry phase and the order of the low-symmetry group. So for a cubic high-symmetry phase, $M = 3, 6, 12$ or 4 for the tetragonal, orthorhombic, monoclinic, respectively, triclinic martensites mentioned above. We denote the stress-free strains of the variants U_ℓ, $\ell = 1, 2, \ldots, M$, and U_0 stands for the stress-free strain of the austenite. The variants can be combined coherently with each other, forming so-called 'twins' of two variants. The resulting structure is then called a 'laminate'.

The mathematical and computational modelling of SMAs represents a tool for the theoretical understanding of phase transition processes in solids. Such an analysis may complement experimental results, predict the response of new materials or facilitate the usage of SMAs in applications. SMAs are genuine 'multi-scale' materials and create a variety of challenges for mathematical modelling. We refer the reader to the literature (Roubíček, 2004) for a survey of a wide menagerie of SMA models ranging from nanoscale to macroscale. In this article, we focus on a mesoscopic model in the framework of continuum mechanics. Beside the macroscopic deformation and its gradient, the model also involves the volume fractions of phases and variants and volume fraction gradients. This seems a fruitful compromise since it allows for the modelling scales of large single crystals or polycrystals.

Let the specimen occupy a domain $\Omega \subset \mathbb{R}^n$. The stress-free parent austenite is a natural state of the material which makes it, in the context of continuum mechanics, a canonical choice for the reference configuration. As usual, $y\colon \Omega \to \mathbb{R}^n$ denotes the 'deformation' and $u\colon \Omega \to \mathbb{R}^n$ the 'displacement', which are related to each other via the identity $y(x) = x + u(x)$, where $x \in \Omega$. Hence, the 'deformation gradient' is $F := \nabla y = \mathbb{I} + \nabla u$. Here, $\mathbb{I} \in \mathbb{R}^{n\times n}$ is the identity matrix and ∇ the gradient operator.

The total stored energy in the bulk occupying, in its reference configuration, the domain Ω is then

$$V(y) := \int_\Omega \varphi(\nabla y(x))\mathrm{d}x. \tag{1.1}$$

A common variational principle in continuum mechanics is the 'minimization' of the stored energy. Due to the coexistence of several variants at low temperature, φ has multiple minima and thus a multi-well character. We consider an isothermal situation with several variants coexisting. Since φ is a multi-well energy density, minimizing sequences of V tend to develop, in general, finer and finer spatial oscillations of their gradients. In other words, the deformation gradient often tends to develop fine spatial oscillations due to lack of quasiconvexity of the stored energy density. (We recall that $\psi\colon \mathbb{R}^{n\times n} \to \mathbb{R}$ is 'quasiconvex' if, for all $A \in \mathbb{R}^{n\times n}$ and all $v \in W_0^{1,\infty}(\Omega;\mathbb{R}^n)$, $\psi(A)|\Omega| \leqslant \int_\Omega \psi(A + \nabla v(x))\mathrm{d}x$. If $\psi \geqslant 0$ is not quasiconvex but measurable and locally bounded, we define its 'quasiconvexification' $Q\psi$ by $Q\psi(A)|\Omega| = \inf_{v\in W_0^{1,\infty}(\Omega;\mathbb{R}^n)} \int_\Omega \psi(A + \nabla v(x))\mathrm{d}x$.) These oscillations are difficult to model in full detail, although some studies in this direction exists (Arndt, 2004). The oscillations correspond to the development of finer and finer microstructures when the stored energy is to be minimized. The minimum of V, under specific boundary conditions for y, is usually not attained in a space of functions. This is a problem of relaxation in the calculus of variations. One possibility is to replace φ by its quasiconvexification $Q\varphi$. However, to calculate $Q\varphi$ is usually extremely difficult and it is not typically known in a closed form. Another possibility is to extend the notion of a solution. 'Young measures' are here an appropriate tool. They are capable of recording, on a mesoscopic level, the limit information of the finer and finer oscillating deformation gradient as we move towards the macroscopic scale. This can be described, for a current macroscopic point $x \in \Omega$, by a probability measure ν_x on the set of deformation gradients, i.e. matrices in $\mathbb{R}^{n\times n}$. See Appendix A for a concise description of Young measures. To describe explicitly, the set of Young measures admissible for our problem (the so-called gradient Young measures, Kinderlehrer & Pedregal, 1994) is equivalent to quasiconvexification of φ. Nevertheless, Young measures have some advantages from the computational point of view, for instance.

1.2 *Plastic variables, gradient terms and dissipative mechanisms*

In many situations, the austenite–martensite phase transformation is connected with plastic effects. In particular, cyclic plasticity may occur, which can negatively affect the performance of the material. Hence, mathematical models including both plasticity of shape memory materials are needed. One such model is discussed here; see Auricchio *et al*. (2007) and Sadjadpour & Bhattacharya (2007) for related phenomenological models. We refer the reader to Carstensen *et al*. (2002) for a model of finite-strain elasto-plasticity where the existence can also be inferred by time discretization.

In order to include plasticity to the model, we assume that elastic properties of the material depend on plastic (internal) variables. In the setting described so far, the deformation y covers both the 'elastic' and the 'plastic' deformation. We employ the multiplicative split $F = F_e F_p$ of the deformation gradient into an elastic part F_e and an irreversible plastic part F_p. The latter belongs to $\mathrm{SL}(n) := \{A \in \mathbb{R}^{n\times n} | \det(A) = 1\}$. In addition to the so-called 'plastic strain' F_p, we consider a vector $p \in \mathbb{R}^m$ of 'hardening variables'. Both F_p and p are internal variables that influence the elasticity. It is to convenient to abbreviate $z = (F_p, p)$. Furthermore, $\lambda\colon \Omega \to \mathbb{R}^{M+1}$ records the volume fraction of austenite and the M variants of martensite at a point $x \in \Omega$ (Mielke & Roubíček, 2003). As mentioned above, the stored energy density of shape memory materials is typically not quasiconvex, which explains why we consider a Young measure ν for the deformation gradient. For the moment, the intuitive interpretation of a Young measure $\nu = \{\nu_x\}_{x\in\Omega}$ as a recording device for the probability to find a phase or variant at the location $x \in \Omega$ suffices; see Appendix A for precise definitions making this intuition rigorous. In summary, we describe the state of the material by $q := (y, \nu, \lambda, F_p, p) = (y, \nu, \lambda, z)$ and denote the set of these q's by $\mathbb{Q}$. A precise definition of $\mathbb{Q}$ tailored to our problem is given in (2.6) below.

1.2.1 *Gradient terms*. Following the approach of Dillon & Kratochvíl (1970) and Gurtin (2000) and others, we work in the framework of so-called 'gradient plasticity', i.e. ∇z enters the problem. As one can regard the volume fraction λ as an internal variable, too, it is natural to include a term involving $\nabla\lambda$ in the energetic contributions as well. This term also serves as a regularization since it ensures compactness. The prominent model for SMAs by Mielke & Roubíček (2003) also introduces this term; as a justification, Mielke & Roubíček (2003) consider the bulk energy associated with the Ericksen–Timoshenko beam

$$\int_\Omega \left[\phi(x, \nabla u) + \epsilon_1 |\nabla^2 u|^2 + \frac{1}{\epsilon_2}|\lambda - \mathcal{L}(x, \nabla u)|^2 + \rho|\nabla\lambda|^2\right] \mathrm{d}x,$$

where ϵ_1 describes the bending rigidity and ϵ_2 describes the deviation of the macroscopic order parameter λ from a microscopic order parameter $\mathcal{L}$ (which is just the projection to the second component in the classic Ericksen–Timoshenko model). Then $\sqrt{\epsilon_2\rho}$ is the internal length scale discussed, e.g. by Ren & Truskinovsky (2000); the model of Mielke & Roubíček (2003) is the asymptotic limit $\epsilon_1 \to 0$ followed by the limit $\epsilon_2 \to 0$. The gradient term survives these limit passages. In simulations, however, it is possible to set the coefficient ρ to zero since the solvability of a finite-dimensional approximation follows even in this case. It is not unreasonable to think of ρ as being smaller than the numerical resolution; then the existence of the infinite-dimensional problem still follows and the term vanishes in numerical computations. We follow this practice in the simulation in Section 4.

We remark that gradient of the volume fraction is also used in phase field models for SMAs and chosen to be 0.0001 in dimensionless units (Shu & Yen, 2008; Artemev *et al*., 2001).

The gradient terms are of phenomenological nature. They reflect the observation that microscopic interactions typically lead to non-local macroscopic terms. For example, dislocations interact in models with plasticity and their microscopic short-range interaction should be captured on the macroscopic level. There are attempts to derive the corresponding macroscopic terms from statistical mechanics (Groma *et al*., 2003; Kratochvíl *et al*., 2007; Kratochvíl & Sedláček, 2008; Kratochvíl *et al*., 2009). Yet, these derivations are very much in their infancy. It is hoped that they will eventually lead to accurate non-local terms in a macroscopic model. At present, however, the modelling is done on a phenomenological basis. The the choice of the gradient terms of the plastic variable and the martensitic volume fraction seems to be reasonable, as they are among the simplest non-local expressions.

In particular, for the volume fraction of the different variants and phases, Mielke & Roubíček (2003) introduce the gradient volume fraction in a purely elastic model of SMAs. We employ the same expression, $\epsilon \int_\Omega \|\nabla\lambda\|^2 \, dx$. A macroscopic justification of this term, due to Mielke & Roubíček (2003), based on the Timoshenko–Ericksen model, is given above. Again, one would like to derive the macroscopic non-local term from microscopic considerations. Yet, this problem seems to be completely open. A well-established class of models of martensites developed from principles of thermodynamics is due to Achenbach & Müller (1982) (see also Achenbach & Müller, 1985 and Müller & Seelecke, 1996). There, the evolution of the number of layers of a phase or variant is considered (equivalently, the fractions of the variants, respectively, phases), which corresponds to an integrated version of the macroscopic variable λ. The microscopic evolution is then determined on the basis of statistical mechanics. It is noteworthy that the microscopic evolution describes only the total fractions, not their local distribution, which is also recorded in λ. A derivation of the macroscopic non-local term describing the phases and variants from microscopic principles seems at least as difficult as the corresponding limit passage for dislocations. We thus use again the simplest possible phenomenological term, $\epsilon \int_\Omega \|\nabla\lambda\|^2 \, dx$, with $\epsilon = 0$ in the simulations, and leave the analysis of more complicated non-linear terms to future research. In particular, the interaction between martensite and dislocations should be addressed in future more refined models.

We consider two kinds of dissipation in our model both of which are 'rate independent'. The first kind is related to the austenite–martensite transformation, respectively, the martensite–martensite transformation and will be characterized by the change of 'volume fraction' in the composition of the material. The second kind is solely related to 'plastic processes' in the material, e.g. to cyclic plasticity. To account for a possible irreversibility, the plastic dissipation may take the value $+\infty$ while the transformation dissipation is taken to be finite.

1.2.2 *Dissipation originating in phase transitions*. In order to describe dissipation due to transformations, we adopt the (to some extent rather simplified) standpoint that the amount of dissipated energy associated with a particular phase transition between an austenite and a martensitic variant or between two martensitic variants can be described by a specific energy density (of the dimension $J/m^3 = Pa$). This viewpoint has been independently adopted in physics (Huo & Müller, 1993; Thamburaja & Anand, 2003; Vivet & Lexcellent, 1998). For an explicit definition of the transformation dissipation, we need to identify the particular phases or phase variants. To this behalf, we define a Lipschitz continuous mapping $\mathcal{L}: \mathbb{R}^{n\times n} \to \triangle$, where

$$\triangle := \left\{ \zeta \in \mathbb{R}^{1+M} \;\middle|\; \zeta_\ell \geqslant 0 \text{ for } \ell = 0, \ldots, M, \text{ and } \sum_{\ell=0}^{M} \zeta_\ell = 1 \right\}$$

is a simplex with $M + 1$ vertices, with M being the number of martensitic variants. Here, $\mathcal{L}$ is related with the material itself and thus has to be frame indifferent. We assume, beside $\zeta_\ell \geqslant 0$ and $\sum_{\ell=0}^{M} \zeta_\ell = 1$, that the coordinate ζ_ℓ of $\mathcal{L}(F)$ takes the value 1 if F is in the ℓth (phase) variant, i.e. F is in a vicinity of ℓth well $\mathrm{SO}(n)U_\ell$ of φ, which can be identified by the stretch tensor $F^\top F$ being close to $U_\ell^\top U_\ell$. If $\mathcal{L}(F)$ is not in any vertex of $\triangle$, then it means that F in the spinodal region where no definite phase or variant is specified. We assume, however, that the wells are sufficiently deep and the phases and variants are geometrically sufficiently far from each other that the tendency for minimization of the stored energy will essentially prevent F to range into the spinodal region. Thus, the concrete form of $\mathcal{L}$ is not important as long as $\mathcal{L}$ enjoys the properties listed above. We remark that $\mathcal{L}$ plays the rôle of what is often called a vector of 'order parameters' or a vector-valued 'internal variable'.

For two states q_1 and q_2, with $q_j = (y_i, \nu_i, \lambda_i, z_i)$ for $j = 1, 2$, we now define the dissipation due to martensitic transformation which 'measures' changes in the volume fraction $\lambda \in L^\infty\left(\Omega; \mathbb{R}^{M+1}\right)$. This dissipation is given by

$$\mathcal{D}_{\mathrm{tr}}(q_1, q_2) := \int_\Omega |\lambda_1(x) - \lambda_2(x)|_{\mathbb{R}^{M+1}} \,\mathrm{d}x, \tag{1.2}$$

where

$$\lambda_j(x) := \int_{\mathbb{R}^{n\times n}} \mathcal{L}(sF_{\mathrm{p}_j}^{-1}(x))\nu_{j,x}(\mathrm{d}s) = \int_{\mathbb{R}^{n\times n}} \mathcal{L}\big(s\big(\mathrm{Cof}\big(F_{\mathrm{p}_j}(x)\big)\big)^\top\big)\nu_{j,x}(\mathrm{d}s), \tag{1.3}$$

where $\mathrm{Cof}\, A := (\det(A))A^{-\top}$ is the so-called cofactor matrix of a regular matrix A.

1.2.3 *Plastic dissipation*. The second source of dissipation is related to temporal changes in the plastic (hardening) variables gathered in $z = (F_{\mathrm{p}}, p)$. We write $Z := \mathrm{SL}(n) \times \mathbb{R}^m \times \mathbb{R}^{n\times m} \times \mathbb{R}^m$ for the Cartesian product of the set of plastic variables with the set of their time derivatives. We write $(z, \dot{z})$ for elements of Z. Let us consider a non-negative function δ: $\Omega \times Z \to \mathbb{R}$ which is positively one-homogeneous in the last variable $\dot{z}$, i.e. $\delta(x, z, \eta\dot{z}) = \eta\delta(x, z, \dot{z})$ for all $\eta \geqslant 0$.The function δ is the Legendre transform of the indicator function of the elasticity domain. So, knowing the yield function and therefore the elasticity domain one can calculate δ. Then the 'dissipation distance' is (Mielke, 2005)

$$D_{\mathrm{p}}(z_1, z_2) = \inf_z \left\{ \int_0^1 \delta(x, z(s), \dot{z}(s))\mathrm{d}s | z: C^1[0, 1] \to Z \text{ with } z(0) = z_1 \text{ and } z(1) = z_2 \right\}.$$

The 'plastic dissipation' is then

$$\mathcal{D}_{\mathrm{p}}(q_1, q_2) := \int_\Omega D_{\mathrm{p}}(x, z_1(x), z_2(x))\mathrm{d}x. \tag{1.4}$$

Consequently, the overall dissipation $\mathcal{D}$ (with values in $[0, +\infty]$, defined on $\mathbb{Q} \times \mathbb{Q}$ with $\mathbb{Q}$ given by (2.6))

$$\mathcal{D}(q_1, q_2) := \mathcal{D}_{\mathrm{p}}(q_1, q_2) + \mathcal{D}_{\mathrm{tr}}(q_1, q_2). \tag{1.5}$$

We point out that the dissipation is not the sum of two independent terms, the one being purely elastic, the other purely plastic, as it appears at first glance. Indeed, the transformation dissipation is coupled to the plastic one since it depends on the plastic term. Hence, the plastic dissipation influences the phase-transition one as can be seen in (1.3). Technical assumptions on the dissipation are stated at the end of Section 2.1.

1.2.4 *Loading and boundary conditions*. In experiments, a specimen occupying the region Ω will be subjected to external loads. In order to simplify our exposition, we consider only dead body forces and surface forces. We assume that we are given two disjoint sets $\Gamma_0, \Gamma_1 \subset \partial\Omega$, where the $(n-1)$-dimensional Hausdorff measure of Γ_0 is positive. We consider Dirichlet boundary conditions $y = y_0$ on Γ_0 for some prescribed (time-independent) mapping y_0. As for the surface forces acting on Γ_1, we define a linear functional

$$L(y) := \int_\Omega f(x) \cdot y(x)\mathrm{d}x + \int_{\Gamma_1} g(x) \cdot y(x)\mathrm{d}S, \tag{1.6}$$

where $f\colon \Omega \to \mathbb{R}^n$ and $g\colon \Gamma_1 \to \mathbb{R}^n$ are the densities of volume and surface forces acting on the material, respectively. Below, we write $L = L(t, y)$ to indicate the possibility of temporally changing forces.

1.3 *Energetic solution*

Combining the previous considerations, we arrive at the energy functional $\mathcal{I}$ of the form

$$\begin{aligned}\mathcal{I}(t,q) := &\int_\Omega \int_{\mathbb{R}^{n\times n}} W(x, sF_\mathrm{p}^{-1}(x), F_\mathrm{p}(x), \nabla F_\mathrm{p}(x), p(x), \nabla p(x))\nu_x(\mathrm{d}s)\mathrm{d}x \\ &+ \varepsilon \|\nabla\lambda\|^2_{L^2(\Omega;\mathbb{R}^{(1+M)\times n})} - L(t, y(t)).\end{aligned} \tag{1.7}$$

So the gradient of the plastic deformation is included, as we work in the realm of strain gradient plasticity (Dillon & Kratochvíl, 1970; Gurtin, 2000). This introduces a gradient (ruling out the formation of ever finer plastic microstructure, by introducing a scale on which the plastic microstructure can exist). Similarly, ever finer elastic microstructures are also not observed in nature. However, here the inclusion of a gradient term in the model is not universally accepted. We thus do not introduce an elastic gradient directly but penalize changes in the volume fraction λ, which combines elastic (transformation) terms with plastic ones. In plasticity, following, e.g. Gurtin (2000), it is not uncommon to restrict the attention to so-called 'separable' materials, i.e.

$$W(x, F_\mathrm{e}, F_\mathrm{p}, \nabla F_\mathrm{p}, p, \nabla p) := \hat{\varphi}(x, F_\mathrm{e}) + W_\mathrm{p}(F_\mathrm{p}, \nabla F_\mathrm{p}, p, \nabla p). \tag{1.8}$$

Thus, the elastic and plastic energy contributions are additively coupled. This concept is also used in linearized elasto-plasticity. However, the analysis of this article does not require this assumption (except for Remark 3.1). Yet, it may be instructive to have separable materials in mind.

It is often convenient to write

$$V(q) := \int_\Omega \int_{\mathbb{R}^{n\times n}} W(x, sF_\mathrm{p}^{-1}(x), F_\mathrm{p}(x), \nabla F_\mathrm{p}, p(x), \nabla p(x))\nu_x(\mathrm{d}s)\mathrm{d}x + \varepsilon \|\nabla\lambda\|^2_{L^2(\Omega;\mathbb{R}^{(1+M)\times n})}. \tag{1.9}$$

We seek to analyse the time evolution of a process $q(t) \in \mathbb{Q}$ during the time interval $[0, T]$; $\mathbb{Q}$ is here the configuration space, whose mathematical definition is given in (2.6) below. The following two properties are key ingredients of the so-called energetic solution introduced by Mielke *et al*. (2002).

(i) Stability inequality: for every $t \in [0, T]$ and every $\tilde{q} \in \mathbb{Q}$, it holds that

$$\mathcal{I}(t, q(t)) \leqslant \mathcal{I}(t, \tilde{q}) + \mathcal{D}(q(t), \tilde{q}). \tag{1.10}$$

(ii) Energy balance: For every $0 \leqslant t \leqslant T$,

$$\mathcal{I}(t, y(t), z(t)) + \mathrm{Var}(\mathcal{D}, q; [0, t]) = \mathcal{I}(0, q(0)) - \int_0^t \dot{L}(\xi, q(\xi))\mathrm{d}\xi, \tag{1.11}$$

where

$$\mathrm{Var}(\mathcal{D}, z; [s, t]) := \sup \left\{ \sum_{j=1}^{N} \mathcal{D}(q(t_{j-1}), q(t_j)) \;\middle|\; \{t_j\}_{j=0}^{N} \text{ is a partition of } [s, t] \right\}$$

is the 'variation' of the dissipation.

DEFINITION 1.1 The mapping q: $[0, T] \to \mathbb{Q}$, with a stable initial datum $q(0) \in \mathbb{Q}$, is an 'energetic solution' to the problem $(\mathcal{I}, \mathcal{D}, L)$ with the energy functional $\mathcal{I}$ as in (1.7), the dissipation $\mathcal{D}$ of (1.5) and the load L as in (1.6) if the stability inequality (1.10) and energy balance (1.11) are satisfied for every $t \in [0, T]$.

2. Mathematical background and assumptions

2.1 *Mathematical framework, assumptions and main result*

We recall $W = W(x, sF_\mathrm{p}^{-1}(x), F_\mathrm{p}(x), \nabla F_\mathrm{p}(x), p(x), \nabla p(x))$; to abbreviate the notation, let us write $A := sF_\mathrm{p}^{-1}$, $G := \nabla F_\mathrm{p}$ and $\pi := \nabla p$. We assume that W satisfies the following requirements:

$$W(x, \cdot) \text{ is continuous for a.e. } x \in \Omega \tag{2.1}$$

$$W(\cdot, A, F_\mathrm{p}, G, p, \pi) \text{ is measurable for all } A, F_\mathrm{p}, G, p, \pi. \tag{2.2}$$

Next, growth conditions: we assume that there are constants $C, c > 0$ and $\alpha, \beta, \omega > 1$ such that

$$C(1 + |A|^\alpha + |F_\mathrm{p}|^\beta + |G|^\beta + |p|^\omega + |\pi|^\omega) \geqslant W(x, A, F_\mathrm{p}, G, p, \pi) \geqslant c(-1 + |A|^\alpha + |F_\mathrm{p}|^\beta + |G|^\beta + |p|^\omega + |\pi|^\omega). \tag{2.3}$$

Thus, the model assumes hardening. Since we work with Young measures, the inclusion of orientation–preservation is currently out of scope for a rigorous analysis. It is reasonable to require convexity in the gradient terms $G = \nabla F_\mathrm{p}$ and $\pi = \nabla p$,

$$W(x, A, F_\mathrm{p}, \cdot, p, \cdot) \text{ is convex for a.e. } x \in \Omega \text{ and every } A, F_\mathrm{p}, p\,. \tag{2.4}$$

In order to simplify the notation, we omit the dependence of W on x. However, we point out that the entire theory developed in this paper applies to spatially inhomogeneous W as well.

In what follows, we suppose that

$$y \in \mathbb{Y}(\Omega; \mathbb{R}^n) := \{y \in W^{1,d}(\Omega; \mathbb{R}^n) | y = y_0 \text{ on } \Gamma_0\}, \tag{2.5}$$

where $\Gamma_0 \subset \partial\Omega$ with a positive surface measure as described in Section 1.2.4. We recall from that Section that $\Gamma_0 \cap \Gamma_1 = \emptyset$ by assumption. Further,

$$\mathbb{P} := \{(F_\mathrm{p}, p) \in W^{1,\beta}(\Omega; \mathbb{R}^{n\times n}) \times W^{1,\omega}(\Omega; \mathbb{R}^m) | F_\mathrm{p}(x) \in \mathrm{SL}(n) \text{ for a.e. } x \in \Omega\}.$$

Then we look for $q \in \mathcal{Q} := \mathbb{Y}(\Omega;\mathbb{R}^n) \times \mathbb{G}(\Omega;\mathbb{R}^n) \times L^\infty\left(\Omega;\mathbb{R}^{M+1}\right) \times \mathbb{P}$ and restrict the space further by imposing the 'admissibility condition'

$$\mathbb{Q} := \{q \in \mathcal{Q} | \lambda = \mathcal{L} \diamond \nu \text{ and } \nabla y = \mathbb{I} \bullet \nu\}, \tag{2.6}$$

where, for almost all $x \in \Omega$, $[\mathcal{L} \diamond \nu](x) := \int_{\mathbb{R}^{n\times n}} \mathcal{L}(sF_\mathrm{p}^{-1}(x))\nu_x(\mathrm{d}s)$ and $[\mathbb{I} \bullet \nu](x) := \int_{\mathbb{R}^{n\times n}} s\nu_x(\mathrm{d}s)$. We need to define the notion of convergence in this space, and do so as follows.

DEFINITION 2.1 Suppose that $\{q_k\}_{k\in\mathbb{N}} \subset \mathbb{Q}$, where $q_k = (y_k, \nu_k, \lambda_k, z_k)$. We say that $q_k \rightharpoonup q := (y, \nu, \lambda, z) \in \mathbb{Q}$ as $k \to \infty$ if $y_k \rightharpoonup y$ in $W^{1,d}(\Omega;\mathbb{R}^n)$, $\nu_k \rightharpoonup^* \nu$ in $L^\infty_\mathrm{w}(\Omega;\mathcal{M}(\mathbb{R}^{n\times n}))$, $\lambda_k \rightharpoonup \lambda$ in $W^{1,2}(\Omega;\mathbb{R}^{M+1})$, $z_k \rightharpoonup z$ in $W^{1,\beta}(\Omega;\mathbb{R}^{n\times n}) \times W^{1,\omega}(\Omega;\mathbb{R}^m)$.

The following lemma shows that the above definition is meaningful.

LEMMA 2.2 Let $\{q_k\}_{k\in\mathbb{N}} \subset \mathbb{Q}$, where $q_k = (y_k, \nu_k, \lambda_k, z_k)$. Let $q_k \rightharpoonup q := (y, \nu, \lambda, z)$. Then $\lambda = \mathcal{L} \diamond \nu$.

Proof. We denote $\Lambda(x) := \int_{\mathbb{R}^{n\times n}} \mathcal{L}(sF_\mathrm{p}(x)^{-1})\nu_x(\mathrm{d}s)$ and estimate for any $f \in L^\infty\left(\Omega;\mathbb{R}^{M+1}\right)$, using (1.3),

$$\left|\int_\Omega (\lambda_k(x) - \Lambda(x)) f(x)\mathrm{d}x\right| \leqslant \int_\Omega \left|\int_{\mathbb{R}^{n\times n}} \mathcal{L}(sF_{\mathrm{p}_k}^{-1}(x))\nu_{k,x}(\mathrm{d}s) - \int_{\mathbb{R}^{n\times n}} \mathcal{L}(sF_\mathrm{p}^{-1}(x))\nu_{k,x}(\mathrm{d}s)\right| |f(x)|\mathrm{d}x$$
$$+ \left|\int_\Omega \left[\int_{\mathbb{R}^{n\times n}} \mathcal{L}(sF_\mathrm{p}^{-1}(x))\nu_{k,x}(\mathrm{d}s) - \int_{\mathbb{R}^{n\times n}} \mathcal{L}(sF_\mathrm{p}^{-1}(x))\nu_x(\mathrm{d}s)\right] f(x)\mathrm{d}x\right|.$$

Due to our assumption $q_k \rightharpoonup q$, we have $z_k \rightharpoonup z$ in $W^{1,\beta}(\Omega;\mathbb{R}^{n\times n})$ and hence $F_{\mathrm{p}_k}^{-1} \to F_\mathrm{p}^{-1}$ strongly in $L^1(\Omega;\mathbb{R}^{n\times n})$. The first term on the right-hand side tends to zero as $k \to \infty$, by Lipschitz continuity of $\mathcal{L}$. The second term on the right-hand side converges to zero by the definition of the weak* convergence of $\{\nu_k\}_{k\in\mathbb{N}}$. Altogether we have that $\lambda_k \to \Lambda$ weakly in $L^1(\Omega;\mathbb{R}^{M+1})$. On the other hand, we assumed that $\lambda_k \to \lambda$ strongly in $L^2(\Omega;\mathbb{R}^{M+1})$ since $q_k \rightharpoonup q$. Hence, $\lambda = \Lambda$. □

In line with related work (Francfort & Mielke, 2006; Mainik & Mielke, 2008), we impose the following conditions on $\mathcal{D}$:

(i) Lower semicontinuity:

$$\mathcal{D}(q, \tilde{q}) \leqslant \liminf_{k\to\infty} \mathcal{D}(q_k, \tilde{q}_k), \tag{2.7}$$

whenever $q_k \rightharpoonup q$ and $\tilde{q}_k \rightharpoonup \tilde{q}$ as $k \to \infty$.

(ii) Positivity: If $\{q_k\}_{k\in\mathbb{N}} \subset \mathbb{Q}$ is bounded and

$$\min\{\mathcal{D}(q_k, q), \mathcal{D}(q, q_k)\} \to 0 \text{ as } k \to \infty \text{ then } q_k \rightharpoonup q. \tag{2.8}$$

Some more assumptions on the dissipation are required to deal with the possibility of the plastic dissipation becoming infinite. We state suitable restrictions for D_p (see Mainik & Mielke, 2008 for similar conditions).

ASSUMPTION 2.3 The plastic dissipation D_p satisfies the following conditions:

1. $D_\mathrm{p}(x, \cdot, \cdot)$: $D(x) \to [0, +\infty)$ is continuous, where

$$D(x) := \{(z_0, z_1) | D_\mathrm{p}(x, z_0, z_1) < +\infty\}. \tag{2.9}$$

2. For every $R > 0$, there is $K > 0$ such that, for almost all $x \in \Omega$, $D_\mathrm{p}(x, z_0, z_1) < K$ if $(z_0, z_1) \in D(x)$ and $|z_0|, |z_0| < R$.
3. There is $v^* \in \mathbb{R}^m$ such that for all $\eta, R > 0$, there is $\rho > 0$ such that for almost every $x \in \Omega$ and every $z, z_0, z_1 \in \mathbb{R}^{n\times n} \times \mathbb{R}^m$:

$$|z - z_0| < \rho \text{ and } (z_0, z_1) \in D(x) \text{ implies } (z, z_1 + (0, \eta v^*)) \in D(x),$$

where $\eta \to 0$ when $\rho \to 0$.

As for the load, we impose the following qualifications:

$$f \in C^1([0,T]; L^{\tilde{d}}(\Omega; \mathbb{R}^d)) \quad \text{with } \tilde{d} \geqslant \begin{cases} \frac{dn}{n-d} & \text{if } 1 \leqslant d < n \\ 1 & \text{else} \end{cases} \quad \text{and} \tag{2.10}$$

$$g \in C^1([0,T]; L^{\hat{d}}(\Gamma_1; \mathbb{R}^d)) \quad \text{with } \hat{d} \geqslant \begin{cases} \frac{nd-d}{nd-n} & \text{if } 1 \leqslant d < n \\ 1 & \text{else.} \end{cases} \tag{2.11}$$

Our main result is the following theorem regarding the existence of an energetic solution.

THEOREM 2.4 Let $\frac{1}{\alpha} + \frac{1}{\beta} \leqslant \frac{1}{d} < \frac{1}{n}$, and let Assumption 2.3, (2.1)–(2.4),(2.7)–(2.11) hold. Then there is a process q: $[0,T] \to \mathbb{Q}$ with $q(t) = (y(t), \nu(t), z(t), \lambda(t))$ such that q is an energetic solution according to Definition 1.1. for a given stable initial condition $q_0 \in \mathbb{Q}$.

The proof of this result relies on approximations by time-discrete (incremental) problems constructed for a given time step. These are minimization problems over spatial variables. Each minimization problem takes into account the solution obtained for the previous time step while the initial condition serves as input for the first minimization problem. Hence, the proof is rather constructive. In addition to Theorem 2.4, we prove various convergence results for the deformation, the martensitic volume fractions and the plastic variables, see Theorem 3.8.

3. Existence of a solution process

3.1 *Incremental problems*

We start the mathematical analysis by defining the set of stable states,

$$\mathcal{S}(t) := \{q \in \mathbb{Q} | \mathcal{I}(t,q) \leqslant \mathcal{I}(t,\tilde{q}) + \mathcal{D}(q,\tilde{q}) \text{ for every } \tilde{q} \in \mathcal{Q}\}; \tag{3.1}$$

let us also define

$$\mathcal{S}_{[0,T]} := \cup_{t\in[0,T]}\{t\} \times \mathcal{S}(t). \tag{3.2}$$

We say that a sequence $\{(t_k, q_k)\}_{k\in\mathbb{N}}$ is 'stable' if $q_k \in \mathcal{S}(t_k)$.

The proof of existence of a rate-independent evolution commonly proceeds via time discretization. Thus, in a first step, a sequence of incremental problems is defined. Let us remind ourselves of the notation $z := (F_\mathrm{p}, p)$. We define a time discretization $0 = t_0 < \cdots < t_n = T$ via a time step $\tau > 0$, chosen in such a way that $N = T/\tau \in \mathbb{N}$. Let an initial state $\mathcal{S}(0) \ni q^0 =: q^0_\tau \in \mathbb{Q}$ be given. For $1 \leqslant k \leqslant N$, we find $q^k_\tau \in \mathbb{Q}$ by solving

$$\text{minimize} \mathcal{I}(t_k, q) + \mathcal{D}(q^{k-1}_\tau, q) \ \text{ subject to } q \in \mathbb{Q}. \tag{3.3}$$

REMARK 3.1 For separable materials (see (1.8)), consider a situation where we know the quasiconvexification $Q\hat{\varphi}$ of $\hat{\varphi}$ explicitly and set $W_{\rm p} = 0$ for simplicity. Then for each point $A \in \mathbb{R}^{n\times n}$, we know a volume fraction $\lambda(A)$ of the austenite and the martensitic variants. This volume fraction, however, does not have to be given uniquely. Then the corresponding time-incremental problem for a given time t_k could read: given λ_{k-1} and z_{k-1} (the volume fraction and plastic variables from the previous time t_{k-1}), minimize in y, z the functional

$$\int_\Omega Q\hat{\varphi}(x, \nabla y(x)F_{\rm p}^{-1}(x))\mathrm{d}x + \int_\Omega |\mathcal{L}(\nabla y(x)F_{\rm p}^{-1}(x)) - \lambda_{k-1}(x)|\mathrm{d}x - L(t_k, y) + \mathcal{D}_{\rm p}(z, z_{k-1}) - L(t_k, y),$$

where we set $\lambda(x) = \mathcal{L}(\nabla y(x)F_{\rm p}^{-1}(x))$. However, as we do not know $Q\hat{\varphi}$ in most cases, we define $\mathcal{D}_{\rm tr}$ using Young measures in (1.2) and the elastic energy would be calculated by $\int_\Omega \int_{\mathbb{R}^{n\times n}} \hat{\varphi}(x, sF_{\rm p}^{-1}(x))\,\nu_x(\mathrm{d}s)\mathrm{d}x$ and we have the relaxed problem

$$\text{minimize}_{y,\nu,z} \int_\Omega \int_{\mathbb{R}^{n\times n}} \hat{\varphi}(x, sF_{\rm p}^{-1}(x))\nu_x(\mathrm{d}s)\mathrm{d}x + \mathcal{D}_{\rm tr}(\lambda, \lambda_{k-1}) + \mathcal{D}_{\rm p}(z, z_{k-1}) - L(t_k, y).$$

Both formulations coincide if $\nu_x = \delta_{\nabla y(x)}$. This motivates our problem framework defined in terms of Young measures.

The existence of a solution to the time step problem (3.3) is ensured by the following lemma.

LEMMA 3.2 (Existence). Let (2.1)–(2.4), (2.7), (2.10) and (2.11) hold. Suppose further $\frac{1}{\alpha} + \frac{1}{\beta} \leqslant \frac{1}{d}$. Then the problem (3.3) has a solution for all $k = 1, \ldots, N = T/\tau$.

Proof. Suppose that $q^{k-1} \in \mathbb{Q}$ is known; let $\{q_j\}_{j\in\mathbb{N}} := \{(y_j, \nu_j, \lambda_j, z_j)\}_j \subset \mathbb{Q}$ be a minimizing sequence for $q \mapsto \mathcal{I}(t_k, q) + \mathcal{D}(q_\tau^{k-1}, q)$.

First, note that $F_{\rm p}^{-1} = (\mathrm{cof}F)^\top$, where 'cof' stands for the cofactor matrix. Suppose that $q_\tau^{k-1} \in \mathbb{Q}$ is known and that $\{q_j\} \subset \mathbb{Q}$ is a minimizing sequence for $q \mapsto \mathcal{I}(t_k, q) + \mathcal{D}(q_\tau^{k-1}, q)$. We use Young's and Hölder's inequalities as in Mainik & Mielke (2008) to obtain the following pointwise inequality for any member of the minimizing sequence (the index j is omitted for simplicity)

$$|FF_{\rm p}^{-1}| \geqslant \frac{|F|}{|F_{\rm p}|} \geqslant r\theta^{r/(r-1)}|F|^{1/r} - (r-1)\theta|F_{\rm p}|^{1/(r-1)}$$

valid for all $r > 1$ and all $\theta > 0$.

Taking into account that $F_{\rm e} = FF_{\rm p}^{-1} \in L^\alpha(\Omega; \mathbb{R}^{n\times n})$, $F_{\rm p} \in L^\beta(\Omega; \mathbb{R}^{n\times n})$ and $F \in L^d(\Omega; \mathbb{R}^{n\times n})$ together with Hölder's inequality, we get for $r := \alpha/d > 1$ and $\frac{1}{b} := \frac{1}{d} - \frac{1}{\alpha} \geqslant \frac{1}{\beta}$

$$\begin{aligned}\left\| FF_{\rm p}^{-1}\right\|^\alpha_{L^\alpha(\Omega;\mathbb{R}^{n\times n})} &\geqslant \frac{\|F\|^\alpha_{L^d(\Omega;\mathbb{R}^{n\times n})}}{\|F_{\rm p}\|^\alpha_{L^\beta(\Omega;\mathbb{R}^{n\times n})}} \\ &\geqslant r\theta^{r/(r-1)}\,\|F\|^d_{L^d(\Omega;\mathbb{R}^{n\times n})} - (r-1)\theta\,\left\|F_{\rm p}\right\|^b_{L^b(\Omega;\mathbb{R}^{n\times n})}.\end{aligned}$$

Using this inequality for θ small enough in the lower bound (2.3) of W gives that $\int_\Omega \int_{\mathbb{R}^{n\times n}} |s|^d \nu_{jx}(\mathrm{d}s)\,\mathrm{d}x < +\infty$ for all j. This together with the Poincaré inequality proves a uniform bound on $\|y_j\|_{W^{1,d}(\Omega;\mathbb{R}^n)}$ for all $j \in \mathbb{N}$. Then, again, (2.3) shows that $\{z_j\}_{j\in\mathbb{N}}$ is bounded in $W^{1,\beta}(\Omega; \mathbb{R}^{n\times n}) \times W^{1,\omega}(\Omega; \mathbb{R}^m)$.

Hence, as $\beta > 1$ and $\omega > 1$, we can extract a weakly converging subsequence (not relabelled) $z_j \rightharpoonup z$ in $W^{1,\beta}(\Omega; \mathbb{R}^{n\times n}) \times W^{1,\omega}(\Omega; \mathbb{R}^m)$, with $z_j = (F_{{\rm p}j}, p_j)$. The strong convergence of $z_j \to z :=$

(F_p, p) in $L^\beta(\Omega; \mathbb{R}^{n\times n}) \times L^\omega(\Omega; \mathbb{R}^m)$ as $j \to \infty$ ensures that $F_\mathrm{p}(x) \in \mathrm{SL}(n)$ almost everywhere. By weak-$\star$ compactness $\nu_j \overset{*}{\rightharpoonup} \nu \in L^\infty_\mathrm{w}(\Omega; \mathcal{M}(\mathbb{R}^{n\times n}))$ as $j \to \infty$; $L^\infty_\mathrm{w}(\Omega; \mathcal{M}(\mathbb{R}^{n\times n}))$ is defined in Appendix A. However, we must show that $\nu \in \mathbb{G}$. Each Young-measure component ν_j, say, of $q_j = (y_j, \nu_j, \lambda_j, F_{\mathrm{p}j}, p_j)$ is generated by a sequence of gradients $\{\nabla y^l_j\}_{l\in\mathbb{N}}$ of maps $y^l_j \in W^{1,d}(\Omega; \mathbb{R}^n)$. As $\{F_{\mathrm{p}j}\}_{j\in\mathbb{N}}$ is bounded in $L^\beta(\Omega; \mathbb{R}^{n\times n})$, we obtain from Hölder's inequality the estimate

$$\|\nabla y^l_j\|_{L^d(\Omega;\mathbb{R}^{n\times n})} \leqslant \|\nabla y^l_j F_{\mathrm{p}j}^{-1}\|_{L^\alpha(\Omega;\mathbb{R}^{n\times n})} \|F_{\mathrm{p}j}\|_{L^\beta(\Omega;\mathbb{R}^{n\times n})} \leqslant C.$$

The first norm on the right-hand side is bounded by Assumption (2.3) so we have that $\{\nabla y^l_j\}_{j,l\in\mathbb{N}}$ is bounded independently of $j, l \in \mathbb{N}$. A diagonalization argument then shows that there is a generating sequence of gradients for η, so that $\eta \in G$. Moreover, Lemma 2.2 ensures that $\lambda = \mathcal{L} \diamond \nu$. The joint convexity of W in the gradient arguments G and π and (2.7) ensure that $\mathcal{I}$ is sequentially weakly lower semicontinuous on $\mathbb{Q}$. The existence of a minimum then follows by the direct method of the calculus of variations. □

3.2 *Interpolation in time*

We now introduce a piecewise constant interpolation q_τ of $q^k_\tau := (y^k_\tau, \nu^k_\tau, \lambda^k_\tau, z^k_\tau)$. Namely, $q_\tau(t) := q^k_\tau$ if $t \in [k\tau, (k+1)\tau)$ and $k = 0, \ldots, N-1 = T/\tau - 1$. Finally, $q_\tau(T) := q^N|_\tau$. Likewise, $L_\tau(t, q) = L(k\tau, q)$ is a piecewise constant interpolation of the load L for suitable piecewise constant q. Analogously, $\mathcal{I}_\tau(t, q) = \mathcal{I}(k\tau, q)$ is a piecewise constant interpolation of $\mathcal{I}$ defined in the same way as L_τ.

PROPOSITION 3.3 (Stability). We make the same assumptions as in Lemma 3.2: let (2.1)–(2.4), (2.7), (2.10) and (2.11) be satisfied and suppose that $\frac{1}{\alpha} + \frac{1}{\beta} \leqslant \frac{1}{d}$. Then the problem (3.3) has a solution $q_\tau(t)$ which is stable, i.e. for all $t \in [0, T]$ and for every $\tilde{q} \in \mathbb{Q}$,

$$\mathcal{I}_\tau(t, q_\tau(t)) \leqslant \mathcal{I}_\tau(t, \tilde{q}) + \mathcal{D}\,(q_\tau(t), \tilde{q})\,. \tag{3.4}$$

Moreover, for all $t_1 \leqslant t_2$ from the set $\{k\tau\}^N_{k=0}$, the following discrete energy inequalities hold if one extends the definition of $q_\tau(t)$ by setting $q_\tau(t) := q_0$ if $t < 0$.

$$\begin{aligned} -\int_{t_1}^{t_2} \dot{L}(t, q_\tau(t-\tau))\mathrm{d}t &\leqslant \mathcal{I}(t_2, q_\tau(t_2)) + \mathrm{Var}(q_\tau, [t_1, t_2]) - \mathcal{I}(t_1, q_\tau(t_1)) \\ &\leqslant -\int_{t_1}^{t_2} \dot{L}(t, q_\tau(t))\mathrm{d}t. \end{aligned} \tag{3.5}$$

Proof. The existence of a solution to (3.3) was proved in Lemma 3.2. The stability estimate (3.4) follows from the minimizing property of q^k_τ and the properties of $\mathcal{D}$. Indeed, by minimality of q^k_τ,

$$\mathcal{I}(k\tau, q^k_\tau) + \mathcal{D}(q^{k-1}_\tau, q^k_\tau) \leqslant \mathcal{I}(k\tau, \tilde{q}) + \mathcal{D}(q^{k-1}_\tau, \tilde{q}), \tag{3.6}$$

which immediately implies that

$$\mathcal{I}(k\tau, q^k_\tau) \leqslant \mathcal{I}(k\tau, \tilde{q}) + \mathcal{D}(q^{k-1}_\tau, \tilde{q}) - \mathcal{D}(q^{k-1}_\tau, q^k_\tau). \tag{3.7}$$

We remark that both dissipative terms satisfy the triangle inequality, $\mathcal{D}_\mathrm{p}(q_1, q_2) + \mathcal{D}_\mathrm{p}(q_2, q_3) \leqslant \mathcal{D}_\mathrm{p}(q_1, q_3)$ and analogously for $\mathcal{D}_\mathrm{tr}$. Thus,

$$\mathcal{D}(q^{k-1}_\tau, \tilde{q}) - \mathcal{D}(q^{k-1}_\tau, q^k_\tau) \leqslant \mathcal{D}(q^k_\tau, \tilde{q}),$$

so that (3.4) follows from (3.7).

Next, we demonstrate the validity of the energy inequality (3.5), using arguments of Mielke *et al.* (2002). For this part, let us test the stability of q_τ^{k-1} with $\tilde{q} := q_\tau^k$. This gives

$$\begin{aligned}\mathcal{I}((k-1)\tau, q_\tau^{k-1}) &\leqslant \mathcal{I}((k-1)\tau, q_\tau^k) + \mathcal{D}(q_\tau^{k-1}, q_\tau^k,) \\ &= \mathcal{I}(k\tau, q_\tau^k) + L(k\tau, q_\tau^k) - L((k-1)\tau, q_\tau^k) + \mathcal{D}(q_\tau^{k-1}, q_\tau^k). \end{aligned} \tag{3.8}$$

Suppose that $0 \leqslant k_1 \leqslant k_2 \leqslant N$ and that $t_1 = k_1\tau$ and $t_2 = k_2\tau$. A summation of (3.8) over $k = k_1+1, \ldots, k_2$ yields

$$\sum_{k=k_1+1}^{k_2} [L((k-1)\tau, q_\tau^k) - L(k\tau, q_\tau^k)] \leqslant \mathcal{I}(k_2\tau, q_\tau^{k_2}) - \mathcal{I}(k_1\tau, q_\tau^{k_1}) + \sum_{k=k_1+1}^{k_2} \mathcal{D}(q_\tau^{k-1}, q_\tau^k). \tag{3.9}$$

We rewrite this inequality in terms of q_τ to see that it is the first inequality in (3.5),

$$\begin{aligned} -\int_{t_1}^{t_2} \dot{L}(t, q_\tau(t-\tau))\mathrm{d}t &\leqslant \mathcal{I}(k_2\tau, q_\tau^{k_2}) - \mathcal{I}(k_1\tau, q_\tau^{k_1}) + \sum_{k=k_1+1}^{k_2} \mathcal{D}(q_\tau^{k-1}, q_\tau^k) \\ &= \mathcal{I}(k_2\tau, q_\tau^{k_2}) - \mathcal{I}(k_1\tau, q_\tau^{k_1}) + \mathrm{Var}(q_\tau, [t_1, t_2]) \end{aligned}$$

(the explicit form of $\mathrm{Var}(q_\tau, [t_1, t_2])$ holds since we consider a step function). To prove the validity of the second inequality in (3.5), we rely on the minimality of q_τ^k when compared with q_τ^{k-1} in (3.6). That is,

$$\mathcal{I}(k\tau, q_\tau^k) + \mathcal{D}(q_\tau^{k-1}, q_\tau^k) \leqslant \mathcal{I}(k\tau, q_\tau^{k-1}) = \mathcal{I}((k-1)\tau, q_\tau^{k-1}) + L((k-1)\tau, q_\tau^{k-1}) - L(k\tau, q_\tau^{k-1}).$$

Similarly as in the previous argument, a summation over $k = k_1+1, \ldots, k_2$ is employed to find that

$$\mathcal{I}(k_2\tau, q_\tau^{k_2}) - \mathcal{I}(k_1\tau, q_\tau^{k_1}) + \sum_{k=k_1+1}^{k_2} \mathcal{D}(q_\tau^{k-1}, q_\tau^k) \leqslant \sum_{k=k_1+1}^{k_2} [L((k-1)\tau, q_\tau^{k-1}) - L(k\tau, q_\tau^{k-1})], \tag{3.10}$$

so that

$$\mathcal{I}(k_2\tau, q_\tau^{k_2}) - \mathcal{I}(k_1\tau, q_\tau^{k_1}) + \mathrm{Var}(q_\tau, [t_1, t_2]) \leqslant -\int_{t_1}^{t_2} \dot{L}(t, q_\tau(t-\tau))\mathrm{d}t,$$

which is the second inequality in (3.5). □

3.3 *Limit passage for vanishing time step*

The next proposition gives the *a priori* bounds needed to pass to the limit as the step size goes to zero.

PROPOSITION 3.4 (*A priori* bounds). Assume that W satisfies the conditions (2.1)–(2.4) and that (2.7), (2.10), (2.11) are satisfied. Let $\frac{1}{\alpha} + \frac{1}{\beta} \leqslant \frac{1}{d}$. Further, suppose that $W^{1,r}(\Omega; \mathbb{R}^d)$ embeds continuously to $L^{r'}(\Omega; \mathbb{R}^d)$ and to $L^{\hat{r}}(\Gamma_1; \mathbb{R}^d)$. Then there exists a constant $\kappa > 0$ such that

$$\|y_\tau\|_{L^\infty(0,T;W^{1,r}(\Omega;\mathbb{R}^d))} < \kappa, \tag{3.11}$$

$$\|(1 \otimes |s|^d) \bullet \nu_\tau\|_{L^\infty(0,T;L^1(\Omega))} < \kappa, \tag{3.12}$$

$$\|\nu_\tau\|_{L^\infty(0,T;L^\infty_{\mathrm{w}}(\Omega;\mathcal{M}(\mathbb{R}^{n\times n})))} < \kappa, \tag{3.13}$$

$$\|\lambda_\tau\|_{L^\infty(0,T;W^{1,2}(\Omega;\mathbb{R}^{M+1}))\cap \mathrm{BV}(0,T;L^1(\Omega;\mathbb{R}^{M+1}))} < \kappa, \tag{3.14}$$

$$\mathrm{Var}(\mathcal{D}, q_\tau; [0,T]) < \kappa, \tag{3.15}$$

and for $\hat{\mathcal{I}}_\tau(t) := \mathcal{I}_\tau(t, q_\tau(t))$,

$$\|\hat{\mathcal{I}}_\tau\|_{\mathrm{BV}(0,T)} < \kappa. \tag{3.16}$$

Proof. We recall from (1.9) that

$$V(q) = \int_\Omega \int_{\mathbb{R}^{n\times n}} W(x, sF_{\mathrm{p}}^{-1}(x), F_{\mathrm{p}}(x), \nabla F_{\mathrm{p}}, p(x), \nabla p(x))\nu_x(\mathrm{d}s)\mathrm{d}x + \varepsilon\, \|\nabla\lambda\|^2_{L^2(\Omega;\mathbb{R}^{(1+M)\times n})}.$$

The growth conditions (2.3) imply that

$$\int_\Omega \int_{\mathbb{R}^{n\times n}} |sF_{\mathrm{p}}^{-1}(x)|^\alpha \nu_x(\mathrm{d}s)\mathrm{d}x + \|F_{\mathrm{p}}\|^\beta_{W^{1,\beta}(\Omega;\mathbb{R}^{n\times n})} + \|p\|^\omega_{W^{1,\beta}(\Omega;\mathbb{R}^m)} \leqslant V(q). \tag{3.17}$$

We use that for some $C > 0$ since $\frac{1}{\alpha} + \frac{1}{\beta} \leqslant \frac{1}{d}$,

$$\begin{aligned}\int_\Omega \Big|\int_{\mathbb{R}^{m\times n}} s\nu_x(\mathrm{d}s)\Big|^d \mathrm{d}x &\leqslant C \int_\Omega \int_{\mathbb{R}^{m\times n}} |s|^d \nu_x(\mathrm{d}s)\mathrm{d}x \\ &\leqslant C\|F_{\mathrm{p}}\|^\beta_{W^{1,\beta}(\Omega;\mathbb{R}^{n\times n})} \int_\Omega \int_{\mathbb{R}^{m\times n}} |sF_{\mathrm{p}}^{-1}(x)|^\alpha \nu_x(\mathrm{d}s)\mathrm{d}x;\end{aligned}$$

this, the fact that $\nabla y(x) = \int_{\mathbb{R}^{n\times n}} s\nu_x(\mathrm{d}s)$ for almost all $x \in \Omega$, and the Poincaré inequality yield together with (3.17)

$$C\|y\|^d_{W^{1,d}(\Omega;\mathbb{R}^n)} + \|F_{\mathrm{p}}\|^\beta_{W^{1,\beta}(\Omega;\mathbb{R}^{n\times n})} + \|p\|^\omega_{W^{1,\beta}(\Omega;\mathbb{R}^m)} \leqslant V(q). \tag{3.18}$$

Since $\mathcal{I} = V - L$ by (1.7), we find from (3.10) for $k_1 = 0$ that

$$V(q_\tau^{k_2}) - L(k_2\tau, q_\tau^{k_2}) - V(q_\tau^0) + L(0, q_\tau^0) \leqslant \sum_{k=1}^{k_2}[L((k-1)\tau, q_\tau^{k-1}) - L(k\tau, q_\tau^{k-1})],$$

which we rewrite as

$$V(q_\tau^{k_2}) \leqslant \sum_{k=1}^{k_2}[L((k-1)\tau, q_\tau^{k-1}) - L(k\tau, q_\tau^{k-1})] + L(k_2\tau, q_\tau^{k_2}) + C. \tag{3.19}$$

Combining this with the estimate (3.18) for $q := q_\tau^{k_2}$, we find for $Y_\tau := \max_{1\leqslant \ell\leqslant N} \|y_\tau^\ell\|^d_{W^{1,d}(\Omega;\mathbb{R}^d)}$ that

$$Y_\tau \leqslant \sum_{k=1}^{k_2}[L((k-1)\tau, q_\tau^{k-1}) - L(k\tau, q_\tau^{k-1})] + C. \tag{3.20}$$

This gives the bound (3.11) since y_τ appears in the load on the right-hand side, but only linearly (see (1.6)); since the power $d > 1$ on the left-hand side is larger, (3.11) follows. With (3.11) at our disposal, we immediately obtain (3.12), and (3.13)–(3.16) are easy (e.g. $\nabla\lambda$ is bounded in $L^2(\Omega, \mathbb{R}^{(M+1)\times n})$, and λ is bounded as a volume fraction; the BV bound comes from its contribution in the dissipation). □

PROPOSITION 3.5 Let $\mathcal{I}$ be weakly sequentially lower semicontinuous. Suppose that for all $(t_*, q_*) \in [0, T] \times \mathbb{Q}$, for all stable sequences $\{(t_k, q_k)\}_{k\in\mathbb{N}}$ with $t_k \to t_*$ and $q_k \rightharpoonup q_*$ in the sense of Definition 2.1, there is a sequence $\{\tilde{q}_k\}_{k\in\mathbb{N}} \subset \mathbb{Q}$ such that for all $\tilde{q} \in \mathbb{Q}$

$$\limsup_{k\to\infty}[\mathcal{I}(t_k, \tilde{q}_k) + \mathcal{D}(q_k, \tilde{q}_k)] \leqslant \mathcal{I}(t_*, \tilde{q}) + \mathcal{D}(q_*, \tilde{q}). \tag{3.21}$$

Then, $\mathcal{I}$ is weakly continuous along stable sequences and $q_* \in \mathcal{S}(t_*)$.

Proof. We follow the proof of Mainik & Mielke (2008, Proposition 4.2). Take $\tilde{q} = q_*$ in (3.21); by stability and then (3.21), we obtain

$$\limsup_{k\to\infty}\mathcal{I}(t_k, q_k) \leqslant \limsup_{k\to\infty}[(\mathcal{I}(t_k, \tilde{q}_k) + \mathcal{D}(q_k, \tilde{q}_k)] \leqslant \mathcal{I}(t_*, \tilde{q}) + \mathcal{D}(q_*, \tilde{q}) = \mathcal{I}(t_*, q_*).$$

We have further

$$\lim_{k\to\infty} \|\mathcal{I}(t_k, q_k) - \mathcal{I}(t_*, q_k)\| = \lim_{k\to\infty} \|L(t_k, q_k) - L(t_*, q_k)\| = 0$$

due to the Assumptions (2.10) and (2.11) on f and g, respectively.

Since $\mathcal{I}$ is weakly lower semicontinuous, it follows that

$$\liminf_{k\to\infty}\mathcal{I}(t_k, q_k) = \liminf_{k\to\infty}[\mathcal{I}(t_k, q_k) - \mathcal{I}(t_*, q_k)] + \liminf_{k\to\infty}\mathcal{I}(t_*, q_k) \geqslant \mathcal{I}(t_*, q_*).$$

This together with (3.21) gives weak continuity of $\mathcal{I}(t_k, q_k) \to \mathcal{I}(t_*, q_*)$. Finally, we have for every $\tilde{q} \in \mathbb{Q}$

$$\mathcal{I}(t_*, q_*) = \lim_{k\to\infty} \mathcal{I}(t_k, q_k) \leqslant \liminf_{k\to\infty}[\mathcal{I}(t_k, \tilde{q}_k) + \mathcal{D}(q_k, \tilde{q}_k)] \leqslant \mathcal{I}(t_*, \tilde{q}) + \mathcal{D}(q_*, \tilde{q}).$$

The arbitrariness of $\tilde{q} \in \mathbb{Q}$ shows the stability of q_*. □

The key point in the existence proof for a rate-independent process is to show the validity of (3.21). Let us suppose for the moment that irreversibility for the plastic process is excluded, i.e. ∞ is not contained in the range of $\mathcal{D}$ defined in (1.4), i.e. $\mathcal{D}_\mathrm{p}\colon \mathbb{Q} \times \mathbb{Q} \to [0, +\infty)$. Then it is sufficient to assume that for $\epsilon > 0$ small enough

$$D_\mathrm{p}(x, z_1, z_2) \leqslant c(x) + C(|F_{\mathrm{p}1}|^{\beta^*-\epsilon} + |F_{\mathrm{p}2}|^{\beta^*-\epsilon} + |p_1|^{\omega^*-\epsilon} + |p_2|^{\omega^*-\epsilon})$$

holds, with

$$\beta^* := \begin{cases} \frac{n\beta}{n-\beta} & \text{if } n > \beta, \\ 1+\xi & \text{else, for some } \xi > 0 \end{cases} \quad \text{and } \omega^* := \begin{cases} \frac{n\omega}{n-\omega} & \text{if } n > \omega, \\ > 1+\xi & \text{else, for some } \xi > 0. \end{cases}$$

Then the compact embedding ensures continuity of $\mathcal{D}_\mathrm{p}$. Similar, $\mathcal{D}_\mathrm{tr}$ is continuous by compactness. Thus, the dissipation $\mathcal{D}$ defined in (1.5) is continuous.

However, we allow irreversibility by including ∞ in the range of $\mathcal{D}_\mathrm{p}$, so that $\mathcal{D}_\mathrm{p}\colon \mathbb{Q}\times\mathbb{Q} \to [0, +\infty]$, and we thus must be more careful. Assumption 2.3 will play a central rôle in the next argument. We recall the notation $z_j := (F_j, p_j) \in \mathbb{R}^{n\times n} \times \mathbb{R}^m$ for $j = 1, 2$ used in that assumption.

PROPOSITION 3.6 Let $\beta, \omega > n$. Let Assumption 2.3 hold. Then (3.21) holds.

Proof. If $\mathcal{D}(q_*, \tilde{q}) = +\infty$ in (3.21), then nothing is to show. So we can assume that $\mathcal{D}_\mathrm{p}(q_*, \tilde{q}) \in \mathbb{R}$; then $(z_*, \tilde{z}) \in D(x)$, with $D(x)$ defined in (2.9). If $q_k \rightharpoonup q_*$ as $k \to \infty$, then due to the compact embedding

$$\rho_k := \|F_{\mathrm{p}k} - F_{\mathrm{p}*}\|_{C(\bar{\Omega};\mathbb{R}^{n\times n})} + \|p_k - p_*\|_{C(\bar{\Omega};\mathbb{R}^m)} \to 0.$$

Thus, there is $R > 0$ such that $\|z_k\| + |z_*| + |\tilde{z}| < R$ if k is large enough. We define $\tilde{z}_k := (\tilde{F}_\mathrm{p}, \tilde{p} + \eta_k v^*)$, where η_k relates to ρ_k as in Assumption 2.3(3). Then $(z_k, \tilde{z}_k) \in D(x)$ by Assumption 2.3(3) with $z_0 := z_*$ and $z_1 := \tilde{z}$. The continuity of D_p (Assumption 2.3(1)) gives the pointwise convergence $D_\mathrm{p}(x, z_k, \tilde{z}_k) \to D_\mathrm{p}(z, z_*, \tilde{z})$. Furthermore, we have $\|z_k\| < R$ and $\|\tilde{z}_k\| < R$ in addition to the property $(z_k, \tilde{z}_k) \in D(x)$ established above. Condition 2 of Assumption 2.3 together with the Lebesgue dominated convergence theorem implies that $\mathcal{D}_\mathrm{p}(q_k, \tilde{q}_k) \to \mathcal{D}_\mathrm{p}(q_*, \tilde{q})$. Further, $\mathcal{D}_\mathrm{tr}$ is continuous by compactness. Thus, the dissipation $\mathcal{D}$ defined in (1.5) is continuous. As for $\mathcal{I}$, Assumptions (2.10) and (2.11) imply that (3.21) is fulfilled with equality. □

The following lemma is taken from Mielke (2005, Theorem 5.21). Let us first denote $\mathbb{X} := L^\beta(\Omega; \mathbb{R}^{n\times n}) \times L^\omega(\Omega; \mathbb{R}^m)$. Note that if (2.7) and (2.8) hold for $\mathcal{D}_\mathrm{p}$ in $\mathbb{P}$, then they hold in $\mathbb{X}$ with the strong convergence in $\mathbb{X}$.

LEMMA 3.7 (Helly for plastic dissipation). Let $\mathcal{D}_\mathrm{p}\colon \mathbb{X} \times \mathbb{X} \to [0, +\infty]$. Let $\mathcal{K}$ be a compact subset of $\mathbb{X}$. Then for every sequence $\{z_k\}_{k\in\mathbb{N}}$, $z_k\colon [0, T] \to \mathcal{K}$ for which $\sup_{k\in\mathbb{N}} \mathrm{Var}(\mathcal{D}_\mathrm{p}, z_k; [0, T]) < C$ with some $C > 0$, there exists a subsequence (not relabelled), a function $z\colon [0, T] \to \mathcal{K}$ and a function $\delta\colon [0, T] \to [0, C]$ such that

1. $\mathrm{Var}(\mathcal{D}_\mathrm{p}, z_k; [0, t]) \to \delta(t)$ for all $t \in [0, T]$,
2. $z_k \to z$ for all $t \in [0, T]$ and
3. $\mathrm{Var}(\mathcal{D}_\mathrm{p}, z; [t_0, t_1]) \leqslant \lim_{t\searrow t_1} \delta(t) - \lim_{t\nearrow t_0} \delta(t)$ for all $0 \leqslant t_0 < t_1 \leqslant T$, with the limits evaluated as $\delta(0) = 0$, respectively, $\delta(T)$ in the cases $t_0 = 0$, respectively, $t_1 = T$.

The assertion of the following theorem includes also the assertion of Theorem 2.4.

THEOREM 3.8 (Existence of a rate-independent process). Let $\frac{1}{\alpha} + \frac{1}{\beta} \leqslant \frac{1}{d} < \frac{1}{n}$. Suppose further that (2.1)–(2.4), (2.7), (2.8) (2.10) and (2.11) hold. Then there exists a process $q\colon [0, T] \to \mathbb{Q}$ with $q(t) = (y(t), \nu(t), z(t), \lambda(t))$ such that q is an energetic solution in the sense of Definition 1.1. The following limit passages are also valid:

(i) for a t-dependent (not relabelled) subsequence, w^*-$\lim \nu_\tau(t) = \nu(t)$ in $L^\infty_\mathrm{w}(\Omega; \mathcal{M}(\mathbb{R}^{n\times n}))$ for all $t \in [0, T]$,
for a t-dependent (not relabelled) subsequence, w-$\lim_{\tau\to 0} y_\tau(t) = y(t)$ in $W^{1,d}(\Omega; \mathbb{R}^n)$ for all $t \in [0, T]$,

(ii) for a (not relabelled) subsequence, w^*-$\lim_{\tau\to 0} \lambda_\tau(t) = \lambda(t)$ in $L^\infty(\Omega; \mathbb{R}^{M+1}) \cap W^{1,2}(\Omega; \mathbb{R}^{M+1})$ for all $t \in [0, T]$ and $\lambda \in \mathrm{BV}([0, T]; L^1(\Omega; \mathbb{R}^L))$,

(iii) for a (not relabelled) subsequence, $\lim_{\tau\to 0} z_\tau(t) = z(t)$ in $\mathbb{X}$ for all $t \in [0, T]$,

(iv) for a (not relabelled) subsequence, $\lim_{\tau\to 0} \mathcal{I}_\tau(t, q_\tau) = \mathcal{I}(t, q(t))$ for all $t \in [0, T]$ and

(v) for a (not relabelled) subsequence, $\lim_{\tau\to 0} \mathrm{Var}(\mathcal{D}, q_\tau; [0, t]) = \mathrm{Var}(\mathcal{D}, q; [0, t])$ for all $t \in [0, T]$.

Proof. The proof is divided into several steps and combines ideas from Francfort & Mielke (2006) and Mielke (2005).

Step 1. The points (i), (ii) and (iii) follow from the *a priori* estimates in Proposition 3.4 and Lemma 3.7 (recall that λ is a volume fraction and thus trivially bounded in L^∞). Altogether, Step 1 implies the existence of the limit $q(t) = (y(t), \nu(t), \lambda(t), z(t))$. Moreover, (3.12) implies that $\int_\Omega \int_{\mathbb{R}^{n\times n}} |s|^d \nu(t)(\mathrm{d}s)\mathrm{d}x < +\infty$ for all $t \in [0, T]$. As $\nabla y_\tau(t, x) = \int_{\mathbb{R}^{n\times n}} s\nu_{\tau,x}(t)(\mathrm{d}s)$ and using Lemma 2.2, we immediately get that $q \in \mathbb{Q}$.

We set $S(t, \tau) := \min_{k\in\mathbb{N}\cup\{0\}}\{k\tau | k\tau \geqslant t\}$. Then $\lim_{\tau\to 0} S(t, \tau) = t$; then $q_\tau(t) := q_\tau(S(t, \tau)) \in \mathcal{S}(S(t, \tau))$. Moreover, by our assumptions on $\mathcal{D}$ (Assumption 2.3 and Proposition 3.6), we know that (3.21) holds. Therefore, $q(t) \in S(t)$, i.e. the limit is stable by Proposition 3.5. Proposition 3.5 also implies (iv).

Step 2. We have $q_\tau(t) = q_\tau(k\tau)$ if $0 \leqslant t - k\tau \leqslant \tau$. Hence, using (3.5) in the first and in the second line, we find that for some $C, C_1 > 0$

$$\begin{aligned}\mathcal{I}(t, q_\tau(t)) + \mathrm{Var}(\mathcal{D}, q_\tau; [0, t]) &\leqslant \mathcal{I}(k\tau, q_\tau(k\tau)) + \mathrm{Var}(\mathcal{D}, q_\tau; [0, k\tau]) + C\tau \\ &\leqslant \mathcal{I}(0, q_\tau(0)) - \int_0^{k\tau} \dot{L}(s, q_\tau(s))\mathrm{d}s + C\tau \\ &\leqslant \mathcal{I}(0, q_\tau(0)) - \int_0^{t} \dot{L}(s, q_\tau(s))\mathrm{d}s + C_1\tau.\end{aligned}$$

Note also, $\theta_\tau(t) := \dot{L}(t, q_\tau)$ is bounded in $L^\infty(0, T)$ by (2.10) and (2.11), so that there is a weak* limit of a subsequence (not relabelled), which we denote θ. We set $\theta_{\mathrm{i}}(t) := \liminf_{\tau\to 0} \theta_\tau(t)$. Further, using Lemma 3.7(i) and the weak lower semicontinuity of the variation, we get in the limit $\tau \to 0$

$$\mathcal{I}(t, q(t)) + \delta(t) + \mathrm{Var}(\mathcal{D}_{\mathrm{tr}}, q; [0, t]) \leqslant \mathcal{I}(0, q(0)) - \int_0^t \theta(s)\mathrm{d}s. \tag{3.22}$$

As $\delta(t) \geqslant \mathrm{Var}(\mathcal{D}_{\mathrm{p}}, q; [0, t])$ by Lemma 3.7 and by Fatou's lemma $\int_0^t \theta(s)\mathrm{d}s \geqslant \int_0^t \theta_{\mathrm{i}}(s)\mathrm{d}s$ for a.e. $t \in [0, T]$, we obtain

$$\mathcal{I}(t, q(t)) + \mathrm{Var}(\mathcal{D}, q; [0, t]) \leqslant \mathcal{I}(0, q(0)) - \int_0^t \theta_{\mathrm{i}}(s)\mathrm{d}s.$$

We observe that $\theta_{\mathrm{i}}(s) = \dot{L}(s, q(s))$. Altogether, we get the upper energy estimate

$$\mathcal{I}(t, q(t)) + \mathrm{Var}(\mathcal{D}, q; [0, t]) \leqslant \mathcal{I}(0, q(0)) - \int_0^t \dot{L}(s, q(s)\mathrm{d}s. \tag{3.23}$$

In order to get the lower estimate, we exploit the fact that $q(t)$ is stable for all $t \in [0, T]$. Take a (possibly non-uniform) partition of a time interval $[t_1, t_2] \subset [0, T]$ such that $t_1 = \vartheta_0 < \vartheta_1 < \vartheta_2 < \vartheta_K = t_2$ such that $\max_i(\vartheta_i - \vartheta_{i-1}) =: \vartheta \to 0$ as $K \to \infty$. We test the stability of $q(\vartheta_{k-1})$ with $q(\vartheta_k)$ for $k = k_1 + 1, \ldots, k_2$. Analogously to (3.9), this yields

$$\sum_{k=1}^{K}[L(\vartheta_{k-1}, q(\vartheta_k)) - L(\vartheta_k, q(\vartheta_k))] \leqslant \mathcal{I}(t_2, q(t_2)) - \mathcal{I}(t_1, q(t_1)) + \sum_{k=1}^{K}\mathcal{D}(q(\vartheta_{k-1}), q(\vartheta_k)). \tag{3.24}$$

Hence,

$$\sum_{k=1}^{K} -\int_{\vartheta_{k-1}}^{\vartheta_k} \dot{L}(s, q(\vartheta_k))\mathrm{d}s \leqslant \mathcal{I}(t_2, q(t_2)) - \mathcal{I}(t_1, q(t_1)) + \mathrm{Var}(\mathcal{D}, q; [t_1, t_2]). \tag{3.25}$$

Finally, we observe that

$$\sum_{k=1}^{K}\int_{\vartheta_{k-1}}^{\vartheta_k}\dot{L}(s,q(\vartheta_k))\mathrm{d}s=\sum_{k=1}^{K}\dot{L}(\vartheta_k,q(\vartheta_k))(\vartheta_k-\vartheta_{k-1})+\sum_{k=1}^{K}\int_{\vartheta_{k-1}}^{\vartheta_k}(\dot{L}(s,q(\vartheta_k))-\dot{L}(\vartheta_k,q(\vartheta_k)))\mathrm{d}s. \tag{3.26}$$

The second term on the right-hand side of (3.26) tends to zero as $\vartheta\to 0$ because the time derivative of external forces is uniformly continuous in time by (2.10) and (2.11). The first term on the right-hand side converges to $\int_{t_1}^{t_2}\dot{L}(s,q(s))\mathrm{d}s$ (Dal Maso *et al.*, 2005, Lemma 4.12). Thus, (3.25) and (3.26) together yield the lower energy bound

$$-\int_{t_1}^{t_2}\dot{L}(s,q(s))\mathrm{d}s\leqslant\mathcal{I}(t_2,q(t_2))-\mathcal{I}(t_1,q(t_1))+\mathrm{Var}(\mathcal{D},q;[t_1,t_2]). \tag{3.27}$$

The upper and lower estimates (3.23) and (3.27) combined yield the energy balance

$$\mathcal{I}(t,q(t))+\mathrm{Var}(\mathcal{D},q;[0,t])=\mathcal{I}(0,q(0))-\int_0^t\dot{L}(s,q(s)\mathrm{d}s. \tag{3.28}$$

Step 3. We obtain (the first inequality relies on (3.27) and the observation $\theta_{\mathrm{i}}(s)=\dot{L}(s,q(s))$ made in Step 2, the second inequality is Lemma 3.7(3), while the third estimate is (3.22))

$$\begin{aligned}\mathcal{I}(0,q(0))-\int_0^t\theta_{\mathrm{i}}(s)\mathrm{d}s&\leqslant\mathcal{I}(t,q(t))+\mathrm{Var}(\mathcal{D},q;[0,t])\\&\leqslant\mathcal{I}(t,q(t))+\delta(t)+\mathrm{Var}(\mathcal{D}_{\mathrm{tr}},q;[0,t])\\&\leqslant\mathcal{I}(0,q(0))-\int_0^t\theta(s)\mathrm{d}s\leqslant\mathcal{I}(0,q(0))-\int_0^t\theta_{\mathrm{i}}(s)\mathrm{d}s.\end{aligned} \tag{3.29}$$

Thus, all inequalities in (3.29) are equalities and consequently, we have shown that (v) holds. □

4. Numerical example

To provide a numerical validation of the model, let us consider the small strain version of the model studied in this article. Thus, with $u\colon\Omega\to\mathbb{R}^n$ being the displacement, let $e=\frac{1}{2}(\nabla u+\nabla u^{\top})$ be the small strain tensor and $E_{\mathrm{p}}\colon\Omega\to\mathbb{R}^{n\times n}$ a plastic strain which is supposed to be trace free. Consider a simple two-phase problem with the energy well corresponding to the first phase positioned at the origin and the well corresponding to the second phase placed at a given symmetric matrix $0\neq\epsilon\in\mathbb{R}^{n\times n}$. If the tensor of elastic constants $\mathbb{C}$ of both phases is considered equal, we define the double well energy density (Kohn, 1991) as

$$\hat{\varphi}(e):=\frac{1}{2}\min\{\mathbb{C}e\colon e,\mathbb{C}(e-\epsilon)\colon(e-\epsilon)+\alpha\},$$

where $\alpha\in\mathbb{R}$ is a vertical shift controlling a mutual position of the bottoms of the two wells and typically depends on the temperature. The colon ':' denotes the dot product on $\mathbb{R}^{n\times n}$. Obviously, $\hat{\varphi}$ is not convex. If $\epsilon=a\otimes b+b\otimes a$ for some vectors $a,b\in\mathbb{R}^n$, the relaxation of this non-convex energy at a given volume fraction λ of the second phase is

$$e\mapsto\frac{1}{2}\mathbb{C}(e-\lambda\epsilon)\colon(e-\lambda\epsilon)+\lambda\alpha$$

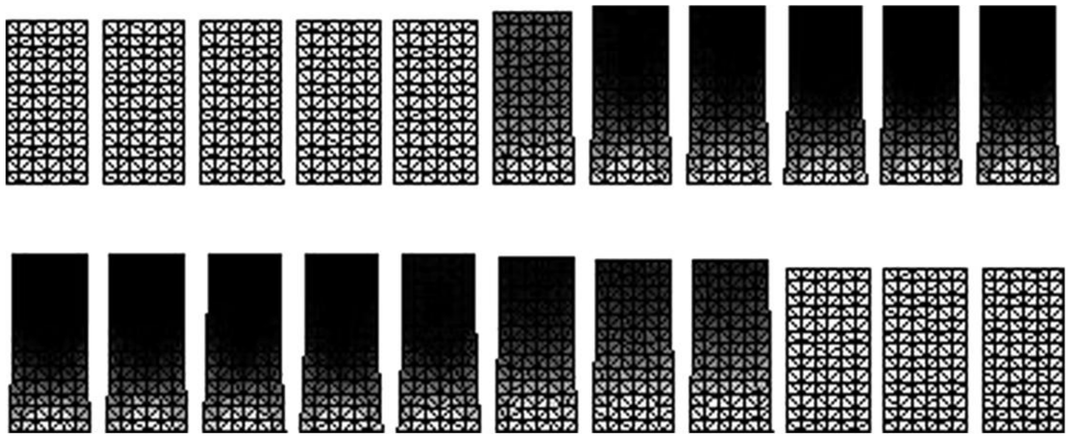

FIG. 2. Evolution of the deformation of the specimen. The load increases from left to right in the upper row and then decreases again (lower row, left to right). White colour denotes austenite, black martensite. The grey scale indicates the volume fraction of the martensite on every element. The displacement is magnified by a factor of two.

as shown by Kohn (1991). In this example, we put $q := (u, \lambda, z)$, where $z := (E_p, p)$ with $p \in \mathbb{R}$ a single hardening variable. Hence, the stored energy functional reads

$$I(q) := \frac{1}{2}\int_\Omega [\mathbb{C}(e - \lambda\epsilon - E_p) : (e - \lambda\epsilon - E_p) + p^2 + \alpha\lambda]\mathrm{d}x - L(u),$$

where L is a linear continuous functional measuring work of external forces. As elsewhere (Carstensen & Plecháč, 2000), the energy dissipation related to the phase transition is simply defined as ($h_1 > 0$)

$$\mathcal{D}_{\mathrm{tr}}(q_1, q_2) := h_1 \int_\Omega \|\lambda_1(x) - \lambda_2(x)\|\mathrm{d}x.$$

As for the plastic dissipation, we follow Mielke (2003) and take for $h_2 > 0$ and $\tilde{h}_2 > 0$

$$\mathcal{D}_p(q_1, q_2) := \begin{cases} \int_\Omega h_2|E_{p_1} - E_{p_2}|\mathrm{d}x & \text{if } p_2 \geqslant p_1 + \tilde{h}_2|E_{p_1} - E_{p_2}|, \\ +\infty & \text{else.} \end{cases}$$

Sending $h_2 \to \infty$ models the problem without plastic dissipation while setting $h_1 = +\infty$ fixes the volume fraction λ and our model simplifies to usual linearized elastoplasticity of the 'mixed' material. Instead of external forces the evolution can be also driven by time-dependent boundary conditions (Mainik & Mielke, 2008; Mielke, 2005).

In computational examples shown in Figs. 2–3, we consider a simple 2D tensile experiment, $n = 2$, with a specimen $\Omega = (0, 1) \times (0, 2)$, over the time interval $[0, 5]$. The specimen is fixed on $(0, 1) \times \{0\}$; a prescribed time-dependent 1-periodic surface force g acts in the vertical direction on $(0, 1) \times \{2\}$, i.e. $g = (0, g_2)$ with

$$g_2(t) = \begin{cases} 5.10^8 t & \text{if } 0 \leqslant t \leqslant 0.5 \\ -5.10^8 t + 5.10^8 & \text{if } 0.5 \leqslant t \leqslant 1. \end{cases}$$

As to the material, we consider a cubic to tetragonal stress-induced transformation with

$$\epsilon := \mathrm{diag}(-0.05, 0.05) = 10^{-2}((5, 5) \otimes (-1, 1) + (-1, 1) \otimes (5, 5))$$

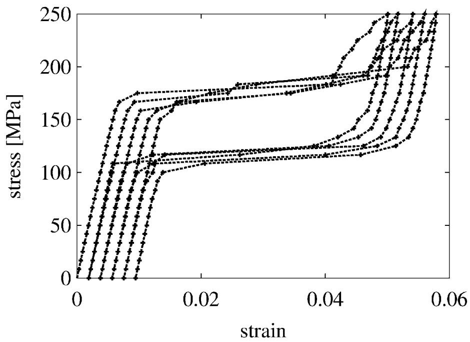

FIG. 3. The 22-component of the stress versus the 22-component of the strain during 6 cycles.

with $h_1 = 8$ MPa, $h_2 = 10$ MPa, $\tilde{h}_2 := 2$ MPa and $\alpha = 5$ MPa. The tensor of elastic constants is, for simplicity, reduced to the Young modulus (10 GPa) and the Poisson ration (0.3). The initial condition is always $\lambda = p = E_{\rm p} = 0$, so that the material is initially in the austenite phase and without any plastic deformation. Computational results are shown in Figs. 2 and 3.

The numerical values for the Poisson ratio and Young's modulus are chosen to be typical for elastic solids; no attempt has been made to fit them to experimental data to get qualitative agreement between simulation and experiment rather than quantitative agreement. This is since a full numerical study of cubic-to-monoclinic transformation in the setting of multiplicative elastoplasticity is a research topic in its own right.

Acknowledgements

This work was initiated during MK's visit of the University of Bath. Its support and hospitality is gratefully acknowledged. We thank Dr. Petr Šittner for providing us with pictures in Figure 1 and to the referees for helpful comments.

Funding

This work was supported by the Royal Society (JP080789); the Ministry of Education, Youth and Sports of the Czech Republic (VZ6840770021 to M.K.); the Academy of Sciences of the Czech Republic (IAA100750802 to M.K.), Grant Agency of the Czech Republic (P201/10/0357 to M.K.) and the Engineering and Physical Sciences Research Council (GR/S99037/1 to J.Z.).

REFERENCES

ACHENBACH, M. & MÜLLER, I. (1982) A model for shape memory. *J. Phys.—Paris*, **43**, 163–167.

ACHENBACH, M. & MÜLLER, I. (1985) Simulation of material behavior of alloys with shape memory. *Arch. Mech.*, **37**, 573–585.

ARNDT, M. (2004) Upscaling from atomistic models to higher order gradient continuum models for crystalline solids. Dissertation, Institute for Numerical Simulation, University of Bonn, Germany.

ARTEMEV, A., JIN, Y. & KHACHATURYAN, A. G. (2001) Three-dimensional phase field model of proper martensitic transformation. *Acta Mater.*, **49**, 1165–1177.

AURICCHIO, F., REALI, A. & STEFANELLI, U. (2007) A three-dimensional model describing stress-induced solid phase transformation with permanent inelasticity. *Int. J. Plast.*, **23**, 207–226. Available at http://www.sciencedirect.com/science/article/B6TWX-4KST3J3-1/2/3d0e05c0a5c11d04300f852a1ecad33a.

BALL, J. M. (1989) A version of the fundamental theorem for Young measures. *PDEs and Continuum Models of Phase Transitions (Nice, 1988)* (M. Rascle, D. Serre & M. Slemrod, eds). Berlin: Springer, pp. 207–215.

BHATTACHARYA, K. (2003) *Microstructure of Martensite. Why It Forms and How It Gives Rise to the Shape-Memory Effect*. Oxford Series on Materials Modelling, vol. 2. New York: Oxford University Press.

CARSTENSEN, C., HACKL, K. & MIELKE, A. (2002) Non-convex potentials and microstructures in finite-strain plasticity. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.*, **458**, 299–317. Available at http://dx.doi.org/10.1098/rspa.2001.0864.

CARSTENSEN, C. & PLECHÁČ, P. (2000) Numerical analysis of compatible phase transitions in elastic solids. *SIAM J. Numer. Anal.*, **37**, 2061–2081 (electronic). Available at http://dx.doi.org/10.1137/S0036142998337697.

DAL MASO, G., FRANCFORT, G. A. & TOADER, R. (2005) Quasistatic crack growth in nonlinear elasticity. *Arch. Ration. Mech. Anal.*, **176**, 165–225.

DELVILLE, R., MALARD, B., PILCH, J., ŠITTNER, P. & SCHRYVERS, D. (2011) Transmission electron microscopy investigation of dislocation slip during superelastic cycling of Ni-Ti wires. *Int. J. Plast.*, **27**, 282–297.

DILLON JR., O. W. & KRATOCHVÍL, J. (1970) A strain gradient theory of plasticity. *Int. J. Solids Struct.*, **6**, 1513–1533.

FRANCFORT, G. & MIELKE, A. (2006) Existence results for a class of rate-independent material models with nonconvex elastic energies. *J. Reine Angew. Math.*, **595**, 55–91.

GROMA, I., CSIKOR, F. F. & ZAISER, M. (2003) Spatial correlations and higher-order gradient terms in a continuum description of dislocation dynamics. *Acta Mater.*, **51**, 1271–1281. Available at http://www.sciencedirect.com/science/article/B6TW8-47VYG72-1/2/b34c96b1d17c0637eb493a339df521fd.

GURTIN, M. E. (2000) On the plasticity of single crystals: free energy, microforces, plastic-strain gradients. *J. Mech. Phys. Solids*, **48**, 989–1036.

HUO, Y. & MÜLLER, I. (1993) Nonequilibrium thermodynamics of pseudoelasticity. *Contin. Mech. Thermodyn.*, **5**, 163–204.

KASTNER, O. & ACKLAND, G. J. (2009) Mesoscale kinetics produces martensitic microstructure. *J. Mech. Phys. Solids*, **57**, 109–121. Available at http://dx.doi.org/10.1016/j.jmps.2008.09.016.

KINDERLEHRER, D. & PEDREGAL, P. (1994) Gradient Young measures generated by sequences in Sobolev spaces. *J. Geom. Anal.*, **4**, 59–90.

KOHN, R. V. (1991) The relaxation of a double-well energy. *Contin. Mech. Thermodyn.*, **3**, 193–236.

KRATOCHVÍL, J., KRUŽÍK, M. & SEDLÁČEK, R. (2007) Statistically based continuum model of misoriented dislocation cell structure formation. *Phys. Rev. B*, **75**, 064104.

KRATOCHVÍL, J., KRUŽÍK, M. & SEDLÁČEK, R. (2009) A model of ultrafine microstructure evolution in materials deformed by high-pressure torsion. *Acta Mater.*, **57**, 739–748.

KRATOCHVÍL, J. & SEDLÁČEK, R. (2008) Statistical foundation of continuum dislocation plasticity. *Phys. Rev. B*, **77**, 134102.

KRUŽÍK, M. & ZIMMER, J. (2010) Evolutionary problems in nonreflexive spaces. *ESAIM Control Optim. Calc. Var.*, **1**, 1–22.

KUNDIN, J., EMMERICH, H. & ZIMMER, J. (2010) Three-dimensional model of martensitic transformations with elasto-plastic effects. *Philos. Mag.*, **90**, 1495–1510.

MAINIK, A. & MIELKE, A. (2008) Global existence for rate-independent gradient plasticity at finite strain, *WIAS Preprint 1299*.

MIELKE, A. (2003) Energetic formulation of multiplicative elasto-plasticity using dissipation distances. *Contin. Mech. Thermodyn.*, **15**, 351–382.

MIELKE, A. (2005) Evolution of rate-independent systems. *Handbook of Differential Equations: Evolutionary Equations*, vol. II. North-Holland, Amsterdam: Elsevier, pp. 461–559.

MIELKE, A. & ROUBÍČEK, T. (2003) A rate-independent model for inelastic behavior of shape-memory alloys. *Multiscale Model. Simul.*, **1**, 571–597 (electronic).

MIELKE, A., THEIL, F. & LEVITAS, V. I. (2002) A variational formulation of rate-independent phase transformations using an extremum principle. *Arch. Ration. Mech. Anal.*, **162**, 137–177.

MÜLLER, I. & SEELECKE, S. (1996) Thermodynamic aspects of shape memory alloys. *Shape Memory Alloys—From Microstructure to Macroscopic Properties* (G. Airoldi ed.). Zurich, Switzerland: Trans Tech Publications.

NOVÁK, V., ŠITTNER, P., PILCH, J. & DELVILLE, R. (2009) Effect of plastic slip on thermomechanical behavior of NiTi polycrystals investigated by micromechanics modelling. *ESOMAT 2009, EDP Sciences*, (P. Šittner, V. Paidar, L. Heller & H. Seiner eds), p. 03009. Available at http://dx.doi.org/10.1051/esomat/200903009.

REN, X. & TRUSKINOVSKY, L. (2000) Finite scale microstructures in nonlocal elasticity. *J. Elasticity*, **59**, 319–355. In recognition of the sixtieth birthday of Roger L. Fosdick (Blacksburg, VA, 1999). Available at http://dx.doi.org/10.1023/A:1011003321453.

ROUBÍČEK, T. (2004) Models of microstructure evolution in shape memory alloys. *Nonlinear Homogenization and its Applications to Composites, Polycrystals and Smart Materials*. NATO Sci. Ser. II Math. Phys. Chem., vol. 170. Dordrecht, (P. Ponte Castañeda, J. Joachim Telega & B. Gambin eds). The Netherlands: Kluwer Academic Publishers, pp. 269–304.

SADJADPOUR, A. & BHATTACHARYA, K. (2007) A micromechanics inspired constitutive model for shape-memory alloys: the one-dimensional case. *Smart Mater. Struct.*, **16**, S51–S62.

SCHONBEK, M. E. (1982) Convergence of solutions to nonlinear dispersive equations. *Commun. Partial Differ. Equ.*, **7**, 959–1000.

SHU, Y. C. & YEN, J. H. (2008) Multivariant model of martensitic microstructure in thin films. *Acta Mater.*, **56**, 3969–3981.

THAMBURAJA, P. & ANAND, L. (2003) Thermo-mechanically coupled superelastic response of initially-textured Ti-Ni sheet. *Acta Mater.*, **51**, 325–338.

VIVET, A. & LEXCELLENT, C. (1998) Micromechanical modelling for tension–compression pseudoelastic behavior of AuCd single crystals. *Eur. Phys. J. Appl. Phys.*, **4**, 125–132. Available at http://dx.doi.org/doi/10.1051/epjap:1998251.

YOUNG, L. C. (1937) Generalized curves and the existence of an attained absolute minimum in the calculus of variations. *C. R. Soc. Sci. Lettres de Varsovie, Cl. III*, **30**, 212–234.

Appendix A. Young measures

We briefly recall the concept of Young measures (Young, 1937) and follow the presentation of Kruží & Zimmer (2010). Young measures describe the limit of a sequence $\{y_k\}_{k\in\mathbb{N}}$ of functions $y_k\colon \Omega \to \mathbb{R}^d$ which converges weakly in $L^q(\Omega;\mathbb{R}^d)$ for $1 \leqslant q < \infty$ or weakly∗ if $q = \infty$. The precise concept is as follows. A 'Young measure' on a bounded domain $\Omega \subset \mathbb{R}^n$ is a weakly* measurable mapping

$$\Omega \to \mathrm{Prob}(\mathbb{R}^d), \quad x \mapsto \nu_x,$$

with values in the probability measures. We recall that a mapping with values in the Radon measures is 'weakly* measurable' if for any $f \in C_0(\mathbb{R}^d)$, the mapping

$$\Omega \to \mathbb{R}, \quad x \mapsto \langle f, \nu_x\rangle := \int_{\mathbb{R}^d} f(s)\nu_x(\mathrm{d}s)$$

is measurable in the usual sense. We denote the set of all Young measures by $\mathcal{Y}(\Omega;\mathbb{R}^d)$.

It is known that $\mathcal{Y}(\Omega;\mathbb{R}^d)$ is a convex subset of $L^\infty_{\mathrm{w}}(\Omega; M(\mathbb{R}^d)) \cong L^1(\Omega; C_0(\mathbb{R}^d))^*$, where $L^\infty_{\mathrm{w}}(\Omega; M(\mathbb{R}^d))$ is the space of weakly* measurable bounded functions. The 'parametrized' Young measure theorem (Schonbek, 1982) states that for every sequence $\{y_k\}_{k\in\mathbb{N}}$ which is bounded in

$L^\infty(\Omega;\mathbb{R}^d)$, there exists a subsequence (denoted by the same indices for notational simplicity) and a Young measure $\nu = \{\nu_x\}_{x\in\Omega} \in \mathcal{Y}(\Omega;\mathbb{R}^d)$ such that for every continuous function $f\colon \mathbb{R}^d \to \mathbb{R}$,

$$f(y_k) \overset{\star}{\rightharpoonup} x \mapsto \langle f, \nu_x\rangle \text{ weakly* in } L^\infty(\Omega), \tag{A.1}$$

where

$$\langle f, \nu_x\rangle := \int_{\mathbb{R}^d} f(s)\nu_x(\mathrm{d}s) \tag{A.2}$$

is the 'expectation' of f. Let $\mathcal{Y}^\infty(\Omega;\mathbb{R}^d)$ denote set of all Young measures that are generated by taking all bounded sequences $\{y_k\}_{k\in\mathbb{N}}$ in $L^\infty(\Omega;\mathbb{R}^d)$.

The above concept is applicable if $\{y_k\}_{k\in\mathbb{N}}$ is uniformly bounded in $L^\infty(\Omega;\mathbb{R}^d)$. If in addition to the uniform bound in $L^\infty(\Omega;\mathbb{R}^d)$, $y_k \rightharpoonup y$ in $L^q(\Omega;\mathbb{R}^d)$ with $1 \leqslant q < \infty$, then $y_k \to y$ if and only if the corresponding Young measure is a Dirac mass, $\nu_x = \delta_{y(x)}$. Non-Dirac Young measures thus record possible oscillations in the limit process.

The assumption that $\{y_k\}_{k\in\mathbb{N}}$ is bounded in $L^\infty(\Omega;\mathbb{R}^d)$ can be relaxed to the assumption of such a bound in $L^q(\Omega;\mathbb{R}^d)$ with $1 < q < \infty$. The parametrized Young measure theorem is then valid under stronger growth conditions on the non-linearity f. The precise formulation has been given by Schonbek (1982, Theorem 2.2) (see also Ball, 1989 for a general formulation of the parametrized Young measure theorem). Namely, for every sequence $\{y_k\}_{k\in\mathbb{N}}$ which is uniformly bounded in $L^q(\Omega;\mathbb{R}^d)$ for some $q > 1$, there exists a subsequence, still indexed by k for notational convenience, and a Young measure $\nu = \{\nu_x\}_{x\in\Omega} \in \mathcal{Y}(\Omega;\mathbb{R}^d)$ such that for every $f \in C(\mathbb{R}^d)$ with

$$\{f(y_k)\}_{k\in\mathbb{N}} \text{ is weakly relatively compact in } L^1(\Omega) \tag{A.3}$$

the following holds in $L^1(\Omega;\mathbb{R}^d)$:

$$f(y_k) \rightharpoonup \langle f, \nu_x\rangle. \tag{A.4}$$

We say that $\{y_k\}_{k\in\mathbb{N}}$ generates ν if (A.4) holds; we denote the set of all Young measures obtained as limits of bounded sequences in $L^q(\Omega;\mathbb{R}^d)$ by $\mathcal{Y}^q(\Omega;\mathbb{R}^d)$. If $\{y_k\}_{k\in\mathbb{N}}$ is bounded in $W^{1,q}(\Omega;\mathbb{R}^d)$, then the set of Young measures generated by subsequences of $\{\nabla y_k\}_{k\in\mathbb{N}}$ will be denoted $\mathbb{G}^q(\Omega;\mathbb{R}^{d\times d})$. In the spirit of (A.1), we extend the energy V to $\bar{V}(\nu) := \int_\Omega \int_{\mathbb{R}^{d\times d}} \varphi(s)\nu_x(\mathrm{d}s)\mathrm{d}x$ for $\nu \in \mathbb{G}^q(\Omega;\mathbb{R}^{d\times d})$.