# Computational Modeling of Size-Dependent Superelasticity in Shape Memory Alloys



by
Lei Qiao
B.S., Peking University (2004)
M.S., Peking University (2007)
S.M., Massachusetts Institute of Technology (2009)

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Aeronautics and Astronautics
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
September 2013



Author .....
Department of Aeronautics and Astronautics
August 22, 2013

Certified by .....
Raúl A. Radovitzky
Professor of Aeronautics and Astronautics
Thesis Supervisor

Committee Member .....
Rohan Abeyaratne
Professor of Mechanical Engineering

Committee Member .....
Lallit Anand
Professor of Mechanical Engineering

Committee Member .....
Christopher A. Schuh
Professor of Materials Science and Engineering

Accepted by .....
Eytan H. Modiano
Professor of Aeronautics and Astronautics
Chair, Graduate Program Committee

# Computational Modeling of Size-Dependent Superelasticity in Shape Memory Alloys

by

Lei Qiao

Submitted to the Department of Aeronautics and Astronautics
on August 22, 2013, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Aeronautics and Astronautics

## Abstract

The superelastic effect in shape memory alloys (SMAs) is attributed to the stress-induced reversible austenitic-martensitic phase transformations. It is characterized by the development of significant strains which are fully recoverable upon unloading, and also characterized by the stress-hysteresis in the loading and unloading cycle which corresponds to the energy dissipated during phase transformations. Recently, experiments have revealed size-dependent effects in the superelastic responses of SMAs at micro- and nanoscales. For instance, the CuAlNi microwires and submicron pillars show a substantially higher capacity for the energy dissipation than that of bulk samples, which offers a significant promise for the applications in protective materials.

In this thesis, a continuum model is developed in order to improve our understanding of size effects in SMAs at small scales. The modeling approach combines classic superelastic models, which use the volume fraction as an internal variable to represent the martensitic phase transformation, with strain gradient plasticity theories. Size effects are incorporated through two internal length scales, an energetic length scale and a dissipative length scale, which correspond to the martensitic volume fraction gradient and its time rate of change, respectively. Introducing the gradient of the martensitic volume fraction leads to coupled macro- and microforce balance equations, where the displacements and the martensitic volume fraction are both independent fields. A variational formulation for the temporally-discretized coupled macro- and microforce balance equations is proposed, as well as a computational framework based on this formulation. A robust and scalable parallel algorithm is implemented within this computational framework, which enables the large-scale three-dimensional study of size effects in SMAs with unprecedented resolution. This modeling and computational framework furnishes, in effect, a versatile tool to analyze a broad range of problems involving size effects in superelasticity with the potential to guide microstructure design and optimization. In particular, the model captures the increase of the stress hysteresis and strain hardening in bulk polycrystalline SMAs for decreasing grain size, as well as the increase of the residual strain for decreasing pillar size in NiTi pillars. The model confirms that constraints like grain boundaries

and the surface Ti oxide layer are responsible for the size-dependent superelasticity in SMAs.

Thesis Supervisor: Raúl A. Radovitzky
Professor of Aeronautics and Astronautics

Committee Member: Rohan Abeyaratne
Professor of Mechanical Engineering

Committee Member: Lallit Anand
Professor of Mechanical Engineering

Committee Member: Christopher A. Schuh
Professor of Materials Science and Engineering

# Acknowledgments

I would like to express my deep gratitude to my advisor, Prof. Raúl Radovitzky, for his constant support, guidance and encouragement since I came to MIT. Throughout the years, he has been a great friend and mentor to me. From motivating this research to improving my presentation skills, his help has been in many aspects.

I would like to thank the thesis committee members: Prof. Rohan Abeyaratne, Prof. Lallit Anand and Prof. Christopher A. Schuh for their invaluable advice, discussion and encouragement throughout this thesis work.

I would like to thank Prof. Ken Kamrin and Prof. Joost Vlassak for reading this thesis. From Prof. Vlassak's lectures at Harvard, I first learned solid mechanics. I would also like to thank Prof. Samuel Allen for reading my thesis proposal, and Prof. Michael Demkowicz for serving as the advisor of my minor field.

I am thankful for the financial support from the U.S. Army through the Institute for Soldier Nanotechnologies.

I am very grateful to my fellow students in the RR group, including Michelle Nyein, Andrew Seagraves, Brandon Talamini, Michael Tupek, Wendy Pino, Abiy Tasissa, Felipe Hernandez, Piotr Fidkowski, Amanda Jason, Li Yu, Srikanti Rupa Avasarala, and Ganesh Gurumurthy, for their friendship, support, and encouragement. I would also like to thank the postdoctoral researchers in the group, including Gauthier Becker, Martin Hautefeuille, Aurélie Jean, Adrian Rosolen, Claudio Pita, Julian Rimoli, Ludovic Noels, Tan Bui, and Antoine Jerusalem. I must especially thank Piotr for initiating SUMMIT, and also thank Aurélie, Martin, and Adrian for continuously developing this software, which I have been used intensively during this work.

I would also like to thank my neighbour, Prof. John Dugundji, for his interest in my research and his care about my family in Belmont.

I would like to thank my parents for supporting and encouraging me to complete this thesis. Last, but not least, I want to thank my wife, Yan Yin, for her love and support. I could not have done this without her.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

The shape memory effect is a phenomenon wherein special materials recover their original shape upon heating to some critical temperature. Alloys with the shape memory effect are called shape memory alloys (SMAs). According to Otsuka and Wayman [80], the shape memory effect was first discovered in Au-Cd alloy by Chang and Read in 1951. Another important milestone is the discovery of the shape memory effect in Ni-Ti alloy by Buehler *et al* in 1963 [16], which eventually led to successful commercialization. The shape memory effect has also been observed in In-Tl, Cu-Zn, Cu-Al-Ni, and others. Besides the shape memory effect, shape memory alloys often exhibit another interesting feature, superelasticity, which refers to the complete recovery of deformation during a mechanical loading and unloading cycle that is far beyond the elastic limit of common metals. Superelasticity is unusual in that despite the complete recovery a certain amount of energy is dissipated through deformation. The unique features of shape memory effect and superelasticity make the applications of SMAs very broad. Examples of applications include antennae and actuators in aerospace engineering, cardiovascular stents and dental braces in biomedical engineering, as well as eyeglass frames, fishing rods, and headbands of headphones in consumer products. A review of SMAs applications can be found in [127].

The mechanism underlying the shape memory and superelastic effects is the so-called martensitic phase transformation, a diffusionless solid to solid transformation [80, 89]. Induced by the stress, temperature or magnetic field, the martensitic phase

transformation occurs as a sudden change in the lattice structure where each atom moves for less than one interatomic spacing. The phase with high order of symmetry in lattice structure (e.g. cubic for Ni-Ti) is called austenite, the stable phase at high temperatures. The phase with relatively low order of symmetry in the lattice structure (e.g. monoclinic for Ni-Ti) is called martensite, the stable phase at low temperatures. In Fig. 1-1, a schematic stress-temperature phase diagram and a stress-loading cycle have been shown in order to demonstrate the superelastic effect. The loading cycle is indicated by the vertical blue line in the phase diagram, which can be described as follows. At temperature $\hat{T} > A_f$, the austenitic phase is stable whereas the martensitic phase is only metastable. When loaded from the stress-free state, the alloy first deforms elastically until the stress reaches a critical value $\sigma^{M_s}$, at which the martensitic phase transformation initiates. Further loading leads to the development of the phase transformation strain until the austenite to martensite transformation (forward) is complete at a stress value $\sigma^{M_f}$. The following deformation is the elastic deformation of the pure martensitic phase. Upon unloading, the alloy initially deforms elastically until the stress reaches another critical value $\sigma^{A_s}$, at which the austenitic phase becomes thermodynamically favorable and the martensite starts to transform back to the austenitic phase. The reverse transformation is complete at a critical stress value $\sigma^{A_f}$, and then the alloy deforms elastically in the austenite until the deformation is fully recovered. As shown in Fig. 1-1, the difference between the stress levels during forward and reverse transformations indicates a certain amount of energy dissipation, i.e. the area encompassed by the stress-strain curve. This energy dissipation is attributed to the creation and motion of the internal austenite-martensite interfaces, as well as martensite-martensite interfaces during phase transformations. It is worth emphasizing that this energy dissipation of SMAs comes with the complete deformation recovery, which makes SMAs promising as protective materials.

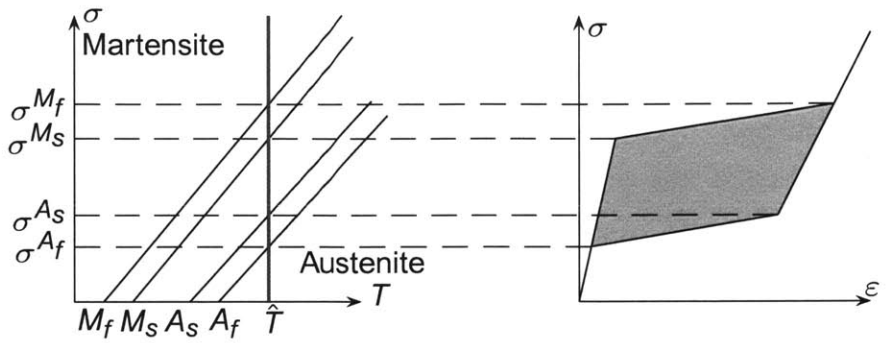


Figure 1-1: Schematic stress-temperature phase diagram and superelastic loading cycle of SMAs.

## 1.1 Experimentally observed size-dependent mechanical responses in SMAs

In response to the rising implementations of micro-devices, researchers have shown increasing interests in the behaviors of SMAs at micro- and nanoscales. The following two questions are the main concerns. First, does the martensitic phase transformation occur at such small scale? Second, if the martensitic phase transformation occurs, is there any size-dependent effect? Experiments have definitely answered the first question. San Juan *et al.* showed that both thermally and stress induced martensitic phase transformations occur in single-crystal Cu-Al-Ni micro- and nanopillars [95]. Ye *et al.* observed the stress induced martensitic phase transformation in single-crystal Ni-Ti nanopillars with diameter less than 200 nm [129]. Recently, Phillips *et al.* observed the thermally induced martensitic phase transformation in free-standing In-Tl nanowires with diameter down to 10 nm [88].

At the same time, some experimental results showed the evidence of the size dependencies of martensitic phase transformations in SMAs. Waitz *et al.* observed the suppression of thermally induced martensitic phase transformation with decreasing grain size in bulk nanocrystalline Ni-Ti SMAs [122]. Frick *et al.* showed that the strain recovered during the loading and unloading cycle diminishes with pillar diameter in the compression tests of single-crystal Ni-Ti pillars [32]. Further study revealed that this trend of losing superelasticity in Ni-Ti nanopillars does not depend on crys-

tal orientation [31]. Ishida and Sato showed that the transformation strain during thermally induced martensitic phase transformation first increases and then decreases for decreasing film thickness in polycrystalline Ni-Ti thin films with average grain size of about 5 $\mu$m [48] . Soul *et al.* studied both strain-rate and size effects in the tension tests of Ni-Ti wires, and they observed that the stress hysteresis is maximized at a specific strain-rate for each wire diameter and is larger in smaller wires at the same strain-rate [103]. For Cu-Al-Ni micro- and nanopillars subject to compressive loading, San Juan *et al.* observed a significant increase in the stress hysteresis between forward and reverse phase transformations for decreasing pillar diameter [96]. Similar effects were observed in oligocrystalline Cu-Al-Ni microwire tension tests by Chen and Schuh [18], as well as in oligocrystalline Cu-Zn-Al microwire tension tests by Ueland *et al.* [117, 116]. It was reported that the strain-hardening rate during phase transformations (or transformation modulus) increases with decreasing grain size in the tension tests of polycrystalline Cu-Zn-Al bars [102]. It was also reported that both the stress hysteresis and the strain-hardening rate increase with decreasing grain size in the tension tests of polycrystalline Cu-Al-Be bars [71], as well as Cu-Al-Mn wires and sheets [109, 108]. Recent reviews of the experimentally observed size-dependent responses of SMAs can be found in [123, 18, 38].

In [18], Chen and Schuh analyzed the experimental evidence of size-dependent martensitic phase transformations in SMAs, and they classified the apparent size-dependent effects into two groups. The first group of size effects results from the thermomechanical coupling. The austenitic and martensitic phase transformations are accompanied by the latent heat release and absorption, as well as the heat release due to the internal friction. If there is not enough time for heat transfer, the accumulated heat will increase the temperature of the material and stabilize the austenitic phase, giving rise to an increase of the critical stress for the successive forward phase transformation, i.e. a hardening effect. This situation is likely to occur in specimens with a small surface to volume ratio, and also in experiments performed with a relatively high strain rate. As a result, size effects in this group are inherently dependent on the ambient temperature, surrounding materials, and the strain rate.

The other group of size effects results from intrinsic properties of the materials. Since the martensitic phase transformation is associated with a change in lattice structure, the grain boundaries are obstacles to the growth of the martensitic phase within each grain. Decreasing the ratio of grain size to specimen size leads to a stricter compatibility requirement and thus inhibits the martensitic phase transformation. In fact, the increase in the critical martensitic transformation stress $\sigma^{M_s}$ and the decrease in the martensite start temperature $M_s$ for decreasing grain size-to-specimen size ratio were reported in [109, 122, 71, 120]. In addition, Chen and Schuh proposed another intrinsic size-dependent effect. They argued that the size-dependent stress-hysteresis observed in single-crystal Cu-Al-Ni compression tests [96] and oligocrystalline microwire tension tests of copper-based SMAs [18, 117, 116] can be attributed to the enhanced rate-independent internal friction during phase transformations.

## 1.2 Previous work on SMA modeling

Two types of continuum models have been proposed: those aiming to describe the microstructures during phase transformations, e.g. the spatial distribution of austenite and martensitic variants in the material; and those focused on simulating macroscopic responses, e.g. the stress-strain relation.

For the work aiming at microstructures, important tools include the Bain matrix or the transformation matrix, which is defined by the deformation mapping from the lattice in austenite to the martensitic variant, and the kinematic compatibility condition, i.e. the requirement for piecewise homogeneous deformation. With these tools and the crystallographic information, researchers successfully predicted various patterns of the austenite-martensite and martensite-martensite interfaces, and also calculated the maximum transformation strain during uniaxial loading [8, 44, 49, 131, 62, 63, 11]. Bhattacharya summarized the efforts along this line in his book [10]. However, there are some difficulties that limit the applications of this microstructural approach. The multi-well potential energy resulting from the crystallographic symmetry is non-convex [10, 119], which causes numerical issues like mesh dependence when solving

the corresponding boundary value problems. Although some efforts have been attempted through relaxing the multi-well energy [6, 70], it is still difficult to simulate the evolution of microstructures and obtain macroscopic stress hysteresis that is comparable with experimental results. It is worth noting that the phase field method is employed to study the microstructure evolution in SMAs, and in particular a gradient term of the phase field is added to the multiwell potential energy for the purpose of regularization [25, 26, 27, 125, 60, 58, 59].

For the work aiming at macroscopic responses, the kinetics of phase transformations is the most important concept. Tanaka *et al.* developed a thermomechanical framework for martensitic phase transformations where the martensite is described by a scalar, the martensitic volume fraction, as a function of stress and temperature. They also applied the framework to study superelastic and shape memory effects in uniaxial loading tests, where exponential hardening is assumed during phase transformations [110]. Brinson improved Tanaka's model by allowing variable elastic stiffness, as well as by introducing two internal variables that enable the separate description of thermally-induced and stress-induced martensitic volume fractions [14]. Abeyaratne and Knowles constructed an explicit tri-linear stress-strain relation, and studied hysteretic responses by describing the phase transformation as a propagating discontinuity in strains [2, 3]. Abeyaratne *et al.* developed a kinetic law for the transition between two martensitic variants under biaxial loading tests of Cu-Al-Ni SMA [1]. In the three-dimensional thermomechanical modeling of SMAs, martensitic phase transformations are usually described by a generalized $J_2$-type plasticity theory, where an extra yield surface is introduced for the reverse phase transformation [64, 13, 7]. Considering numerical simulations, Brinson and Lammering implemented the finite element calculations for Brinson's one-dimensional model and simulated uniaxial loading tests [15]. Auricchio *et al.* implemented the three-dimensional finite element calculations for the generalized plasticity theory of SMAs, and simulated superelastic behaviors in four-point and three-point bending tests [7]. Qidwai and Lagoudas evaluated the numerical implementations of thermomechanical SMA constitutive models using return mapping algorithms [91]. Reese and Christ developed

a three-dimensional finite deformation SMA constitutive model and implemented the finite element calculations [94, 19]. For a thorough review of the macroscopic thermo-mechanical modeling of SMAs, the reader is referred to the recent book of Lagoudas [54]. In order to account for the orientation-dependent responses of single-crystal and polycrystalline SMAs, researchers proposed crystal plasticity-like models where slip systems are replaced by the prescribed martensitic phase transformation systems. Thamburaja and Anand developed a polycrystalline SMA model to study the texture effect where 24 transformation systems of Ni-Ti SMA are considered [113]. Later, this model was extended by Anand and Gurtin to include 192 transformation systems as well as thermal effects [4]. Some other examples of crystalline SMA models can be found in [101, 30, 37, 51, 85, 100]. It is important to note that introducing multiple transformation systems requires nontrivial constitutive updates to determine the active systems and their volume fraction changes. The computationally intensive constitutive updates limit either the number of transformation systems that can be considered or the simulation scale like the number of elements in a finite element mesh. Patoor *et al.* reviewed the modeling work of single-crystal SMAs in [85], and Lagoudas *et al.* reviewed the corresponding work of polycrystalline SMAs in [53].

Atomistic approaches have also been employed to study martensitic phase transformations in binary SMAs like Fe-Ni [23, 73], Ni-Al [61, 82, 84, 115], Ni-Mn [45], and Ni-Ti [72, 133, 134, 39]. With tools of molecular dynamics (MD) and density function theory (DFT), researchers were able to investigate lattice structural changes as well as the effect of free surface and various defects (dislocations, grain boundaries) during martensitic phase transformations of SMAs. Through MD simulations of thermally and stress induced martensitic phase transformations in Ni-Al SMA, Li *et al.* showed that grain boundaries are not favorable for the martensite nucleation and they even hinder the martensite growth [61]. Hildebrand and Abeyaratne investigated the kinetics of detwinning in Ni-Mn SMA using MD simulations, from which they obtained an explicit formula for the continuum kinetic relation of detwinning [45]. Mutter and Nielaba studied thermally-induced austenitic phase transformation in Ni-Ti nanoparticles, where they observed through MD simulations that the austenitic phase

transformation initiates at the surface and then propagate into the interior of particles, and they also showed that the austenitic transformation temperature decreases with decreasing particle size [72]. Ni *et al.* showed that the martensitic phase transformation temperature decreases with decreasing grain size in nano-grained Fe-Ni SMA [73]. It is important to note that although atomistic simulations can provide insights of structural changes in SMAs, macroscopic responses like the stress-strain relation obtained usually cannot be quantitatively compared with experimental observations due to the limited simulation timescale [134] and the accuracy of interatomic potentials. A recent review of atomistic approaches for SMAs can be found in [52].

So far, there are few models that incorporate size-dependent effects of SMAs. Sun and He proposed a two-dimensional strain gradient viscoelastic model to study the grain-size dependence of the superelasticity in bulk nanocrystalline Ni-Ti SMAs [107]. In their work, grain boundaries are assumed to be finite-thickness layers that do not participate in martensitic phase transformations, and the strain gradient is introduced to help regularize the multi-well strain energy and avoid numerical issues. Their model captures the experimentally observed decrease in the stress hysteresis with decreasing nanograin size. Petryk *et al.* developed a model of evolving microstructures, which considers explicitly the evolution of the austenite-martensite interface and the twin boundaries [86, 87, 106]. In their work, a fraction (close to one) of the interfacial energy stored during loading is assumed to be dissipated as the interfaces diminish during unloading, and a single type of laminated microstructure is assumed to expand within each grain. Their model is able to describe the increase in the stress hysteresis for decreasing grain size. Waitz *et al.* investigated the suppression of thermally-induced martensitic phase transformation in NiTi nanograins [121]. With the nanograin modeled as an inclusion comprising twinned martensite in austenite matrix, they calculated various energy contributions including the strain energy and chemical energy, and obtained an energy barrier that increases for decreasing grain size. Their model successfully predicts an experimentally observed critical grain size, 50 nm, under which the thermally-induced martensitic phase transformation is completely suppressed. As mentioned in the previous paragraph, there

are atomistic simulations that reveal size-dependent transformation temperatures in nanoscale SMAs [72, 73]. Unfortunately, there is no experimental result that can be compared with these simulations since the atomistic simulations were necessarily performed with extremely small sizes, e.g. nanoparticles with diameter between 4 and 17 nm [72] and nanograins with grain size between 5 to 15 nm [73].

It is worth to note that there has been rich work on the study of size effects in plastic deformation [104, 29, 105, 126, 38] and also the development of size-dependent plasticity theories [28, 34, 46, 5, 41, 42, 40]. Anand *et al.* proposed a general framework to incorporate size effects in plasticity theories by combining a virtual power principle and restrictions of thermodynamics [5, 41, 42, 56, 57]. In their framework, the gradient of the plastic strain and its time rate of change are introduced to account for the contribution from non-uniform plastic deformation to the free energy and the energy dissipation, respectively. Associated with the gradients, various internal length scales are included, which enable the description of the size-dependent yield strength and flow stress. There were a couple of attempts to combine gradient plasticity theories with superelasticity, however the purposes are for the regularization of numerical methods rather than size effects [114, 112, 22].

## 1.3 Thesis objectives and approach

As can be seen, there has been significant progress on SMA modeling. However, only few models incorporate size effects, and there is in general a lack of three-dimensional formulation for modeling size effects of SMAs. There has also been rick work on the development of size-dependent plasticity theories, however there has not been any attempt to adopt these theories in modeling size effects of SMAs. The overarching goal of this thesis is to develop a computational framework for modeling size effects of SMAs in three dimensions based on a sound formulation incorporating size dependencies. Specific objectives are: 1) to develop a superelastic model for the size-dependent strain hardening and stress hysteresis during phase transformations; 2) to develop a large-scale three-dimensional computational framework for the simulations

of size-dependent mechanical responses of SMAs; 3) to study the mechanism of size effects due to internal and external constraints of phase transformations, and link microscale mechanisms to macroscopic responses.

In order to achieve these objectives, I pursued a continuum approach by extending the local thermomechanical formulation of superelasticity based on the martensitic volume fraction and flow rules [13], and combining it with the gradient plasticity theories of Anand *et al.* [5, 41, 42, 56, 57]. In this approach, both the displacements and the martensitic volume fraction are treated as primary independent fields. To account for the additional contribution from nonuniform phase transformations, the gradient of martensitic volume fraction and its time rate of change are introduced in the free energy and energy dissipation, respectively. Through the gradient terms, both an energetic length scale and a dissipative length scale are included in the model. We showed that this approach has the following benefits. First, the martensitic volume fraction as an independent field facilitates the representation of constraints for phase transformations. Secondly, a rich array of size effects can be described through the two internal length scales. Thirdly, as a continuum approach, this model is suitable for large-scale three-dimensional simulations, which enable the investigation of size-dependent responses in SMAs with unprecedented resolution.

We initially developed a one-dimensional nonlocal superelastic model, and investigated the effects of the two internal length scales. The model was then applied to simulate the size-dependent stress hysteresis in the compression tests of Cu-Al-Ni pillars. The three-dimensional nonlocal superelastic model was developed in both small strain and finite deformation. In order to solve the tightly coupled governing equations resulting from the nonlocal superelastic model, a variational formulation for the incremental problem was proposed as well as a computational framework based on this formulation. A robust and scalable solver, parallel dynamic relaxation method, was adapted within the computational framework, and was shown to be more efficient than staggered Newton methods for large scale problems. The computational model was then applied to the study of the grain boundary constraint effect in polycrystalline SMAs as well as the surface Ti oxide effect in single crystal NiTi pillars.

As a separate effort, analytical solutions were investigated in order to better understand the model responses. For the model with only a dissipative length scale (Appendix A), a minimization conjecture on the normalized plastic strain rate in [5] was linked to the variational incremental formulation presented in this thesis, and the existence of continuous minimizers was discussed. For the model with only an energetic length scale (Appendix B), an analytical stress-strain relation was derived for stress-controlled uniaxial loadings.

# Chapter 2

# One dimensional nonlocal superelastic model

Recently, San Juan *et al.* reported experimental observations of the superelastic effect in Cu-13.7Al-5Ni (wt%)[93] micro- and nanopillars subjected to compressive loading. Their observations exhibit a clear size dependence in damping capacity upon unloading [95, 96]. More specifically, their uniaxial compression tests on [001]-oriented Cu-Al-Ni single crystals show that the hysteresis loop in the stress-strain curve for a nanopillar is significantly larger than that for a bulk single crystal. In order to simulate this size-dependent effect, a one-dimensional nonlocal superelastic model is developed in this chapter following the gradient plasticity theories [41, 5, 40]. Two internal length scales, an energetic length scale $\ell_{\mathrm{e}}$ and a dissipative length scale $\ell_{\mathrm{d}}$, are introduced in the free energy and the dissipation rate respectively, leading to gradient terms on the martensitic volume fraction and its time rate of change. The formulation leads to a coupled set of partial differential equations of macroscopic equilibrium and micro-force balance, whose unknowns are the spatial distribution of the displacement and the martensitic volume fraction. The model responses are investigated with focus on the effects of the internal length scales. The model is then applied to simulate the pillar compression tests and compared with the experimental results.

## 2.1 Formulation

Consider a pillar with height $h$ subject to the stress-induced martensitic phase transformation under isothermal condition at temperature $T$. The martensitic volume fraction is represented by $\xi$. For small strains,

$$\varepsilon = u_{,x} = \varepsilon^{\mathrm{e}} + \xi\bar{\varepsilon}^{\mathrm{t}} \ , \tag{2.1}$$

where $u$ is the displacement, $\varepsilon$ is the total strain, $\varepsilon^{\mathrm{e}}$ is the elastic strain, and $\bar{\varepsilon}^{\mathrm{t}}$ is the maximum transformation strain, which is a material constant.

The free energy per unit volume comprises an elastic, a chemical and a nonlocal term

$$\psi = \frac{1}{2}E(\xi)(\varepsilon - \xi\bar{\varepsilon}^{\mathrm{t}})^2 - \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}})\xi + \frac{1}{2}S_0{\ell_{\mathrm{e}}}^2(\xi_{,x})^2 \ , \tag{2.2}$$

where $E(\xi) = \frac{E_a E_m}{E_m + \xi(E_a - E_m)}$ is the effective Young's modulus [66, 65], $E_a$ and $E_m$ are the Young's moduli in austenite and martensite respectively, $T_{\mathrm{eq}}$ is the equilibrium temperature between the two phases in the stress-free state, $\Delta s_{\mathrm{eq}}$ is the austenite to martensite transformation entropy, and $T$ is the temperature at which the experiments are performed. The nonlocal term can be viewed as the interface energy between the two phases. $S_0$ is a model parameter with the dimension of stress and $\ell_{\mathrm{e}}$ is an internal (energetic) length scale.

The introduction of a gradient term on the volume fraction in the free energy results in a separate (micro-force) equilibrium equation where the volume fraction is the primary unknown. Consider any segment of the pillar $\{x \in [x_1, x_2] |\ 0 \leq x_1 < x_2 \leq h\}$. The internal power in this segment is defined as

$$\mathrm{P}^{\mathrm{int}}(\dot{\varepsilon}^{\mathrm{e}}, \dot{\xi}) = \int_{x_1}^{x_2} \sigma\dot{\varepsilon}^{\mathrm{e}} + k\dot{\xi} + k^{\mathrm{nl}}\dot{\xi}_{,x}\, dx \ , \tag{2.3}$$

where $\dot{\bullet}$ is the rate of variable $\bullet$, $\sigma$ is the stress, $k$ and $k^{\mathrm{nl}}$ are the work-conjugates to the volume fraction $\xi$ and its gradient $\xi_{,x}$ respectively. The external power expended

on this segment is defined as

$$\mathrm{P}^{\mathrm{ext}}(\dot{u},\dot{\xi}) = (\hat{t}\dot{u} + \hat{k}\dot{\xi})|_{x_1}^{x_2} \,, \tag{2.4}$$

where $\hat{t}$ and $\hat{k}$ are, respectively, the applied boundary traction and volume fraction force conjugate. At any fixed time $\tau$, the principle of virtual power requires

$$\mathrm{P}^{\mathrm{int}}(\dot{\bar{\varepsilon}}^{\mathrm{e}},\dot{\bar{\xi}}) = \mathrm{P}^{\mathrm{ext}}(\dot{\tilde{u}},\dot{\bar{\xi}}) \tag{2.5}$$

for any generalized virtual velocity $(\dot{\tilde{u}},\dot{\bar{\varepsilon}}^{\mathrm{e}},\dot{\bar{\xi}})$ satisfying the kinematic requirement $\dot{\tilde{u}}_{,x} = \dot{\bar{\varepsilon}}^{\mathrm{e}} + \dot{\bar{\xi}}\bar{\varepsilon}^{\mathrm{t}}$. Integration by parts leads to the variational statement

$$\begin{aligned} 0 = &- \int_{x_1}^{x_2} \sigma_{,x}\dot{\tilde{u}}\, dx + \int_{x_1}^{x_2} (k - \sigma\bar{\varepsilon}^{\mathrm{t}} - k^{\mathrm{nl}}_{,x})\dot{\bar{\xi}}\, dx \\ &+ [(\sigma - \hat{t})\dot{\tilde{u}}]|_{x_1}^{x_2} + [(k^{\mathrm{nl}} - \hat{k})\dot{\bar{\xi}}]|_{x_1}^{x_2} \quad , \end{aligned}$$

which yields

$$\sigma_{,x} = 0 \,, \tag{2.6}$$

$$k - \sigma\bar{\varepsilon}^{\mathrm{t}} - k^{\mathrm{nl}}_{,x} = 0 \,, \tag{2.7}$$

for $x \in (x_1, x_2)$, and $\sigma(x_i,\tau) = \hat{t}(x_i,\tau)$, $k^{\mathrm{nl}}(x_i,\tau) = \hat{k}(x_i,\tau)$, $i = 1,2$. Equations (2.6, 2.7) are valid for any $x_1, x_2$ in the admissible range, and in particular, for the whole pillar $(x_1 = 0, x_2 = h)$.

Thermodynamic restrictions require that the temporal increase in the free energy should not be greater than the external power expended on the material, i.e.

$$\dot{\int_{x_1}^{x_2}} \psi\, dx \leq \mathrm{P}^{\mathrm{ext}}(\dot{u},\dot{\xi}) \,. \tag{2.8}$$

From Eqn (2.5) it then follows that

$$\dot{\psi} - \sigma\dot{\varepsilon}^{\mathrm{e}} - k\dot{\xi} - k^{\mathrm{nl}}\dot{\xi}_{,x} \leq 0 \tag{2.9}$$

for $x \in (0, h)$. By applying the temporal derivative to the free energy density in Eqn (2.2) and assuming the elastic response

$$\sigma = E(\xi)(\varepsilon - \xi\bar{\varepsilon}^{\mathrm{t}}) \,, \tag{2.10}$$

Eqn (2.9) is reduced to

$$\begin{aligned} 0 \leq & \left[k - \frac{1}{2}\frac{\partial E}{\partial \xi}(\varepsilon - \xi\bar{\varepsilon}^{\mathrm{t}})^2 + \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}})\right]\dot{\xi} \\ & + (k^{\mathrm{nl}} - S_0\ell_{\mathrm{e}}^2\xi_{,x})\dot{\xi}_{,x} = \mathcal{D} \,, \end{aligned} \tag{2.11}$$

where $\mathcal{D}$ is the rate of energy dissipated per unit volume. Inspired by the strain gradient plasticity theories [5, 41, 40], $k$ and $k^{\mathrm{nl}}$ are defined as follows

$$k = \frac{1}{2}\frac{\partial E}{\partial \xi}(\varepsilon - \xi\bar{\varepsilon}^{\mathrm{t}})^2 - \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) + \frac{Y\dot{\xi}}{\sqrt{(\dot{\xi})^2 + \ell_{\mathrm{d}}^2(\dot{\xi}_{,x})^2}} \,, \tag{2.12}$$

$$k^{\mathrm{nl}} = S_0\ell_{\mathrm{e}}^2\xi_{,x} + \frac{Y\ell_{\mathrm{d}}^2\dot{\xi}_{,x}}{\sqrt{(\dot{\xi})^2 + \ell_{\mathrm{d}}^2(\dot{\xi}_{,x})^2}} \,, \tag{2.13}$$

where $Y$ is a model parameter with the dimension of stress, and $\ell_{\mathrm{d}}$ is an internal (dissipative) length scale, which defines the influence of the nonuniform distribution of $\dot{\xi}$ on the dissipation. Indeed, it has been shown in [18] that in small samples of SMA, the surfaces are likely pinning points for the transformation, which tend to suppress the rate of transformation near them, relative to bulk regions away from the surfaces. This provides a possible underlying mechanism for a gradient in $\dot{\xi}$, which in turn gives rise to the dissipative length scale $\ell_{\mathrm{d}}$. The dissipation function (per unit volume) then takes the form

$$\mathcal{D} = Y\sqrt{(\dot{\xi})^2 + \ell_{\mathrm{d}}^2(\dot{\xi}_{,x})^2} \,, \tag{2.14}$$

which is nonnegative as required by Eqn (2.11). It is clear that a nonuniform distribution of $\dot{\xi}$ and larger $\ell_{\mathrm{d}}$ leads to more dissipation.

The martensitic phase transformation occurs when the thermodynamic driving force associated with the volume fraction reaches some critical value and stays at that value until the transformation is complete [85, 54, 4]. Inserting Eqs. (2.12, 2.13) in Eqn (2.7), we obtain

$$\begin{aligned}&\sigma\bar{\varepsilon}^{\mathrm{t}} - \frac{1}{2}\frac{\partial E}{\partial \xi}(\varepsilon - \xi\bar{\varepsilon}^{\mathrm{t}})^2 + \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) + S_0\ell_{\mathrm{e}}^2\xi_{,xx} \\ =&\frac{Y\dot{\xi}}{\sqrt{(\dot{\xi})^2 + \ell_{\mathrm{d}}^2(\dot{\xi}_{,x})^2}} - \frac{\partial}{\partial x}\left(\frac{Y\ell_{\mathrm{d}}^2\dot{\xi}_{,x}}{\sqrt{(\dot{\xi})^2 + \ell_{\mathrm{d}}^2(\dot{\xi}_{,x})^2}}\right),\end{aligned} \tag{2.15}$$

which governs the evolution of the volume fraction $\xi$. It should be noted that Eqn (2.15) degenerates to $\sigma\bar{\varepsilon}^{\mathrm{t}} + \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) = \mathrm{sign}(\dot{\xi})Y$, for $\dot{\xi} \neq 0$, which are the constraints during phase transformation in the local model [4], if $\xi$ and $\dot{\xi}$ are uniform or both internal length scales $\ell_{\mathrm{e}}$ and $\ell_{\mathrm{d}}$ are zeros, and $\frac{\partial E}{\partial \xi} = 0$.

## 2.2 Effects of energetic and dissipative length scales

Replacing the stress $\sigma$ from Eqn (2.10) in Eqs. (2.6, 2.15) leads to two coupled partial differential equations governing the displacement $u(x,t)$ and the volume fraction distribution $\xi(x,t)$ with suitable initial and boundary conditions. In our experimental tests, the pillars are assumed to be initially in a stress-free austenitic phase, i.e. $u(x,0) = 0$, $\xi(x,0) = 0$ for $x \in [0,h]$. On the boundary, $u(0,t) = 0$, $u(h,t) = \hat{u}(t)$, where $\hat{u}(t)$ is the prescribed displacement, while $\xi(0,t) = \xi(h,t) = 0$ which assumes that the ends of the pillar are obstacles to the martensitic phase transformation. This fully specifies the initial boundary value problem. The resulting equations are solved using a finite element discretization.

The basic model response to compressive loading and unloading cycles is explored for the following parameter values: $E_a = 10$ GPa, $E_m = 15$ GPa, $\bar{\varepsilon}^{\mathrm{t}} = -0.04$, $\Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) = -4$ MPa, $S_0 = 0.1$ GPa, $Y = 1$ MPa, $h = 1$ m. Figure 2-1 shows a stress-strain cycle fixing $\ell_{\mathrm{d}} = 0$ and varying $\frac{\ell_{\mathrm{e}}}{h}$. The solid line corresponds to $\frac{\ell_{\mathrm{e}}}{h} = 0$ (local model) and exhibits the typical superelastic response of bulk single crystal

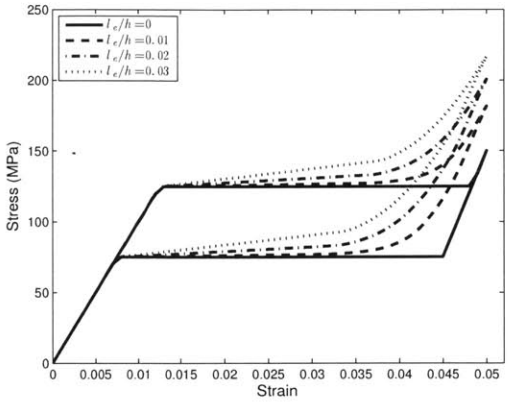


Figure 2-1: Stress $\sigma$ vs macroscopic strain $\hat{u}(t)/h$ for $\ell_e/h = 0, 0.01, 0.02, 0.03, \ell_d = 0$.

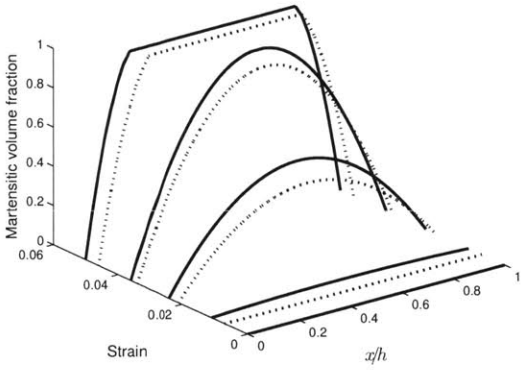


Figure 2-2: Evolution of the martensitic volume fraction for the $\ell_e/h = 0.03, \ell_d = 0$ case. Solid lines are used for loading, dashed for unloading.

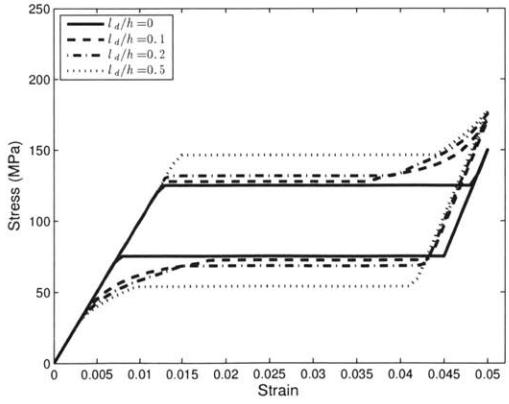


Figure 2-3: Stress $\sigma$ vs macroscopic strain $\hat{u}(t)/h$ for $\ell_d/h = 0, 0.1, 0.2, 0.5, \ell_e = 0$.

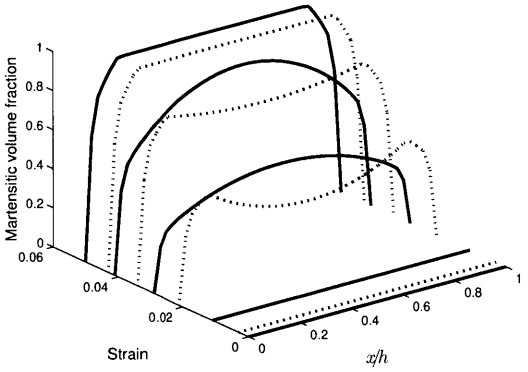


Figure 2-4: Evolution of the martensitic volume fraction for the $\ell_d/h = 0.5, \ell_e = 0$ case. Solid lines are used for loading, dashed for unloading.

SMAs. For increasing $\frac{\ell_e}{h}$, the phase transformation stage exhibits increased hardening, while the critical stress for the forward transformation and the energy dissipation are not affected. During unloading, the reverse transformation starts earlier but ends at the same point. The evolution of the martensitic volume fraction is plotted in Fig. 2-2 for the case $\frac{\ell_e}{h} = 0.03$. Because of the boundary constraints, the distribution of the martensitic volume fraction along the pillar is nonuniform during the phase transformation. This nonuniformity is responsible for the smooth transition in the stress-strain curve at the end of the forward transformation and at the beginning of the reverse transformation, in contrast with the sharp changes exhibited by the case of $\frac{\ell_e}{h} = 0$. Figure 2-3 compares the role of the dissipative length scale $\ell_d$ fixing $\ell_e = 0$. As $\frac{\ell_d}{h}$ increases, the gap between the critical stresses for the forward and the reverse transformation also increases, resulting in increased energy dissipation. Figure 2-4 demonstrates the evolution of the martensitic volume fraction for the case $\frac{\ell_d}{h} = 0.5$. It differs significantly from Figure 2-2, specifically during the unloading part, where at first the reverse transformation occurs everywhere in the pillar, which leads to a sharp change in slope in the stress-strain curve as in the case of $\frac{\ell_d}{h} = 0$. Close to the end of the reverse transformation, a pure austenitic domain first appears in the center of the pillar and then gradually expands toward the specimen edges, leading to a smooth segment in the stress-strain curve.

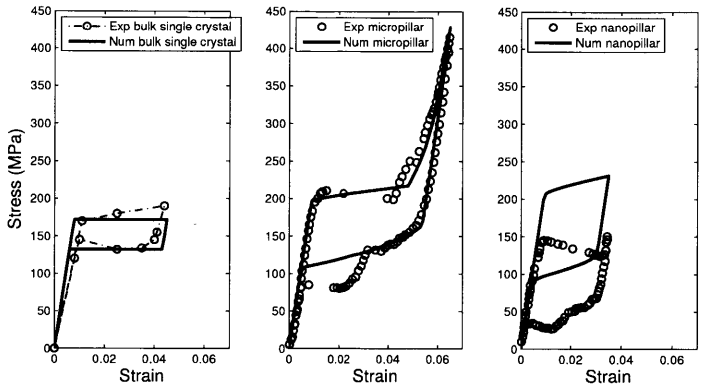


Figure 2-5: Simulation of single crystal Cu-Al-Ni compression tests in comparison with experimental results [95, 96].

## 2.3 Simulation of CuAlNi nano-pillar compression tests

Subsequently, we explore the model's ability to describe the experimentally observed response of single crystals for three different specimen sizes: (i) bulk single crystal ($h$ = 9 mm), (ii) micropillar ($h$ = 5.1 $\mu$m, diameter 1.7 $\mu$m), and (iii) nanopillar ($h$ = 3.8 $\mu$m, diameter 0.9 $\mu$m) [95, 96]. The Young's modulus of the austenitic phase, $E_a$ = 22.1 GPa, was obtained from the measurement in [96], $E_m$ = 23.5 GPa is extracted from the slope of the initial unloading part of the stress-strain curve in the case (ii), $\bar{\varepsilon}^{\mathrm{t}} = -0.05$ is obtained from the calculation in [131]. Other model parameters are calibrated to the case (ii), which furnishes the following values: $\Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) = -7.6$ MPa, $S_0$ = 0.22 GPa, $Y$ = 1 MPa, $\ell_{\mathrm{e}}$ = 0.1 $\mu$m, and $\ell_{\mathrm{d}}$ = 3.5 $\mu$m. Figure 2-5 shows the computed (solid) and experimental (circles) stress-strain curves.The model captures a number of features of the response, including the elastic loading and unloading in the two phases, the hardening during the forward and reverse transformation, and the size of the hysteresis loop (dissipation). Considering that temperature changes associated with the transformation were not experimentally available, the thermal stress was held fixed at the calibrated value. This explains the discrepancy in the stress levels predicted for the remaining cases. Regarding the

negative slope in the experiment in the case (iii), we note that there are explanations available in the literature for some superelastic materials [66], and in the present case we believe this is an artifact of the mechanical test apparatus, which operates in a condition that is neither exactly load- nor displacement-controlled.

In summary, we presented a nonlocal superelastic model for single-crystal SMAs including both an energetic and a dissipative length scales. The agreement with experimental observations suggests that the size-dependent effects in the hardening and energy dissipation of single-crystal Cu-Al-Ni SMAs can be attributed to the nonuniform evolution of the martensitic phase arising during the deformation.

# Chapter 3

# Three dimensional nonlocal superelastic model

In this chapter, a three-dimensional nonlocal superelastic model is developed in both small strain and finite deformation formulations, following the strain-gradient plasticity theories by Anand *et al.* [5, 56, 57]. The coupled governing equations resulting from the nonlocal superelastic model are then discretized in time, leading to the formulation of the incremental problem, which is then restated in a variational form and fully discretized. A general algorithm is proposed to solve the fully-discretized incremental problem. A specific solver, parallel dynamic relaxation method, is presented in detail and compared with other solvers. Applications of the three-dimensional nonlocal superelastic model will be presented in the next chapter.

## 3.1 Small strain formulation

Assume that the SMA solid occupies the volume $B \subset \mathbb{R}^3$. At each material point $\mathbf{x} \in B$, the displacement field is denoted as $\mathbf{u}$, and the strain tensor is

$$\boldsymbol{\varepsilon} = \frac{1}{2}[\nabla \mathbf{u} + (\nabla \mathbf{u})^{\tau}] \,, \tag{3.1}$$

where $\nabla$ denotes spatial derivatives and $()^\tau$ denotes the transpose. The strain tensor can be decomposed into an elastic part $\boldsymbol{\varepsilon}^{\mathrm{e}}$ and a phase transformation part $\boldsymbol{\varepsilon}^{\mathrm{t}}$, i.e.

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^{\mathrm{e}} + \boldsymbol{\varepsilon}^{\mathrm{t}} \ . \tag{3.2}$$

The evolution of the phase transformation strain tensor is assumed to follow the relation

$$\dot{\boldsymbol{\varepsilon}}^{\mathrm{t}} = \dot{\xi}\boldsymbol{\Lambda} \ , \tag{3.3}$$

where $\boldsymbol{\Lambda}$ is the phase transformation flow direction. Following Boyd *et al.* [13], we assume that $\boldsymbol{\Lambda}$ takes the following form:

$$\boldsymbol{\Lambda} = \begin{cases} \sqrt{\frac{3}{2}}\bar{\varepsilon}^{\mathrm{t}} \frac{\boldsymbol{\sigma}^{\mathrm{dev}}}{\|\boldsymbol{\sigma}^{\mathrm{dev}}\|} \ ; & \dot{\xi} > 0 \\ \sqrt{\frac{3}{2}}\bar{\varepsilon}^{t} \frac{\boldsymbol{\varepsilon}^{\mathrm{t,r}}}{\|\boldsymbol{\varepsilon}^{\mathrm{t,r}}\|} \ ; & \dot{\xi} < 0 \end{cases} \tag{3.4}$$

where $\boldsymbol{\sigma}^{\mathrm{dev}} = \boldsymbol{\sigma} - \frac{1}{3}\mathrm{trace}(\boldsymbol{\sigma})\mathbf{I}$ is the deviatoric part of the stress tensor, $\bar{\varepsilon}^{\mathrm{t}}$ is the maximum transformation strain, $\boldsymbol{\varepsilon}^{\mathrm{t,r}}$ is the phase transformation strain tensor upon unloading, and $\|\cdot\|$ denotes Frobenius norm. The flow direction above differs from isotropic $J_2$ plasticity theories in the reverse transformation part, and the specific definition ensures that the phase transformation strain accumulated during the forward phase transformation will diminish during the reverse phase transformation. In [114], the flow direction contains an additional factor to account for the tension and compression asymmetry, which we do not pursue in this work for simplicity.

## Free energy

The free energy per unit volume comprises the elastic, chemical, hardening and nonlocal terms, i.e.

$$\psi(\boldsymbol{\varepsilon}^{\mathrm{e}}, \xi, \nabla\xi) = \frac{1}{2}(\mathscr{C}\boldsymbol{\varepsilon}^{\mathrm{e}}) : \boldsymbol{\varepsilon}^{\mathrm{e}} - \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}})\xi + \frac{1}{2}H^{\mathrm{t}}\xi^2 + \frac{1}{2}S_0\ell_{\mathrm{e}}^2\|\nabla\xi\|^2 \ . \tag{3.5}$$

In this free energy expression, $\xi$ is the martensitic volume fraction. The elastic moduli $\mathscr{C}$ are defined as $\mathscr{C} = (1-\xi)\mathscr{C}^A + \xi\mathscr{C}^M$ with $\mathscr{C}^A$ and $\mathscr{C}^M$ the elastic moduli of the pure austenite and martensite, respectively. $T_{\text{eq}}$ is the equilibrium temperature between the two phases in the stress-free state, $\Delta s_{\text{eq}}$ is the austenite to martensite transformation entropy, and $T$ is the temperature at which experiments are performed. The hardening coefficient $H^{\text{t}}$ with dimensions of stress accounts for the classic hardening effect during phase transformations. $S_0$ is a model parameter with dimensions of stress, and $\ell_{\text{e}}$ is an internal (energetic) length scale.

## Governing equations

Introducing the gradient of the martensitic volume fraction in the free energy (3.5) leads to an extra governing equation besides the classic force balance equation. Principle of virtual power is thus invoked to derive the governing equations. We define the following work conjugates: $\boldsymbol{\sigma}$ with respect to $\dot{\boldsymbol{\varepsilon}}^{\text{e}}$, $k$ with respect to $\dot{\xi}$, and $\mathbf{k}^{\text{nl}}$ with respect to $\nabla\dot{\xi}$ . The internal power expended in any sub-domain $V \subset B$ can be expressed as follows

$$\mathcal{P}^{\text{int}}(\dot{\boldsymbol{\varepsilon}}^{\text{e}}, \dot{\xi}) = \int_V \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}}^{\text{e}} + k\dot{\xi} + \mathbf{k}^{\text{nl}} \cdot \nabla\dot{\xi} \, dV \, . \tag{3.6}$$

The external power expended on this sub-domain can be expressed as

$$\mathcal{P}^{\text{ext}}(\dot{\mathbf{u}}, \dot{\xi}) = \int_{\partial V} \hat{\mathbf{t}} \cdot \dot{\mathbf{u}} + \hat{k}\dot{\xi} \, dS \, , \tag{3.7}$$

where $\hat{\mathbf{t}}$ is the applied traction, and $\hat{k}$ is the applied work-conjugate to $\dot{\xi}$.

The principle of virtual power states that the internal power equals the external power, i.e.

$$\mathcal{P}^{\text{int}}(\tilde{\boldsymbol{\varepsilon}}^{\text{e}}, \tilde{\xi}) = \mathcal{P}^{\text{ext}}(\tilde{\mathbf{u}}, \tilde{\xi}) \, . \tag{3.8}$$

for any generalized velocity $(\tilde{\boldsymbol{\varepsilon}}^{\mathrm{e}}, \tilde{\xi}, \tilde{\mathbf{u}})$ subject to the following kinematic requirements

$$\tilde{\boldsymbol{\varepsilon}} = \tilde{\boldsymbol{\varepsilon}}^{\mathrm{e}} + \tilde{\xi}\boldsymbol{\Lambda} \,, \tag{3.9}$$

$$\tilde{\boldsymbol{\varepsilon}} = \frac{1}{2}[\nabla\tilde{\mathbf{u}} + (\nabla\tilde{\mathbf{u}})^{\tau}] \,. \tag{3.10}$$

With these kinematic requirements, the internal power in Eqn (3.8) can be reformulated through integration by parts

$$\begin{aligned}
\mathcal{P}^{\mathrm{int}} &= \int_V \boldsymbol{\sigma} : \tilde{\boldsymbol{\varepsilon}}^{\mathrm{e}} + k\tilde{\xi} + \mathbf{k}^{\mathrm{nl}} \cdot \nabla\tilde{\xi} \, dV \\
&= \int_V \boldsymbol{\sigma} : (\tilde{\boldsymbol{\varepsilon}} - \tilde{\xi}\boldsymbol{\Lambda}) + k\tilde{\xi} - \operatorname{div}\mathbf{k}^{\mathrm{nl}}\tilde{\xi} \, dV + \int_{\partial V} \mathbf{k}^{\mathrm{nl}} \cdot \mathbf{n}\tilde{\xi} \, dS \\
&= \int_V -\operatorname{div}(\boldsymbol{\sigma}) \cdot \tilde{\mathbf{u}} + \left(-\boldsymbol{\sigma} : \boldsymbol{\Lambda} + k - \operatorname{div}\mathbf{k}^{\mathrm{nl}}\right)\tilde{\xi} \, dV + \int_{\partial V} (\boldsymbol{\sigma} \cdot \mathbf{n}) \cdot \tilde{\mathbf{u}} + (\mathbf{k}^{\mathrm{nl}} \cdot \mathbf{n})\tilde{\xi} \, dS,
\end{aligned} \tag{3.11}$$

where $\mathbf{n}$ is the unit outer normal to the surface $\partial V$, and $\boldsymbol{\sigma}$ is assumed to be symmetric. Due to the arbitrariness of the generalized velocity, the principle of virtual power leads to the classic force balance equation

$$\operatorname{div}(\boldsymbol{\sigma}) = \mathbf{0} \,, \tag{3.12}$$

and the micro-force balance equation

$$\boldsymbol{\sigma} : \boldsymbol{\Lambda} - k + \operatorname{div}(\mathbf{k}^{\mathrm{nl}}) = 0 \tag{3.13}$$

in the body, and also two boundary conditions:

$$\boldsymbol{\sigma} \cdot \mathbf{n} = \hat{\mathbf{t}} \,; \tag{3.14}$$

$$\mathbf{k}^{\mathrm{nl}} \cdot \mathbf{n} = \hat{k} \,. \tag{3.15}$$

## Constitutive relations

Thermodynamic restrictions require that the increment of the free energy should not be greater than the external power expended, i.e.

$$\overline{\int_V \psi \, dV}^{\cdot} \leq \mathcal{P}^{\text{ext}} \, . \tag{3.16}$$

With the expression of the free energy rate

$$\dot{\psi} = (\mathscr{C}\boldsymbol{\varepsilon}^{\text{e}}) : \dot{\boldsymbol{\varepsilon}}^{\text{e}} + \left[\frac{1}{2}(\mathscr{C}_{,\xi}\boldsymbol{\varepsilon}^{\text{e}}) : \boldsymbol{\varepsilon}^{\text{e}} - \Delta s_{\text{eq}}(T - T_{\text{eq}}) + H^{\text{t}}\xi\right]\dot{\xi} + S_0\ell_{\text{e}}^2\nabla\xi\cdot\nabla\dot{\xi} \, , \tag{3.17}$$

Eqn (3.16) and the principle of virtual power (3.8) lead to the following inequality

$$\begin{aligned}&\int_V (\mathscr{C}\boldsymbol{\varepsilon}^{\text{e}}) : \dot{\boldsymbol{\varepsilon}}^{\text{e}} + \left[\frac{1}{2}(\mathscr{C}_{,\xi}\boldsymbol{\varepsilon}^{\text{e}}) : \boldsymbol{\varepsilon}^{\text{e}} - \Delta s_{\text{eq}}(T - T_{\text{eq}}) + H^{\text{t}}\xi\right]\dot{\xi} + S_0\ell_{\text{e}}^2\nabla\xi\cdot\nabla\dot{\xi} \, dV \\ \leq &\int_V \boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}}^{\text{e}} + k\dot{\xi} + \mathbf{k}^{\text{nl}}\cdot\nabla\dot{\xi} \, dV \, ,\end{aligned}$$

which then gives rise to

$$\begin{aligned}0 \leq\ & (\boldsymbol{\sigma} - \mathscr{C}\boldsymbol{\varepsilon}^{\text{e}}) : \dot{\boldsymbol{\varepsilon}}^{\text{e}} \\ &+ \left[k - \frac{1}{2}(\mathscr{C}_{,\xi}\boldsymbol{\varepsilon}^{\text{e}}) : \boldsymbol{\varepsilon}^{\text{e}} + \Delta s_{\text{eq}}(T - T_{\text{eq}}) - H^{\text{t}}\xi\right]\dot{\xi} \\ &+ (\mathbf{k}^{\text{nl}} - S_0\ell_{\text{e}}^2\nabla\xi)\cdot\nabla\dot{\xi} \, .\end{aligned} \tag{3.18}$$

Eqn (3.18) imposes thermodynamic restrictions on the constitutive relations. In order to satisfy Eqn (3.18), constitutive relations for the work conjugates $\boldsymbol{\sigma}$, $k$ and $\mathbf{k}^{\text{nl}}$ are defined as follows:

Stress $\boldsymbol{\sigma}$ obeys Hooke's law

$$\boldsymbol{\sigma} = \mathscr{C}\boldsymbol{\varepsilon}^{\text{e}} \, . \tag{3.19}$$

$k$ as the work conjugate to $\dot{\xi}$ is defined as

$$k = \frac{1}{2}(\mathscr{C}_{,\xi}\boldsymbol{\varepsilon}^{\text{e}}) : \boldsymbol{\varepsilon}^{\text{e}} - \Delta s_{\text{eq}}(T - T_{\text{eq}}) + H^{\text{t}}\xi + \frac{Y\dot{\xi}}{\sqrt{(\dot{\xi})^2 + \ell_{\text{d}}^2\|\nabla\dot{\xi}\|^2}} \, , \tag{3.20}$$

where $\ell_{\mathrm{d}}$ is an internal (dissipative) length scale, and $Y$ with dimensions of stress is the dissipative resistance to phase transformations.

$\mathbf{k}^{\mathrm{nl}}$ as the work conjugate to $\nabla\dot{\xi}$ is defined as

$$\mathbf{k}^{\mathrm{nl}} = S_0\ell_{\mathrm{e}}^2\nabla\xi + \frac{Y\ell_{\mathrm{d}}^2\nabla\dot{\xi}}{\sqrt{(\dot{\xi})^2 + \ell_{\mathrm{d}}^2\|\nabla\dot{\xi}\|^2}} \,. \tag{3.21}$$

With the newly defined constitutive relations, the energy dissipation rate per unit volume can be expressed as

$$\begin{aligned}\mathcal{D} &= \left[k - \frac{1}{2}(\mathscr{C}_{,\xi}\boldsymbol{\varepsilon}^{\mathrm{e}}) : \boldsymbol{\varepsilon}^{\mathrm{e}} + \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) - H^{\mathrm{t}}\xi\right]\dot{\xi} + (\mathbf{k}^{\mathrm{nl}} - S_0\ell_{\mathrm{e}}^2\nabla\xi)\cdot\nabla\dot{\xi} \\ &= Y\sqrt{(\dot{\xi})^2 + \ell_{\mathrm{d}}^2\|\nabla\dot{\xi}\|^2}\,,\end{aligned} \tag{3.22}$$

which is always non-negative.

## Microforce balance equation revisited

Using constitutive relations (3.20) and (3.21), the micro-force balance equation (3.13) can be rewritten as

$$\begin{aligned}&\boldsymbol{\sigma} : \boldsymbol{\Lambda} - \frac{1}{2}(\mathscr{C}_{,\xi}\boldsymbol{\varepsilon}^{\mathrm{e}}) : \boldsymbol{\varepsilon}^{\mathrm{e}} + \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) - H^{\mathrm{t}}\xi + S_0\ell_{\mathrm{e}}^2\mathrm{div}(\nabla\xi) \\ =&\frac{Y\dot{\xi}}{\sqrt{(\dot{\xi})^2 + \ell_{\mathrm{d}}^2\|\nabla\dot{\xi}\|^2}} - \mathrm{div}\left(\frac{Y\ell_{\mathrm{d}}^2\nabla\dot{\xi}}{\sqrt{(\dot{\xi})^2 + \ell_{\mathrm{d}}^2\|\nabla\dot{\xi}\|^2}}\right),\end{aligned} \tag{3.23}$$

which is in general nonlinear due to the presence of the dissipative length scale $\ell_{\mathrm{d}}$. If both internal length scales are zeros, Eqn (3.23) becomes

$$\boldsymbol{\sigma} : \boldsymbol{\Lambda} - \frac{1}{2}(\mathscr{C}_{,\xi}\boldsymbol{\varepsilon}^{\mathrm{e}}) : \boldsymbol{\varepsilon}^{\mathrm{e}} + \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) - H^{\mathrm{t}}\xi = \mathrm{sign}(\dot{\xi})Y \,, \tag{3.24}$$

which represents phase transformation conditions in classic superelastic models [54]. Therefore, the nonlocal superelastic model presented can be viewed as a nonlocal extension of classic superelastic models (local theory) in such a way that the martensitic volume fraction $\xi$, an internal variable in the local theory, becomes an independent

variable whose evolution is governed by the partial differential equation (3.23).

## 3.2 Finite deformation formulation

The finite deformation version of the nonlocal superelastic model uses the multiplicative decomposition of deformation gradient and a logarithmic strain measure. The derivation follows the same approach as the one in the small strain formulation.

Assume that the deformation gradient $\mathbf{F} = \nabla\mathbf{u} + \mathbf{I}$ can be decomposed multiplicatively into

$$\mathbf{F} = \mathbf{F}^{\mathrm{e}}\mathbf{F}^{\mathrm{t}} \,, \tag{3.25}$$

where $\mathbf{F}^{\mathrm{e}}$ and $\mathbf{F}^{\mathrm{t}}$ are the elastic and the phase transformation deformation gradients, respectively. The evolution of $\mathbf{F}^{\mathrm{t}}$ is assumed to follow the relation

$$\dot{\mathbf{F}}^{\mathrm{t}} = (\dot{\xi}\boldsymbol{\Lambda})\mathbf{F}^{\mathrm{t}} \,, \tag{3.26}$$

where $\boldsymbol{\Lambda}$ denotes the flow direction tensor, which will be defined later. The right Cauchy-Green deformation tensor $\mathbf{C}$, the phase transformation deformation tensor $\mathbf{C}^{\mathrm{t}}$, and the elastic deformation tensor $\mathbf{C}^{\mathrm{e}}$ are expressed as

$$\mathbf{C} = \mathbf{F}^{\mathrm{T}}\mathbf{F} \,, \quad \mathbf{C}^{\mathrm{t}} = \mathbf{F}^{\mathrm{t}\mathrm{T}}\mathbf{F}^{\mathrm{t}} \,, \text{ and } \mathbf{C}^{\mathrm{e}} = \mathbf{F}^{\mathrm{e}\mathrm{T}}\mathbf{F}^{\mathrm{e}} = \mathbf{F}^{\mathrm{t}-\mathrm{T}}\mathbf{C}\mathbf{F}^{\mathrm{t}-1} \,. \tag{3.27}$$

Applying polar decomposition to the deformation gradients leads to

$$\mathbf{F}^{\mathrm{e}} = \mathbf{R}^{\mathrm{e}}\mathbf{U}^{\mathrm{e}} \,, \text{ and } \quad \mathbf{F}^{\mathrm{t}} = \mathbf{R}^{\mathrm{t}}\mathbf{U}^{\mathrm{t}} \,, \tag{3.28}$$

where $\mathbf{R}^{\mathrm{e}}$ and $\mathbf{R}^{\mathrm{t}}$ are rotations, while $\mathbf{U}^{\mathrm{e}}$ and $\mathbf{U}^{\mathrm{t}}$ are positive-definite stretch tensors. Assume that $\{\lambda_i^{\mathrm{e}}, \mathbf{r}_i^{\mathrm{e}}\}$ and $\{\lambda_i^{\mathrm{t}}, \mathbf{r}_i^{\mathrm{t}}\}$ are the eigenvalue and right eigenvector pairs of $\mathbf{U}^{\mathrm{e}}$ and $\mathbf{U}^{\mathrm{t}}$, respectively. Then the logarithmic elastic strain tensor is defined as

$$\mathbf{E}^{\mathrm{e}} = \frac{1}{2}\log\mathbf{C}^{\mathrm{e}} = \sum_{i=1}^{3}\log(\lambda_i^{\mathrm{e}})\mathbf{r}_i^{\mathrm{e}} \otimes \mathbf{r}_i^{\mathrm{e}} \,, \tag{3.29}$$

and the logarithmic phase transformation strain tensor is defined as

$$\mathbf{E}^{\mathrm{t}} = \frac{1}{2}\log \mathbf{C}^{\mathrm{t}} = \sum_{i=1}^{3} \log(\lambda_i^{\mathrm{t}})\mathbf{r}_i^{\mathrm{t}} \otimes \mathbf{r}_i^{\mathrm{t}} \,. \tag{3.30}$$

## Free energy

The free energy per unit volume consists of elastic, chemical, hardening and nonlocal terms

$$\psi = \frac{1}{2}(\mathscr{C}\mathbf{E}^{\mathrm{e}}) : \mathbf{E}^{\mathrm{e}} - \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}})\xi + \frac{1}{2}H^{\mathrm{t}}(\xi)^2 + \frac{1}{2}S_0\ell_{\mathrm{e}}^2\|\nabla\xi\|^2 \,. \tag{3.31}$$

The elastic moduli $\mathscr{C}$ are the arithmetic average of the two phases, i.e. $\mathscr{C} = (1-\xi)\mathscr{C}^A + \xi\mathscr{C}^M$ with $\xi$, the volume fraction of martensite. $T_{\mathrm{eq}}$ is the equilibrium temperature between the two phases in the stress-free state, $\Delta s_{\mathrm{eq}}$ is the transformation entropy from austenite to martensite at $T_{\mathrm{eq}}$, and $T$ is the temperature at which experiments are performed. The hardening parameter $H^{\mathrm{t}}$ with dimensions of stress is used to describe the classic strain-hardening during phase transformation. $S_0$ is a model parameter with dimensions of stress, and $\ell_{\mathrm{e}}$ is an internal (energetic) length scale.

## Governing equations

Introducing the gradient of $\xi$ in the free energy leads to an additional governing partial differential equation (micro-force balance) associated with the field $\xi$. Principle of virtual power is thus invoked to determine the governing equations. Consider any sub-domain $V$ of the solid. The internal power expended can be expressed as

$$\mathcal{P}^{\mathrm{int}}(\dot{\mathbf{E}}^{\mathrm{e}}, \dot{\xi}) = \int_V \mathbf{S}^{\mathrm{e}} : \dot{\mathbf{E}}^{\mathrm{e}} + k\dot{\xi} + \mathbf{k}^{\mathrm{nl}} \cdot \nabla\dot{\xi}\, dV \tag{3.32}$$

where $\mathbf{S}^{\mathrm{e}}$, $k$, $\mathbf{k}^{\mathrm{nl}}$ are the work-conjugates to $\dot{\mathbf{E}}^{\mathrm{e}}$, $\dot{\xi}$, and $\nabla\dot{\xi}$, respectively.

The external power expended on this sub-domain can be expressed as

$$\mathcal{P}^{\text{ext}}(\dot{\mathbf{u}}, \dot{\xi}) = \int_{\partial V} \hat{\mathbf{t}} \cdot \dot{\mathbf{u}} + \hat{k}\dot{\xi}\, dS \tag{3.33}$$

where $\hat{\mathbf{t}}$ and $\hat{k}$ are the traction and micro-traction, respectively.

The principle of virtual power states that the internal power equals the external power

$$\mathcal{P}^{\text{int}}(\tilde{\mathbf{E}}^{\text{e}}, \tilde{\xi}) = \mathcal{P}^{\text{ext}}(\tilde{u}, \tilde{\xi}) \tag{3.34}$$

for any generalized velocity $(\tilde{u}, \tilde{\xi}, \tilde{\mathbf{E}}^{\text{e}})$ subject to the kinematic requirements:

$$\tilde{\mathbf{F}} = \nabla\tilde{u} = \tilde{\mathbf{F}}^{\text{e}}\mathbf{F}^{\text{t}} + \mathbf{F}^{\text{e}}\tilde{\mathbf{F}}^{\text{t}}\ , \quad \tilde{\mathbf{F}}^{\text{t}} = (\tilde{\xi}\boldsymbol{\Lambda})\mathbf{F}^{\text{t}}\ , \tag{3.35}$$

$$\tilde{\mathbf{E}}^{\text{e}} = \frac{\partial \mathbf{E}^{\text{e}}}{\partial \mathbf{C}^{\text{e}}} : \tilde{\mathbf{C}}^{\text{e}} = \frac{\partial \mathbf{E}^{\text{e}}}{\partial \mathbf{C}^{\text{e}}} : (\tilde{\mathbf{F}}^{\text{e}\tau}\mathbf{F}^{\text{e}} + \mathbf{F}^{\text{e}\tau}\tilde{\mathbf{F}}^{\text{e}})\ . \tag{3.36}$$

In order to simplify the derivation, we further define

$$\tilde{\mathbf{L}}^{\text{t}} = \tilde{\mathbf{F}}^{\text{t}}\mathbf{F}^{\text{t}^{-1}}\ , \quad \tilde{\mathbf{C}} = \tilde{\mathbf{F}}^{\tau}\mathbf{F} + \mathbf{F}^{\tau}\tilde{\mathbf{F}}\ . \tag{3.37}$$

The first term in the internal power (3.34) can be rewritten as follows:

$$\begin{aligned}
\mathbf{S}^{\text{e}} : \tilde{\mathbf{E}}^{\text{e}} &= \mathbf{S}^{\text{e}} : (\frac{\partial \mathbf{E}^{\text{e}}}{\partial \mathbf{C}^{\text{e}}} : \tilde{\mathbf{C}}^{\text{e}}) \\
&= \mathbf{S}^{\text{e}} : \left[\frac{\partial \mathbf{E}^{\text{e}}}{\partial \mathbf{C}^{\text{e}}} : \left(-\tilde{\mathbf{L}}^{\text{t}^{\tau}}\mathbf{C}^{\text{e}} + \mathbf{F}^{\text{t}^{-\tau}}\tilde{\mathbf{C}}\mathbf{F}^{\text{t}^{-1}} - \mathbf{C}^{\text{e}}\tilde{\mathbf{L}}^{\text{t}}\right)\right] \\
&= \mathbf{F}^{\text{t}^{-1}}(2\mathbf{S}^{\text{e}} : \frac{\partial \mathbf{E}^{\text{e}}}{\partial \mathbf{C}^{\text{e}}})\mathbf{F}^{\text{t}^{-\tau}} : \frac{1}{2}\tilde{\mathbf{C}} - 2\left[\mathbf{C}^{\text{e}}(\mathbf{S}^{\text{e}} : \frac{\partial \mathbf{E}^{\text{e}}}{\partial \mathbf{C}^{\text{e}}})\right] : \tilde{\mathbf{L}}^{\text{t}} \\
&= \mathbf{F}^{\text{e}}(2\mathbf{S}^{\text{e}} : \frac{\partial \mathbf{E}^{\text{e}}}{\partial \mathbf{C}^{\text{e}}})\mathbf{F}^{\text{t}^{-\tau}} : \tilde{\mathbf{F}} - \left[\mathbf{C}^{\text{e}}(2\mathbf{S}^{\text{e}} : \frac{\partial \mathbf{E}^{\text{e}}}{\partial \mathbf{C}^{\text{e}}})\right] : \boldsymbol{\Lambda}\tilde{\xi}\ .
\end{aligned} \tag{3.38}$$

The internal power in (3.34) can be then reformulated using integration by parts

as follows:

$$\begin{aligned}\mathcal{P}^{\text{int}}(\tilde{\mathbf{E}}^{\text{e}},\tilde{\xi}) &= \int_V \mathbf{F}^{\text{e}}(2\mathbf{S}^{\text{e}} : \frac{\partial \mathbf{E}^{\text{e}}}{\partial \mathbf{C}^{\text{e}}})\mathbf{F}^{\text{t}^{-\tau}} : \tilde{\mathbf{F}} - \left[\mathbf{C}^{\text{e}}(2\mathbf{S}^{\text{e}} : \frac{\partial \mathbf{E}^{\text{e}}}{\partial \mathbf{C}^{\text{e}}})\right] : \mathbf{\Lambda}\tilde{\xi}\, dV \\ &\quad + \int_V k\tilde{\xi} + \mathbf{k}^{\text{nl}} \cdot \nabla\tilde{\xi}\, dV \\ &= \int_V -\text{div}\left[\mathbf{F}^{\text{e}}(2\mathbf{S}^{\text{e}} : \frac{\partial \mathbf{E}^{\text{e}}}{\partial \mathbf{C}^{\text{e}}})\mathbf{F}^{\text{t}^{-\tau}}\right] \cdot \tilde{\mathbf{u}}\, dV \\ &\quad + \int_V \left\{ - \left[\mathbf{C}^{\text{e}}(2\mathbf{S}^{\text{e}} : \frac{\partial \mathbf{E}^{\text{e}}}{\partial \mathbf{C}^{\text{e}}})\right] : \mathbf{\Lambda} + k - \text{div}(\mathbf{k}^{\text{nl}}) \right\} \tilde{\xi}\, dV \\ &\quad + \int_{\partial V} \left\{ \left[\mathbf{F}^{\text{e}}(2\mathbf{S}^{\text{e}} : \frac{\partial \mathbf{E}^{\text{e}}}{\partial \mathbf{C}^{\text{e}}})\mathbf{F}^{\text{t}^{-\tau}}\right] \cdot \mathbf{n} \right\} \cdot \tilde{\mathbf{u}} + \mathbf{k}^{\text{nl}} \cdot \mathbf{n}\tilde{\xi}\, dS\,, \end{aligned} \tag{3.39}$$

where $\mathbf{n}$ is the unit outer normal to the surface $\partial V$.

Two stress tensors can be defined to simplify the notation in Eqn (3.39): first the work conjugate to $\dot{\mathbf{F}}$,

$$\mathbf{P} = \mathbf{F}^{\text{e}}(2\mathbf{S}^{\text{e}} : \frac{\partial \mathbf{E}^{\text{e}}}{\partial \mathbf{C}^{\text{e}}})\mathbf{F}^{\text{t}^{-\tau}}\,, \tag{3.40}$$

and secondly

$$\mathbf{M}^{\text{e}} = \mathbf{C}^{\text{e}}(2\mathbf{S}^{\text{e}} : \frac{\partial \mathbf{E}^{\text{e}}}{\partial \mathbf{C}^{\text{e}}})\,, \tag{3.41}$$

which provides the driving force for phase transformations.

Equation (3.39) and the principle of virtual power (3.34) result in two partial differential equations in the body, i.e. the macroforce balance equation

$$\text{div}(\mathbf{P}) = \mathbf{0}\,, \tag{3.42}$$

and the microforce balance equation

$$\mathbf{M}^{\text{e}} : \mathbf{\Lambda} - k + \text{div}(\mathbf{k}^{\text{nl}}) = 0\,, \tag{3.43}$$

as well as two boundary conditions on the surface:

$$\mathbf{P} \cdot \mathbf{n} = \hat{\mathbf{t}} \quad \text{and} \quad \mathbf{k}^{\text{nl}} \cdot \mathbf{n} = \hat{k}\,. \tag{3.44}$$

## Constitutive relations

Thermodynamic restrictions require that the temporal increase in the free energy must be no greater than the external power expended, i.e.

$$\dot{\overline{\int_V \psi \, dV}} \leq \mathcal{P}^{ext} \, . \tag{3.45}$$

With the rate of free energy per unit volume

$$\dot{\psi} = \frac{\partial \psi}{\partial \mathbf{E}^{\mathrm{e}}} : \dot{\mathbf{E}}^{\mathrm{e}} + \frac{\partial \psi}{\partial \xi}\dot{\xi} + \frac{\partial \psi}{\partial \nabla \xi} \cdot \nabla \dot{\xi} \, , \tag{3.46}$$

and the principle of virtual power (3.34), Eqn (3.45) then leads to

$$0 \leq (\mathbf{S}^{\mathrm{e}} - \frac{\partial \psi}{\partial \mathbf{E}^{\mathrm{e}}}) : \dot{\mathbf{E}}^{\mathrm{e}} + (k - \frac{\partial \psi}{\partial \xi})\dot{\xi} + (\mathbf{k}^{nl} - \frac{\partial \psi}{\partial \nabla \xi}) \cdot \nabla \dot{\xi} \, . \tag{3.47}$$

In order to satisfy the restriction of Eqn (3.47), constitutive relations for the work-conjugates are defined as follows:

$\mathbf{S}^{\mathrm{e}}$ as the work conjugate to $\dot{\mathbf{E}}^{\mathrm{e}}$ obeys Hooke's law

$$\mathbf{S}^{\mathrm{e}} = \frac{\partial \psi}{\partial \mathbf{E}^{\mathrm{e}}} = \mathscr{C}\mathbf{E}^{\mathrm{e}} \, . \tag{3.48}$$

$k$ as the work conjugate to $\dot{\xi}$ is defined as

$$\begin{aligned} k &= \frac{\partial \psi}{\partial \xi} + \frac{Y\dot{\xi}}{\sqrt{(\dot{\xi})^2 + (\ell_{\mathrm{d}})^2 \|\nabla \dot{\xi}\|^2}} \\ &= \frac{1}{2}(\mathscr{C}_{,\xi}\mathbf{E}^{\mathrm{e}}) : \mathbf{E}^{\mathrm{e}} - \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) + H^{\mathrm{t}}\xi + \frac{Y\dot{\xi}}{\sqrt{(\dot{\xi})^2 + (\ell_{\mathrm{d}})^2 \|\nabla \dot{\xi}\|^2}} \, , \end{aligned} \tag{3.49}$$

where $\ell_{\mathrm{d}}$ is an internal (dissipative) length scale, and $Y$ with dimensions of stress is the dissipative resistance to phase transformations.

$\mathbf{k}^{\mathrm{nl}}$ as the work conjugate to $\nabla\dot{\xi}$ is defined as

$$\begin{aligned}\mathbf{k}^{\mathrm{nl}} &= \frac{\partial\psi}{\partial\nabla\xi} + \frac{Y\ell_{\mathrm{d}}^2\nabla\dot{\xi}}{\sqrt{(\dot{\xi})^2 + (\ell_{\mathrm{d}})^2\|\nabla\dot{\xi}\|^2}} \\ &= S_0\ell_{\mathrm{e}}^2\nabla\xi + \frac{Y\ell_{\mathrm{d}}^2\nabla\dot{\xi}}{\sqrt{(\dot{\xi})^2 + (\ell_{\mathrm{d}})^2\|\nabla\dot{\xi}\|^2}} \,. \end{aligned} \tag{3.50}$$

Using Eqn (3.48), (3.49), and (3.50), the energy dissipation rate per unit volume can be expressed as

$$\begin{aligned}\mathcal{D} &= (\mathbf{S}^{\mathrm{e}} - \frac{\partial\psi}{\partial\mathbf{E}^{\mathrm{e}}}) : \dot{\mathbf{E}}^{\mathrm{e}} + (k - \frac{\partial\psi}{\partial\xi})\dot{\xi} + (\mathbf{k}^{\mathrm{nl}} - \frac{\partial\psi}{\partial\nabla\xi}) \cdot \nabla\dot{\xi} \\ &= Y\sqrt{(\dot{\xi})^2 + \ell_{\mathrm{d}}^2\|\nabla\dot{\xi}\|^2} \,, \end{aligned} \tag{3.51}$$

which is always non-negative.

The expressions of stresses $\mathbf{P}$ and $\mathbf{M}^{\mathrm{e}}$, Eqn (3.40) and (3.41), involve derivatives of the logarithmic elastic strain $\mathbf{E}^{\mathrm{e}}$, i.e.

$$\frac{\partial\mathbf{E}^{\mathrm{e}}}{\partial\mathbf{C}^{\mathrm{e}}} = \frac{1}{2}\frac{\partial\log\mathbf{C}^{\mathrm{e}}}{\partial\mathbf{C}^{\mathrm{e}}} \,, \tag{3.52}$$

which can be computed straightforwardly using the explicit formula in [77].

When elastic responses are isotropic, $\mathbf{P}$ and $\mathbf{M}^{\mathrm{e}}$ can be obtained without the calculation of derivatives of the logarithmic elastic strain. In this case, it can be seen from Hooke's law (3.48) that $\mathbf{S}^{\mathrm{e}}$ is symmetric and it has the same right eigenvectors $\{\mathbf{r}_i^{\mathrm{e}}\}$ as $\mathbf{E}^{\mathrm{e}}$ and $\mathbf{C}^{\mathrm{e}}$, i.e. $\mathbf{S}^{\mathrm{e}}$ can be written as

$$\mathbf{S}^{\mathrm{e}} = \sum_{i=1}^{3} s_i^{\mathrm{e}}\mathbf{r}_i^{\mathrm{e}} \otimes \mathbf{r}_i^{\mathrm{e}} \tag{3.53}$$

with eigenvalues $s_i^{\mathrm{e}}$, for $i = 1, 2, 3$. Applying the formula for derivatives of the logarithmic mapping in [77] to the symmetric tensor $\mathbf{C}^{\mathrm{e}}$ (3.27) results in

$$\frac{\partial\log\mathbf{C}^{\mathrm{e}}}{\partial\mathbf{C}^{\mathrm{e}}} = \sum_{i,j=1}^{3} g_{ij}(\mathbf{r}_i^{\mathrm{e}} \otimes \mathbf{r}_j^{\mathrm{e}}) \otimes (\mathbf{r}_i^{\mathrm{e}} \otimes \mathbf{r}_j^{\mathrm{e}}) \,, \tag{3.54}$$

where coefficients $g_{ij}$ are defined as

$$g_{ij} = \begin{cases} \frac{\log(c_i^e)-\log(c_j^e)}{c_i^e-c_j^e} , & \text{if } c_i^e \neq c_j^e , \\ \frac{1}{c_i^e} , & \text{otherwise ;} \end{cases} \tag{3.55}$$

with $c_i^e = (\lambda_i^e)^2$, for $i = 1, 2, 3$, the eigenvalues of $\mathbf{C}^e$. Therefore, it can be deduced that

$$\begin{aligned} \mathbf{S}^e : \frac{\partial \log \mathbf{C}^e}{\partial \mathbf{C}^e} &= \sum_{i,j,k=1}^{3} (s_k^e \mathbf{r}_k^e \otimes \mathbf{r}_k^e) : [g_{ij}(\mathbf{r}_i^e \otimes \mathbf{r}_j^e) \otimes (\mathbf{r}_i^e \otimes \mathbf{r}_j^e)] \\ &= \sum_{i,j,k=1}^{3} s_k^e g_{ij} \underbrace{[(\mathbf{r}_k^e \otimes \mathbf{r}_k^e) : (\mathbf{r}_i^e \otimes \mathbf{r}_j^e)]}_{=\delta_{ik}\delta_{jk}} \mathbf{r}_i^e \otimes \mathbf{r}_j^e \\ &= \sum_{k=1}^{3} \frac{s_k^e}{c_k^e} \mathbf{r}_k^e \otimes \mathbf{r}_k^e . \end{aligned} \tag{3.56}$$

It follows immediately that $\mathbf{M}^e$ (3.41) can be simplified as

$$\begin{aligned} \mathbf{M}^e &= \sum_{i,k=1}^{3} (c_i^e \mathbf{r}_i^e \otimes \mathbf{r}_i^e)(\frac{s_k^e}{c_k^e} \mathbf{r}_k^e \otimes \mathbf{r}_k^e) \\ &= \sum_{k=1}^{3} s_k^e \mathbf{r}_k^e \otimes \mathbf{r}_k^e = \mathbf{S}^e . \end{aligned} \tag{3.57}$$

As a result, the expression for $\mathbf{P}$ can be rewritten as

$$\mathbf{P} = \mathbf{F}^{e-\tau} \mathbf{S}^e \mathbf{F}^{t-\tau} , \tag{3.58}$$

which does not contain the derivatives of $\mathbf{E}^e$ either.

| Parameters | Variables |
|---|---|
| elastic moduli (A, M) | $\mathscr{C}^A$, $\mathscr{C}^M$ |
| maximum transformation strain | $\bar{\varepsilon}^{\mathrm{t}}$ |
| transformation resistance | $Y$ |
| temperature | $T$ |
| equilibrium temperature | $T_{\mathrm{eq}}$ |
| transformation entropy (A$\rightarrow$ M) | $\Delta s_{\mathrm{eq}}$ |
| hardening coefficient | $H^{\mathrm{t}}$ |
| interfacial energy coefficient | $S_0$ |
| energetic length scale | $\ell_{\mathrm{e}}$ |
| dissipative length scale | $\ell_{\mathrm{d}}$ |

Table 3.1: Parameters of the nonlocal superelastic model.

## Microforce balance equation revisited

With the newly derived constitutive equations, the micro-force balance equation (3.43) can be rewritten as

$$\begin{aligned}&\mathbf{M}^{\mathrm{e}} : \mathbf{\Lambda} - \frac{1}{2}(\mathscr{C}_{,\xi}\mathbf{E}^{\mathrm{e}}) : \mathbf{E}^{e} + \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) - H^{\mathrm{t}}\xi + S_0\ell_{\mathrm{e}}^2(\nabla\cdot\nabla)\xi \\ =&\frac{Y\dot{\xi}}{\sqrt{(\dot{\xi})^2 + \ell_{\mathrm{d}}^2\|\nabla\dot{\xi}\|^2}} - \mathrm{div}\left(\frac{\mathrm{Y}\ell_{\mathrm{d}}^2\nabla\dot{\xi}}{\sqrt{(\dot{\xi})^2 + \ell_{\mathrm{d}}^2\|\nabla\dot{\xi}\|^2}}\right) .\end{aligned} \tag{3.59}$$

Similarly to Eqn (3.4), the flow direction tensor $\mathbf{\Lambda}$ is assumed to take the following expression:

$$\mathbf{\Lambda} = \begin{cases} \sqrt{\frac{3}{2}}\bar{\varepsilon}^{\mathrm{t}}\frac{\mathbf{M}^{\mathrm{edev}}}{\|\mathbf{M}^{\mathrm{edev}}\|}, & \text{for } \dot{\xi} > 0\ , \\ \sqrt{\frac{3}{2}}\bar{\varepsilon}^{\mathrm{t}}\frac{\mathbf{E}^{\mathrm{t,r}}}{\|\mathbf{E}^{\mathrm{t,r}}\|}, & \text{for } \dot{\xi} < 0\ , \end{cases} \tag{3.60}$$

where $\mathbf{M}^{\mathrm{edev}}$ is the deviatoric part of $\mathbf{M}^{\mathrm{e}}$, $\bar{\varepsilon}^{\mathrm{t}}$ is the maximum transformation strain, and $\mathbf{E}^{\mathrm{t,r}}$ is the logarithmic phase transformation strain tensor upon unloading.

## Model parameters

The finite deformation formulation shares the same set of model parameters with the small strain formulation. In Table 3.1, the parameters of the nonlocal superelastic model are summarized. It is also interesting to note the strains and stresses corre-

| | Finite deformation | Small strain |
|---|---|---|
| total strain | $\mathbf{E}$ | $\boldsymbol{\varepsilon}$ |
| elastic strain | $\mathbf{E}^{\mathrm{e}}$ | $\boldsymbol{\varepsilon}^{\mathrm{e}}$ |
| phase transformation strain | $\mathbf{E}^{\mathrm{t}}$ | $\boldsymbol{\varepsilon}^{\mathrm{t}}$ |
| work conjugate to elastic strain | $\mathbf{S}^{\mathrm{e}}$ | $\boldsymbol{\sigma}$ |
| stress in macroforce balance equation | $\mathbf{P}$ | $\boldsymbol{\sigma}$ |
| stress in microforce balance equation | $\mathbf{M}^{\mathrm{e}}$ | $\boldsymbol{\sigma}$ |

Table 3.2: Strains and stresses correspondences between finite deformation and small strain formulations.

spondences between the small strain and finite deformation formulations, which are listed in Table 3.2.

## 3.3 Numerical discretization

In this section, the coupled governing equations from the nonlocal model are first discretized in time. Based on the temporal discretization, a variational incremental problem is formulated in order to guide numerical applications. The macro- and micro-force balance equations are characterized as the Euler-Lagrange equations of an incremental functional, and the solution fields are characterized as the minimizer subject to kinematic constraints. Later, spatial discretization is presented, followed by the discussion of the fully discretized formulation. For simplicity, small strain formulation is adopted in the derivations, nonetheless it is straightforward to extend the derivations to finite deformation formulation with the correspondences presented in Table (3.2).

### 3.3.1 Time discretization and variational incremental problem

The superelastic response considered in this thesis is quasistatic, and therefore time $t$ is not a variable. However, the martensitic phase transformation depends on the loading history, and it is necessary to break the loading history into sequential load

increments, and solve the discretized macro- and micro-force balance equations for the increment of the displacement and martensitic volume fraction at each load increment. It is also worth noting that the micro-force balance equation (3.23) is rate-independent in the sense that multiplying the rate of martensitic volume fraction $\dot{\xi}$ by a positive scalar does not change the equation. Nonetheless, the increment of martensitic volume fraction for each load increment will be determined by solving the corresponding boundary value problem. In the following derivation, load increments are described by pseudo time intervals $[t^{(n)}, t^{(n+1)}]$, for $n = 0, 1, 2, \ldots$ . And $t^{(0)}$ denotes the initial state. For a general load increment $[t^{(n)}, t^{(n+1)}]$, variables evaluated at the beginning and the end of this load increment are denoted as $\bullet^{(n)}$ and $\bullet^{(n+1)}$, respectively.

**Time discretization of variables**

1. Increment of displacement is $\Delta u_i := u_i^{(n+1)} - u_i^{(n)}$ , and consequently $\Delta u_{i,j} = u_{i,j}^{(n+1)} - u_{i,j}^{(n)}$.

2. Increment of martensitic volume fraction is $\Delta\xi := \xi^{(n+1)} - \xi^{(n)}$ , and consequently $\Delta\xi_{,i} = \xi_{,i}^{(n+1)} - \xi_{,i}^{(n)}$.

3. Increment of total strain is $\Delta\varepsilon_{ij} := \varepsilon_{ij}^{(n+1)} - \varepsilon_{ij}^{(n)} = \frac{1}{2}(\Delta u_{i,j} + \Delta u_{j,i})$.

4. Increment of inelastic strain is $\Delta\varepsilon_{ij}^{\mathrm{t}} := \varepsilon^{\mathrm{t}(n+1)}_{ij} - \varepsilon^{\mathrm{t}(n)}_{ij}$ , and according to Eqn (3.3), it is also linked to the flow direction tensor through

$$\Delta\varepsilon_{ij}^{\mathrm{t}} = \Delta\xi\Lambda_{ij} \,. \tag{3.61}$$

5. Increment of elastic strain is $\Delta\varepsilon_{ij}^{\mathrm{e}} := \varepsilon^{\mathrm{e}(n+1)}_{ij} - \varepsilon^{\mathrm{e}(n)}_{ij} = \Delta\varepsilon_{ij} - \Delta\varepsilon_{ij}^{\mathrm{t}}$ .

6. The energy dissipation per unit volume during this increment is

$$\mathcal{D}^{\Delta t} := Y\sqrt{(\Delta\xi)^2 + \ell_{\mathrm{d}}^2(\Delta\xi_{,k}\Delta\xi_{,k})} \,, \tag{3.62}$$

which approximates $\int_{t^{(n)}}^{t^{(n+1)}} \mathcal{D}\, d\tau = \int_{t^{(n)}}^{t^{(n+1)}} Y\sqrt{\dot{\xi}^2(x_i,\tau) + \ell_{\mathrm{d}}^2\|\nabla\dot{\xi}(x_i,\tau)\|^2}\, d\tau$ .

7. Increment of the free energy per unit volume is

$$\begin{aligned}\psi^{\Delta t} &:= \psi^{(n+1)} - \psi^{(n)} \\ &= \frac{1}{2}\mathscr{C}_{ijkl}^{(n+1)}\varepsilon_{ij}^{\mathrm{e}(n+1)}\varepsilon_{kl}^{\mathrm{e}(n+1)} - \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}})\xi^{(n+1)} \\ &\quad + \frac{1}{2}H^{\mathrm{t}}(\xi^{(n+1)})^2 + \frac{1}{2}S_0\ell_e^2\|\nabla\xi^{(n+1)}\|^2 - \psi^{(n)}\end{aligned} \tag{3.63}$$

with $\mathscr{C}^{(n+1)} = (1 - \xi^{(n+1)})\mathscr{C}^A + \xi^{(n+1)}\mathscr{C}^M$.

8. Flow direction tensor $\Lambda_{ij}$ according to Eqn (3.4) is expressed as

$$\Lambda_{ij} = \begin{cases} \sqrt{\frac{3}{2}}\bar{\varepsilon}^{\mathrm{t}}\frac{\tilde{\sigma}_{ij}^{\mathrm{dev}}}{\|\tilde{\boldsymbol{\sigma}}^{\mathrm{dev}}\|}\,, & \text{for } \Delta\xi > 0\,, \\ \sqrt{\frac{3}{2}}\bar{\varepsilon}^{\mathrm{t}}\frac{\varepsilon^{\mathrm{t}}{}_{ij}^{(n)}}{\|\boldsymbol{\varepsilon}^{\mathrm{t}(n)}\|}\,, & \text{for } \Delta\xi < 0\,, \end{cases} \tag{3.64}$$

where $\tilde{\boldsymbol{\sigma}}$ is defined as

$$\tilde{\sigma}_{ij} := \mathscr{C}_{ijkl}^{(n+1)}(\varepsilon_{kl}^{(n+1)} - \varepsilon^{\mathrm{t}}{}_{kl}^{(n)})\,. \tag{3.65}$$

For isotropic elastic responses, the deviatoric part of $\tilde{\boldsymbol{\sigma}}$ and that of $\boldsymbol{\sigma}^{(n+1)}$ are collinear.

In the constitutive updates of conventional macroscopic superelastic models [13, 54], the increment of martensitic volume fraction and the flow direction tensor are determined locally at integration points using yield conditions, provided a strain input. In the gradient superelastic model presented, the martensitic volume fraction is a primary unknown variable, so the flow direction tensor, Eqn (3.64), depends on both strain (therefore displacement) and martensitic volume fraction fields, which causes some implementation difficulties. First, the dependency on the martensitic volume fraction can be problematic in the computation since the flow direction changes abruptly as $\Delta\xi$ changes sign. Second, there is no yield condition in the presence of the dissipative length scale, because the flow resistance depends on $\Delta\xi$ and its gradient, which are unknown at $t^{(n)}$. It is possible to enforce Eqn (3.64) weakly as a variational result. Details of this

discussion are provided in Appendix C.4. Nonetheless, in order to avoid the computational burden, we solve the macro-force balance equation completely at the beginning of this increment, use the stress obtained to determine the flow direction tensor at each integration point, and fix the flow direction tensor in the rest of the solving process. The flow direction tensor in computation is determined as follows: given the elastic trial stress $\hat{\boldsymbol{\sigma}}$,

$$\Lambda_{ij} = \begin{cases} \sqrt{\frac{3}{2}}\bar{\varepsilon}^{\mathrm{t}} \frac{\hat{\sigma}_{ij}^{\mathrm{dev}}}{\|\hat{\boldsymbol{\sigma}}^{\mathrm{dev}}\|} , & \text{if} \quad \|\hat{\boldsymbol{\sigma}}^{\mathrm{dev}}\| \geq \|\boldsymbol{\sigma}^{(n)\mathrm{dev}}\| \text{ or } \|\boldsymbol{\varepsilon}^{\mathrm{t}(n)}\| = 0 , \\ \sqrt{\frac{3}{2}}\bar{\varepsilon}^{\mathrm{t}} \frac{\varepsilon_{ij}^{\mathrm{t}(n)}}{\|\boldsymbol{\varepsilon}^{\mathrm{t}(n)}\|} , & \text{otherwise} . \end{cases} \tag{3.66}$$

It is worth noting that the determination of the flow direction tensor doesn't necessarily lead to phase transformation along that direction, because the increment of martensitic volume fraction can be zero, which yields an elastic response.

In summary, for a general load increment, $\mathbf{u}^{(n)}$, $\xi^{(n)}$, $\boldsymbol{\varepsilon}^{(n)}$, $\boldsymbol{\varepsilon}^{\mathrm{t}(n)}$, and $\boldsymbol{\sigma}^{(n)}$ are assumed known. The flow direction tensor $\boldsymbol{\Lambda}$ is obtained by solving the macro-force balance equation with $\xi = \xi^{(n)}$ and applying Eqn (3.66). The increments of the primary fields, $\Delta\mathbf{u}$ and $\Delta\xi$ are the unknowns that will be resolved by solving the incremental boundary value problem. After the solution is obtained, $\mathbf{u}^{(n+1)}$, $\xi^{(n+1)}$, $\boldsymbol{\varepsilon}^{(n+1)}$, $\boldsymbol{\varepsilon}^{\mathrm{t}(n+1)}$, and $\boldsymbol{\sigma}^{(n+1)}$ will be determined, and the simulation can be advanced to the next load increment.

### Temporally discretized governing equations

The governing equations can be written in temporally discretized form using the identities above. The macroforce balance equation (3.12) evaluated at $t^{(n+1)}$ reads

$$0 = \sigma_{ij,j}^{(n+1)} = \frac{\partial}{\partial x_j}(\mathscr{C}_{ijkl}^{(n+1)} \varepsilon_{kl}^{\mathrm{e}(n+1)}) , \tag{3.67}$$

while the microforce balance equation (3.23) evaluated at $t^{(n+1)}$ reads

$$0 = -\sigma_{ij}^{(n+1)}\Lambda_{ij} + \frac{1}{2}\frac{\partial \mathscr{C}_{ijkl}^{(n+1)}}{\partial \xi}\varepsilon^{\mathrm{e}(n+1)}_{ij}\varepsilon^{\mathrm{e}(n+1)}_{kl} - \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) + H^{\mathrm{t}}\xi^{(n+1)} - S_0\ell_e^2\xi^{(n+1)}_{,jj} + \frac{Y\Delta\xi}{\sqrt{(\Delta\xi)^2 + \ell_{\mathrm{d}}^2(\Delta\xi_{,k}\Delta\xi_{,k})}} - \frac{\partial}{\partial x_j}\left(\frac{Y\ell_{\mathrm{d}}^2\Delta\xi_{,j}}{\sqrt{(\Delta\xi)^2 + \ell_{\mathrm{d}}^2(\Delta\xi_{,k}\Delta\xi_{,k})}}\right) . \tag{3.68}$$

It is interesting to see that those temporally discretized equations can be rewritten compactly using the definitions of the energy dissipation (3.62) and the free energy increment (3.63):

$$0 = \frac{\partial}{\partial x_j}\left(\frac{\partial \psi^{\Delta t}}{\partial \Delta\varepsilon^{\mathrm{e}}_{ij}}\right) , \tag{3.69}$$

$$0 = \frac{\partial \psi^{\Delta t}}{\partial \Delta\xi} - \frac{\partial}{\partial x_j}\left(\frac{\partial \psi^{\Delta t}}{\partial \Delta\xi_{,j}}\right) + \frac{\partial \mathcal{D}^{\Delta t}}{\partial \Delta\xi} - \frac{\partial}{\partial x_j}\left(\frac{\partial \mathcal{D}^{\Delta t}}{\partial \Delta\xi_{,j}}\right) . \tag{3.70}$$

Using the symmetry of the Cauchy stress tensor, it can be seen that

$$\begin{aligned}
\frac{\partial}{\partial x_j}\left(\frac{\partial \psi^{\Delta t}}{\partial \Delta u_{i,j}}\right) &= \frac{\partial}{\partial x_j}\left[\frac{1}{2}\times 2\times(\mathscr{C}_{klpq}^{(n+1)}\varepsilon_{pq}^{e(n+1)})\frac{\partial \varepsilon^{e(n+1)}_{kl}}{\partial \Delta u_{i,j}}\right] \\
&= \frac{\partial}{\partial x_j}\left[\sigma_{kl}^{(n+1)}\frac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})\right] \\
&= \sigma_{ij,j}^{(n+1)} \\
&= \frac{\partial}{\partial x_j}\left(\frac{\partial \psi^{\Delta t}}{\partial \Delta\varepsilon^{\mathrm{e}}_{ij}}\right) .
\end{aligned}$$

Therefore Eqn (3.69) can be rewritten as

$$0 = \frac{\partial}{\partial x_j}\left(\frac{\partial \psi^{\Delta t}}{\partial \Delta u_{i,j}}\right) . \tag{3.71}$$

The structure of Eqn (3.71) and Eqn (3.70) suggests that they could be Euler-Lagrange equations of a certain functional with the following integrand

$$F^{\Delta t} := \psi^{\Delta t} + \mathcal{D}^{\Delta t} . \tag{3.72}$$

### Incremental Functional

Assume that macro-traction $\hat{t}_i^{(n+1)}$ and micro-traction $\hat{k}^{(n+1)}$ are prescribed on $\partial B_{N_u}$ and $\partial B_{N_\xi}$ of the surface, respectively. Inspired by the structure of the temporally discretized governing equations, we define the following incremental functional

$$\mathcal{J}^{\Delta t}(\Delta u_i, \Delta\xi) := \int_B F^{\Delta t}\, dV - \int_{\partial B_{N_u}} \hat{t}_i^{(n+1)} \Delta u_i\, dS - \int_{\partial B_{N_\xi}} \hat{k}^{(n+1)} \Delta\xi\, dS\ , \tag{3.73}$$

where the integrand $F^{\Delta t}$ (3.72) is defined as the summation of the free energy increment and the energy dissipation during the time increment, which is quadratic in $\Delta u_{i,j}$, and nonlinear in $\Delta\xi$ and $\Delta\xi_{,i}$ due to the formulation of the energy dissipation.

### Statement of the incremental problem

Assume $\hat{u}_i$ and $\hat{\xi}$ are the prescribed displacement and martensitic volume fraction on $\partial B_{D_u}$ and $\partial B_{D_\xi}$ of the surface, respectively. The incremental problem can be stated as: to find the increment of the displacement $\Delta u_i$ and the increment of martensitic volume fraction $\Delta\xi$ that minimize the incremental functional defined in Eqn (3.73), and satisfy the Dirichlet boundary conditions:

$$\Delta u_i = \hat{u}_i^{(n+1)} - u_i^{(n)} \qquad \text{on } \partial B_{D_u}\ , \tag{3.74}$$

$$\Delta\xi = \hat{\xi}^{(n+1)} - \xi^{(n)} \qquad \text{on } \partial B_{D_\xi}\ , \tag{3.75}$$

as well as the constraints for martensitic volume fraction:

$$-\Delta\xi - \xi^{(n)} \le 0\ , \tag{3.76}$$

$$\Delta\xi + \xi^{(n)} - 1 \le 0\ . \tag{3.77}$$

Considering the constraints in Eqn (3.76) and Eqn (3.77), we introduce two multipliers $\lambda^{\mathrm{l}}$ and $\lambda^{\mathrm{u}}$ and construct the following Lagrange functional

$$\mathcal{L}(\Delta u_i, \Delta\xi, \lambda^{\mathrm{l}}, \lambda^{\mathrm{u}}) := \mathcal{J}^{\Delta t}(\Delta u_i, \Delta\xi) + \int_B \lambda^{\mathrm{l}}(-\Delta\xi - \xi^{(n)}) + \lambda^{\mathrm{u}}(\Delta\xi + \xi^{(n)} - 1)\, dV\ . \tag{3.78}$$

At a local minimum, multipliers satisfy the following complementary conditions:

$$\lambda^{\mathrm{l}} \geq 0, \qquad \lambda^{\mathrm{l}}(-\Delta\xi - \xi^{(n)}) = 0 \,, \tag{3.79}$$

$$\lambda^{\mathrm{u}} \geq 0, \qquad \lambda^{\mathrm{u}}(\Delta\xi + \xi^{(n)} - 1) = 0 \,. \tag{3.80}$$

The variation of $\mathcal{L}$ with respect to the field $\Delta u_i$ leads to

$$-\frac{\partial}{\partial x_j}\left(\frac{\partial F^{\Delta t}}{\partial \Delta u_{i,j}}\right) = -\frac{\partial}{\partial x_j}\left(\frac{\partial \psi^{\Delta t}}{\partial \Delta u_{i,j}}\right) = 0 \,, \qquad \text{in } B \,, \tag{3.81}$$

which is no other than the temporally discretized macro-force balance equation (3.71), and also

$$\frac{\partial F^{\Delta t}}{\partial \Delta u_{i,j}} n_j = \frac{\partial \psi^{\Delta t}}{\partial \Delta u_{i,j}} n_j = \sigma_{ij}^{(n+1)} n_j = \hat{t}_j^{(n+1)} \,, \qquad \text{on } \partial B_{N_u} \,, \tag{3.82}$$

i.e. the macro-force balance on $\partial B_{N_u}$.

The variation of $\mathcal{L}$ with respect to the field $\Delta\xi$ leads to

$$\frac{\partial F^{\Delta t}}{\partial \Delta\xi} - \lambda^{\mathrm{l}} + \lambda^{\mathrm{u}} - \frac{\partial}{\partial x_j}\left(\frac{\partial F^{\Delta t}}{\partial \Delta\xi_{,j}}\right) = 0 \,, \qquad \text{in } B \,, \tag{3.83}$$

and also

$$\frac{\partial F^{\Delta t}}{\partial \Delta\xi_{,j}} n_j = \left(\frac{\partial \psi^{\Delta t}}{\partial \Delta\xi_{,j}} + \frac{\partial \mathcal{D}^{\Delta t}}{\partial \Delta\xi_{,j}}\right) n_j = \hat{k}^{(n+1)} \,, \qquad \text{on } \partial B_{N_\xi} \,, \tag{3.84}$$

i.e. the micro-force balance on $\partial B_{N_\xi}$. Eqn (3.83) can be rewritten as

$$\frac{\partial \psi^{\Delta t}}{\partial \Delta\xi} - \frac{\partial}{\partial x_j}\left(\frac{\partial \psi^{\Delta t}}{\partial \Delta\xi_{,j}}\right) + \frac{\partial \mathcal{D}^{\Delta t}}{\partial \Delta\xi} - \frac{\partial}{\partial x_j}\left(\frac{\partial \mathcal{D}^{\Delta t}}{\partial \Delta\xi_{,j}}\right) = \lambda^{\mathrm{l}} - \lambda^{\mathrm{u}} \,. \tag{3.85}$$

Using the complementary equations, Eqn (3.79) and (3.80), it can be seen that

$$\lambda^{\mathrm{l}} - \lambda^{\mathrm{u}} = \begin{cases} 0 \,, & \text{if} \quad -\xi^{(n)} < \Delta\xi < 1 - \xi^{(n)} \,, \\ \lambda^{\mathrm{l}} \geq 0 \,, & \text{if} \quad \Delta\xi = -\xi^{(n)} \,, \\ -\lambda^{\mathrm{u}} \leq 0 \,, & \text{if} \quad \Delta\xi = 1 - \xi^{(n)} \,. \end{cases} \tag{3.86}$$

It follows immediately that the temporally discretized micro-force balance equation, Eqn (3.68), is a special case where both the lower and upper constraints on the

martensitic volume fraction are inactive, and when one of the constraints is active, an inequality instead of the equality is obtained.

In [92, 78], Radovitzky and Ortiz derived a local variational formulation of constitutive updates for a wide class of materials. The variational formulation gives rise to a symmetric tangent and is also taken as a basis for the error estimation and mesh adaptivity. In the multi-variant superelastic model of [51, 100], the constitutive updates were performed through minimizing the local sum of the free energy increment and the energy dissipation. In [68, 69], Miehe *et al.* developed a multi-field incremental variational framework for gradient extended models starting from general expressions of the free energy density and the dissipation potential, which are assumed to depend on the first order derivatives of the internal variables from local theories. The incremental problem presented here falls into the two-field formulation in their framework. A special benefit of the variational incremental formulation is that it helps define the convergence criteria at the points where bound constraints are active. In the strain gradient plasticity theory [5], Anand *et al.* proposed a minimum principle for the normalized plastic strain rate, which we find can be linked to our two-field incremental variational formulation (Appendix A).

### 3.3.2 Full discretization: time and spatial

#### Spatial discretization

Assume that domain $B$ is triangulated as $B_h$ by a set of elements $V_h^e$, $e = 1, ..., N^{\text{el}}$, where $N^{\text{el}}$ is the number of elements, i.e. $B \approx B_h = \bigcup_{e=1,...,N^{\text{el}}} V_h^e$. Denote $\varphi_a(x)$ as the nodal shape function at node $a$, for $a = 1, ..., N^{\text{node}}$, where $N^{\text{node}}$ is the number of nodes. For simplicity, same discretization is chosen for the displacement and the martensitic volume fraction fields. Denote $u_{ia}$ and $\xi_a$ as the nodal values of the displacement field and the martensitic volume fraction field, respectively. Then these fields can be represented as $u_i(x) \approx u_{hi}(x) = \sum_{a=1}^{N^{\text{node}}} u_{ia}\varphi_a(x)$, and $\xi(x) \approx \xi_h(x) =$

$\sum_{a=1}^{N^{\text{node}}} \xi_{ha}\varphi_a(x)$, and consequently the increment of these fields can be expressed as [1]

$$\Delta u_{hi}(x) = \Delta u_{ia}\varphi_a(x), \tag{3.87}$$

$$\Delta \xi_h(x) = \Delta \xi_a \varphi_a(x) \ . \tag{3.88}$$

**Statement of fully discretized incremental problem**

Assume that $\partial B_{hD_u}$, $\partial B_{hD_\xi}$, $\partial B_{hN_u}$, and $\partial B_{hN_\xi}$ are the correspondences to $\partial B_{D_u}$, $\partial B_{D_\xi}$, $\partial B_{N_u}$, and $\partial B_{N_\xi}$, respectively. The fully discretized incremental problem can be stated as: to find the increment of the displacement field $\Delta u_{hi}$ and the increment of the martensitic volume fraction field $\Delta \xi_h$ that minimize the following function

$$\mathcal{J}_h^{\Delta t}(\Delta u_{ia}, \Delta \xi_a) := \int_{B_h} F^{\Delta t}(\Delta u_{hi}, \Delta \xi_h)\, dV - \int_{\partial B_{hN_u}} \hat{t}_i^{(n+1)} \Delta u_{hi}\, dS - \int_{\partial B_{hN_\xi}} \hat{k}^{(n+1)} \Delta \xi_h\, dS \ . \tag{3.89}$$

and satisfy the boundary conditions:

$$\Delta u_{ia} = \hat{u}_{ia}^{(n+1)} - u_{ia}^{(n)} \qquad \text{on } \partial B_{hD_u} \ , \tag{3.90}$$

$$\Delta \xi_a = \hat{\xi}_a^{(n+1)} - \xi_a^{(n)} \qquad \text{on } \partial B_{hD_\xi} \ , \tag{3.91}$$

as well as the constraints of the martensitic volume fraction:

$$-\Delta \xi_a - \xi_a^{(n)} \leq 0 \ , \tag{3.92}$$

$$\Delta \xi_a + \xi_a^{(n)} - 1 \leq 0 \ . \tag{3.93}$$

The incremental functional in Eqn (3.73) now becomes a function of two real vectors. The dependence on $\Delta u_{ia}$ alone is quadratic, whereas the dependence on $\Delta \xi_a$ is nonlinear due to the formulation of the elastic moduli and the energy dissipation.

Considering the bound constraints of the martensitic volume fraction, we intro-

[1] In order to simplify the notation, summation convention will be used when it does not cause confusion.

duce two multiplier vectors, $\lambda_a^{\mathrm{l}}$ and $\lambda_a^{\mathrm{u}}$, and construct the following Lagrange function

$$\mathcal{L}_h(\Delta u_{ia}, \Delta\xi_a, \lambda_a^{\mathrm{l}}, \lambda_a^{\mathrm{u}}) := \mathcal{J}_h^{\Delta t}(\Delta u_{ia}, \Delta\xi_a) + \sum_b \left(\lambda_b^{\mathrm{l}}(-\Delta\xi_b - \xi_b^{(n)}) + \lambda_b^{\mathrm{u}}(\Delta\xi_b + \xi_b^{(n)} - 1)\right) . \quad (3.94)$$

At a local minimum, the multipliers satisfy the following complementary conditions:

$$\lambda_a^{\mathrm{l}} \geq 0, \qquad \lambda_a^{\mathrm{l}}(-\Delta\xi_a - \xi_a^{(n)}) = 0 , \quad \forall a , \quad (3.95)$$

$$\lambda_a^{\mathrm{u}} \geq 0, \qquad \lambda_a^{\mathrm{u}}(\Delta\xi_a + \xi_a^{(n)} - 1) = 0 , \quad \forall a , \quad (3.96)$$

**First order derivatives**

The partial derivative of $\mathcal{L}_h$ with respect to $\Delta u_{ia}$ reads

$$\begin{aligned}
\frac{\partial \mathcal{L}_h}{\partial \Delta u_{ia}} &= \frac{\partial \mathcal{J}_h^{\Delta t}}{\partial \Delta u_{ia}} \\
&= \int_{B_h} C_{mnpq}^{(n+1)} \varepsilon_{pq}^{e(n+1)} \frac{\partial \varepsilon_{mn}^{e(n+1)}}{\partial \Delta u_{ia}} \, dV - \int_{\partial B_{h N_u}} \hat{t}_i^{(n+1)} \varphi_a \, dS \\
&= \int_{B_h} C_{mnpq}^{(n+1)} \varepsilon_{pq}^{e(n+1)} \frac{\partial}{\partial \Delta u_{ia}} \left(\frac{1}{2}(\Delta u_{hm,n} + \Delta u_{hn,m})\right) dV - \int_{\partial B_{h N_u}} \hat{t}_i^{(n+1)} \varphi_a \, dS \\
&= \int_{B_h} C_{mnpq}^{(n+1)} \varepsilon_{pq}^{e(n+1)} \frac{\partial}{\partial \Delta u_{ia}} \left(\frac{1}{2}(\Delta u_{mb}\varphi_{b,n} + \Delta u_{nb}\varphi_{b,m})\right) dV - \int_{\partial B_{h N_u}} \hat{t}_i^{(n+1)} \varphi_a \, dS \\
&= \int_{B_h} \sigma_{mn}^{(n+1)} \left(\frac{1}{2}(\delta_{im}\delta_{ab}\varphi_{b,n} + \delta_{in}\delta_{ab}\varphi_{b,m})\right) dV - \int_{\partial B_{h N_u}} \hat{t}_i^{(n+1)} \varphi_a \, dS \\
&= \int_{B_h} \frac{1}{2}(\sigma_{in}^{(n+1)} \varphi_{a,n} + \sigma_{mi}^{(n+1)} \varphi_{a,m}) \, dV - \int_{\partial B_{h N_u}} \hat{t}_i^{(n+1)} \varphi_a \, dS \\
&= \int_{B_h} \sigma_{ij}^{(n+1)} \varphi_{a,j} \, dV - \int_{\partial B_{h N_u}} \hat{t}_i^{(n+1)} \varphi_a \, dS , \qquad (3.97)
\end{aligned}$$

where $\delta_{ij}$ takes value of one for $i = j$, and zero for $i \neq j$; the symmetry of the Cauchy stress tensor is also used in the derivation.

The partial derivative of $\mathcal{L}_h$ with respect to $\Delta\xi_a$ reads

$$\frac{\partial \mathcal{L}_h}{\partial \Delta\xi_a} = \frac{\partial \mathcal{J}_h^{\Delta t}}{\partial \Delta\xi_a} - \lambda_a^{\mathrm{l}} + \lambda_a^{\mathrm{u}} , \quad (3.98)$$

where

$$\frac{\partial \mathcal{J}_h^{\Delta t}}{\partial \Delta \xi_a} = \int_{B_h} \left( -\sigma_{ij}^{(n+1)} \Lambda_{ij} + \frac{1}{2} \frac{\partial C_{ijkl}^{(n+1)}}{\partial \xi} \varepsilon_{ij}^{\mathrm{e}(n+1)} \varepsilon_{kl}^{\mathrm{e}(n+1)} - \Delta s_{\mathrm{eq}} (T - T_{\mathrm{eq}}) + H^{\mathrm{t}} \xi^{(n+1)} \right) \varphi_a$$
$$+ S_0 \ell_e^2 \xi_{,k}^{(n+1)} \varphi_{a,k} + Y \frac{(\Delta \xi_c \varphi_c) \varphi_a + \ell_{\mathrm{d}}^2 (\Delta \xi_c \varphi_{c,k}) \varphi_{a,k}}{\sqrt{(\Delta \xi_b \varphi_b)^2 + \ell_{\mathrm{d}}^2 (\Delta \xi_b \varphi_{b,k})(\Delta \xi_d \varphi_{d,k})}} \, dV$$
$$- \int_{\partial B_{h N_\xi}} \hat{k}^{(n+1)} \varphi_a \, dS \, . \tag{3.99}$$

**First order necessary conditions for minimizer**

A local minimizer of the function (3.89) has to satisfy the following necessary conditions [9]:

$$\frac{\partial \mathcal{L}_h}{\partial \Delta u_{ia}} = 0, \tag{3.100}$$
$$\frac{\partial \mathcal{L}_h}{\partial \Delta \xi_a} = 0, \tag{3.101}$$

and the complementary conditions for the multiplier vectors, Eqn (3.95) and Eqn (3.96). These necessary conditions can be rewritten as

$$\frac{\partial \mathcal{J}_h^{\Delta t}}{\partial \Delta u_{ia}} = 0, \tag{3.102}$$
$$\frac{\partial \mathcal{J}_h^{\Delta t}}{\partial \Delta \xi_a} \begin{cases} = 0 \, , & \text{if } -\xi_a^{(n)} < \Delta \xi_a < 1 - \xi_a^{(n)} \, , \\ \geq 0 \, , & \text{if } \Delta \xi_a = -\xi_a^{(n)} \, , \\ \leq 0 \, , & \text{if } \Delta \xi_a = 1 - \xi_a^{(n)} \, . \end{cases} \tag{3.103}$$

**Second-order derivatives**

It is straightforward to derive the second-order partial derivatives of the Lagrange function $\mathcal{L}_h$ based on the first-order derivatives, Eqn (3.97) and (3.98). The second-order partial derivative with respect to the increment of the displacement field can

be derived as follows

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}_h}{\partial \Delta u_{ia} \partial \Delta u_{kb}} &= \frac{\partial^2 \mathcal{J}_h^{\Delta t}}{\partial \Delta u_{ia} \partial \Delta u_{kb}} \\
&= \int_{B_h} \frac{\partial \sigma_{ij}^{(n+1)}}{\partial \Delta u_{kb}} \varphi_{a,j} \, dV \\
&= \int_{B_h} C_{ijmn}^{(n+1)} \frac{\partial \varepsilon_{mn}^{\mathrm{e}(n+1)}}{\partial \Delta u_{kb}} \varphi_{a,j} \, dV \\
&= \int_{B_h} C_{ijmn}^{(n+1)} \frac{\partial}{\partial \Delta u_{kb}} \left[ \frac{1}{2} (\Delta u_{mc} \varphi_{c,n} + \Delta u_{nc} \varphi_{c,m}) \right] \varphi_{a,j} \, dV \\
&= \int_{B_h} C_{ijkn}^{(n+1)} \varphi_{a,j} \varphi_{b,n} \, dV \, . \qquad (3.104)
\end{aligned}$$

The mixed partial derivative can be derived as follows

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}_h}{\partial \Delta u_{ia} \partial \Delta \xi_b} &= \frac{\partial^2 \mathcal{J}_h^{\Delta t}}{\partial \Delta u_{ia} \partial \Delta \xi_b} \\
&= \int_{B_h} \frac{\partial \sigma_{ij}^{(n+1)}}{\partial \Delta \xi_b} \varphi_{a,j} \, dV \\
&= \int_{B_h} (C_{ijmn}^{(n+1)} \frac{\partial \varepsilon_{mn}^{\mathrm{e}(n+1)}}{\partial \Delta \xi_b} + \frac{\partial C_{ijmn}^{(n+1)}}{\partial \Delta \xi_b} \varepsilon_{mn}^{e(n+1)}) \varphi_{a,j} \, dV \\
&= \int_{B_h} (-C_{ijmn}^{(n+1)} \Lambda_{mn} + \frac{\partial C_{ijmn}^{(n+1)}}{\partial \xi} \varepsilon_{mn}^{e(n+1)}) \varphi_b \varphi_{a,j} \, dV \, . \qquad (3.105)
\end{aligned}$$

The second-order partial derivative with respect to the increment of the martensitic volume fraction can be derived as follows

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}_h}{\partial \Delta \xi_a \partial \Delta \xi_b} &= \frac{\partial^2 \mathcal{J}_h^{\Delta t}}{\partial \Delta \xi_a \partial \Delta \xi_b} \\
&= \int_{B_h} \left( C_{ijkl}^{(n+1)} \Lambda_{ij} \Lambda_{kl} - 2 \frac{\partial C_{ijkl}^{(n+1)}}{\partial \xi} \varepsilon_{ij}^{\mathrm{e}(n+1)} \Lambda_{kl} + \frac{1}{2} \frac{\partial^2 C_{ijkl}^{(n+1)}}{\partial \xi^2} \varepsilon_{ij}^{\mathrm{e}(n+1)} \varepsilon_{kl}^{\mathrm{e}(n+1)} \right) \varphi_b \varphi_a \\
&\quad + H^{\mathrm{t}} \varphi_a \varphi_b + S_0 \ell_e^2 \varphi_{a,k} \varphi_{b,k} + \frac{Y(\varphi_b \varphi_a + \ell_{\mathrm{d}}^2 \varphi_{b,k} \varphi_{a,k})}{\left[ (\Delta \xi_h)^2 + \ell_{\mathrm{d}}^2 (\Delta \xi_{h,k} \Delta \xi_{h,k}) \right]^{\frac{1}{2}}} \\
&\quad - \frac{Y(\Delta \xi_h \varphi_a + \ell_{\mathrm{d}}^2 \Delta \xi_{h,k} \varphi_{a,k})(\Delta \xi_h \varphi_b + \ell_{\mathrm{d}}^2 \Delta \xi_{h,j} \varphi_{b,j})}{\left[ (\Delta \xi_h)^2 + \ell_{\mathrm{d}}^2 (\Delta \xi_{h,k} \Delta \xi_{h,k}) \right]^{\frac{3}{2}}} \, dV \, , \qquad (3.106)
\end{aligned}$$

where $\Delta\xi_h$ is not expanded in order to simplify the notation. The Hessian matrix takes the following blockwise form

$$\begin{bmatrix} \frac{\partial^2 \mathcal{L}_h}{\partial \Delta u_{ia} \partial \Delta u_{kb}} & \frac{\partial^2 \mathcal{L}_h}{\partial \Delta u_{ia} \partial \Delta \xi_q} \\ \frac{\partial^2 \mathcal{L}_h}{\partial \Delta \xi_p \partial \Delta u_{kb}} & \frac{\partial^2 \mathcal{L}_h}{\partial \Delta \xi_p \partial \Delta \xi_q} \end{bmatrix} . \tag{3.107}$$

The positive definiteness of the Hessian matrix is crucial for numerical applications. In general, the Hessian matrix, Eqn (3.107), may not be positive definite since the dependence of the elastic moduli on the martensitic volume fraction can introduce the nonconvexity. In practice, we can simplify the elastic moduli as constant during the increment, i.e.

$$C_{ijkl} \approx C_{ijkl}(\xi^{(n)}) . \tag{3.108}$$

Under this simplification, we can test the positive definiteness of the Hessian matrix with any pair of testing fields $v_{hi} = v_{ia}\varphi_a$, and $\eta_h = \eta_p \varphi_p$ as follows:

$$\begin{aligned} & [v_{ia}, \eta_p] \begin{bmatrix} \frac{\partial^2 \mathcal{L}_h}{\partial \Delta u_{ia} \partial \Delta u_{kb}} & \frac{\partial^2 \mathcal{L}_h}{\partial \Delta u_{ia} \partial \Delta \xi_q} \\ \frac{\partial^2 \mathcal{L}_h}{\partial \Delta \xi_p \partial \Delta u_{kb}} & \frac{\partial^2 \mathcal{L}_h}{\partial \Delta \xi_p \partial \Delta \xi_q} \end{bmatrix} \begin{bmatrix} v_{kb} \\ \eta_q \end{bmatrix} \\ = & \int_{B_h} C_{ijkl}(v_{hi,j} - \Lambda_{ij}\eta_h)(v_{hk,l} - \Lambda_{kl}\eta_h)\, dV \\ & + \int_{B_h} H^{\mathrm{t}}(\eta_h)^2 + S_0 \ell_e^2 (\eta_{h,k}\eta_{h,k})\, dV \\ & + \int_{B_h} \frac{Y(\eta_h^2 + \ell_{\mathrm{d}}^2 \eta_{h,k}\eta_{h,k})}{\left[(\Delta\xi_h)^2 + \ell_{\mathrm{d}}^2(\Delta\xi_{h,k}\Delta\xi_{h,k})\right]^{\frac{1}{2}}} dV \\ & - \int_{B_h} \frac{Y(\Delta\xi_h\eta_h + \ell_{\mathrm{d}}^2 \Delta\xi_{h,k}\eta_{h,k})^2}{\left[(\Delta\xi_h)^2 + \ell_{\mathrm{d}}^2(\Delta\xi_{h,k}\Delta\xi_{h,k})\right]^{\frac{3}{2}}} dV . \end{aligned} \tag{3.109}$$

The Hessian matrix is positive definite if the right hand side of Eqn (3.109) is greater than zero. First, the symmetry and positive definiteness assumptions of the elastic moduli [47] lead to

$$C_{ijkl}(v_{hi,j} - \Lambda_{ij}\eta_h)(v_{hk,l} - \Lambda_{kl}\eta_h) \geq 0 .$$

Secondly, by applying Cauchy-Schwarz inequality, we can also obtain

$$(\eta_h^2 + \ell_{\mathrm{d}}^2 \eta_{h,k}\eta_{h,k})(\Delta\xi_h^2 + \ell_{\mathrm{d}}^2(\Delta\xi_{h,k}\Delta\xi_{h,k})) - (\eta_h\Delta\xi_h + \ell_{\mathrm{d}}^2\eta_{h,k}\Delta\xi_{h,k})^2 \geq 0 \,.$$

It then follows that

$$\begin{aligned} & [v_{ia}, \eta_p] \begin{bmatrix} \frac{\partial^2 \mathcal{L}_h}{\partial \Delta u_{ia} \partial \Delta u_{kb}} & \frac{\partial^2 \mathcal{L}_h}{\partial \Delta u_{ia} \partial \Delta \xi_q} \\ \frac{\partial^2 \mathcal{L}_h}{\partial \Delta \xi_p \partial \Delta u_{kb}} & \frac{\partial^2 \mathcal{L}_h}{\partial \Delta \xi_p \partial \Delta \xi_q} \end{bmatrix} \begin{bmatrix} v_{kb} \\ \eta_q \end{bmatrix} \\ \geq & \int_{B_h} H^{\mathrm{t}}(\eta_h)^2 + S_0 \ell_e^2 (\eta_{h,k}\eta_{h,k})\, dV \,. \end{aligned} \tag{3.110}$$

Since the parameter group $S_0\ell_e^2$ is greater than zero, it is obvious that the Hessian matrix is positive definite for $H^{\mathrm{t}} \geq 0$, i.e. when the phase transformation does not exhibit softening, and the Hessian matrix is strictly positive definite for $H^{\mathrm{t}} > 0$, i.e. when the phase transformation exhibits hardening. While for $H^{\mathrm{t}} < 0$, the Hessian matrix is positive definite if the inequality

$$H^{\mathrm{t}} + S_0\ell_e^2 \frac{\int_{B_h} \eta_{h,k}\eta_{h,k}\, dV}{\int_{B_h} \eta_h^2\, dV} \geq 0 \tag{3.111}$$

holds for any test field $\eta_h$ subject to $\int_{B_h} \eta_h^2\, dx > 0$ and the boundary condition $\eta_h = 0$ on $\partial B_{hD_\xi}$. In the case of $\partial B_{hD_\xi} = \partial B_h$, the infimum of the fraction in Eqn (3.111)

$$\inf \frac{\int_{B_h} \eta_{h,k}\eta_{h,k}\, dV}{\int_{B_h} \eta_h^2\, dV} \tag{3.112}$$

is the finite element approximation to the smallest eigenvalue $\lambda$ of the following problem:

$$\begin{cases} \eta_{,kk} = \lambda\eta & \text{in} \quad B_h \,, \\ \eta = 0 & \text{on} \quad \partial B_h \,. \end{cases} \tag{3.113}$$

In one dimension, it can be shown that the smallest eigenvalue of problem (3.113) is $(\frac{\pi}{L})^2$ with $L$ the size of the domain. In general, the infimum depends on the domain $B_h$, the finite element discretization, and also $\partial B_{hD_\xi}$ where the Dirichlet boundary condition is applied. When this infimum is strictly greater than zero, the inequality

(3.111) will hold for sufficiently large $S_0\ell_e^2$, and the Hessian matrix will be positive definite.

### Connection to the terminology in FEM for mechanical problems

It is straightforward to find the correspondences of the derivatives of the incremental function to the common terminology in finite element methods for mechanical problems, e.g. the internal force, the external force, the residual force, and the stiffness matrix.

The internal, external, and residual forces associated with the macro-force balance equation can be expressed as

$$f_{u_{ia}}^{\text{int}} := \int_{B_h} \sigma_{ij}^{(n+1)} \varphi_{a,j} \, dV \ , \tag{3.114}$$

$$f_{u_{ia}}^{\text{ext}} := \int_{\partial B_{h N_u}} \hat{t}_i^{(n+1)} \varphi_a \, dS \ , \tag{3.115}$$

$$r_{u_{ia}} := f_{u_{ia}}^{ext} - f_{u_{ia}}^{int} \ , \tag{3.116}$$

while the internal, external, and residual forces associated with the micro-force balance equation can be expressed as

$$\begin{aligned} f_{\xi_a}^{\text{int}} := \int_{B_h} \Bigg( & -\sigma_{ij}^{(n+1)} \Lambda_{ij} + \frac{1}{2} \frac{\partial C_{ijkl}^{(n+1)}}{\partial \xi} \varepsilon^{\text{e}(n+1)}_{ij} \varepsilon^{\text{e}(n+1)}_{kl} - \Delta s_{\text{eq}} (T - T_{\text{eq}}) + H^{\text{t}} \xi^{(n+1)} \Bigg) \varphi_a \\ & + S_0 \ell_e^2 \xi_{,k}^{(n+1)} \varphi_{a,k} + Y \frac{(\Delta \xi_c \varphi_c) \varphi_a + \ell_{\text{d}}^2 (\Delta \xi_c \varphi_{c,k}) \varphi_{a,k}}{\left[ (\Delta \xi_b \varphi_b)^2 + \ell_{\text{d}}^2 (\Delta \xi_b \varphi_{b,k}) (\Delta \xi_d \varphi_{d,k}) \right]^{\frac{1}{2}}} \, dV \ , \end{aligned} \tag{3.117}$$

$$f_{\xi_a}^{\text{ext}} := \int_{\partial B_{h N_\xi}} \hat{k}^{(n+1)} \varphi_a \, dS \ , \tag{3.118}$$

$$r_{\xi_a} := f_{\xi_a}^{\text{ext}} - f_{\xi_a}^{\text{int}} \ . \tag{3.119}$$

It follows immediately that the residual forces are the negative gradients of the incremental function (3.89), i.e.

$$r_{u_{ia}} = -\frac{\partial \mathcal{J}_h^{\Delta t}}{\partial \Delta u_{ia}} \ , \tag{3.120}$$

and

$$r_{\xi_p} = -\frac{\partial \mathcal{J}_h^{\Delta t}}{\partial \Delta \xi_p} \,. \tag{3.121}$$

The stiffness matrix is defined as the partial derivative of the internal force with respect to the unknown field. In this two-fields coupled problem, it can be expressed as

$$\mathbf{K} = \begin{bmatrix} K_{ia\,kb} & K_{ia\,q} \\ K_{p\,kb} & K_{p\,q} \end{bmatrix} \,, \tag{3.122}$$

and the components are defined as

$$K_{ia\,kb} := \frac{\partial f^{\text{int}}_{u_{ia}}}{\partial \Delta u_{kb}} = \frac{\partial^2 \mathcal{J}_h^{\Delta t}}{\partial \Delta u_{ia} \partial \Delta u_{kb}} \,, \tag{3.123}$$

$$K_{ia\,q} := \frac{\partial f^{\text{int}}_{u_{ia}}}{\partial \Delta \xi_q} = \frac{\partial^2 \mathcal{J}_h^{\Delta t}}{\partial \Delta u_{ia} \partial \Delta \xi_q} \,, \tag{3.124}$$

$$K_{p\,kb} := \frac{\partial f^{\text{int}}_{\xi_p}}{\partial \Delta u_{kb}} = \frac{\partial^2 \mathcal{J}_h^{\Delta t}}{\partial \Delta u_{kb} \partial \Delta \xi_p} \,, \tag{3.125}$$

$$K_{p\,q} := \frac{\partial f^{\text{int}}_{\xi_p}}{\partial \Delta \xi_q} = \frac{\partial^2 \mathcal{J}_h^{\Delta t}}{\partial \Delta \xi_p \partial \Delta \xi_q} \,. \tag{3.126}$$

It follows immediately that the stiffness matrix is the Hessian matrix of the incremental function 3.89).

Due to the bound constraints for $\Delta\xi$, we can not expect $r_{\xi_p} = 0$ for the minimizer of the incremental function. Instead, the previous discussion on the first order optimal conditions suggests that the minimizer $(\Delta u^*_{ia}, \Delta \xi^*_p)$ satisfies

$$r_{\xi_q}(\Delta u^*_{ia}, \Delta \xi^*_p) = -\frac{\partial \mathcal{J}_h^{\Delta t}}{\partial \Delta \xi_q}(\Delta u^*_{ia}, \Delta \xi^*_p) \begin{cases} = 0 \,, & \text{if } -\xi_q^{(n)} < \Delta \xi^*_q < 1 - \xi_q^{(n)} \,, \\ \leq 0 \,, & \text{if } \Delta \xi^*_q = -\xi_q^{(n)} \,, \\ \geq 0 \,, & \text{if } \Delta \xi^*_q = 1 - \xi_q^{(n)} \,. \end{cases} \tag{3.127}$$

**Comments on the regularization**

There are terms in the micro-force balance equation (3.23) including $(\dot{\xi})^2 + \ell_{\mathrm{d}}^2 \|\nabla\dot{\xi}\|^2$ in the denominator. When $|\dot{\xi}| = \ell_{\mathrm{d}}\|\nabla\dot{\xi}\| = 0$, the micro-force balance equation is not well defined. Similarly, when $\Delta\xi = \ell_{\mathrm{d}}\Delta\xi_{,k}\Delta\xi_{,k} = 0$, derivatives of the incremental function, Eqn (3.99) and (3.106), are not well defined. Therefore the incremental function (3.89) is not smooth. In order to use numerical methods that require continuous derivatives of the incremental function, we add a small positive constant $\beta$ (typically $10^{-8}$) in the expression of the energy dissipation rate, i.e.

$$\mathcal{D}^{\Delta t} \approx Y\left[(\Delta\xi)^2 + \ell_{\mathrm{d}}^2\Delta\xi_{,k}\Delta\xi_{,k} + \beta\right]^{\frac{1}{2}} , \tag{3.128}$$

so the incremental function becomes smooth. Some other schemes for regularization can be found in a recent book by Han and Reddy [43].

### 3.3.3 A staggered algorithm for fully discretized incremental problem

The fully discretized incremental problem has been formulated as a nonlinear constrained minimization problem of two vectors, $\Delta u_{ia}$ and $\Delta\xi_a$. In numerical simulations, $\Delta u_{ia}$ and $\Delta\xi_a$ that satisfy the first order necessary conditions, Eqn (3.102) and (3.103), the Dirichlet boundary conditions, Eqn (3.90) and (3.91), and the bound constraints, Eqn (3.92) and (3.93), are accepted as the solution to the incremental problem.

Using the simplification of elastic moduli, Eqn (3.108), and the regularization of energy dissipation, Eqn (3.128), the fully discretized incremental function can be

expressed as

$$\mathcal{J}_h^{\Delta t}(\Delta u_{ia}, \Delta\xi_a) = \int_{B_h} \frac{1}{2}\mathscr{C}_{ijkl}(\varepsilon_{kl} - \Delta\xi_h\Lambda_{kl} - \varepsilon_{kl}^{\mathrm{t}(n)})(\varepsilon_{ij} - \Delta\xi_h\Lambda_{ij} - \varepsilon_{ij}^{\mathrm{t}(n)})\, dV$$
$$+ \int_{B_h} -\Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}})\Delta\xi_h + \frac{1}{2}H^{\mathrm{t}}(\xi_h^{(n)} + \Delta\xi_h)^2\, dV$$
$$+ \int_{B_h} \frac{1}{2}S_0(\ell_{\mathrm{e}})^2\|\nabla(\xi_h^{(n)} + \Delta\xi_h)\|^2 + Y\sqrt{(\Delta\xi_h)^2 + \ell_{\mathrm{d}}^2\Delta\xi_{h,k}\Delta\xi_{h,k} + \beta}\, dV$$
$$- \int_{\partial B_{h N_u}} \hat{t}_i^{(n+1)}\Delta u_{hi}\, dS - \int_{\partial B_{h N_\xi}} \hat{k}^{(n+1)}\Delta\xi_h\, dS \tag{3.129}$$

with $\Delta\xi_h = \Delta\xi_a\varphi_a$, $\Delta u_{hi} = \Delta u_{ia}\varphi_a$ and $\varepsilon_{ij} = \varepsilon_{ij}^{(n)} + \frac{1}{2}(\Delta u_{ia}\varphi_{a,j} + \Delta u_{ja}\varphi_{a,i})$.

As a result, the weak forms of the two governing equations, Eqn (3.102) and (3.103), can be rewritten as follows: for macro-force balance, the residual force must be zero, i.e.

$$r_{u_{ia}} = -\int_{B_h} \mathscr{C}_{ijkl}(\varepsilon_{kl} - \Delta\xi_h\Lambda_{kl} - \varepsilon_{kl}^{\mathrm{t}(n)})\varphi_{a,j}\, dV + \int_{\partial B_{h N_u}} \hat{t}_i^{(n+1)}\varphi_a\, dS = 0\ ; \tag{3.130}$$

while for micro-force balance, the residual force is expressed as

$$r_{\xi_a} = -\int_{B_h} \left[-\mathscr{C}_{ijkl}(\varepsilon_{kl} - \Delta\xi_h\Lambda_{kl} - \varepsilon_{kl}^{\mathrm{t}(n)})\Lambda_{ij} - \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) + H^{\mathrm{t}}(\xi_h^{(n)} + \Delta\xi_h)\right]\varphi_a$$
$$+ S_0\ell_e^2(\xi_{h,k}^{(n)} + \Delta\xi_{h,k})\varphi_{a,k} + Y\frac{\Delta\xi_h\varphi_a + \ell_{\mathrm{d}}^2\Delta\xi_{h,k}\varphi_{a,k}}{\sqrt{(\Delta\xi_h)^2 + \ell_{\mathrm{d}}^2\Delta\xi_{h,k}\Delta\xi_{h,k} + \beta}}\, dV$$
$$+ \int_{\partial B_{h N_\xi}} \hat{k}^{(n+1)}\varphi_a\, dS\ , \tag{3.131}$$

which must satisfy

$$r_{\xi_a}\begin{cases} = 0\ , & \text{if } -\xi_a^{(n)} < \Delta\xi_a < 1 - \xi_a^{(n)}\ , \\ \leq 0\ , & \text{if } \Delta\xi_a = -\xi_a^{(n)}\ , \\ \geq 0\ , & \text{if } \Delta\xi_a = 1 - \xi_a^{(n)}\ . \end{cases} \tag{3.132}$$

Coupled terms in the weak forms of the macro-force and micro-force balance

equations are highlighted in blue. It can be seen that these two equations are loosely coupled in the following way: the macro-force balance equation requires the inelastic strain $\boldsymbol{\varepsilon}^{\mathrm{t}} = \Delta\xi_h \boldsymbol{\Lambda} + \boldsymbol{\varepsilon}^{\mathrm{t}(n)}$ to form the stress and also the martensitic volume fraction to form the elastic moduli $\mathscr{C} = \mathscr{C}(\xi^{(n)})$, while the micro-force balance equation needs the total strain $\boldsymbol{\varepsilon}$ to form the stress $\boldsymbol{\sigma} = \mathscr{C}(\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\mathrm{t}})$, and to provide a driving force for phase transformations. It can also be seen that the bound constraints only act on the micro-force balance equation, and the two equations have quite different stiffness matrices, i.e. Eqn (3.104) and Eqn (3.106). Therefore, it is natural to consider a staggered algorithm to solve these coupled equations.

A staggered algorithm for the fully discretized incremental problem includes the following two steps:

1. Elastic predictor:

   (a) apply displacement boundary conditions, Eqn (3.90);

   (b) solve the macro-force balance equation $r_{u_{ia}} = 0$ (3.130) with $\Delta\xi_a = 0$;

   (c) obtain stress $\hat{\boldsymbol{\sigma}}$, strain $\boldsymbol{\varepsilon}$; and determine the flow direction tensor $\boldsymbol{\Lambda}$ using Eqn (3.66).

2. Iterative solutions of micro-force balance and macro-force balance equations until convergence is achieved [2]:

| micro-force balance | macro-force balance |
|---|---|
| receive $\boldsymbol{\varepsilon}$ | receive $\boldsymbol{\varepsilon}^{\mathrm{t}}$ |
| solve Eqn (3.131), (3.132) for $\Delta\xi_a$ | solve Eqn (3.130) for $\Delta u_{ia}$ |
| send $\boldsymbol{\varepsilon}^{\mathrm{t}} := \Delta\xi_h \boldsymbol{\Lambda} + \boldsymbol{\varepsilon}^{\mathrm{t}(n)}$ | send $\boldsymbol{\varepsilon}$ |

The iteration above is described as *block coordinate descent* or *nonlinear Gauss-Seidel* method in nonlinear programming. Convergence is guaranteed if the objective function is smooth and the solution to each sub-problem is uniquely obtained during

[2] A typical convergence criterion taken in the computation is $\max_a |\Delta\xi_a^{k+1} - \Delta\xi_a^k| \leq 10^{-6}$.

iteration [9]. In our case, the objective function $\mathcal{J}_h^{\Delta}$, Eqn (3.129), is smooth, and it is also convex for $H^{\mathrm{t}} \geq 0$, so the convergence is guaranteed.

We have implemented a Newton-Raphson method with line search and also a parallel dynamic relaxation method to solve the sub-problems. The latter is found to be more efficient in both time and memory consumption, and therefore suitable for large scale three-dimensional problems. The parallel dynamic relaxation method is presented in detail in the following section, while a comparison of the solvers is provided at the end of this chapter.

### 3.3.4 Parallel dynamic relaxation method

Dynamic relaxation method is an iterative method for solving static problems, especially those in structural mechanics [81, 83, 118]. The idea of this method is that a static solution can be viewed as the steady state of a dynamic response. For instance, a solid structure is in equilibrium when the residual force $\mathbf{r}$ is zero under a certain displacement field $\mathbf{u}$, i.e.

$$\mathbf{0} = \mathbf{r}(\mathbf{u}) \,, \tag{3.133}$$

where $\mathbf{u}$ is the static solution. This solution can be viewed as the steady state of the following dynamic response

$$\rho\ddot{\mathbf{u}} + c\rho\dot{\mathbf{u}} = \mathbf{r}(\mathbf{u}) \,, \tag{3.134}$$

at which the acceleration and the velocity are both zeros. In Eqn (3.134), $\rho$ is the artificial material density, and $c > 0$ is the damping coefficient. Once the displacement boundary conditions are applied, the residual force is out of balance, and therefore it generates a stress wave. As a result of the viscosity, the total energy of the dynamic system will continuously decrease during the stress wave propagation until the acceleration and the velocity become zeros, i.e. a static solution is achieved. In practice, mass lumping techniques and the explicit time integration (explicit Newmark) [47] are employed to simulate the dynamic response.

The spatial discretization of Eqn (3.134) can be written as

$$\mathbf{Ma} + c\mathbf{Mv} = \mathbf{r}(\mathbf{u}) \tag{3.135}$$

where $\mathbf{M}$ is the diagonal mass matrix from mass lumping, $\mathbf{a}$ and $\mathbf{v}$ are the acceleration and velocity nodal fields.

The time integration follows the explicit Newmark algorithm. More specifically, assuming that the time-step is $\tau$, a general temporal increment includes three steps to obtain $\{\mathbf{u}^{(k+1)}, \mathbf{v}^{(k+1)}, \mathbf{a}^{(k+1)}\}$ from $\{\mathbf{u}^{(k)}, \mathbf{v}^{(k)}, \mathbf{a}^{(k)}\}$ :
Predictor[3]:

$$\mathbf{u}^{(k+1)} := \mathbf{u}^{(k)} + \tau\mathbf{v}^{(k)} + \frac{1}{2}(\tau)^2\mathbf{a}^{(k)} , \tag{3.136}$$

$$\tilde{\mathbf{v}} := \mathbf{v}^{(k)} + \frac{1}{2}\tau\mathbf{a}^{(k)} . \tag{3.137}$$

Calculation of the residual force $\mathbf{r}(\mathbf{u}^{(k+1)})$.
Corrector[4]:

$$\mathbf{a}^{(k+1)} := (\mathbf{M}^{-1}\mathbf{r} - c\tilde{\mathbf{v}})/(1 + \frac{1}{2}c\tau) , \tag{3.138}$$

$$\mathbf{v}^{(k+1)} := \tilde{\mathbf{v}} + \frac{1}{2}\tau\mathbf{a}^{(k+1)} . \tag{3.139}$$

The dynamic relaxation method for solving Eqn (3.133) can be stated as follows

1. Set the damping coefficient $c$, time-step $\tau$ and mass matrix $\mathbf{M}$. Set the tolerance $\epsilon_0$, $\epsilon$, and the maximum number of iteration $k^{\max}$. Set $k := 0$.

2. Apply boundary conditions and initialize $\mathbf{u}^{(k)}$. Set $\mathbf{v}^{(k)} := \mathbf{0}$, $\mathbf{a}^{(k)} := \mathbf{0}$.

3. Predictor.

4. Compute the residual force, and record $\|\mathbf{r}^{(k)}\|$. Check the convergence as follows:

[3]In practice, the intermediate velocity vector $\tilde{\mathbf{v}}$ can be stored in $\mathbf{v}^{(k+1)}$.
[4]It is possible to compute the acceleration on the fly without allocating memory. The acceleration vector is remained here to fit into the general framework of simulating dynamic responses.

If $\|\mathbf{r}^{(0)}\| < \epsilon_0$ or $\|\mathbf{r}^{(k)}\| \leq \epsilon\|\mathbf{r}^{(0)}\|$, exit with convergence;

else if $k = k^{\max}$, exit without convergence;

else continue.

5. Corrector. $k := k + 1$. Go to 3.

### Adaptive dynamic relaxation method

In order to accelerate the convergence, it is desirable to choose a time-step as large as possible while keeping the simulation stable. For linear problems, the residual force can be formulated as $\mathbf{r}(\mathbf{u}) = \mathbf{f}^{\text{ext}} - \mathbf{K}\mathbf{u}$, where the stiffness matrix $\mathbf{K}$ is assumed to be symmetric and positive definite. The stability condition for the explicit Newmark algorithm can be written as [47]

$$\tau\sqrt{\lambda^{\max}(\mathbf{M}^{-1}\mathbf{K})} \leq 2\,, \tag{3.140}$$

where $\lambda^{\max}()$ denotes the maximum eigenvalue. Since the dynamic response is only a solution strategy, the density (and correspondingly the mass matrix) and the damping coefficient do not need to represent the physical quantities. By estimating the convergence rate, Underwood [118] suggested the following expression of the damping coefficient

$$c \approx 2\sqrt{\lambda^{\min}(\mathbf{M}^{-1}\mathbf{K})}\,, \tag{3.141}$$

where $\lambda^{min}()$ denotes the minimum eigenvalue. Inequality (3.140) and Eqn (3.141) bring up two relations between $\tau$, $c$ and $\mathbf{M}$, which leaves a freedom to specify one of them. It is popular to define a constant time-step, e.g. $\tau := 1$, and then determine the diagonal mass matrix $\mathbf{M}$ to fulfill the stability requirement by invoking *Gerschgorin circle theorem* [36]:

**Theorem 3.3.1.** Assume matrix $\mathbf{A} = [A_{ij}] \in \mathbb{C}^{n\times n}$, and $\lambda(\mathbf{A})$ is the set of eigenvalues of $\mathbf{A}$. Then

$$\lambda(\mathbf{A}) \subset \bigcup_{i=1}^{n} G_i(\mathbf{A}) \tag{3.142}$$

$$\text{with } G_i(\mathbf{A}) = \left\{ z \in \mathbb{C} : |z - A_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^{n} |A_{ij}| \right\}, \quad \text{for } i = 1, 2, ..., n \ .$$

A straightforward implementation of this theorem leads to

$$\lambda^{\max}(\mathbf{M}^{-1}\mathbf{K}) \leq \max_{1 \leq i \leq \text{ndof}} \sum_{j=1}^{\text{ndof}} M_{ii}^{-1} |K_{ij}| \ , \tag{3.143}$$

where ndof is the size of the mass matrix. As a result, a diagonal mass matrix $\hat{\mathbf{M}}$ with the diagonal entry defined by

$$\hat{M}_{ii}(\mathbf{K}, \tau) := \frac{\tau^2}{4} \sum_{j=1}^{\text{ndof}} |K_{ij}| \ , \tag{3.144}$$

for $i$=1, 2, ..., ndof, satisfies the stability condition (3.140), and in this case $\lambda^{\max} \leq \frac{4}{\tau^2}$. The minimum eigenvalue in Eqn (3.141) can be estimated using Rayleigh quotient with the current state of $\mathbf{u}$ [118], i.e.

$$\lambda^{\min}(\mathbf{M}^{-1}\mathbf{K}) \approx \frac{\mathbf{u}^\tau \mathbf{K} \mathbf{u}}{\mathbf{u}^\tau \mathbf{M} \mathbf{u}} \ , \tag{3.145}$$

assuming $\mathbf{u}^\tau \mathbf{M} \mathbf{u} > 0$. Consequently, the damping coefficient can be defined as

$$\hat{c}(\mathbf{M}, \mathbf{K}) := \begin{cases} 2\sqrt{\frac{\mathbf{u}^\tau \mathbf{K} \mathbf{u}}{\mathbf{u}^\tau \mathbf{M} \mathbf{u}}} \ , & \text{if } \mathbf{u}^\tau \mathbf{M} \mathbf{u} > 0 \text{ and } \mathbf{u}^\tau \mathbf{K} \mathbf{u} > 0 \ , \\ 0 \ , & \text{otherwise} \ . \end{cases} \tag{3.146}$$

The mass matrix (3.144) and the damping coefficient (3.146) both require the knowledge of the stiffness matrix, however assembling the stiffness is computationally intensive and the storage of the stiffness matrix is memory consuming. In order to avoid this burden, only the diagonal entries of the stiffness are considered, and they can be obtained through the finite difference method, i.e.

$$\hat{K}_{ij} := \begin{cases} \left[ f_i^{\text{int}}(\mathbf{u}^{(k+1)}) - f_i^{\text{int}}(\mathbf{u}^{(k)}) \right] / (\tilde{v}_i \tau) \ , & i = j \ , \\ 0 \ , & i \neq j \ , \end{cases} \tag{3.147}$$

where $\tilde{\mathbf{v}}$ is the intermediate velocity vector in the Predictor step. In the absence of external force, i.e. no body force or surface traction, $\mathbf{r} = -\mathbf{f}^{int}$, and it is equivalent to write Eqn (3.147) as

$$\hat{K}_{ii} := \begin{cases} \left[-r_i(\mathbf{u}^{(k+1)}) + r_i(\mathbf{u}^{(k)})\right] / (\tilde{v}_i \tau) \,, & i = j \,, \\ 0 \,, & i \neq j \,, \end{cases} \tag{3.148}$$

which can save the storage for the internal force.

The discussion above leads to an adaptive dynamic relaxation method:

1. Set the tolerance $\epsilon_0$, $\epsilon$, and the maximum number of iteration $k^{\max}$. Set $k := 0$, $\tau := 1$.

2. Apply boundary conditions and initialize $\mathbf{u}^{(k)}$. Set $\mathbf{v}^{(k)} := \mathbf{0}$, $\mathbf{a}^{(k)} := \mathbf{0}$.

3. Predictor.

4. Compute the residual force, and record $\|\mathbf{r}^{(k)}\|$. Check the convergence as follows:

   If $\|\mathbf{r}^{(0)}\| < \epsilon_0$ or $\|\mathbf{r}^{(k)}\| \leq \epsilon \|\mathbf{r}^{(0)}\|$, exit with convergence;

   else if $k = k^{\max}$, exit without convergence;

   else continue.

5. Calculate the mass matrix, $\hat{\mathbf{M}}(\hat{\mathbf{K}}, 1.1\tau)$, using Eqn (3.144), and the damping coefficient $\hat{c}(\hat{\mathbf{M}}, \hat{\mathbf{K}})$ (3.146) using $\mathbf{u}^{(k+1)}$ and stiffness (3.147). If $\hat{c} \geq 4/\tau$, set $\hat{c} := 3.8/\tau$. [5]

6. Corrector. $k := k + 1$. Go to 3.

The algorithm above has an issue that the mass matrix for the first step can be undetermined because in the computation of the stiffness (3.147) the velocity as a denominator is initially zero. In practice, if the stable time-step can be easily

[5] Following the numerical treatments by Underwood [118], the mass matrix is evaluated at $1.1\tau$ instead of $\tau$ to secure the stability, and $\hat{c}$ is truncated since the sequence $c \approx 2\sqrt{\lambda^{\min}} \leq 2\sqrt{\lambda^{\max}} \leq 4/\tau$ may be violated by the finite difference approximation of the stiffness matrix.

estimated from the material model and the finite element mesh size, this stable time-step and the lumped mass matrix based on the physical density can be used instead. If it is not the case, the assembled stiffness can be used to ensure the stability under constant time-step.

There are quite a few benefits to use dynamic relaxation method. First, the explicit formulation makes it easy to program. Secondly, the vectorized formulation makes the parallelization straightforward [74, 75, 76], reduces the memory consumption, and increases the efficiency of computation as vectorized operations can be optimized by contemporary compilers and CPUs. Thirdly, it is convenient to treat the local kinematic constraints like contacts or bound constraints, which can be directly applied in the Predictor step. There is no need for specific treatments like the penalty or the line search to keep the current state of $\mathbf{u}$ admissible.

### Parallelization of the dynamic relaxation method

It is convenient to parallelize the adaptive dynamic relaxation method and speed up the computation. Based on the formulation of the adaptive dynamic relaxation method, we have tried two ways to implement the parallelization. One way is to use open multiprocessing (OpenMP), which requires small effort from programmers while leaves most of work to compilers. In this way, the data structure does not need to change, and the sequential code can be kept. Instructions (macros) are inserted whenever there are intensive vectorized operations like looping nodal fields in the Predictor and Corrector steps, so the compiler knows the work can be distributed over multiple processors of the computer. The sequential code may need to be adjusted slightly to take full advantage of this technique, e.g. changing the sequence of multi-loops. The major limitation of this OpenMP approach is that the code only runs on a shared memory machine, and the number of CPUs and the memory size impose the threshold on the benefit from parallelization. A piece of code using OpenMP for the Predictor step is posted below:

```
#ifdef WITH_OPENMP
```

```
#include <omp.h>
#endif

#ifdef WITH_OPENMP
  #pragma omp parallel for
#endif
  for (int i = 0; i < nodes; ++i) {
    for (int j = 0; j < dim; ++j) {
      if (_bc(i,j) == BC_NEUMANN) {
        u(i,j) += dt*v(i,j)+0.5*dt*dt*a(i,j);
        v(i,j) += 0.5*dt*a(i,j);
      }
      else if (_bc(i,j) == BC_DIRICHLET) {
        u(i,j) = _forces(i,j);
        v(i,j) = 0.0;
      }
    }
  }
```

Another way is to use Message Passing Interface (MPI), which requires more effort from programmers than that in the previous approach. In this approach, the data storage and the computation load can be distributed over a large number of computers, which could significantly increase the size of the problems that can be solved. In the finite element code (SUMMIT) developed by our group, the original finite element mesh for the undeformed solid is partitioned into sub-domains, i.e. partitions. Examples of partitions are shown in Fig (3-1). Each processor takes one partition and constructs locally the necessary computational data, e.g. $\mathbf{u}, \mathbf{v}, \mathbf{a}, \mathbf{M}$. The dynamic relaxation algorithm is executed locally unless it requires information from other partitions. More specifically, in the adaptive dynamic relaxation algorithm, the initialization (Step 1, 2), predictor (Step 3), and corrector (Step 6) can be completed in each partition without the information from other partitions, whereas Step 4 and 5 require the communication with neighboring partitions or even all other partitions. For nodal data like the diagonal mass matrix and the residual force, the contribution of each partition is subject to summation. The communication for the summation of nodal data is as follows: each partition sends out the locally assembled value at its boundary nodes to adjacent partitions (neighbors), receive the values from

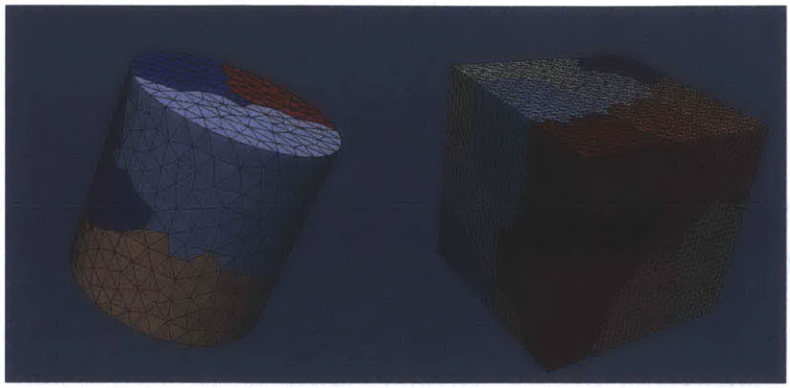

Figure 3-1: Partitioned meshes: cylinder with 4500 elements is partitioned for 4 processes, while cube with 196608 elements is partitioned for 16 processes. Different colors indicate different processes in the parallel finite element computation.

its neighbors, and add them to its local value. For the calculation of the damping coefficient and the norm of the residual force, quantities like $\mathbf{u}^\tau\mathbf{K}\mathbf{u}$, $\mathbf{u}^\tau\mathbf{M}\mathbf{u}$ and $\mathbf{r}^\tau\mathbf{r}$ from each partition must be summarized and broadcast with special care to remove duplicated counts of the nodes at the interfaces between partitions.[6]

### Dynamic relaxation method for the coupled problem

The coupled macro-force balance equation (mech) and micro-force balance equation (mpt) are solved using the staggered algorithm described in Section 3.3.3. Parameters for the two sub-problems are specified as follows:

(mech): The physical density of the material $\rho$ is used in calculating the lumped mass matrix, and the stable time-step is estimated using the finite element mesh size and the critical wave speed of the material, e.g. $\tau = h/\sqrt{\frac{\lambda+2\mu}{\rho}}$ for linear elastic responses, where $\lambda$ and $\mu$ are the Lamé constants, and $h$ is the minimum inner radius of the elements in the mesh. The damping coefficient is estimated using the approximated stiffness from finite difference.

(mpt): The stable time-step is fixed as $\tau = 1$, and the mass matrix is calculated using the assembled stiffness, which could be updated after a period of iterations in

[6]If the physical stable time-step is in use, it has to be synchronized using the minimum value of the stable time-steps from all partitions. If the mass matrix is based on the locally assembled stiffness, it can be shown through the triangle inequality that the stability condition (3.140) is valid after the communication.

order to reduce the computational expense. The damping coefficient is estimated using the approximated stiffness from finite difference.

The (mech) sub-problem is solved as follows:

1. Set the tolerance $\epsilon_0$, $\epsilon$, and the maximum number of iteration $k^{\max}$. Set $k := 0$. Calculate $\tau$, and the lumped mass matrix.

2. Compute residuals:

   Access the inelastic strain that is produced by (mpt). Compute the residual force, and record $\|\mathbf{r}^{\text{mech}(k)}\|$.

3. Check the convergence as follows:

   If $\|\mathbf{r}^{\text{mech}(k)}\| \leq \epsilon \|\mathbf{r}^{\text{mech}(0)}\|$ or $\|\mathbf{r}^{\text{mech}(0)}\| \leq \epsilon_0$ holds, exit with convergence;

   else if $k = k^{\max}$, exit without convergence;

   else continue.

4. Corrector.

5. Predictor.

6. $k := k + 1$. Go to 2.

The (mpt) sub-problem is solved as follows:

1. Set the tolerance $\epsilon_0$, $\epsilon$, and the maximum number of iteration $k^{\max}$. Set $k := 0$. $\Delta\boldsymbol{\xi}^{(k)} := 0$. Denote the current nodal field of the martensitic volume fraction $\boldsymbol{\xi}$ as $\boldsymbol{\xi}^{\text{prev}}$. Set $\tau := 1$. Calculate the mass matrix with assembled stiffness, and set a period $T$ for updating the mass matrix.

2. Compute residuals:

   Access the total strain that is produced by (mech). Recalculate the mass matrix if $n = \text{round}(n/T) \times T$. Compute the residual force and record

$$\|\mathbf{r}^{\text{mpt}(k)}\|_\star := \left[\sum_{j \in \mathscr{A}} (r_j^{\text{mpt}(k)})^2\right]^{\frac{1}{2}}, \tag{3.149}$$

where $\mathscr{A}$ is a set of nodal indices such that for any $j \in \mathscr{A}$, $\Delta\xi_j^{(k)}(= u_j^{\mathrm{mpt}(k)})$ satisfies one of the following conditions: a) $-\xi_j^{\mathrm{prev}} < \Delta\xi_j^{(k)} < 1 - \xi_j^{\mathrm{prev}}$; b) $\Delta\xi_j^{(k)} \leq -\xi_j^{\mathrm{prev}}$, and $r_j^{\mathrm{mpt}(k)} > 0$; c) $\Delta\xi_j^{(k)} \geq 1 - \xi_j^{\mathrm{prev}}$, and $r_j^{\mathrm{mpt}(k)} < 0$.

3. Check the convergence as follows:

   If $\|\mathbf{r}^{\mathrm{mpt}(0)}\|_\star \leq \epsilon_0$ or $\|\mathbf{r}^{\mathrm{mpt}(k)}\|_\star \leq \epsilon\|\mathbf{r}^{\mathrm{mpt}(0)}\|_\star$ holds, exit with convergence;

   else if $k = k^{\max}$, exit without convergence;

   else continue.

4. Corrector.

5. Predictor; and apply the constraints as follows:

   $\Delta\xi_j^{(k)} := -\xi_j^{\mathrm{prev}}$, $v_j^{\mathrm{mpt}(k)} := 0$ and $a_j^{\mathrm{mpt}(k)} := 0$, if $\Delta\xi_j^{(k)} < -\xi_j^{\mathrm{prev}}$ ;

   $\Delta\xi_j^{(k)} := 1 - \xi_j^{\mathrm{prev}}$, $v_j^{\mathrm{mpt}(k)} := 0$ and $a_j^{\mathrm{mpt}(k)} := 0$, if $\Delta\xi_j^{(k)} > 1 - \xi_j^{\mathrm{prev}}$.

6. $k := k + 1$. Go to 2.

In the computation, parameters take the following typical values: $\epsilon_0 = \epsilon = 10^{-8}$, $T = 100$, and $k^{\max}$ equals the degrees of freedom.

### Convergence study of dynamic relaxation method

We first investigate the convergence of the dynamic relaxation method for different sizes of the load increment. Consider the uniaxial compression of a pillar with height $h = 1$ m, and radius $r = 0.1$ m. The martensitic volume fraction at both ends is confined as $\xi = 0$. The displacement at the bottom of the pillar is fixed, while at the top, $u_3$ is prescribed such that $u_3/h$ decreases to $-0.05$ and then increases to 0 with load increments: $\Delta u_3/h = -0.002, -0.001, -0.0005$ in three tests, respectively. Due to the symmetry, only quarter of the pillar is considered, and 3360 quadratic tetrahedron element is used for discretization. In Fig (3-2), the nominal stress strain responses from the three tests are shown. It can be seen that the three tests have almost identical stress strain responses until the nominal strain reaches about 0.038.

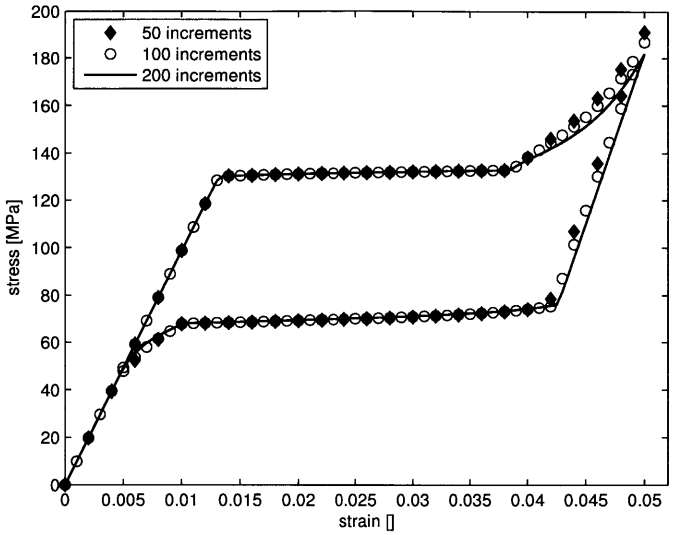


Figure 3-2: Stress-strain relations from simulations with different sizes of load increment.

Results with larger increment size then exhibit larger stiffness for the rest of loading and the elastic unloading steps. Nonetheless, the reverse transformation occurs at the same stress value, and the results converge for the rest of the unloading steps.

We then investigate the convergence of the dynamic relaxation method for spatial refinement. Consider the uniaxial compression of a bar[7] with height $h = 1$ m. The cross-section is a square with side length 0.2 m. The martensitic volume fraction at both ends is confined as $\xi = 0$. The displacement at the bottom of the bar is fixed, while at the top, $u_3$ is prescribed such that $u_3/h$ decreases to $-0.05$ and then increases to 0 with load increment $\Delta u_3/h = -0.001$. Due to the symmetry, only quarter of the bar is considered. In Fig (3-3), the nominal stress strain responses from the tests with 188, 600, and 3536 quadratic tetrahedron elements are shown. These responses are almost identical, which confirms the convergence of the method with respect to the spatial refinement.

Material parameters for these convergence tests are $E^A = 10$ GPa, $E^M = 15$ GPa, $\nu^A = \nu^M = 0.3$, $Y = 1$ MPa, $\Delta s_{\text{eq}}(T - T_{\text{eq}}) = -4$ MPa, $H^t = 0$ MPa, $S_0 = 0.1$ GPa,

[7]A bar instead of a pillar is used in this test to avoid the difference caused by the resolution of the cross-section.

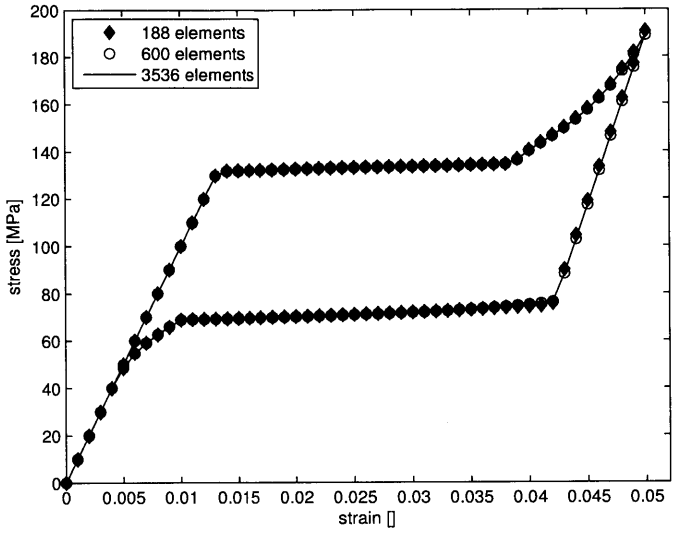


Figure 3-3: Stress-strain relations from simulations with different number of elements.

$\ell_e = 0.01$ m, $\ell_d = 0.2$ m.

### 3.3.5 Comparison of solvers

Since the superelastic loading procedure is considered quasi-static, the static solution to the coupled macro- and micro-force balance equations has to be obtained for each load increment.

The code in our group originally contained only a serial Gaussian elimination (serial GE) solver for mechanical problems (macro-force balance equation) using Cholesky decomposition and skyline storage. In order to take the full advantage of the multi-core processors in the current computers, we have tried to use a third-party solver, the Parallel Direct Sparse Solver (PARDISO) from Intel Math Kernel Library [98], and we have also implemented a parallel dynamic relaxation (DR) solver based on OpenMP. Details of the dynamic relaxation solver can be found in Section 3.3.4. These three solvers have been tested on the following mechanical problem.

**Test 1.** Assume that a cube with side length 1 m is confined at the bottom and stretched at the top: $u_i|_{x_3=0} = 0$ for $i = 1, 2, 3$; $u_i|_{x_3=1} = 0$ for $i = 1, 2$, and $u_3|_{x_3=1} = 0.1$ m. The material model is Neohookean extended to the compressible

range. The strain energy function is $w = (\frac{1}{2}\hat{\lambda}\log J - \hat{\mu})\log J + \frac{1}{2}\hat{\mu}(I_1 - 3)$, where $J = \lambda_1\lambda_2\lambda_3$ and $I_1 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2$ with $\lambda_i$ the eigenvalues of the right stretch tensor. Model parameters take the following values: $\hat{\lambda} = 122$ GPa, and $\hat{\mu} = 81$ GPa in the tests. The cube is discretized with second order tetrahedron elements. The convergence criterion is that the norm of the residual force is $10^{-8}$ of the original value. Platform of the test is: (CPU) 2 Intel Xeon X5550@ 2.67 GHz (8 cores in total); (memory) 24 GB; (operation system) Ubuntu 10.04, Linux 2.6.32; (compiler) gcc 4.4.3.

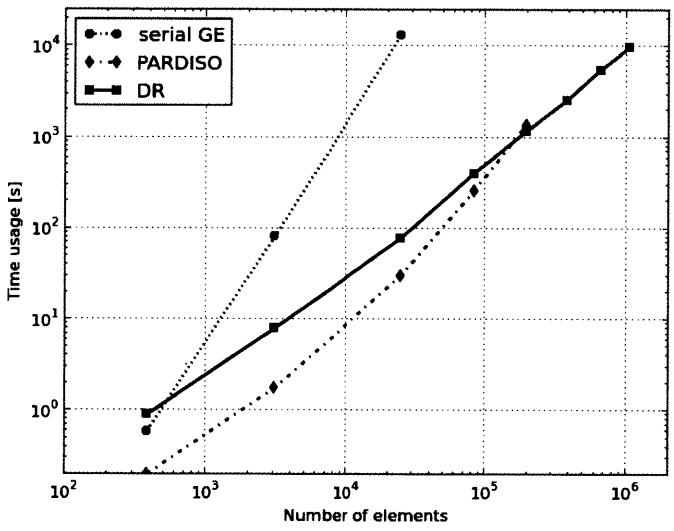


Figure 3-4: Time usage versus the number of elements for serial GE, PARDISO, dynamic relaxation solvers in Test 1.

The time usage and memory consumption as a function of the number of elements have been plotted in Fig (3-4) and (3-5), respectively. It is clear that the serial GE solver is not efficient in either the time or the memory. PARDISO is fastest up to the mesh size about 200K elements where it is surpassed by the DR solver. Both the serial GE and PARDISO solvers consume significantly more memory than the DR solver. For the largest mesh sizes, only DR solver works as its memory request does not exceed the limit of the computer. From this simple test, it has been shown that the DR method is promising for large-scale simulations. Nonetheless, the performance of the DR method using OpenMP is still limited by the number of cores and the memory

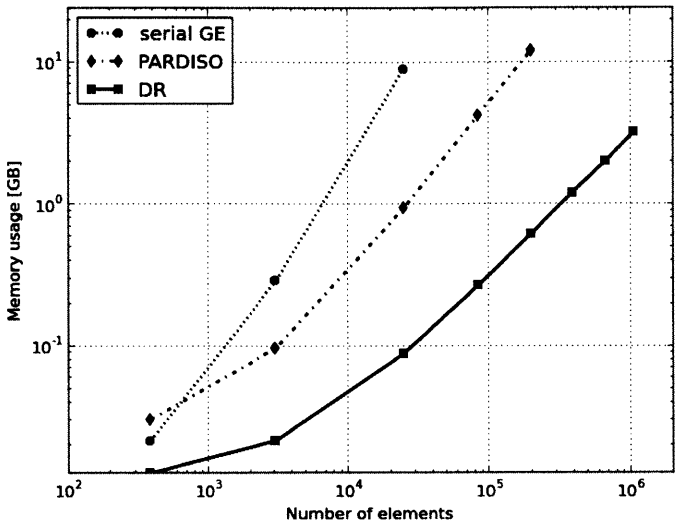


Figure 3-5: Memory usage versus the number of elements for serial GE, PARDISO, dynamic relaxation solvers in Test 1.

size in a shared-memory machine. Therefore, we have eventually implemented the distributed-memory version of the DR method using MPI (Section 3.3.4) so that the program can run on multiple computers.

We have also tested the scalability of the dynamic relaxation method using MPI for the two-field coupled problems that contain Dirichlet boundary conditions for the micro-force balance equation.

**Test 2.** Assume that two tubes with length 1 m are subject to the following boundary conditions: at one end ($x_3 = 0$ m), $\xi = 0$ and $u_i = 0$ for $i = 1, 2, 3$; at the other end ($x_3 = 1$ m), $\xi = 0$, $u_i = 0$ for $i = 1, 2$, and $u_3$ is increased to 0.05 m and then decreased to 0 m in increments of 0.001 m, i.e. 100 increments in total for the complete loading cycle. The material response is modeled by the small strain nonlocal superelasticity with the following parameters: $E^A = E^M = 10$ GPa, $\nu^A = \nu^M = 0.3$, $\bar{\varepsilon}^t = 0.04$, $\Delta s_{\text{eq}}(T - T_{\text{eq}}) = -4$ MPa, $Y = 1$ MPa, $H^t = 0$ MPa, $S_0 = 0.1$ GPa, $\ell_e = 0.03$ m, and $\ell_d = 0.1$ m. Tube A has an exterior radius 0.5 m and an interior radius 0.4 m, which is discretized using 16320 linear tetrahedron elements. Tube B has the same exterior radius but a larger interior radius 0.45 m, which is then discretized using 163200 linear tetrahedron elements. The platform of this test is

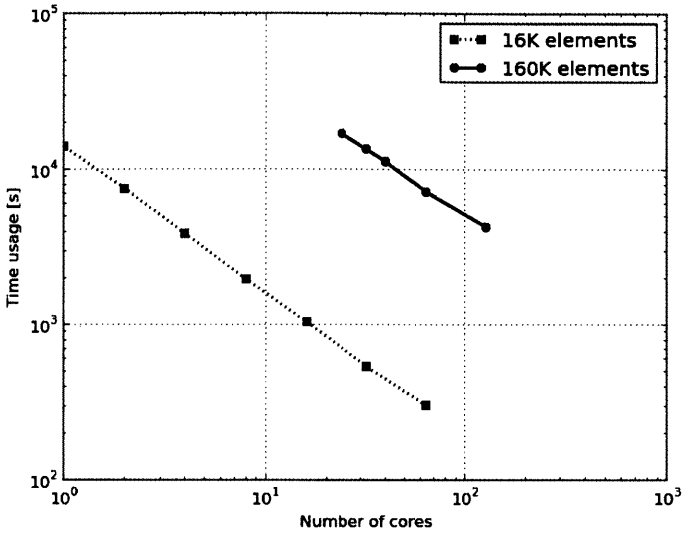


Figure 3-6: Time usage versus number of cores for dynamic relaxation method using MPI in Test 2.

our group's cluster, where each computational node contains two Intel Xeon E5520@ 2.27GHz (8 cores) and 24 GB memory, the operation system is CentOS Linux 2.6.18, and the compiler is gcc 4.1.2.

In Fig (3-6), the time usage of the test has been plotted against the number of cores in use. It can be seen that the time usage decreases as the number of cores increases in both cases. For Tube A with 16K elements, the time usage for the simulation with $k$ cores, $t^k$, can be fitted into the following scaling relation with linear regression

$$t^k \approx 1.4 \times 10^4 \, k^{-0.93} [\mathrm{s}] \, , \tag{3.150}$$

while for Tube B with 160K elements, it follows the scaling relation

$$t^k \approx 2.4 \times 10^5 \, k^{-0.83} [\mathrm{s}] \, . \tag{3.151}$$

We have also compared the dynamic relaxation method using MPI and a staggered Newton method for a two-field coupled problem. In the staggered Newton method, the macro- and micro-force balance equations are completely solved one by one at each

| | total 30 increments [s] | average of initial 12 increments [s] | average of subsequent 18 increments [s] |
|---|---|---|---|
| staggered Newton | 39589.2 | 202.8 | 2064.2 |
| DR 64 cores | 21404.9 | 225.4 | 1038.9 |
| DR 128 cores | 11998.2 | 121.4 | 585.7 |

Table 3.3: Time usage for the staggered Newton method and the dynamic relaxation method in Test 3.

iteration, and the iteration continues until the difference of the martensitic volume fraction between two consecutive iterations is significantly small. The Newton method with Armijo line search is implemented to solve the nonlinear micro-force balance equation, and PARDISO is used as a basic linear solver. The solution of the micro-force balance equation is truncated in order to satisfy the bound constraints of the martensitic volume fraction.

**Test 3.** Tube B from **Test 2** is discretized using 163200 quadratic tetrahedron elements. Material parameters and boundary conditions remain the same expect the prescribed displacement $u_3$, which is increased to 0.03 m in increments of 0.001 m, i.e. 30 increments in total. The test platform is our group's cluster, which has been described in **Test 2**. The staggered Newton method utilizes OpenMP with only eight cores, a limitation imposed by the machine.

As can be seen in Table (3.3), the staggered Newton method takes about twice as much total time as the DR method with 64 cores (DR-64) does. At the initial elastic loading (first 12 increments), the staggered Newton method takes about the same average time as DR-64 does for a single step. However, at the subsequent steps that involve the phase transformation and coupling, it takes about twice as much time as DR-64 does for a single step. The DR method with 128 cores takes about half of the time that DR-64 does, which indicates the scalability of the DR method.

# Chapter 4

# Applications of the nonlocal superelastic model

## 4.1 Grain size dependence of the stress hysteresis and the strain hardening in polycrystalline SMAs

It has been reported that both the stress hysteresis and the strain-hardening rate during phase transformations (or transformation modulus) increase with decreasing grain size in the tension tests of polycrystalline Cu-Al-Be bars [71] and polycrystalline Cu-Al-Mn wires [109]. As the first application of the three-dimensional nonlocal superelastic model, we study these grain-size dependencies of the superelasticity in polycrystalline SMAs.

### 4.1.1 Description of the numerical simulations

We have made the following assumptions of the polycrystalline SMA model: 1) The geometry of grains is described as a truncated-octahedron. 2) The growth of grains is not considered. 3) Grain boundaries are obstacles to martensitic phase transformations, and remain in the austenitic phase. In other words, the martensitic phase

transformation is isolated, and Dirichlet boundary conditions are assumed for the micro-force balance equation within each grain. 4) Elastic and superelastic responses within each grain are isotropic.

It has to be admitted that the polycrystalline model is very idealized with the assumptions above. Important crystalline responses including the anisotropy of the elastic moduli and the orientation dependence of phase transformations are not considered. The polycrystalline model is in fact a solid that consists of a set of truncated-octahedra subject to isolated martensitic phase transformations. The surface energy is not considered in the nonlocal superelastic model. The outer boundary of the material excluding grain boundaries is treated as traction free for the micro-force balance equation. Despite these limitations, we can investigate the grain boundary constraining effects in polycrystalline SMAs with the two-field formulation of the nonlocal superelastic model.

### Grain modeled by truncated-octahedron

Following [132], the grain is described as a truncated-octahedron, which can be generated from a regular octahedron by removing six square pyramids, one from each vertex. In Fig (4-1) and (4-2), we plot the top and side views of the parent octahedron and the truncated-octahedron. Assume that the side length of the parent octahedron is $3a$. The truncated-octahedron then has the following geometry: it has 32 edges of length $a$, and 14 faces (8 hexagons and 6 squares); its surface area is

$$A = 6 \times a^2 + 8 \times \frac{3\sqrt{3}}{2} a^2 = (6 + 12\sqrt{3}) a^2 \approx 27 a^2 \ ; \tag{4.1}$$

its volume is

$$V = 2 \times \frac{1}{3} (3a)^2 \frac{\sqrt{2}}{2} (3a) - 6 \times \frac{1}{3} a^2 \frac{2}{2} a = 8\sqrt{2} a^2 \approx 11 a^3 \ ; \tag{4.2}$$

the surface to volume ratio is about $2.4/a$, and the grain size $d$ equals $\sqrt{10}a$ ($\approx 3.2a$).

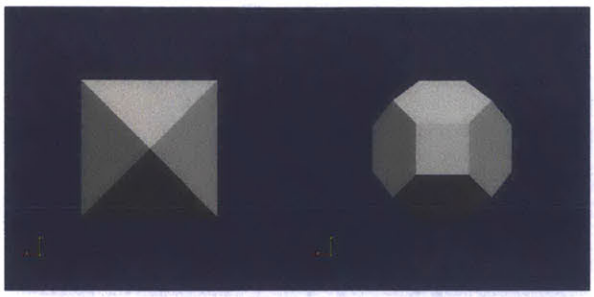

Figure 4-1: Top view of the octahedron and the truncated-octahedron.

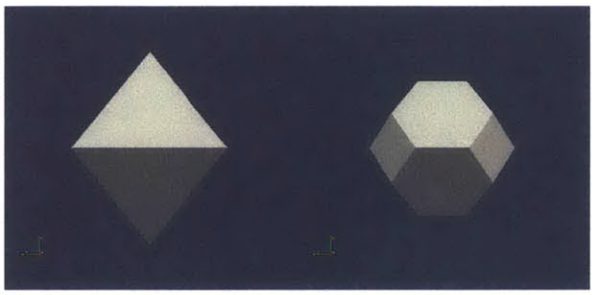

Figure 4-2: Side view of the octahedron and the truncated-octahedron.

**Grain in finite element mesh**

One benefit of the truncated-octahedron grain model is that a larger number of grains can be assembled seamlessly in a finite element mesh. More specifically, the truncated-octahedron can be easily embedded into a cube in a finite element mesh. This cube is then taken as a unit cell and replicated in space, which leads to a set of coherent grains. Figure (4-3) shows a unit cell and the truncated-octahedron that sits in it. The cube contains a complete truncated-octahedron at center, and eight incomplete truncated-octahedra, one from a corner. Grain boundaries are highlighted by the red color, which consist of the surface of the complete truncated-octahedron, and the interfaces between the incomplete truncated-octahedra. Assume that the side length of the cube is $e$. Then the edge length of the truncated-octahedron is $a = \frac{\sqrt{2}}{4}e$, and the grain size is $\sqrt{10}a = \frac{\sqrt{5}}{2}e \approx 1.1e$. Therefore, the grain size $d$ in the finite element mesh is about the same as the side length of the unit cell.

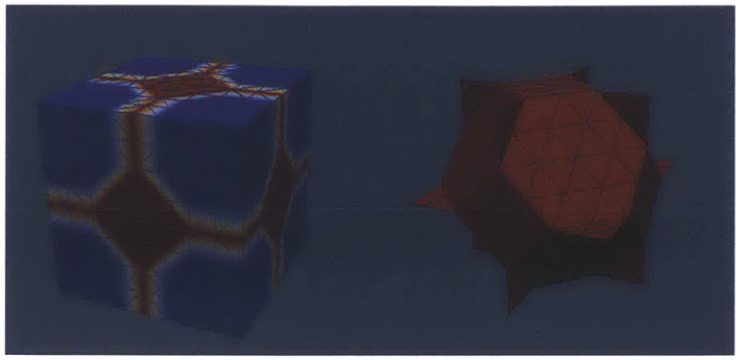

Figure 4-3: Left: a unit cell in finite element model. Right: grain boundaries of the truncated-octahedron that sits in the unit cell, and eight incomplete octahedra, one from a corner.

#### Setup of finite element simulations

A series of uniaxial stretch tests have been performed to investigate the role of grain boundary constraints in the superelastic responses of polycrystalline SMAs. The solid under concern is a cube with side length $D$. The uniaxial stretch is applied along $x_3$ direction. The boundary conditions are defined as follows: $u_i(x_1, x_2, 0) = 0$ for $i = 1, 2, 3$; $u_i(x_1, x_2, D) = 0$ for $i = 1, 2$; and $u_3(x_1, x_2, D) = \hat{u}_3$. The displacement $\hat{u}_3$ is applied in increments of $0.001D$, and it first increases up to $0.05D$ and then decreases to zero, which leads to a maximum strain 5%. $\xi = 0$ is specified at grain boundaries. The material parameters are $E^A = E^M = 10$ GPa, $\nu^A = \nu^M = 0.3$, $\bar{\varepsilon}^{\mathrm{t}} = 0.04$, $S_0 = 0.01E^A$, $\Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) = -4$ MPa, and $Y = 1$ MPa. The classic hardening term is ignored ($H^{\mathrm{t}} = 0$ MPa), so the strain-hardening effect is due to the energetic length scale and nonlocal effect.

### 4.1.2 Results and discussion

#### Evolution of stress and martensitic volume fraction from a simulation

In Fig (4-4), we plot the stress along the loading direction ($P_{33}$) and the martensitic volume fraction from a finite element simulation with $D = 1$ m, grain size $d = 0.28$ m, $\ell_{\mathrm{e}} = 0.01$ m, $\ell_{\mathrm{d}} = 0.1$ m at several representative load increments. The load increments correspond to the macroscopic strain values 1.5%, 5% during loading,

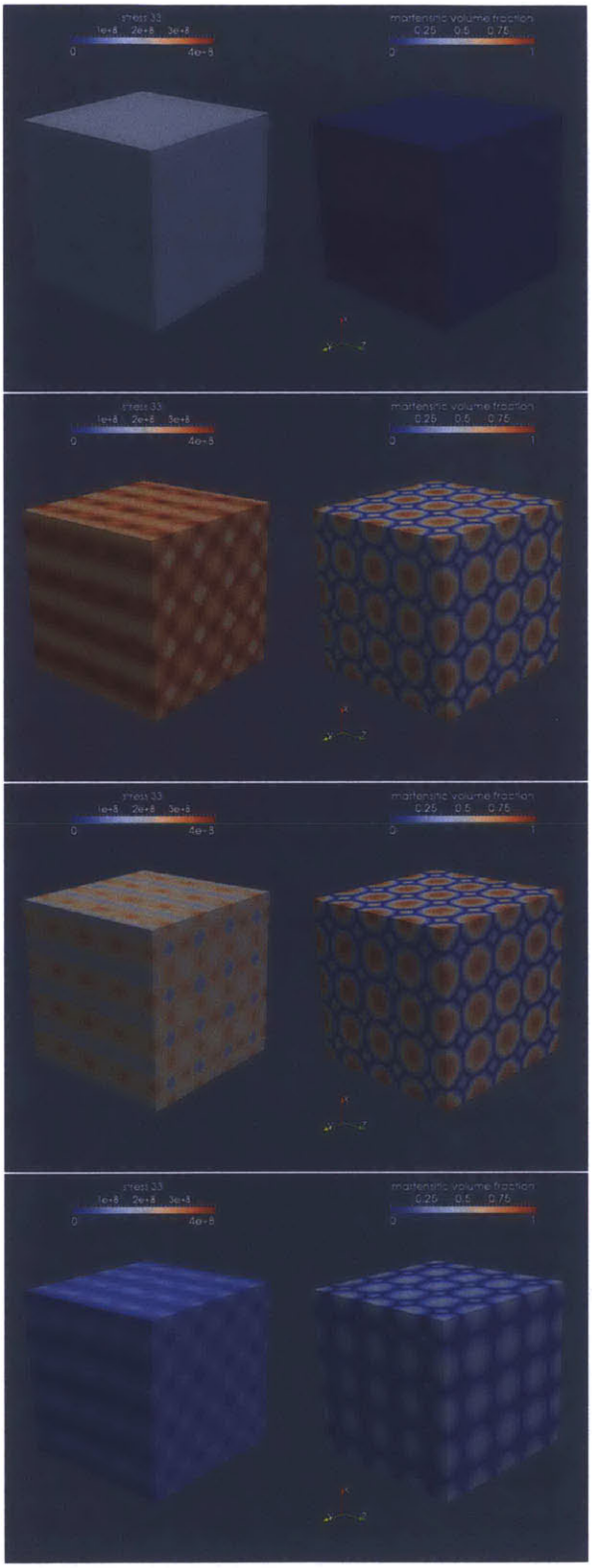


Figure 4-4: (Left) stress $P_{33}$ and (right) the martensitic volume fraction from finite element simulation at load increments 15, 50, 60, 85 (from top to bottom). ($D = 1$ m, $d = 0.28$ m, $\ell_e = 0.01$ m, $\ell_d = 0.1$ m)

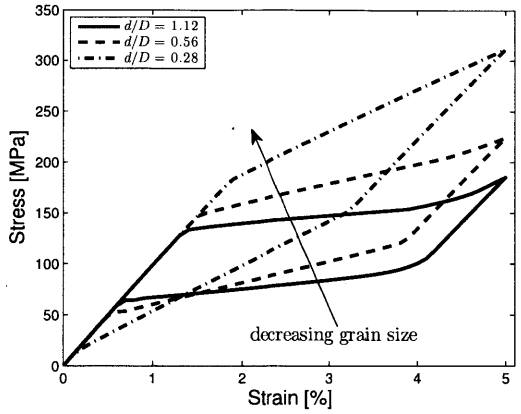


Figure 4-5: Macroscopic stress-strain relations for different grain sizes. ($D = 1$ m, $\ell_e = 0.01$ m, $\ell_d = 0.1$ m)

and 4%, 1.5% during unloading. It can be seen that at strain 1.5% during loading, the stress is uniform and the phase transformation has not happened yet. At strain 5% during loading, it is clear that the martensitic phase transformation has already happened, and it gets pinned at grain boundaries (blue color in the figure), which corresponds to our setup of the simulations. The stress is no longer uniform since the constraint on the martensitic volume fraction leads to zero phase transformation strains at grain boundaries, which distort the stress fields. At strain 4% during unloading, the martensitic volume fraction remains unchanged, and the solid responds elastically. Further unloading activates the reverse phase transformation, as shown in the last row of the figure. The macroscopic stress-strain relation corresponding to this simulation is plotted as the dashdot line in Fig (4-5).

**Effect of grain size $d$**

We first study the effect of the grain size, $d$, while keeping the solid size $D$ and the internal length scales $\{\ell_e, \ell_d\}$ fixed. Figure (4-5) shows the stress-strain curves from the simulations with three different grain sizes, i.e. $d/D$ =1.12, 0.56, and 0.28. It can be seen that the critical stress for the forward phase transformation increases for decreasing grain size; the stress-hysteresis between the forward and reverse phase transformations and the strain-hardening rate during phase transformations also in-

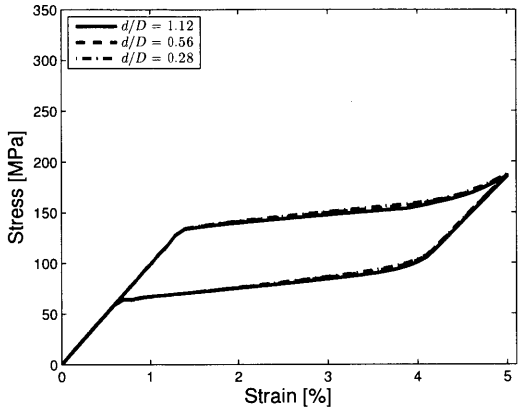


Figure 4-6: Stress strain relations showing the effect of $d/D$ with fixed $\{\ell_e/d, \ell_d/d\}$. $(10\ell_e/d = \ell_d/d = 0.089)$

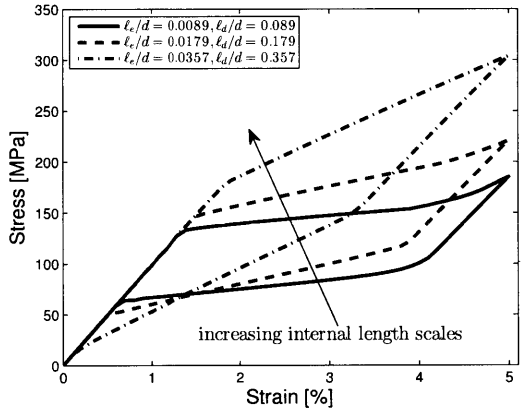


Figure 4-7: Stress strain relations showing the effect of $\{\ell_e/d, \ell_d/d\}$ with fixed $d/D$. $(d/D = 1.12)$

crease for decreasing grain size. Therefore the simulations have reproduced the size-dependent superelastic responses that were observed in the Cu-based polycrystalline SMAs [109, 71]. Nonetheless, from the modeling perspective, it is not clear whether the size effect originates from the relative grain size $d/D$ or the internal constraints $\{\ell_e/d, \ell_d/d\}$, since both of them change as the grain size $d$ changes.

### Parametric study of $d/D$, and $\{\ell_e/d, \ell_d/d\}$

In order to understand the contributions from the relative grain size $d/D$ and the internal constraints $\{\ell_e/d, \ell_d/d\}$ of the nonlocal superelastic model, another two sets of

simulations have been performed. In one set, $\{\ell_e/d, \ell_d/d\}$ is fixed, while $d/D$ changes; in the other set, $d/D$ is fixed, while $\{\ell_e/d, \ell_d/d\}$ changes. In Fig (4-6), the stress-strain curves from simulations with three different values of $d/D$ have been shown, where $\{\ell_e/d, \ell_d/d\}$ is fixed in these simulations. It can be seen that the difference between these curves is negligible, which implies that the relative grain size alone will not affect the macroscopic stress-strain relation. In Fig (4-7), the stress-strain curves from simulations with three different values of $\{\ell_e/d, \ell_d/d\}$ have been shown, where $d/D$ is fixed in these simulations. It can be seen that the stress-strain curves replicate the results shown in Fig (4-5), which means that the internal constraints for each grain are really the cause of the size-dependent superelastic responses in the polycrystalline SMA models.

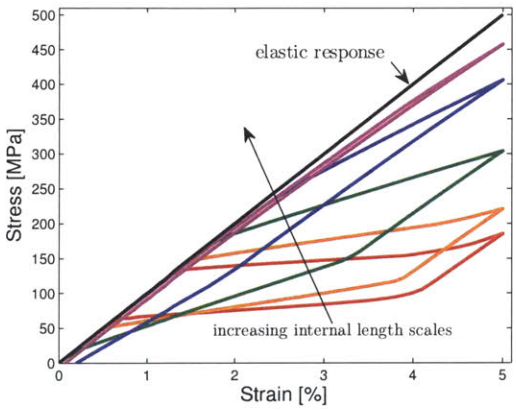


Figure 4-8: Stress-strain curves showing the saturated effect of $\{\ell_e/d, \ell_d/d\}$ with fixed $d/D$. $(d/D = 1.12)$

We have further investigated the effect of internal constraints. Fig (4-8) shows the stress-strain curves from simulations with five different values of $\{\ell_e/d, \ell_d/d\}$. It can be seen that for increasing $\{\ell_e/d, \ell_d/d\}$, the stress hysteresis first clearly increases but then decreases to almost zero, whereas the strain-hardening rate increases monotonically and approximates the purely elastic response. In Fig (4-9), we have shown quantitatively the stress hysteresis and the strain-hardening rate during phase transformations as a function of $\{\ell_e/d, \ell_d/d\}$. The stress hysteresis is measured at the strain value 2%. It can be seen that the strain-hardening rate increases monotonically,

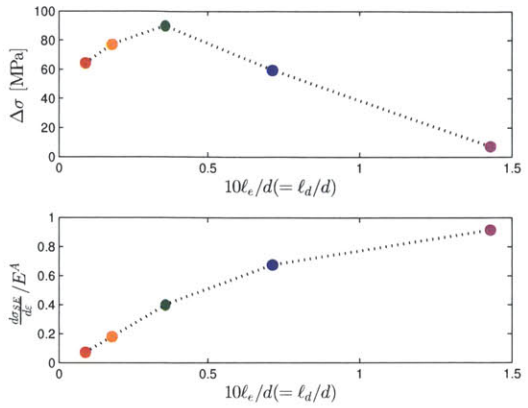


Figure 4-9: Stress hysteresis and strain-hardening rate as a function of $\{\ell_e/d, \ell_d/d\}$ with fixed $d/D$. $(d/D = 1.12)$

and eventually approaches the value of Young's modulus for increasing these ratios. The stress hysteresis first increases but then decreases to zero as these ratios become significantly large, which implies that the grain size dependence can be saturated for very small grain sizes, and the material responds almost elastically throughout the loading cycle. In reality, this limiting response is unlikely to occur since the plasticity or fracture may have already played an important role once the stress is beyond the elastic limit.

## 4.2 Loss of superelasticity in NiTi pillar compression tests

Compression tests on single crystal Ti-50.9 at.%Ni pillars have shown that the strain recovered during unloading diminishes with the pillar diameter and is suppressed for pillars with diameter smaller than about 200 nm [32]. Further study has revealed that this trend of losing superelasticity at small pillar sizes does not depend on the crystal orientations [31]. Different explanations of this size-dependent behavior have been proposed. In-situ compression tests have provided evidence of stress-induced martensitic phase transformation in NiTi pillars with diameter below 200 nm [129], which rules out the possible explanation that stress-induced martensitic phase trans-

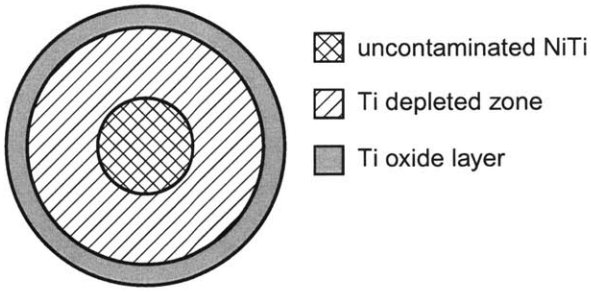


Figure 4-10: Schematic cross-section of a NiTi pillar with Ti oxide layer and Ti-depleted zone.

formation may be absent at this length scale. Focused ion beam (FIB) machining that is commonly used to prepare these small pillars will leave a $Ga^+$ implanted outer layer about 10 nm in thickness, and it is hypothesized that this damaged layer could significantly affect the mechanical response of NiTi pillars when this outer layer thickness is comparable to the pillar diameters [32]. Unfortunately, the mechanical properties of this $Ga^+$ implanted NiTi layer are not available, which prevents further quantitative investigation. Another explanation comes from the consideration of the surface Ti oxide layer [32, 97], which has been shown to constrain the thermally induced martensitic phase transformation in thin films [48, 33]. This Ti oxide layer about 15 nm in thickness does not participate in the phase transformation, and also creates a Ti-depleted zone about 50 nm in thickness [48, 124, 97], which has limited ability for the phase transformation since the increase in the Ni content stabilizes the austenitic phase [80, 111]. For very small pillars, Fig (4-10), the fixed-thickness Ti oxide layer and Ti-depleted zone take most of the pillar volume, and the suppression of superelasticity can be expected. In this section, we attempt to provide a model-based quantitative study on how this Ti oxide layer affects the mechanical behavior of NiTi pillars under compression, giving special emphasis to the size-dependent incomplete strain recovery observed experimentally.

There have been a few papers addressing incomplete strain recovery for SMAs: Yu et al. incorporated plasticity in their austenite model at high temperature when slip becomes active [130]; Yan et al. incorporated plasticity in martensite to study the stabilization of martensite due to slip [128]; Lagoudas et al. modeled the saturation

of residual strain under cyclic loading, where the plastic strain rate was assumed to be proportional to the rate of detwinned martensitic volume fraction [12, 55].

The modeling approach that we have adopted treats the NiTi pillars as a composite material comprising a uncontaminated NiTi core, and an external Ti oxide layer. We propose a nonlocal superelastic model for the NiTi core, and an elastoplastic model for the Ti oxide layer. Through Voigt-average analysis and finite element simulations, these models are used to investigate the quantitative influence of the Ti oxide layer on the mechanical responses of NiTi pillars under cyclic compression loading. The simulation results show that the plastic deformation in the Ti oxide layer constrains the recovery of deformation in the whole pillar, and the effect becomes severe with diminishing pillar size. The agreement with experimental results suggests that the size-dependent strain recovery and the loss of superelasticity in small pillars are likely to be associated with the plastic deformation in the Ti oxide layer.

### 4.2.1 Description of the NiTi-$TiO_2$ composite model

The NiTi pillar consists of a Ti oxide layer (mainly $TiO_2$ [20, 48]), a Ti-depleted zone and an uncontaminated NiTi SMA core, Fig (4-10). The Ti-depleted zone is expected to behave as a smooth transition from Ti oxide to NiTi SMA. Due to the lack of material properties for this region, we investigate the two bounding cases in which the Ti-depleted zone is either full NiTi or full $TiO_2$. The $TiO_2$ layer has, respectively, a thickness of 15 and 65 nm. Material models for the NiTi SMA and $TiO_2$ will be discussed in the following subsections.

#### NiTi SMA – nonlocal superelasticity

We assume for simplicity isotropic responses for both elastic and superelastic effects. Specifically, we ignore the dependency of the elastic moduli, the critical stresses for phase transformation, the maximum phase transformation strain, and the phase transformation strain-hardening on crystal orientation. For definiteness, we calibrate our model parameters to one specific composition and orientation. In our model, dis-

placement $\mathbf{u}$ and martensitic volume fraction $\xi$ are the two primary unknown fields, and the formulation follows the small strain nonlocal superelastic model presented in Section 3.1. As there is no clear evidence of the size-dependent stress hysteresis in experiments, we exclude the dependence of the dissipative length scale $\ell_{\mathrm{d}}$ for this application. Therefore, the macro-force balance equation reads

$$\nabla \cdot \boldsymbol{\sigma} = 0 \,, \tag{4.3}$$

and the micro-force balance equation reads

$$\begin{aligned} Y \mathrm{sign}(\dot{\xi}) = \boldsymbol{\sigma} : \Lambda^{\mathrm{t}} - \frac{1}{2}(\frac{\partial \mathscr{C}}{\partial \xi} : \boldsymbol{\varepsilon}^{\mathrm{e}}) : \boldsymbol{\varepsilon}^{e} + \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) \\ - H^{\mathrm{t}}\xi + S_0 \ell_{\mathrm{e}}^2 (\nabla \cdot \nabla)\xi \,. \end{aligned} \tag{4.4}$$

**Ti oxide – plasticity**

The $TiO_2$ layer is modeled as isotropic elastoplastic. The decomposition of the total strain tensor now reads

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^{\mathrm{e}} + \boldsymbol{\varepsilon}^{\mathrm{p}} \,, \tag{4.5}$$

where $\boldsymbol{\varepsilon}^{\mathrm{p}}$ is the plastic strain tensor. The evolution of $\boldsymbol{\varepsilon}^{\mathrm{p}}$ follows the flow rule

$$\dot{\boldsymbol{\varepsilon}}^{\mathrm{p}} = \dot{\varepsilon}^{\mathrm{p}} \boldsymbol{\Lambda}^{\mathrm{p}} \,, \tag{4.6}$$

where $\dot{\varepsilon}^{\mathrm{p}}$ denotes the equivalent plastic strain rate, and $\boldsymbol{\Lambda}^{\mathrm{p}}$ is the plastic flow direction, which takes the normality rule

$$\boldsymbol{\Lambda}^{\mathrm{p}} = \sqrt{\frac{3}{2}} \frac{\boldsymbol{\sigma}^{\mathrm{dev}}}{\|\boldsymbol{\sigma}^{\mathrm{dev}}\|} \,. \tag{4.7}$$

The constitutive relations include Hooke's law,

$$\boldsymbol{\sigma} = \mathscr{C}^{\mathrm{O}} : \boldsymbol{\varepsilon}^{\mathrm{e}} \,. \tag{4.8}$$

where $\mathscr{C}^{\mathrm{O}}$ is the elastic moduli of $TiO_2$, and the conventional $J_2$ plastic yield condition,

$$\sqrt{\frac{3}{2}}\|\boldsymbol{\sigma}^{\mathrm{dev}}\| - \bar{\sigma}_{\mathrm{y}} = 0 \,, \tag{4.9}$$

where $\bar{\sigma}_y$ is the compressive yield strength.

The plastic hardening of $TiO_2$ is ignored because it is expected to be much smaller than the strain-hardening rate of NiTi SMA. The fixed-thickness $TiO_2$ layer is supposed to dominate in the small pillars, while it has been observed that pillars with diameter smaller than 200 nm exhibit less strain-hardening than pillars with larger diameters, and the 162 nm [210] oriented pillar even shows a perfect plateau [31].

### Model parameters

The values of the SMA model parameters are determined for [111] oriented Ti-50.9at%Ni, for which the size dependence of the strain recovery is observed [32]. The elastic moduli are taken from the estimation of the corresponding bulk material with the austenite Young's modulus $E^A = 59$ GPa, and the austenite Poisson's ratio $\nu^A = 0.3$ [32]. The elastic properties of martensite are assumed to be the same as those of austenite, i.e. $E^M = E^A = 59$ GPa, $\nu^M = \nu^A = 0.3$. The equilibrium temperature $T_{\mathrm{eq}} = 200$ K, and the transformation entropy $\Delta s_{\mathrm{eq}}$ = -4.05 $\mathrm{J{\cdot}mol^{-1}{\cdot}K^{-1}}/(a_0^3 N_A)$ = -0.245 $\mathrm{MPa{\cdot}K^{-1}}$ are obtained from [111], where $a_0 = 0.3015$ nm is the lattice parameter of the austenite NiTi at room temperature [79], and $N_A$ is the Avogadro constant. The maximum transformation strain $\bar{\varepsilon}^{\mathrm{t}} = 0.036$ is obtained from [99]. The transformation resistance, $Y$, is calculated through the 1D degenerate case of the micro-force balance equation, Eq. (4.4),

$$Y\,\mathrm{sign}(\dot{\xi}) = \sigma\bar{\varepsilon}^{\mathrm{t}} + \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) - H^{\mathrm{t}}\xi \,, \tag{4.10}$$

where $\sigma$ is the stress along the loading direction. At $T = 298$ K, a stress value 800 MPa has been reported as the point in which the forward martensitic phase transformation initiates [99, 31]. $Y = 4.6$ MPa is then obtained by applying the values of the

parameters above. The hardening coefficient $H^{\mathrm{t}}$ is derived from the experimental strain-hardening rate in the following way. From Eq. (4.10), one obtains $\frac{\partial \sigma}{\partial \varepsilon}\bar{\varepsilon}^{\mathrm{t}} - H^{\mathrm{t}}\frac{\partial \xi}{\partial \varepsilon} = 0$ by taking the derivative with respect to the total strain $\varepsilon$. From Hooke's law, one obtains $\frac{\partial \sigma}{\partial \varepsilon} = E(1 - \frac{\partial \xi}{\partial \varepsilon}\bar{\varepsilon}^{\mathrm{t}})$ with assumption $E = E^A = E^M$. Combining these two equations leads to $H^{\mathrm{t}} = \frac{\partial \sigma}{\partial \varepsilon}(\bar{\varepsilon}^{\mathrm{t}})^2/(1 - \frac{1}{E}\frac{\partial \sigma}{\partial \varepsilon})$. By replacing $\frac{\partial \sigma}{\partial \varepsilon}$ with the experimentally reported value 20 GPa [31], $H^{\mathrm{t}} = 39.2$ MPa is obtained.

The group of parameters $S_0\ell_{\mathrm{e}}^2$ has the effect to enhance the strain-hardening rate for nonuniform phase transformations [90]. In this study, the values, $S_0\ell_{\mathrm{e}}^2 = 0.01$ $\mathrm{nm}^2 E^A$ and 1 $\mathrm{nm}^2 E^A$, will be adopted to study this effect.

Material parameters for $TiO_2$ including the Young's modulus $E^{\mathrm{O}} = 287$ GPa, the Poisson's ratio $\nu^{\mathrm{O}} = 0.268$, and the compressive yield strength $\bar{\sigma}_{\mathrm{y}} = 3$ GPa are obtained from [17].

### Composite Voigt-average model

In the analysis of composite materials, Voigt average, which assumes uniform strains, is commonly used to estimate the stiffness and the stresses. In this work, we also employ it to analyze the response of the composite NiTi/$TiO_2$ pillars. Consider a NiTi pillar with diameter $D$ that contains a $TiO_2$ layer with thickness $t^{\mathrm{O}}$. The strain along the loading direction $\varepsilon$ is assumed identical in the two materials. Given a strain history, the stress along the loading direction within each material, $\sigma^{\mathrm{NiTi}}$ and $\sigma^{\mathrm{O}}$, can be calculated independently using its constitutive relations. Figure (4-11) plots the stress-strain curves of NiTi SMA and $TiO_2$ during a compressive loading cycle with maximum strain 3%. Complete strain recovery and stress hysteresis in the strain-loading cycle can be observed in the response of NiTi SMA. For $TiO_2$, one can observe the typical strain-cycle response for an elastic perfectly-plastic material leading to a residual stress when the strain goes back to zero. The reaction force from the pillar cross-section, $f$, is the sum of the reaction forces from the two materials, i.e.

$$f = \pi\left(\frac{D}{2} - t^{\mathrm{O}}\right)^2 \sigma^{\mathrm{NiTi}} + \pi\left[\left(\frac{D}{2}\right)^2 - \left(\frac{D}{2} - t^{\mathrm{O}}\right)^2\right]\sigma^{\mathrm{O}}\,. \tag{4.11}$$

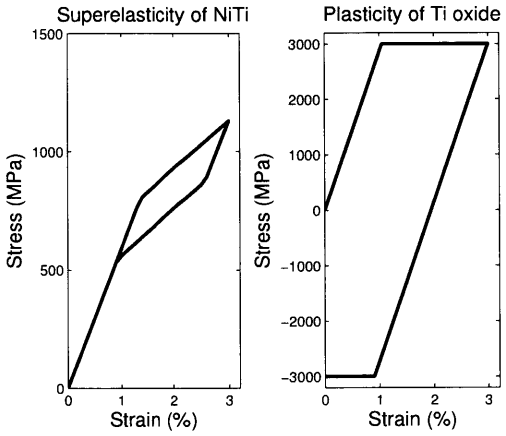


Figure 4-11: Compressive stress-strain curves of NiTi SMA and $TiO_2$ under a uniaxial loading cycle with maximum strain 3%.

And the average stress response of the composite can be obtained as follows

$$\sigma = \frac{f}{\pi \left(\frac{D}{2}\right)^2} = w\sigma^{\mathrm{NiTi}} + (1-w)\sigma^{\mathrm{O}} \,, \tag{4.12}$$

where the weight $w$ is defined as

$$w = \left[1 - 2\left(\frac{t^{\mathrm{O}}}{D}\right)\right]^2 \,. \tag{4.13}$$

The results using this model are shown in Section 4.2.2.

### Finite element simulation

The composite model presented in the previous section does not consider the interaction between the $TiO_2$ layer and the NiTi core, and in particular ignores the constraint from the $TiO_2$ layer on the martensitic phase transformation in NiTi SMA. In addition, due to the locality of the constitutive models for the $TiO_2$ plasticity and the SMA superelasticity, the homogenized approach can only capture size effects through the volume ratio of the two components but will be insensitive to a change of the spatial scale.

In order to explore the role of the interaction between the two components including gradient effects at the $TiO_2$-NiTi interface produced by the internal constraint to

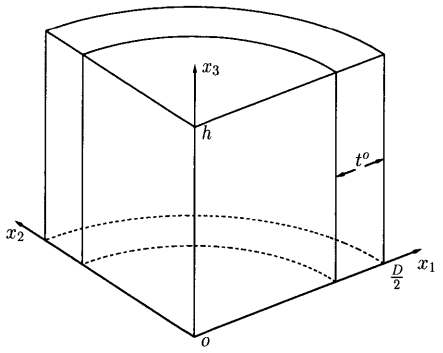


Figure 4-12: Quarter pillar for finite element calculations.

the phase transformation, three-dimensional finite element calculations are performed using the full nonlocal SMA model. The pillar is modeled as a cylinder of diameter $D$ and height $h$. Due to symmetry, only a quarter of the pillar is considered in the computation, Fig (4-12). In reality, the top surface is also covered by the $TiO_2$ layer, which could significantly affect the mechanical response if the aspect ratio $h/D$ is small. It has been reported that the aspect ratio of all samples ranges between 1.6 and 3.9 [32], although no such information for individual pillar is provided. For simplicity, we ignore this top $TiO_2$ layer, and focus on the size effect related to changes in the diameter. A fixed pillar height $h = 100$ nm is then assumed for all the pillars in the finite element simulations. Due to the gradient terms, the micro-force balance equation for NiTi SMA, Eq. (3.59), is a partial differential equation of the martensitic volume fraction, and is coupled with the macro-force balance equation, Eq. (4.3). With proper boundary conditions, these two equations for NiTi SMA, and the governing equation for $TiO_2$ (same as Eq. (4.3)), complete the formulation of the pillar compression test boundary value problem. A finite element discretization with a staggered coupled scheme is used to approximate the resulting coupled macro- and micro-force balance equations in weak form.

### 4.2.2 Results and discussion

For both the composite Voigt-average and the finite element models, the experiments are simulated as follows. Since both the superelasticity and the plasticity are history-

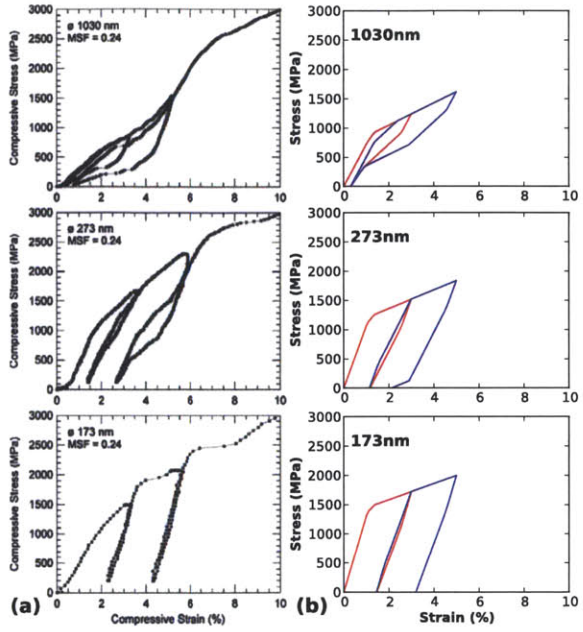


Figure 4-13: Compressive stress-strain curves from (a) experiments [31] and (b) Voigt-average analysis. Red and blue colors indicate the first and second loading cycles, respectively.

dependent, the strain history is applied in increments of 0.1%, and at each strain increment the constitutive models are integrated numerically. Following the experimental conditions, the strain is first increased to -3%, and then decreased until the reaction force becomes zero. The pillar is then reloaded to -5% strain, and unloaded until the reaction force becomes zero again. The evolution of the stress, the martensitic volume fraction (NiTi SMA) and the plastic strain ($TiO_2$) are recorded during the entire procedure. The strain history is applied at a constant temperature $T = 298$ K.

### Voigt-average model

In Fig (4-13), the stress-strain curves from the composite Voigt-average model are compared to the experimental results for pillars with diameter 1030, 273, and 173 nm. In these calculations, the $TiO_2$ layer thickness $t^{\mathrm{O}}$ is taken as 15 nm. The simulations reproduce some important features of the experimental results. First, the residual strain at the end of the first loading cycle increases significantly as the

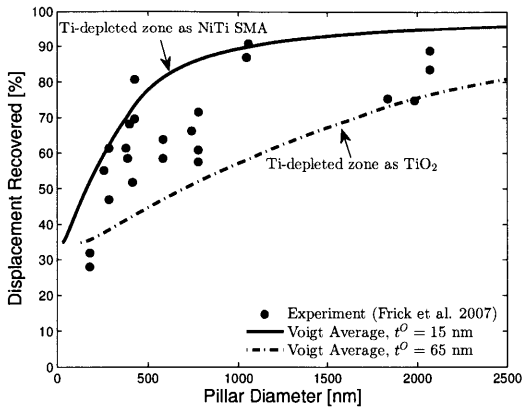


Figure 4-14: Comparison of experiments and Voigt-average of two extreme $TiO_2$ thicknesses for the displacement recovery during the first loading cycle.

pillar diameter decreases. For the smallest diameter (173 nm) there is essentially no strain recovery except for the elastic response, which indicates that in this case the superelastic effect is suppressed. By contrast, the 1030 nm pillar almost completely recovers its deformation. The medium-size pillar (273 nm) shows an intermediate response between these two limits with some strain recovery. It can also be observed that the stress hysteresis between the intermediate unloading and reloading clearly decreases as the pillar diameter decreases.

Fig (4-14) shows a summary of the experimentally-observed displacement recovery as a function of pillar diameter as well as the predictions from the Voigt model for two extreme $TiO_2$ layer thicknesses. Two values of the $TiO_2$ layer thickness, 15 and 65 nm, are used as described in Section 4.2.1. It can be seen that the model captures the decrease in the displacement recovery for decreasing pillar size and that the extreme cases in which the Ti-depleted zone is considered as full NiTi and $TiO_2$ provide nice bounds for the experimental values. For very small pillar diameters, the two $TiO_2$ layer thicknesses considered give an identical limit value of the recovered displacement $\bar{\sigma}_y/(E^O 3\%) \approx 34.8\%$, which represents the response of pure $TiO_2$.

In order to gain more insight into the model response, in Fig (4-15) we plot the stress and strain history experienced by each material component as well as the macroscopic average value as a function of load increment for the case of the

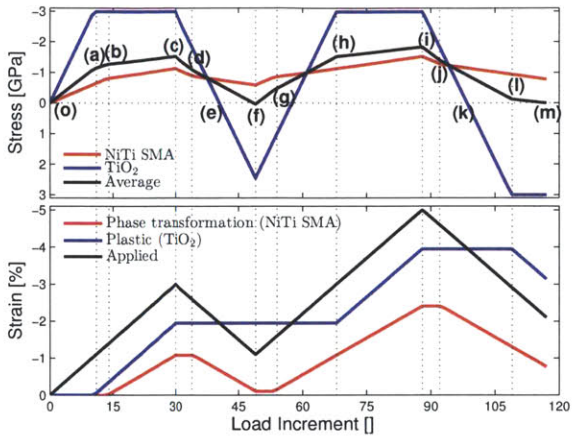


Figure 4-15: Evolution of the stresses (top) and the strains (bottom) for the 273 nm pillar from Voigt-average analysis.

273 nm pillar with $TiO_2$ thickness 15 nm. Singular points in the load history have been identified with letters to facilitate the discussion. During the first thirty load increments in which the applied strain is increased up to -3%, we can first observe the elastic loading up to point (a) ( -1% applied strain) when $TiO_2$ starts to yield plastically, followed by the onset of transformation in NiTi SMA at (b) ( -1.4%). Continued loading promotes the development of the phase transformation strain and the plastic strain until (c) where the applied strain reaches the prescribed maximum. At point (c) when unloading begins both components experience elastic unloading until (d) when the NiTi SMA starts the reverse phase transformation. It is worth noting that at point (e) during the elastic unloading the stress in $TiO_2$ vanishes before the average stress does and becomes tensile with further decrease of the applied strain. At (f), the average stress reaches zero and the first loading cycle is complete with a residual strain of about -1.1%; residual stresses of about 2.5 GPa (tensile) and -0.6 GPa (compressive) remain in $TiO_2$ and NiTi SMA, respectively; the residual plastic strain in $TiO_2$ is about -2%, whereas the residual phase transformation strain in NiTi SMA is about -0.1%. During the second loading cycle the applied strain is increased from its residual value to -5% (i). There is first elastic reloading up to (g) where the forward phase transformation begins, whereas $TiO_2$ continues to load elastically up to (h) where plastic yielding starts again. Both the phase transformation strain and

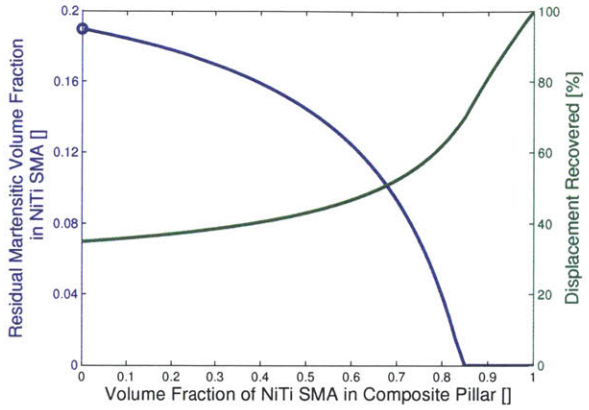


Figure 4-16: Residual martensitic volume fraction in NiTi SMA and the percentage of displacement recovery at the end of first loading cycle.

the plastic strain continue developing until the applied strain reaches the prescribed maximum at (i). Subsequent unloading from (i) proceeds elastically until the reverse phase transformation of NiTi SMA starts at (j). At (k), the stress in $TiO_2$ becomes tensile as in the first loading cycle and grows to the point (l) where plastic yielding under tension starts. At (m), the average stress eventually decreases to zero, and the second loading cycle is complete with a residual strain about -2.1%; residual stresses of about 3 GPa (tensile) and -0.78 GPa (compressive) remain in $TiO_2$ and NiTi SMA, respectively; the residual plastic strain in $TiO_2$ is about -3.2%, whereas the residual phase transformation strain in NiTi SMA is about -0.8%.

Further insights can be obtained from the Voigt-average model. For example, Fig (4-16) shows the residual martensitic volume fraction in NiTi SMA and the displacement recovery at the end of the first loading cycle as a function of the volume fraction of NiTi SMA in the composite pillar, i.e. $w$ defined in Eq. (4.13). It can be seen that the residual martensitic volume fraction decreases as $w$ increases, and eventually vanishes at $w = 0.85$. For a fixed $TiO_2$ thickness, it means that the stress-induced martensite does not fully transform back to the austenite in small pillars; the amount of the residual martensite decreases with increasing pillar size; and the reverse transformation will be complete for pillars with the volume fraction of NiTi SMA above 0.85. It can also be seen that the displacement recovery increases monotonically with

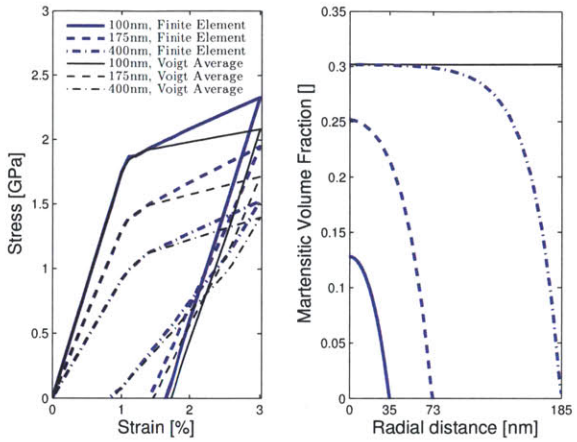


Figure 4-17: Compressive stress-strain curves of 100, 175, 400 nm pillars during the first loading cycle, and the distribution of martensitic volume fraction along pillar radius when first loaded to 3% strain.

$w$. When $w = 0$, the displacement recovery equals $\bar{\sigma}_{\mathrm{y}}/(E^{\mathrm{O}}3\%)$, which represents the pure $TiO_2$ response. When $w = 1$, the displacement recovery is 100%, which represents the pure NiTi SMA response. The curve is steepest for $w$ above 0.8, which indicates that the displacement recovery is most sensitive within this range.

### Finite element calculations

The following boundary conditions are adopted to simulate the pillar compression tests: $u_i = 0$ at $x_i = 0$ for $i = 1, 2, 3$, and $u_3 = \hat{u}_3$ at $x_3 = h$. The displacement $\hat{u}_3$ is prescribed to match the strain history in the experiments. At the $TiO_2$-NiTi interface ($(x_1^2 + x_2^2)^{\frac{1}{2}} = D/2 - t^{\mathrm{O}}$), we constrain the martensitic phase transformation with the boundary condition in the micro-force balance equation (3.43) by setting the martensite volume fraction $\xi = 0$.

In Fig (4-17), stress-strain curves extracted from finite element simulations for three pillar diameters with $TiO_2$ thickness of 15 nm and the nonlocal energetic coefficient $S_0\ell_{\mathrm{e}}^2 = 1\,\mathrm{nm}^2 E^A$ are compared with the Voigt-average results. It can be seen that the finite element model has also captured the feature of increasing residual strain for decreasing pillar size, and the residual strains predicted are very close to the Voigt-average results. It is also clear in both the finite element and Voigt-average

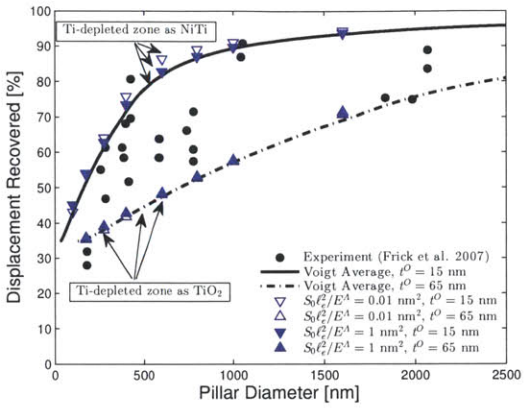


Figure 4-18: Displacement recovery during the first loading cycle extracted from finite element simulations in comparison with experiments and Voigt-average results.

results, that the apparent elastic modulus and the yield stress increase for decreasing pillar size. This can be attributed to the increasing proportion of $TiO_2$, whose Young's modulus and yield strength are larger than the Young's modulus and the critical stress of the NiTi SMA, respectively. The finite element results also show an enhanced strain-hardening compared to the Voigt-average results, as expected from the nonlocal SMA model [90]. In Fig (4-17), we also plot the martensitic volume fraction along the radial direction at the maximum applied strain. It can be seen that the Voigt model has predicted an identical value about 0.3 for the three pillar sizes, whereas the martensitic volume fraction predicted by the finite element model decreases, and the relative area of influence of the $TiO_2$-NiTi interface expands for decreasing pillar size.

In Fig (4-18), we summarize the displacement recovery at the end of the first loading cycle predicted by the finite element model for a wide range of pillar diameters, and compare it with the experiments and the predictions of the Voigt-average model. It can be seen that the displacement recovery predicted by the finite element model for the two representative values of $S_0\ell_e^2$ is very close to and sometimes even coincides with the corresponding prediction of the Voigt-average model. Since the nonlocal energy and the interaction between the NiTi SMA and $TiO_2$ are not considered in the Voigt-average model, the match suggests that these two factors have a negligible

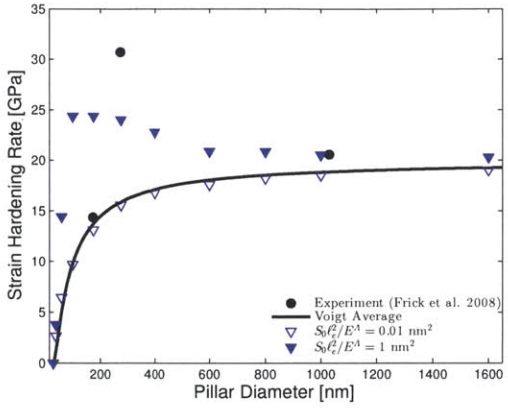


Figure 4-19: Strain-hardening rate during the phase transformation extracted from finite element simulations in comparison with experiments and Voigt-average results.

impact on the amount of the displacement recovery.

In [31], the experimental stress-strain curves have shown that the strain-hardening rate during the phase transformation is highest for medium-size pillars with diameter between 200 and 400 nm, and the differences in the strain-hardening rate are attributed to the taper shape of the individual pillar. However, this explanation has not been further quantified. In this study, we proceed to interpret the experimental observations with our model, which suggests a pillar-size dependence on the strain-hardening rate. In Fig (4-19), the strain-hardening rate extracted from the finite element simulations with $TiO_2$ thickness of 15 nm and two representative values of $S_0\ell_e^2$ are compared with those extracted from the experiments and the Voigt-average results. For the finite element simulations with $S_0\ell_e^2/E^A = 1\,\mathrm{nm}^2$, it can be seen that starting from large pillar sizes, the strain-hardening rate first increases for decreasing pillar size, and at about 200 nm it starts to decrease with further decrease in the pillar size, which is consistent with the experimental observations. It has been shown in [90] that for pure SMA, the hardening effect increases for decreasing pillar size due to the nonlocal term in the free energy and the constraint of phase transformations. However, because of the presence of the $TiO_2$ layer, in smaller and smaller pillars, the strain-hardening rate eventually drops as it approaches the perfect plastic response. For the finite element simulations with much smaller $S_0\ell_e^2$, the enhancement

of the strain-hardening rate due to the nonlocal energy is negligible, and the result simply coincides with the prediction from the Voigt-average model, which decreases monotonically with decreasing pillar size.

### Conclusions

In this work, we have proposed an approach to model the NiTi nanopillars subject to cyclic compressive loadings. The NiTi pillars have been treated as a composite material comprising a NiTi SMA core, and a $TiO_2$ outer layer, whose thickness is assumed to be fixed regardless of pillar sizes. A nonlocal superelastic material model with the gradient of the martensitic volume fraction in the free energy has been used for NiTi SMA, and an elastoplastic material model has been used for $TiO_2$. Composite Voigt-average analysis and finite element calculations have been performed to study the role of the $TiO_2$ layer in the cyclic compression tests of NiTi pillars.

Both Voigt-average and finite element simulations have captured the experimental observation of the loss of superelasticity in the small pillars. It has been shown that the plastic deformation in the $TiO_2$ layer prevents the complete strain recovery of the pillar during unloading, an effect that is more noticeable for smaller pillar sizes, i.e. as the $TiO_2$ layer takes more of the pillar volume. This results in the increase of both the residual strain and the residual martensitic volume fraction for decreasing pillar size.

The finite element simulations have also provided an explanation of the experimentally observed size dependence on the strain-hardening rate during the phase transformation, where the strain-hardening rate first increases and then decreases with decreasing pillar size. In large pillars, where NiTi SMA occupies most of the volume, the nonlocal energy together with the confinement from the $TiO_2$ layer on the phase transformation causes the increase of the strain-hardening rate for decreasing pillar size. In very small pillars, where the $TiO_2$ layer occupies relatively more volume, the response approximates the perfect plasticity, leading to the drop of the strain-hardening rate.

# Chapter 5

# Conclusions

## 5.1 Thesis contributions

In this thesis, we have developed a nonlocal continuum model to study size effects in the superelasticity of SMAs. The modeling approach combines classic superelastic models [13] with strain gradient plasticity theories [5, 56, 57]. Both the displacements and the martensitic volume fraction are considered as independent fields. Associated with the martensitic volume fraction gradient in the free energy and the rate of the martensitic volume fraction gradient in the dissipation function, two internal length scales, the energetic and the dissipative length scales, are incorporated in the model, which allow the description of size dependent stress-strain responses.

The model responses have been investigated in detail with focus on the effects of the two internal length scales in the one-dimensional case. It has been shown that the strain-hardening rate increases as the energetic length scale increases, and the stress-hysteresis increases as the dissipative length scale increases. The one-dimensional model has been applied to simulate the compression tests of Cu-Al-Ni pillars, where the increase in the stress hysteresis for deceasing pillar size has been successfully captured.

We have derived a variational incremental formulation for the coupled macro- and micro-force balance equations that result from the nonlocal model. The variational formulation has been shown to help define the convergence criteria for the numerical

simulations. Based on the variational formulation, a computational framework is formulated and implemented. In particular, a robust and scalable iterative solver (parallel dynamic relaxation) has been implemented, which enables the large-scale three-dimensional study of the size effects in SMAs with unprecedented resolution.

By applying this computational model, we have explored different examples of the size-dependent superelastic responses in SMAs. We have captured the increase of the stress hysteresis and the strain-hardening rate for deceasing grain size, and attributed them to the confining effect of grain boundaries in polycrystalline SMAs. We have also captured the loss of superelasticity in NiTi pillar compression tests using Voigt-average analysis and finite element simulations. The increase of the residual strain after unloading has been attributed to the plastic deformation of the $TiO_2$ outer layer. Finite element simulations have suggested a size-dependent strain hardening effect, which can not be seen from the Voigt-average analysis since it does not consider the interaction between NiTi and the $TiO_2$ layer. In sum, the computational model has confirmed the influence of constraints like the grain boundaries and the surface Ti oxide layer on the size-dependent superelastic responses.

As a separate effort, we have conducted analytical study of the nonlocal superelastic model. In particular, we have linked the minimization conjecture on the normalized plastic strain rate in [5] to the two-field variational incremental formulation in this thesis. We have shown theoretically for $\alpha = 1$ and numerically for $\alpha = 2$ that there is no continuous minimizer for the minimization conjecture with effective plastic strain rate $\dot{E}_{\mathrm{p}} = \left[|\dot{\varepsilon}^{\mathrm{p}}|^{\alpha} + |\ell_{\mathrm{d}}\dot{\varepsilon}^{\mathrm{p}}_{,x}|^{\alpha}\right]^{\frac{1}{\alpha}}$. In addition, a one-dimensional analytical stress-strain relation has been derived for the stress-controlled uniaxial loading cycles in the absence of the dissipative length scale.

## 5.2 Future work

The nonlocal superelastic model describes the martensitic phase transformation by a scalar, the martensitic volume fraction, and the orientation dependencies of the martensitic phase transformation are not incorporated. The model parameters in

Table 3.1 are only calibrated into specific orientations and loading directions in the simulations of single-crystal compression tests. This suggests an immediate extension of the current work, i.e. to develop a nonlocal superelastic model for single crystalline SMAs using crystallographic information. However, this extension requires significant more experimental data like the anisotropic elastic moduli, transformation systems, and the critical transformation stresses. A further extension could be to apply the single crystal model to the study of size-dependent polycrystalline responses, and especially to improve the setting of conditions for the phase transformation at grain boundaries.

Another extension could be to investigate the size-dependent shape memory effect by introducing the temperature as an independent field. A further extension along this line could be to quantify the contribution from the strain rate effect and the constraining effect like grain boundaries to the size dependencies of SMAs.

# Appendix A

# Discussion on the strain gradient plasticity formulation

In this chapter, the minimum principle proposed in Anand *et al.* [5] is linked to the variational incremental formulation in Section 3.3.1. The existence of continuous minimizers is studied for two formulations of the effective plastic strain rate. All the discussions are based on the benchmark problem: shearing of a constrained infinite layer. For simplicity, the energetic length scale $\ell_{\mathrm{e}}$ is assumed to be zero throughout this chapter.

Consider an infinite layer with thickness $h$ aligned with $y$ direction subject to shearing along $x$ direction. Denote the displacement along $x$ direction as $u$, and assume that the shear strain $\gamma$ can be decomposed as follows

$$\gamma = u_{,y} = \gamma^{\mathrm{e}} + \gamma^{\mathrm{p}} \,, \tag{A.1}$$

where $\gamma^{\mathrm{e}}$ and $\gamma^{\mathrm{p}}$ are the elastic and plastic shear strains, respectively. Assume further that the free energy density contains only the elastic part, i.e.

$$\psi(\gamma^{\mathrm{e}}) = \frac{1}{2}\mu(\gamma^{\mathrm{e}})^2 \,. \tag{A.2}$$

where $\mu$ is the shear modulus. Define $\tau$, $k$, and $k^{\mathrm{nl}}$ as the work conjugates to $\gamma^{\mathrm{e}}$,

$\gamma^{\mathrm{p}}$, and $\gamma^{\mathrm{p}}_{,x}$, respectively. The following governing equations can be derived from the principle of virtual power:

$$\tau_{,y} = 0 \,, \tag{A.3}$$

$$k - \tau - k^{\mathrm{nl}}_{,y} = 0 \,, \tag{A.4}$$

where the first equation (macro-force balance) implies that the shear stress $\tau$ is uniform through the layer thickness. By imposing the free energy imbalance requirement, Anand *et al.* [5, 41] defined the following constitutive relations

$$\tau = \mu\gamma^{\mathrm{e}} \,, \tag{A.5}$$

$$k = \tau^0_Y \frac{\dot{\gamma}^{\mathrm{p}}}{\dot{E}^{\mathrm{p}}} \,, \tag{A.6}$$

$$k^{\mathrm{nl}} = \ell^2_{\mathrm{d}} \tau^0_Y \frac{\dot{\gamma}^{\mathrm{p}}_{,y}}{\dot{E}^{\mathrm{p}}} \,, \tag{A.7}$$

where $\tau^0_Y$ is the yield (shear) strength, and $\dot{E}^{\mathrm{p}}$ is the effective plastic strain rate, which is defined as

$$\dot{E}^{\mathrm{p}} = \left[(\dot{\gamma}^{\mathrm{p}})^2 + (\ell_{\mathrm{d}}\dot{\gamma}^{\mathrm{p}}_{,y})^2\right]^{\frac{1}{2}} \tag{A.8}$$

with $\ell_{\mathrm{d}}$ the dissipative length scale. Resulting from the constitutive relations, the energy dissipation rate can be expressed as

$$\mathcal{D} = k\dot{\gamma}^{\mathrm{p}} + k^{\mathrm{nl}}\dot{\gamma}^{\mathrm{p}}_{,y} = \tau^0_Y \dot{E}^{\mathrm{p}} \,, \tag{A.9}$$

which is always non-negative, and the micro-force balance equation (A.4) can be rewritten as

$$\frac{\dot{\gamma}^{\mathrm{p}}}{\dot{E}^{\mathrm{p}}} - \ell^2_{\mathrm{d}} \frac{\partial}{\partial y}\left(\frac{\dot{\gamma}^{\mathrm{p}}_{,y}}{\dot{E}^{\mathrm{p}}}\right) = \tau \,. \tag{A.10}$$

## A.1 Minimization conjecture and variational incremental problem

Anand *et al.* [5] proposed a conjecture that the problem

$$\text{Inf } \mathscr{F}(\eta) := \frac{\tau_Y^0}{h} \int_0^h \left[\eta^2 + (\ell_{\mathrm{d}}|\eta_{,y}|)^2\right]^{\frac{1}{2}} \, dy \tag{A.11}$$

with $\eta$ subject to $\eta(0) = \eta(h) = 0$ and $\frac{1}{h}\int_0^h \eta \, dy = 1$, has a minimum, and the minimizing field $\eta^*$ satisfies Eqn (A.10) with the flow stress $\tau = \mathscr{F}(\eta^*)$.

We propose a minimization principle based on the incremental functional of the coupled fields, and link it to the conjecture above. After time discretization, the incremental problem for a generic increment $[t^{(n)}, t^{(n+1)}]$ can be stated as

$$\text{Inf } \mathscr{J}(\Delta u, \Delta\gamma^{\mathrm{p}}) := \int_0^h \frac{\mu}{2}(\Delta u_{,y} - \Delta\gamma^{\mathrm{p}} + u_{,y}^{(n)} - \gamma^{\mathrm{p}(n)})^2 + \tau_Y^0 \left[(\Delta\gamma^{\mathrm{p}})^2 + (\ell_{\mathrm{d}}|\Delta\gamma_{,y}^{\mathrm{p}}|)^2\right]^{\frac{1}{2}} \, dy \tag{A.12}$$

with $\Delta u$ and $\Delta\gamma^{\mathrm{p}}$ subject to the boundary conditions

$$\Delta u(0) = -\Delta L/2 \,, \quad \Delta u(h) = \Delta L/2 \,, \tag{A.13}$$

$$\Delta\gamma^{\mathrm{p}}(0) = \Delta\gamma^{\mathrm{p}}(h) = 0 \,, \tag{A.14}$$

where $\Delta L$ is the applied displacement increment. The Euler-Lagrange equations of the incremental functional $\mathscr{J}$ can be expressed as

$$\Delta u_{,yy} - \Delta\gamma_{,y}^{\mathrm{p}} + u_{,yy}^{(n)} - \gamma_{,y}^{\mathrm{p}(n)} = 0 \,, \tag{A.15}$$

$$-\mu(\Delta u_{,y} - \Delta\gamma^{\mathrm{p}} + u_{,y}^{(n)} - \gamma^{\mathrm{p}(n)}) + \tau_Y^0 \frac{\Delta\gamma^{\mathrm{p}}}{\Delta E^p} - \tau_Y^0 \ell_{\mathrm{d}}^2 \frac{\partial}{\partial y}\left(\frac{\Delta\gamma_{,y}^{\mathrm{p}}}{\Delta E^p}\right) = 0 \,, \tag{A.16}$$

where $\Delta E^p := \left[(\Delta\gamma^{\mathrm{p}})^2 + (\ell_{\mathrm{d}}|\Delta\gamma_{,y}^{\mathrm{p}}|)^2\right]^{\frac{1}{2}}$ .

Assume that the two-field problem has a minimum with minimizing fields ($\Delta u^*$,

$\Delta\gamma^{p*}$). Then the first Euler-Lagrange equation implies that the flow stress defined as

$$\tau^* := \mu(\Delta u^*_{,y} - \Delta\gamma^{p*} + u^{(n)}_{,y} - \gamma^{p(n)}) \tag{A.17}$$

is uniform through the layer thickness, and the second Euler-Lagrange equation states that

$$\tau^* = \tau^0_Y \frac{\Delta\gamma^{p*}}{\Delta E^{p*}} - \tau^0_Y \ell^2_d \frac{\partial}{\partial y}\left(\frac{\Delta\gamma^{p*}_{,y}}{\Delta E^{p*}}\right) \tag{A.18}$$

with $\Delta E^{p*} := \left[(\Delta\gamma^{p*})^2 + (\ell_d|\Delta\gamma^{p*}_{,y}|)^2\right]^{\frac{1}{2}}$. Using the boundary conditions and integrating by parts, we can obtain the following important equality

$$\tau^* \int_0^h \Delta\gamma^{p*}\, dy = \tau^0_Y \int_0^h \Delta E^{p*}\, dy\,. \tag{A.19}$$

On the one hand, if $(\Delta u^*, \Delta\gamma^{p*})$ is known, we can define the following field

$$\zeta := \frac{h}{\int_0^h \Delta\gamma^{p*}\, dy}\Delta\gamma^{p*}\,. \tag{A.20}$$

It is straightforward to show that $\zeta$ satisfies $\zeta(0) = \zeta(h) = 0$, $\frac{1}{h}\int_0^h \zeta\, dy = 1$, $\tau^* = \mathscr{F}(\zeta)$, and $(\zeta, \tau^*)$ satisfies the micro-force balance equation (A.10). Therefore field $\zeta$ satisfies both the kinematic constraints and the Euler-Lagrange equation for the minimization problem (A.11).

On the other hand, if minimizer $\eta^*$ for the problem (A.11) is known, we can define the plastic strain increment

$$\varphi := \lambda\eta^*\,, \tag{A.21}$$

with the scalar multiplier defined as

$$\lambda := \frac{1}{h}\left(u^{(n)}(h) - u^{(n)}(0) + \Delta L - \int_0^h \frac{\mathscr{F}(\eta^*)}{\mu} + \gamma^{p(n)}\, dy\right)\,; \tag{A.22}$$

and also the displacement increment

$$v := \int_0^y \frac{\mathscr{F}(\eta^*)}{\mu} + \gamma^{p(n)} + \varphi\, dy + u^{(n)}(0) - \frac{\Delta L}{2} - u^{(n)}(y)\,. \tag{A.23}$$

It is straightforward to verify that $v$ and $\varphi$ satisfy the boundary conditions and the Euler-Lagrange equations for the two-field minimization problem (A.12). It then follows immediately that

$$\mu\left(v_{,y} + u^{(n)}_{,y} - \varphi - \gamma^{\mathrm{p}(n)}\right) = \mathscr{F}(\eta^*) \,. \tag{A.24}$$

which means that the flow stress for the two-field problem is also $\mathscr{F}(\eta^*)$.

As a result, if the minimizing fields exist and are sufficiently characterized by the Euler-Lagrange equations in each problem, the solution to the two-field problem can be obtained by solving the one-field problem and then utilizing the displacement boundary conditions. In particular, the flow stress in the two-field problem (A.17) is independent of the displacement boundary conditions, which implies that the stress-elongation curve from the two-field problem will be a perfect plateau after plastic yielding, and the yield strength is simply the minimum of the one-field minimization problem.

## A.2 Effective plastic strain rate and existence of minimizer

The effective plastic strain rate (A.8) can be generalized as

$$\dot{E}^{\mathrm{p}} = \left[(\dot{\gamma}^{\mathrm{p}})^{\alpha} + (\ell_{\mathrm{d}}|\dot{\gamma}^{\mathrm{p}}_{,y}|)^{\alpha}\right]^{\frac{1}{\alpha}} \tag{A.25}$$

with parameter $\alpha > 0$.

For $\alpha = 1$, the constitutive equations can be expressed as

$$k = \tau_Y^0 \mathrm{sgn}(\dot{\gamma}^{\mathrm{p}}) \,, \tag{A.26}$$

$$k^{\mathrm{nl}} = \ell_{\mathrm{d}} \tau_Y^0 \mathrm{sgn}(\dot{\gamma}^{\mathrm{p}}_{,y}) \,, \tag{A.27}$$

which are associated with the rate-independent limit of the strain-gradient isotropic

viscoplastic theory by Lele and Anand [56, 57]. Similarly to Problem (A.11), we propose the following minimization problem:

$$\text{Inf } \mathscr{G}(\eta) := \frac{\tau_Y^0}{h} \int_0^h |\eta| + \ell_\text{d} |\eta_{,y}| \, dy \tag{A.28}$$

with $\eta$ subject to $\eta(0) = \eta(h) = 0$ and $\frac{1}{h}\int_0^h \eta \, dy = 1$. This problem has an infimum $\tau_Y^0(1+2\ell_\text{d}/h)$, but does not have any continuous minimizer, which is proved as follows.

For any continuous function $\eta$, the kinematic requirements implies that $\eta^{\max} := \max_{y\in[0,h]} \eta > 1$. Assume that $\eta$ achieves the maximum value $\eta^{\max}$ at point $\hat{y}$. Then

$$\begin{aligned}
\mathscr{G}(\eta) &= \frac{\tau_Y^0}{h} \int_0^h |\eta| \, dy + \frac{\tau_Y^0 \ell_\text{d}}{h} \int_0^h |\eta_{,y}| \, dy \\
&\geq \frac{\tau_Y^0}{h} \int_0^h \eta \, dy + \frac{\tau_Y^0 \ell_\text{d}}{h} \left( \int_0^{\hat{y}} \eta_{,y} \, dy + \int_{\hat{y}}^h -\eta_{,y} \, dy \right) \\
&= \tau_Y^0 + \frac{\tau_Y^0 \ell_\text{d}}{h} (\eta^{\max} + \eta^{\max}) \\
&> \tau_Y^0 (1 + \frac{2\ell_\text{d}}{h}) \, ,
\end{aligned} \tag{A.29}$$

which leads to a lower bound of the infimum:

$$\text{Inf } \mathscr{G}(\eta) \geq \tau_Y^0 (1 + \frac{2\ell_\text{d}}{h}) \, . \tag{A.30}$$

We can construct a series of continuous functions

$$\eta_n(y) := \begin{cases} \frac{n\bar{\eta}}{h} y \, , & 0 \leq y \leq \frac{h}{n} \\ \bar{\eta} \, , & \frac{h}{n} < y < h(1 - \frac{2}{n}) \\ \frac{n\bar{\eta}}{h}(h - y) \, , & h(1 - \frac{2}{n}) \leq y \leq h \end{cases} \tag{A.31}$$

for any integer $n \geq 2$, where $\bar{\eta} := \frac{1}{(1-\frac{1}{n})}$ is also the maximum value of function $\eta_n$. It is straightforward to verify that $\eta_n$ with $n \geq 2$ satisfies the kinematic requirements and also

$$\mathscr{G}(\eta_n) = \tau_Y^0 [1 + \frac{2\ell_\text{d}}{h(1 - \frac{1}{n})}] \, , \tag{A.32}$$

which leads to an upper bound of the infimum:

$$\text{Inf}\, \mathscr{G}(\eta) \leq \lim_{n \to +\infty} \mathscr{G}(\eta_n) = \tau_Y^0 (1 + \frac{2\ell_\text{d}}{h}) \,. \tag{A.33}$$

The lower bound, Eqn (A.30), and the upper bound, Eqn. (A.33) are identical, which means that $\tau_Y^0(1 + \frac{2\ell_\text{d}}{h})$ is the infimum of the minimization problem. However, this infimum can not be achieved by any continuous function because of the strict inequality, Eqn (A.29).

For $\alpha = 2$, Anand *et al.* [5] obtained an upper bound for the infimum of the one-field constrained minimization problem (A.11):

$$\text{Inf}\, \mathscr{F}(\eta) \leq \tau_Y^0 (1 + \frac{2\ell_\text{d}}{h}) \,. \tag{A.34}$$

Consider the existence of the minimizer. The integrand of functional $\mathscr{F}$ satisfies

$$\left[\eta^2 + (\ell_\text{d} |\eta_{,y}|)^2\right]^{\frac{1}{2}} \geq \ell_\text{d} |\eta_{,y}| \,, \tag{A.35}$$

so $\mathscr{F}$ is coercive in Sobolev space $W^{1,1}$ [24]. Nonetheless, this space is not reflexive, and the classic existence theorem of the direct methods in the calculus of variation is not applicable [21, 24]. In [21], Dacorogna showed that there is no continuous minimizer for the same functional with $\eta$ subject to $\eta(0) = 0$, and $\eta(h) = 1$. Further theoretical study of the existence of the minimizer is not pursued in this thesis. Instead, we consider a numerical approximation of the problem using the finite difference method described in [35]. The interval $[0, h]$ is divided into $n + 1$ equal subintervals by points

$$y_k = k\Delta y \,, \quad \text{for } k = 0, ..., n+1, \tag{A.36}$$

with $\Delta y = h/(n+1)$. The function $\eta(y)$ is approximated by a piece-wise linear function taking value $\eta_k$ at each turning point $y_k$. Then Problem (A.11) is approximated

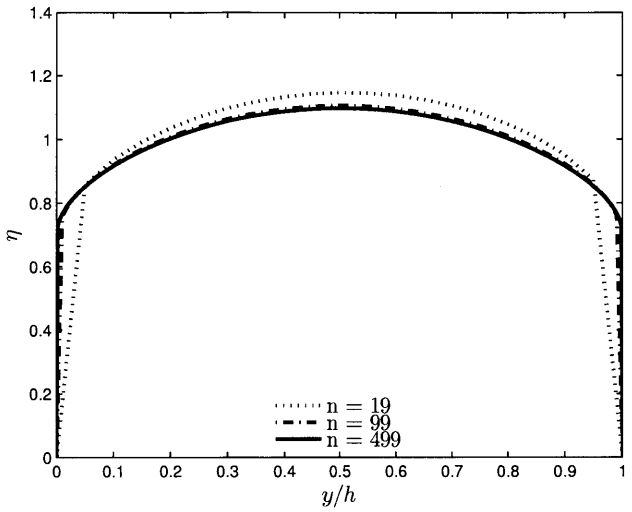


Figure A-1: Profile of the minimizer for $\ell_d/h = 1$.

by the following finite dimensional optimization problem:

$$\text{Inf } \mathscr{F}_h(\eta_0, \eta_1, ..., \eta_{n+1}) := \tau_Y^0 \sum_{k=0}^{n} \sqrt{(\eta_k \Delta y)^2 + \ell_d^2 (\eta_{k+1} - \eta_k)^2}\ , \tag{A.37}$$

with $(\eta_0, \eta_1, ..., \eta_{n+1})$ subject to $\eta_0 = \eta_{n+1} = 0$ and $\sum_{k=0}^{n} \eta_k \Delta y = 1$. The discretized problem is then solved by MATLAB constrained minimization function 'fmincon' [67] with the gradient of $\mathscr{F}_h$ provided. $h = 1$ is assumed in the numerical tests.

In Fig (A-1), the profiles of the minimizer from three numerical approximations are plotted for $\ell_d/h = 1$. It is clear that at $y = 0$ and $y = h$, there are sharp boundary layers, and the thickness of these boundary layers decreases as $n$ increases. In the case of $n = 499$, i.e. 500 subintervals, the boundary layers almost vanish, leading to jumps of the function value at the two ends. These numerical results suggest that the original problem (A.11) does not have any continuous minimizer. Meanwhile, the minimum of the numerical problem is convergent as the number of degrees of freedom $(n)$ increases, which can be seen in Fig (A-2).

In Fig (A-3), profiles of the minimizer are plotted for different dissipative length scales. It can be seen that as $\ell_d/h$ increases, the jump of the function value at the two ends increases, whereas the total variation of the function decreases. In Fig (A-4),

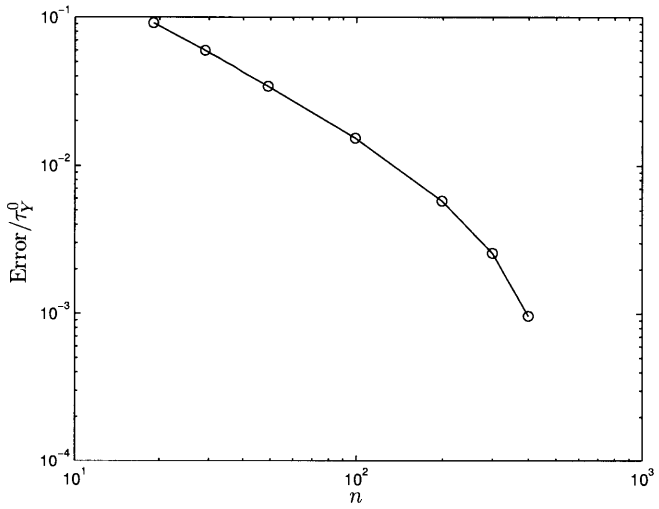


Figure A-2: Difference between the minimum obtained with $n+1$ subintervals and the minimum obtained with 500 subintervals. ($\ell_{\mathrm{d}}/h = 1$)

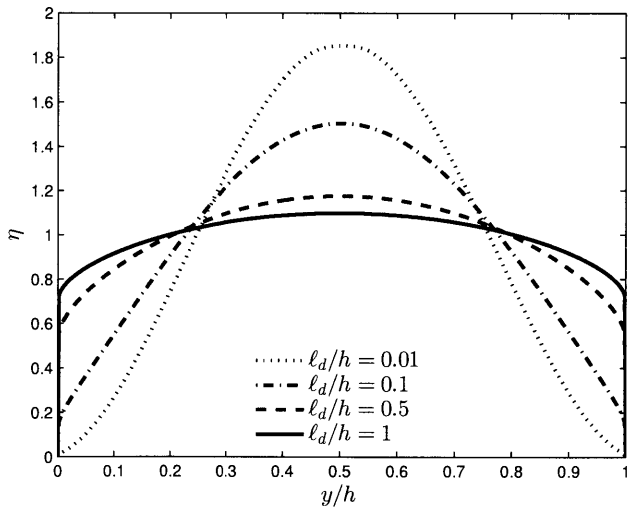


Figure A-3: Profile of the minimizer for different values of $\ell_{\mathrm{d}}/h$. (n=499)

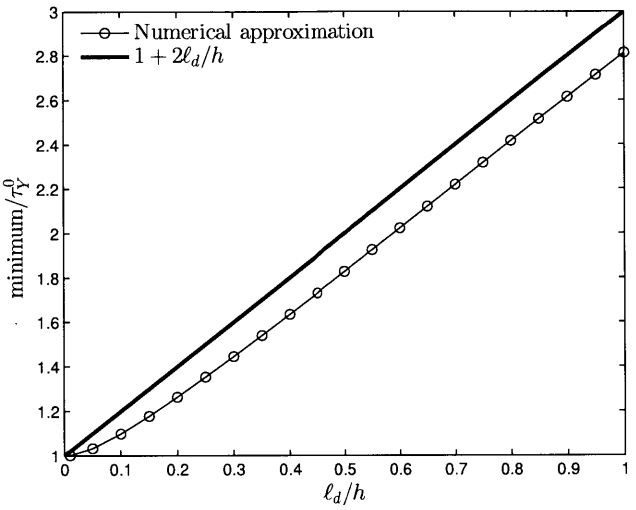


Figure A-4: Minimum of Problem (A.37) as a function of $\ell_d/h$ compared with the upper bound $1 + 2\ell_d/h$. (n=499)

the minimum of Problem (A.37) as a function of $\ell_d/h$ is compared with the upper bound $1 + 2\ell_d/h$ obtained in [5]. It can be seen that this upper bound is not tight for the range of $\ell_d/h$ shown, and the numerical results suggest that the infimum of the original problem (A.11) is a nonlinear function of $\ell_d/h$.

# Appendix B

# One dimensional analytical solution for energetic hardening

When the one-dimensional nonlocal superelastic model contains only the energetic length scale, it is possible to obtain an analytical expression for the stress-strain relation. To simply the problem, we further assume that the elastic moduli are the same for the austenite and the martensite.

## B.1 Uniaxial tension of a single crystal

The macro-force balance equation, Eqn (2.6), implies that the stress $\sigma$ is uniform. Since the dissipative length scale is zero, the micro-force balance equation, Eqn (2.15), can be written as

$$\sigma\bar{\varepsilon}^{\mathrm{t}} - B_\theta + S_0\ell_{\mathrm{e}}^2\xi_{,xx} = Y\mathrm{sgn}(\dot{\xi}) \ , \tag{B.1}$$

where $B_\theta := -\Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}})$ is defined to simplify the notation. Eqn (B.1) can be viewed as a nonlocal yield condition, where the threshold is $\pm Y$ for the forward and reverse phase transformations, respectively. Assume that the length of this single crystal is $h$, and that the ends of the crystal are obstacles for phase transformations, i.e.

$$\xi(0) = \xi(h) = 0. \tag{B.2}$$

Also, the martensitic volume fraction should satisfy the constraints

$$0 \leq \xi \leq 1 \,. \tag{B.3}$$

The strain $\varepsilon = \sigma/E + \xi\bar{\varepsilon}^{\mathrm{t}}$ is not uniform during the phase transformation, as the phase transformation strain $\xi\bar{\varepsilon}^{\mathrm{t}}$ is pinned at the ends. Therefore, we define the following average strain

$$\tilde{\varepsilon} := \frac{1}{h}\int_0^h \varepsilon \, dx \,, \tag{B.4}$$

and explore the relation between $\sigma$ and $\tilde{\varepsilon}$. It can be seen that the following equation

$$\sigma = E\left(\tilde{\varepsilon} - \frac{\bar{\varepsilon}^{\mathrm{t}}}{h}\int_0^h \xi \, dx\right) \tag{B.5}$$

holds by applying the definition (B.4) and integrating Hooke's law $\sigma = E(\varepsilon - \xi\bar{\varepsilon}^{\mathrm{t}})$ over the length of the crystal.

Consider a stress-controlled loading cycle where the applied stress $\sigma$ first increases to a prescribed maximum value, and then decreases to zero. At first, the material deforms elastically until the forward phase transformation occurs, i.e.

$$\tilde{\varepsilon} = \frac{\sigma}{E} \,, \tag{B.6}$$

for $0 \leq \sigma \leq \frac{Y+B_\theta}{\bar{\varepsilon}^{\mathrm{t}}}$.

Continued loading promotes the development of the phase transformation strain. Before the bound constraint $\xi \leq 1$ is active, the distribution of the martensitic volume fraction can be found by solving the ordinary differential equation

$$\sigma\bar{\varepsilon}^{\mathrm{t}} - B_\theta + S_0\ell_{\mathrm{e}}^2\xi_{,xx} = Y \tag{B.7}$$

with boundary conditions $\xi(0) = \xi(h) = 0$. It is straightforward to obtain that

$$\xi = \frac{Y + B_\theta - \sigma\bar{\varepsilon}^{\mathrm{t}}}{2S_0\ell_{\mathrm{e}}^2}\left[(x - \frac{h}{2})^2 - \frac{h^2}{4}\right] \,, \tag{B.8}$$

and $\xi$ takes its maximum $\frac{(-Y-B_\theta+\sigma\bar{\varepsilon}^{\mathrm{t}})h^2}{8S_0\ell_{\mathrm{e}}^2}$ at $h/2$. As a result,

$$\tilde{\varepsilon} = \frac{\sigma}{E} - \frac{h^2(Y+B_\theta-\sigma\bar{\varepsilon}^{\mathrm{t}})}{12S_0\ell_{\mathrm{e}}^2}\bar{\varepsilon}^{\mathrm{t}} \ , \tag{B.9}$$

for $\frac{Y+B_\theta}{\bar{\varepsilon}^{\mathrm{t}}} < \sigma \leq \frac{Y+B_\theta+8S_0\ell_{\mathrm{e}}^2/h^2}{\bar{\varepsilon}^{\mathrm{t}}}$. At this stage, the strain-hardening rate is constant, which can be expressed as

$$\frac{\partial\sigma}{\partial\tilde{\varepsilon}} = \left[\frac{1}{E} + \frac{1}{12S_0}(\frac{h\bar{\varepsilon}^{\mathrm{t}}}{\ell_e})^2\right]^{-1} \ . \tag{B.10}$$

Continued loading activates the bound constraint $\xi \leq 1$, and the distribution of the martensitic volume fraction can be separated into three sections: the left boundary layer $[0,p]$ and the right boundary layer $[h-p,h]$ with some point $p < h/2$ to be determined, and the flat interior $[p,h-p]$ where $\xi$ equals 1. The martensitic volume fraction within the left boundary layer can be obtained by solving the ordinary differential equation B.7 with the boundary conditions $\xi(0) = 0$ and $\xi'(p) = 0$, where the smoothness of $\xi$, or equivalently the continuity of the microstress $k^{\mathrm{nl}} = S_0\ell_{\mathrm{e}}^2\xi_{,x}$ (2.13), is assumed at $p$. It is then straightforward to obtain

$$\xi = \frac{Y+B_\theta-\sigma\bar{\varepsilon}^{\mathrm{t}}}{2S_0\ell_{\mathrm{e}}^2}\left[(x-p)^2-p^2\right] \ , \tag{B.11}$$

for $x \in [0,p]$. It can also be seen that $\xi$ takes its maximum $\frac{(-Y-B_\theta+\sigma\bar{\varepsilon}^{\mathrm{t}})p^2}{2S_0\ell_{\mathrm{e}}^2}$ at $p$. In order to determine $p$, the continuity of $\xi$ is applied, i.e. $\xi(p) = 1$, which leads to

$$p = \left(\frac{2S_0\ell_{\mathrm{e}}^2}{-Y-B_\theta+\sigma\bar{\varepsilon}^{\mathrm{t}}}\right)^{\frac{1}{2}} \ . \tag{B.12}$$

In sum, we obtain

$$\tilde{\varepsilon} = \frac{\sigma}{E} + \bar{\varepsilon}^{\mathrm{t}}\left[1 - \frac{2}{3h}\left(\frac{2S_0\ell_{\mathrm{e}}^2}{-Y-B_\theta+\sigma\bar{\varepsilon}^{\mathrm{t}}}\right)^{\frac{1}{2}}\right] \ , \tag{B.13}$$

for $\sigma > \frac{Y+B_\theta+8S_0\ell_{\mathrm{e}}^2/h^2}{\bar{\varepsilon}^{\mathrm{t}}}$. At this stage, the strain-hardening rate increases as the applied stress increases, and approximates the Young's modulus $E$. Assume that $\tilde{\varepsilon} = \tilde{\varepsilon}^{\max}$

when the stress reaches the prescribed maximum $\sigma^{\max}$.

Upon decreasing the stress, the material deforms elastically until the reverse transformation occurs, i.e.

$$\tilde{\varepsilon} = \tilde{\varepsilon}^{\max} + \frac{\sigma - \sigma^{\max}}{E} \tag{B.14}$$

for $-2Y/\bar{\varepsilon}^{\mathrm{t}} + \sigma^{\max} \leq \sigma < \sigma^{\max}$, where the ending point of this stage is determined by applying the nonlocal yield condition, Eqn (B.1).

The reverse phase transformation starts initially in the two boundary layers $[0, q]$ and $[h-q, h]$ with some point $q < h/2$. Within the boundary layer $[0, q]$, the martensitic volume fraction can be obtained by solving the ordinary differential equation

$$\sigma\bar{\varepsilon}^{\mathrm{t}} - B_\theta + S_0\ell_{\mathrm{e}}^2\xi_{,xx} = -Y \tag{B.15}$$

with boundary conditions $\xi(0) = 0$ and $\xi'(q) = 0$. It is straightforward to obtain the solution

$$\xi = \frac{-Y + B_\theta - \sigma\bar{\varepsilon}^{\mathrm{t}}}{2S_0\ell_{\mathrm{e}}^2}\left[(x-q)^2 - q^2\right] , \tag{B.16}$$

for $x \in [0, q]$. It can also be seen that $\xi$ takes its maximum $\frac{Y - B_\theta + \sigma\bar{\varepsilon}^{\mathrm{t}} q^2}{2S_0\ell_{\mathrm{e}}^2}$ at $q$. Applying the continuity of $\xi$ at $q$ ($\xi(q) = 1$) leads to

$$q = \left(\frac{2S_0\ell_{\mathrm{e}}^2}{Y - B_\theta + \sigma\bar{\varepsilon}^{\mathrm{t}}}\right)^{\frac{1}{2}} . \tag{B.17}$$

In sum, we obtain

$$\tilde{\varepsilon} = \frac{\sigma}{E} + \bar{\varepsilon}^{\mathrm{t}}\left[1 - \frac{2}{3h}\left(\frac{2S_0\ell_{\mathrm{e}}^2}{Y - B_\theta + \sigma\bar{\varepsilon}^{\mathrm{t}}}\right)^{\frac{1}{2}}\right] , \tag{B.18}$$

for $\frac{-Y + B_\theta + 8S_0(\frac{\ell_{\mathrm{e}}}{h})^2}{\bar{\varepsilon}^{\mathrm{t}}} \leq \sigma < -2Y/\bar{\varepsilon}^{\mathrm{t}} + \sigma^{\max}$, where the ending point of this stage is determined by applying $q = h/2$ in Eqn (B.17).

Further loading promotes the reverse phase transformation in the entire crystal. At this stage, the martensitic volume fraction can be obtained by solving Eqn (B.15)

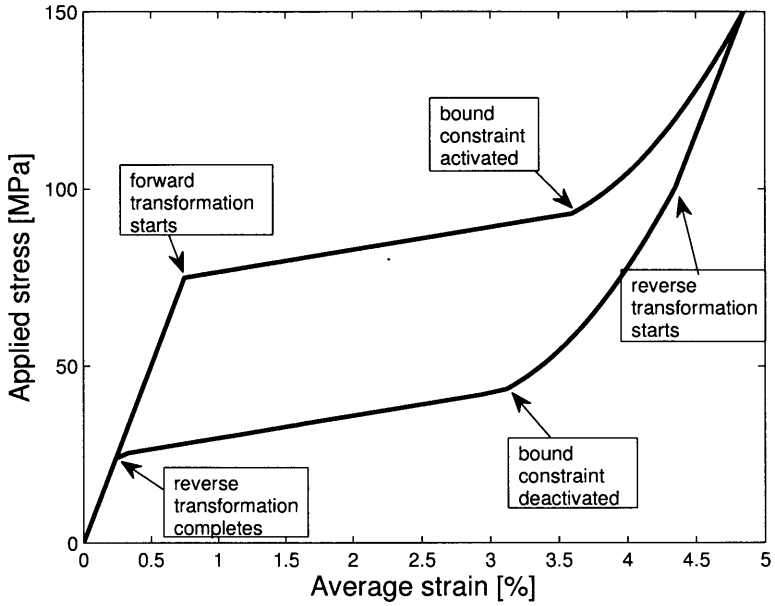


Figure B-1: Analytical stress-strain relation for $\ell_e/h = 0.03$.

with boundary conditions $\xi(0) = \xi(h) = 0$. It is straightforward to obtain

$$\xi = \frac{-Y + B_\theta - \sigma\bar{\varepsilon}^t}{2S_0\ell_e^2}\left[(x - \frac{h}{2})^2 - \frac{h^2}{4}\right] . \tag{B.19}$$

And consequently, we obtain

$$\tilde{\varepsilon} = \frac{\sigma}{E} - \frac{h^2(-Y + B_\theta - \sigma\bar{\varepsilon}^t)}{12S_0\ell_e^2}\bar{\varepsilon}^t , \tag{B.20}$$

for $\frac{-Y+B_\theta}{\bar{\varepsilon}^t} \leq \sigma < \frac{-Y+B_\theta+8S_0(\frac{\ell_e}{h})^2}{\bar{\varepsilon}^t}$. The reverse phase transformation is complete at the end of this stage. Denote the average strain at this ending point as $\tilde{\varepsilon}^r$.

The last part of the loading cycle is the elastic unloading in the austenite:

$$\tilde{\varepsilon} = \tilde{\varepsilon}^r + \frac{\sigma - \frac{-Y+B_\theta}{\bar{\varepsilon}^t}}{E} , \tag{B.21}$$

for $0 \leq \sigma < \frac{-Y+B_\theta}{\bar{\varepsilon}^t}$.

It can be seen from Eqn (B.9), (B.13), (B.18) and (B.20) that the energetic length scale $\ell_e$ appears in the stress-strain relations in a dimensionless form $\ell_e/h$, which

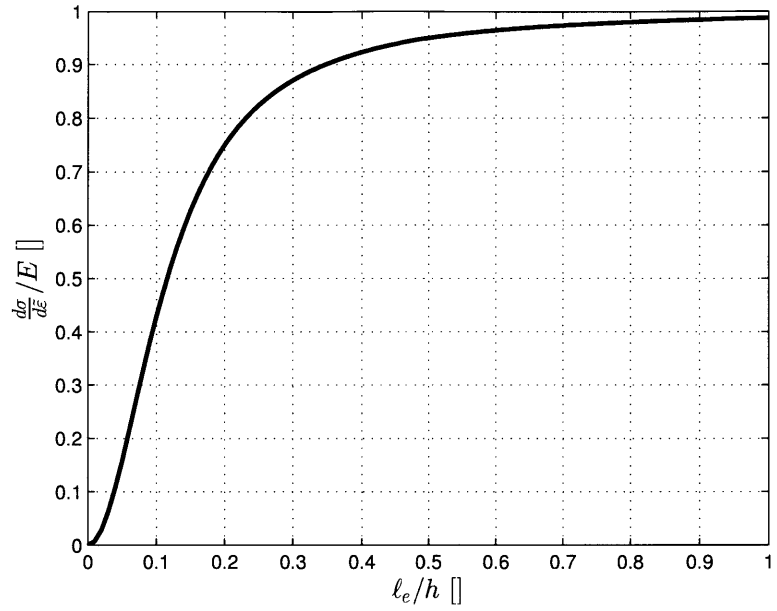


Figure B-2: Analytical strain-hardening rate versus energetic length scale, Eqn (B.10).

implies that it affects the mechanical response only through the ratio $\ell_e/h$. In Fig (B-1), we have plotted an example of the stress-strain curve using the expressions just derived, and marked the turning points for different stages in the loading cycle. In Fig (B-2), we plot the strain-hardening rate during the phase transformation as a function of the energetic length scale using Eqn (B.10). It can be seen that $\frac{d\sigma}{d\bar{\varepsilon}}$ increases monotonically from zero to the limit $E$ for increasing $\ell_e/h$, and this effect is most significant for $\ell_e/h < 0.2$. In the calculations, model parameters take the following values: $E = 10$ GPa, $S_0 = 0.1$ GPa, $\bar{\varepsilon}^t = 0.04$, $B_\theta = 2$ MPa, $Y = 1$ MPa.

## B.2 Uniaxial tension of a chain of grains

Using the results of the single crystal, it is straightforward to derive the stress-strain relation for a chain of grains. Assume that a polycrystal contains a chain of $n$ grains. Assume further that for $i = 1, ..., n$, the length of $i$th grain is $h^{(i)}$, and the material parameters of $i$th grain are $E^{(i)}$, $S_0^{(i)}$, $\ell_e^{(i)}$, $\bar{\varepsilon}^{t(i)}$, $B_\theta^{(i)}$, and $Y^{(i)}$. Given a stress-controlled loading history, the average strain of each grain $\bar{\varepsilon}^{(i)}$ can be obtained independently

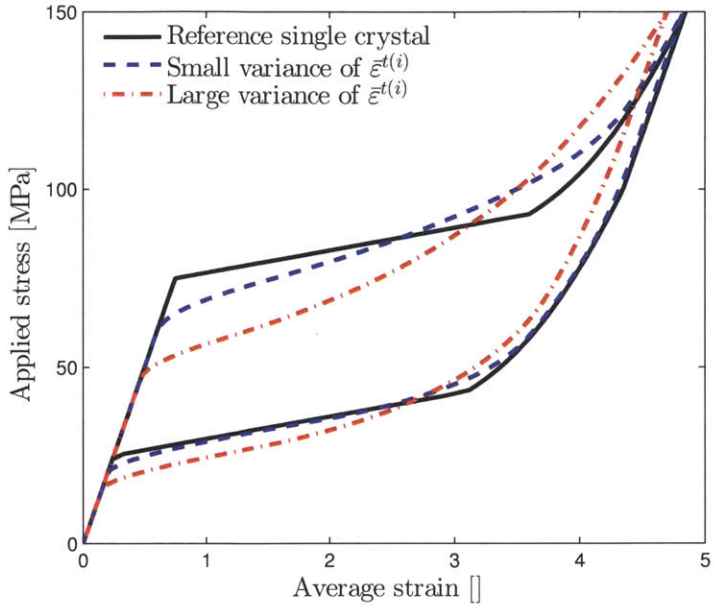


Figure B-3: Stress-strain relation for a chain of 1000 grains with varying maximum transformation strain. $\bar{\varepsilon}^{t(i)}/\bar{\varepsilon}^{t} \in [0.75, 1.25]$ for small variance, and $\bar{\varepsilon}^{t(i)}/\bar{\varepsilon}^{t} \in [0.4, 1.6]$ for large variance.

using the results from the previous section since the condition $\xi = 0$ at the grain boundaries isolates the phase transformation within each grain. As a result, the average strain of the polycrystal reads

$$\tilde{\varepsilon} = \frac{1}{\sum_{i=1}^{n} h^{(i)}} \sum_{i=1}^{n} \tilde{\varepsilon}^{(i)} h^{(i)} \ . \tag{B.22}$$

Taking the previous single-crystal example as a reference, we first investigate the effect of distributed maximum transformation strains. Assume that

$$\bar{\varepsilon}^{t(i)} = \bar{\varepsilon}^{\mathrm{t}} \times \alpha^{i} \ , \tag{B.23}$$

where $\alpha^{i}$ is a random number following the uniform distribution. Two cases are considered. In the small variance case, $\alpha^{i}$ takes its value in $[0.75, 1.25]$, while in the large variance case, $\alpha^{i}$ takes its value in $[0.4, 1.6]$. In Fig (B-3), we plot the stress strain

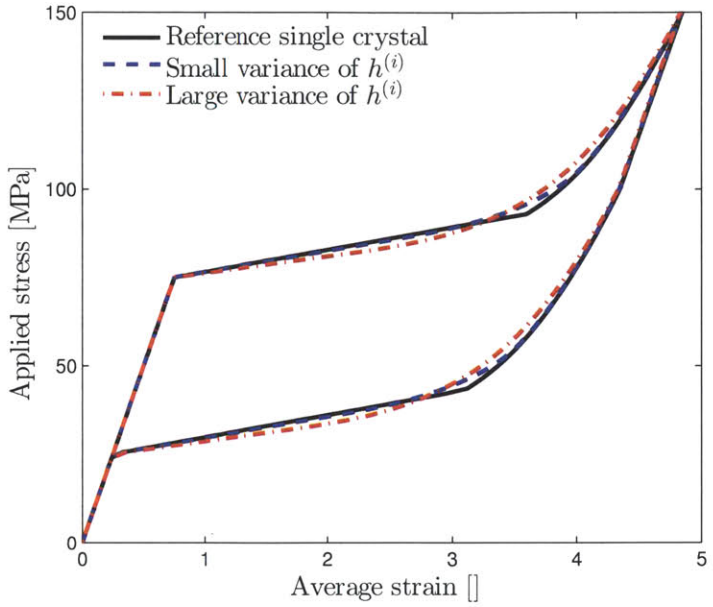


Figure B-4: Stress strain relation for a chain of 1000 grains with varying grain lengths. $h^{(i)}/h \in [0.75, 1.25]$ for small variance, and $h^{(i)}/h \in [0.4, 1.6]$ for large variance.

curve for a polycrystal with 1000 grains[1] in comparison with the reference single-crystal result. It can be observed that the critical stress for the forward transformation is lower in the polycrystal cases. This can be explained as follows. Since the critical stress for each grain is $\sigma^{M_s(i)} = \frac{Y+B_\theta}{\bar{\varepsilon}^{t(i)}}$, the apparent low values for the polycrystal cases can be attributed to the grain with largest $\bar{\varepsilon}^{t(i)}$ that yields first. The critical stresses are $0.8 \times \frac{Y+B_\theta}{\bar{\varepsilon}^t}$ for the small variance case, and $0.625 \times \frac{Y+B_\theta}{\bar{\varepsilon}^t}$ for the large variance case. It can also be seen that during the forward transformation polycrystal cases show a much larger strain-hardening rate than the reference single crystal, whereas the difference is not significant for the reverse transformation.

We also investigate the effect of distributed grain lengths. Assume that

$$h^{(i)} = h \times \alpha^i \,, \tag{B.24}$$

where $\alpha^i$ is a random number following the uniform distribution. Two cases are

[1]The number of grains is sufficiently large to ensure the convergence. A small number of grains will lead to stochastic results.

considered. $\alpha^i$ takes its value in $[0.75, 1.25]$ for the small variance case, and in $[0.4, 1.6]$ for the large variance case, respectively. In Fig (B-4), we plot the stress strain curve for a polycrystal with 1000 grains in comparison to the result of the reference single crystal. It can be observed that the polycrystal results are very close to the result of the reference single crystal except that the polycrystal results are smooth at the turning points where the bound constraint $\xi \leq 1$ is activated and deactivated.

# Appendix C

# Variational incremental formulation

In this chapter, the rate-independent isotropic plasticity, strain gradient plasticity, and gradient superelasticity are presented in a variational incremental formulation. In this formulation, the displacement increment, effective plastic strain increment, and the flow direction are considered as unknown variables. The equilibrium (macroforce balance equation), the normality rule, yield condition and Kuhn-Tucker conditions (for local model and gradient models with only energetic length scales) will be derived as variational results. When gradient models contain the dissipative length scale, the variational result corresponding to the effective plastic strain increment can be interpreted as a microforce balance equation rather than a yield condition. The derivation follows the approach of Simo and Hughes [50] for local plasticity, and Miehe [69] for gradient plasticity. It is worth noting that a mathematical framework has been proposed by Han and Reddy for local and gradient plasticity models using variational inequalities [43].

## C.1 Rate-independent isotropic plasticity

The functional that we want to minimize reads:

$$\mathcal{J}(\Delta\mathbf{u}, \Delta\gamma^{\mathrm{p}}, \mathbf{\Lambda}) = \int_V \psi^{\Delta} + \mathcal{D}^{\Delta}\, dV\ , \tag{C.1}$$

where the traction on the outer surface is not considered for simplicity. The free energy increment is

$$\begin{aligned}\psi^{\Delta} = &\frac{1}{2}(\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\mathrm{p}(n)} - \Delta\gamma^{\mathrm{p}}\mathbf{\Lambda}) : \mathcal{C} : (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\mathrm{p}(n)} - \Delta\gamma^{\mathrm{p}}\mathbf{\Lambda}) \\ &- \frac{1}{2}(\boldsymbol{\varepsilon}^{(n)} - \boldsymbol{\varepsilon}^{\mathrm{p}(n)}) : \mathcal{C} : (\boldsymbol{\varepsilon}^{(n)} - \boldsymbol{\varepsilon}^{\mathrm{p}(n)})\ ,\end{aligned} \tag{C.2}$$

and the energy dissipation during this increment is

$$\mathcal{D}^{\Delta} = \bar{\tau}^{\mathrm{Y}}\Delta\gamma^{\mathrm{p}}\ , \tag{C.3}$$

where $\bar{\tau}^{\mathrm{Y}}$ is the yield strength. The variables of the functional are subject to the constraints: $\Delta\gamma^{\mathrm{p}} \geq 0$,[1] $\mathbf{\Lambda} : \mathbf{\Lambda} = \frac{3}{2}$, and $\mathrm{trace}(\mathbf{\Lambda}) = 0$. Also the displacement increment $\Delta\mathbf{u}$ satisfies Dirichlet boundary conditions on $\partial_{D_u}V$.

Corresponding to the inequality and equality constraints, we introduce three multipliers $\lambda_i$, $i = 1, 2, 3$, and formulate the Lagrange functional

$$\mathcal{L}(\Delta\mathbf{u}, \Delta\gamma^{\mathrm{p}}, \mathbf{\Lambda}, \lambda_1, \lambda_2, \lambda_3) = \mathcal{J} + \int_V \lambda_1(-\Delta\gamma^{\mathrm{p}}) + \lambda_2(\mathbf{\Lambda} : \mathbf{\Lambda} - \frac{3}{2}) + \lambda_3\mathrm{trace}(\mathbf{\Lambda})\, dV\ . \tag{C.4}$$

At a local minimum, the multipliers satisfy the following complementary condition

$$\lambda_1(-\Delta\gamma^{\mathrm{p}}) = 0 \quad \text{and} \quad \lambda_1 \geq 0\ . \tag{C.5}$$

[1] In this case, constraint $\Delta\gamma^{\mathrm{p}} \geq 0$ is equivalent to the non-negative dissipation requirement: $\mathcal{D}^{\Delta} \geq 0$.

The variation of $\mathcal{L}$ with respect to $\Delta\mathbf{u}$ leads to

$$\mathrm{Div}\,\boldsymbol{\sigma}^{(n+1)} = 0 \tag{C.6}$$

with $\boldsymbol{\sigma}^{(n+1)} := \mathcal{C} : (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\mathrm{p}(n)} - \Delta\gamma^{\mathrm{p}}\boldsymbol{\Lambda})$ .

The variation of $\mathcal{L}$ with respect to $\boldsymbol{\Lambda}$ leads to

$$-\boldsymbol{\sigma}^{(n+1)}\Delta\gamma^{\mathrm{p}} + 2\lambda_2\boldsymbol{\Lambda} + \lambda_3\mathbf{I} = \mathbf{0} \,. \tag{C.7}$$

Taking the trace of this equation and using the constraint $\mathrm{trace}(\boldsymbol{\Lambda}) = 0$ lead to

$$\lambda_3 = \frac{1}{3}\Delta\gamma^{\mathrm{p}}\mathrm{trace}(\boldsymbol{\sigma}^{(n+1)}) \,.$$

Insert the value of $\lambda_3$ in Eqn (C.7), and then we can see that $\boldsymbol{\Lambda}$ is collinear with the deviatoric part of $\boldsymbol{\sigma}^{(n+1)}$, $\boldsymbol{\sigma}_0^{(n+1)}$. Because of the equality constraint $\boldsymbol{\Lambda} : \boldsymbol{\Lambda} = \frac{3}{2}$, this variation eventually gives rise to the normality rule

$$\boldsymbol{\Lambda} = \sqrt{\frac{3}{2}}\frac{\boldsymbol{\sigma}_0^{(n+1)}}{\|\boldsymbol{\sigma}_0^{(n+1)}\|} \,. \tag{C.8}$$

The variation of $\mathcal{L}$ with respect to $\Delta\gamma^{\mathrm{p}}$ leads to

$$-\boldsymbol{\sigma}^{(n+1)} : \boldsymbol{\Lambda} + \bar{\tau}^{\mathrm{Y}} - \lambda_1 = 0 \,. \tag{C.9}$$

Eqn (C.9), (C.5) and (C.8) lead to the yield condition

$$-\lambda_1 = \boldsymbol{\sigma}^{(n+1)} : \boldsymbol{\Lambda} - \bar{\tau}^{\mathrm{Y}} = \sqrt{\frac{3}{2}}\|\boldsymbol{\sigma}_0^{(n+1)}\| - \bar{\tau}^{\mathrm{Y}} \leq 0 \,. \tag{C.10}$$

and the Kuhn-Tucker condition

$$(\sqrt{\frac{3}{2}}\|\boldsymbol{\sigma}_0^{(n+1)}\| - \bar{\tau}^{\mathrm{Y}})\Delta\gamma^{p} = 0 \,. \tag{C.11}$$

It is worth noting that the yield condition is an algebraic equation rather than

a partial differential equation, and therefore it is not necessary to treat the effective plastic strain increment as a primary unknown in computation.

## C.2 Rate-independent gradient plasticity with energetic length scale

In this case, the free energy increment is

$$\psi^{\Delta} = \frac{1}{2}(\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\mathrm{p}(n)} - \Delta\gamma^{\mathrm{p}}\boldsymbol{\Lambda}) : \mathcal{C} : (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\mathrm{p}(n)} - \Delta\gamma^{\mathrm{p}}\boldsymbol{\Lambda}) + \frac{1}{2}S_0\ell_{\mathrm{e}}^2\|\nabla\gamma^{\mathrm{p}(n+1)}\|^2 - \frac{1}{2}(\boldsymbol{\varepsilon}^{(n)} - \boldsymbol{\varepsilon}^{\mathrm{p}(n)}) : \mathcal{C} : (\boldsymbol{\varepsilon}^{(n)} - \boldsymbol{\varepsilon}^{\mathrm{p}(n)}) - \frac{1}{2}S_0\ell_{\mathrm{e}}^2\|\nabla\gamma^{\mathrm{p}(n)}\|^2 \,. \tag{C.12}$$

Following the same process, we can define three multipliers and construct a Lagrange functional. The first two variational results are the same as those in the local plasticity model, while the variation of $\mathcal{L}$ with respect to $\Delta\gamma^{\mathrm{p}}$ now leads to

$$-\boldsymbol{\sigma}^{(n+1)} : \boldsymbol{\Lambda} - S_0\ell_{\mathrm{e}}^2\nabla^2\gamma^{\mathrm{p}(n+1)} + \bar{\tau}^{\mathrm{Y}} - \lambda_1 = 0 \,, \tag{C.13}$$

which can be rewritten as

$$\begin{aligned} -\lambda_1 &= \boldsymbol{\sigma}^{(n+1)} : \boldsymbol{\Lambda} + S_0\ell_{\mathrm{e}}^2\nabla^2\gamma^{\mathrm{p}(n+1)} - \bar{\tau}^{\mathrm{Y}} \\ &= \sqrt{\frac{3}{2}}\|\boldsymbol{\sigma}_0^{(n+1)}\| + S_0\ell_{\mathrm{e}}^2\nabla^2\gamma^{\mathrm{p}(n+1)} - \bar{\tau}^{\mathrm{Y}} \,. \end{aligned} \tag{C.14}$$

Using Eqn (C.5), we can obtain

$$\sqrt{\frac{3}{2}}\|\boldsymbol{\sigma}_0^{(n+1)}\| + S_0\ell_{\mathrm{e}}^2\nabla^2\gamma^{\mathrm{p}(n+1)} - \bar{\tau}^{\mathrm{Y}} \leq 0 \,, \tag{C.15}$$

$$(\sqrt{\frac{3}{2}}\|\boldsymbol{\sigma}_0^{(n+1)}\| + S_0\ell_{\mathrm{e}}^2\nabla^2\gamma^{\mathrm{p}(n+1)} - \bar{\tau}^{\mathrm{Y}})\Delta\gamma^{p} = 0 \,. \tag{C.16}$$

Inequality (C.15) includes the Laplacian term, $S_0\ell_{\mathrm{e}}^2\nabla^2\gamma^{\mathrm{p}(n+1)}$, and thus can be viewed as a nonlocal extension of the yield condition (C.10). Plastic deformation is possible

if the elastic predictor stress $\boldsymbol{\sigma}^{\mathrm{pre}}$ satisfies

$$\sqrt{\frac{3}{2}}\|\boldsymbol{\sigma}_0^{\mathrm{pre}}\| + S_0\ell_{\mathrm{e}}^2\nabla^2\gamma^{\mathrm{p}(n)} \geq \bar{\tau}^{\mathrm{Y}} \ . \tag{C.17}$$

## C.3 Rate-independent gradient plasticity with energetic and dissipative length scales

In this case, the free energy increment is the same as the purely energetic case, i.e.

$$\begin{aligned}\psi^{\Delta} = &\frac{1}{2}(\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\mathrm{p}(n)} - \Delta\gamma^{\mathrm{p}}\boldsymbol{\Lambda}) : \mathcal{C} : (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\mathrm{p}(n)} - \Delta\gamma^{\mathrm{p}}\boldsymbol{\Lambda}) + \frac{1}{2}S_0\ell_{\mathrm{e}}^2\|\nabla\gamma^{\mathrm{p}(n+1)}\|^2 \\ &- \frac{1}{2}(\boldsymbol{\varepsilon}^{(n)} - \boldsymbol{\varepsilon}^{\mathrm{p}(n)}) : \mathcal{C} : (\boldsymbol{\varepsilon}^{(n)} - \boldsymbol{\varepsilon}^{\mathrm{p}(n)}) - \frac{1}{2}S_0\ell_{\mathrm{e}}^2\|\nabla\gamma^{\mathrm{p}(n)}\|^2 \ ,\end{aligned} \tag{C.18}$$

while the energy dissipation during this increment is

$$\mathcal{D}^{\Delta} = \bar{\tau}^{\mathrm{Y}}\sqrt{(\Delta\gamma^{\mathrm{p}})^2 + \ell_{\mathrm{d}}^2\Delta\gamma^{\mathrm{p}}_{,K}\Delta\gamma^{\mathrm{p}}_{,K}} \ . \tag{C.19}$$

In order to simplify the notation, we define an auxiliary variable

$$d^{\mathrm{p}} = \sqrt{(\Delta\gamma^{\mathrm{p}})^2 + \ell_{\mathrm{d}}^2\Delta\gamma^{\mathrm{p}}_{,K}\Delta\gamma^{\mathrm{p}}_{,K}} \ . \tag{C.20}$$

Following the same process as described in the local plasticity model, we can define three multipliers and construct a Lagrange functional. The first two variational results are the same as those in the local case, while the variation of $\mathcal{L}$ with respect to $\Delta\gamma^{\mathrm{p}}$ now leads to

$$0 = -\boldsymbol{\sigma}^{(n+1)} : \boldsymbol{\Lambda} - S_0\ell_{\mathrm{e}}^2\nabla^2\gamma^{\mathrm{p}(n+1)} + \bar{\tau}^{\mathrm{Y}}\left[\frac{\Delta\gamma^{\mathrm{p}}}{d^{\mathrm{p}}} - \frac{\partial}{\partial x_J}\left(\frac{\ell_{\mathrm{d}}^2\Delta\gamma^{\mathrm{p}}_{,J}}{d^{\mathrm{p}}}\right)\right] - \lambda_1 \ , \tag{C.21}$$

which can be rewritten as

$$-\lambda_1 = \boldsymbol{\sigma}^{(n+1)} : \boldsymbol{\Lambda} + S_0 \ell_{\mathrm{e}}^2 \nabla^2 \gamma^{\mathrm{p}(n+1)} - \bar{\tau}^{\mathrm{Y}} \left[ \frac{\Delta\gamma^{\mathrm{p}}}{d^{\mathrm{p}}} - \frac{\partial}{\partial x_J} \left( \frac{\ell_{\mathrm{d}}^2 \Delta\gamma^{\mathrm{p}}_{,J}}{d^{\mathrm{p}}} \right) \right]$$
$$= \sqrt{\frac{3}{2}} \| \boldsymbol{\sigma}_0^{(n+1)} \| + S_0 \ell_{\mathrm{e}}^2 \nabla^2 \gamma^{\mathrm{p}(n+1)} - \bar{\tau}^{\mathrm{Y}} \left[ \frac{\Delta\gamma^{\mathrm{p}}}{d^{\mathrm{p}}} - \frac{\partial}{\partial x_J} \left( \frac{\ell_{\mathrm{d}}^2 \Delta\gamma^{\mathrm{p}}_{,J}}{d^{\mathrm{p}}} \right) \right] . \qquad \text{(C.22)}$$

Using Eqn (C.5), we can obtain

$$\sqrt{\frac{3}{2}} \| \boldsymbol{\sigma}_0^{(n+1)} \| + S_0 \ell_{\mathrm{e}}^2 \nabla^2 \gamma^{\mathrm{p}(n+1)} - \bar{\tau}^{\mathrm{Y}} \left[ \frac{\Delta\gamma^{\mathrm{p}}}{d^{\mathrm{p}}} - \frac{\partial}{\partial x_J} \left( \frac{\ell_{\mathrm{d}}^2 \Delta\gamma^{\mathrm{p}}_{,J}}{d^{\mathrm{p}}} \right) \right] \leq 0 , \qquad \text{(C.23)}$$

$$\left\{ \sqrt{\frac{3}{2}} \| \boldsymbol{\sigma}_0^{(n+1)} \| + S_0 \ell_{\mathrm{e}}^2 \nabla^2 \gamma^{\mathrm{p}(n+1)} - \bar{\tau}^{\mathrm{Y}} \left[ \frac{\Delta\gamma^{\mathrm{p}}}{d^{\mathrm{p}}} - \frac{\partial}{\partial x_J} \left( \frac{\ell_{\mathrm{d}}^2 \Delta\gamma^{\mathrm{p}}_{,J}}{d^{\mathrm{p}}} \right) \right] \right\} \Delta\gamma^{\mathrm{p}} = 0 . \qquad \text{(C.24)}$$

Unlike the purely energetic case, inequality (C.23) cannot be viewed as a yield condition since $\lim_{\Delta\gamma^{\mathrm{p}} \to 0} \frac{\Delta\gamma^{\mathrm{p}}}{d^{\mathrm{p}}} - \frac{\partial}{\partial x_J} \left( \frac{\ell_{\mathrm{d}}^2 \Delta\gamma^{\mathrm{p}}_{,J}}{d^{\mathrm{p}}} \right)$ is unknown at the beginning of this load increment.

It is worth noting that for gradient plasticity models, the increment of effective plastic strain is treated as a primary variable, its value will be an output of the calculation, and Eqn (C.23) or the nonlocal yield condition (C.15) for purely energetic case will be satisfied weakly.

## C.4 Gradient superelasticity

Macroscopic superelasticity models contain different transformation directions for forward and reverse transformations [13]:

$$\boldsymbol{\Lambda} = \begin{cases} \sqrt{\frac{3}{2}} \bar{\varepsilon}^{\mathrm{t}} \frac{\boldsymbol{\sigma}_0}{\|\boldsymbol{\sigma}_0\|} , & \Delta\xi > 0 , \\ \sqrt{\frac{3}{2}} \bar{\varepsilon}^{\mathrm{t}} \frac{\boldsymbol{\varepsilon}^{\mathrm{t}(n)}}{\|\boldsymbol{\varepsilon}^{\mathrm{t}(n)}\|} , & \Delta\xi < 0 . \end{cases} \qquad \text{(C.25)}$$

The transformation direction tensor for the reverse transformation is defined by the transformation strain tensor at the beginning of this load increment. This special treatment ensures that the transformation strain accumulated during the forward transformation will vanish with decreasing martensitic volume fraction. It is worth

noting that for mesoscopic superelasticity models, the transformation direction for each transformation system is prescribed by crystallographic theories, and remains the same for both forward and reverse transformations.

It is possible to derive a variational principle that incorporates the two transformation direction tensors in Eqn (C.25). For this purpose, we follow the technique from Thamburaja [112], where the increment of martensitic volume fraction is decomposed into a forward part $\Delta\xi^+$ and a reverse part $\Delta\xi^-$, i.e.

$$\Delta\xi = \Delta\xi^+ + \Delta\xi^- \tag{C.26}$$

with $\Delta\xi^+ \geq 0$ and $\Delta\xi^- \leq 0$. The increment of phase transformation strain tensor is then expressed as

$$\Delta\boldsymbol{\varepsilon}^{\mathrm{t}} = \Delta\xi^+\boldsymbol{\Lambda}^{\mathrm{f}} + \Delta\xi^-\boldsymbol{\Lambda}^{\mathrm{r}} \,, \tag{C.27}$$

where the direction $\boldsymbol{\Lambda}^{\mathrm{f}}$ will be the outcome of the variational principle, and the direction $\boldsymbol{\Lambda}^{\mathrm{r}} = \sqrt{\frac{3}{2}}\bar{\varepsilon}^{\mathrm{t}}\frac{\boldsymbol{\varepsilon}^{\mathrm{t}(n)}}{\|\boldsymbol{\varepsilon}^{\mathrm{t}(n)}\|}$ is prescribed.

The functional that we want to minimize reads

$$\mathcal{J}(\Delta\mathbf{u}, \Delta\xi^+, \Delta\xi^-, \boldsymbol{\Lambda}^{\mathrm{f}}) = \int_V \psi^\Delta + \mathcal{D}^\Delta \, dV \,. \tag{C.28}$$

The increment of free energy is

$$\begin{aligned}\psi^\Delta = {} & \frac{1}{2}(\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\mathrm{t}(n)} - \Delta\boldsymbol{\varepsilon}^{\mathrm{t}}) : \mathcal{C} : (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\mathrm{t}(n)} - \Delta\boldsymbol{\varepsilon}^{\mathrm{t}}) - \frac{1}{2}(\boldsymbol{\varepsilon}^{(n)} - \boldsymbol{\varepsilon}^{\mathrm{t}(n)}) : \mathcal{C} : (\boldsymbol{\varepsilon}^{(n)} - \boldsymbol{\varepsilon}^{\mathrm{t}(n)}) \\ & - \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}})\xi^{(n+1)} + \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}})\xi^{(n)} \\ & + \frac{1}{2}S_0\ell_{\mathrm{e}}^2\|\nabla\xi^{(n+1)}\|^2 - \frac{1}{2}S_0\ell_{\mathrm{e}}^2\|\nabla\xi^{(n)}\|^2 \,,\end{aligned} \tag{C.29}$$

where elastic moduli are assumed to be identical for the two phases, and the classic hardening term is ignored in order to simplify the derivation.

The energy dissipation during this increment is

$$\mathcal{D}^\Delta = Y\sqrt{(\Delta\xi)^2 + \ell_{\mathrm{d}}^2\Delta\xi_{,K}\Delta\xi_{,K}} \,. \tag{C.30}$$

In order to simplify the notation, we define an auxiliary variable

$$d^{\mathrm{t}} = \sqrt{(\Delta\xi)^2 + \ell_{\mathrm{d}}^2 \Delta\xi_{,K}\Delta\xi_{,K}} \,. \tag{C.31}$$

The variables of the incremental functional are subject to the following constraints:

$$\Delta\xi^+ \geq 0 \,, \tag{C.32}$$

$$\Delta\xi^- \leq 0 \,, \tag{C.33}$$

$$\Delta\xi^+ + \Delta\xi^- \leq 1 - \xi^{(n)} \,, \tag{C.34}$$

$$\Delta\xi^+ + \Delta\xi^- \geq -\xi^{(n)} \,, \tag{C.35}$$

$$\boldsymbol{\Lambda}^{\mathrm{f}} : \boldsymbol{\Lambda}^{\mathrm{f}} = \frac{3}{2}(\bar{\varepsilon}^t)^2 \,, \tag{C.36}$$

$$\mathrm{trace}(\boldsymbol{\Lambda}^{\mathrm{f}}) = 0 \,. \tag{C.37}$$

Corresponding to these constraints, we introduce six Lagrange multipliers, $\lambda_i$, for $i = 1, 2, ..., 6$, and construct the Lagrange functional

$$\begin{aligned}
&\mathcal{L}(\Delta\mathbf{u}, \Delta\xi^+, \Delta\xi^-, \boldsymbol{\Lambda}^{\mathrm{f}}, \lambda_i) \\
&= \mathcal{J} + \int_V \lambda_1(-\Delta\xi^+) + \lambda_2\Delta\xi^- + \lambda_3(\Delta\xi^+ + \Delta\xi^- + \xi^{(n)} - 1)\, dV \\
&\quad + \int_V \lambda_4(-\Delta\xi^+ - \Delta\xi^- - \xi^{(n)}) + \lambda_5(\boldsymbol{\Lambda}^{\mathrm{f}} : \boldsymbol{\Lambda}^{\mathrm{f}} - \frac{3}{2}(\bar{\varepsilon}^t)^2) + \lambda_6 \mathrm{trace}(\boldsymbol{\Lambda}^{\mathrm{f}})\, dV \,.
\end{aligned} \tag{C.38}$$

At a local minimum, the multipliers satisfy

$$\lambda_1(-\Delta\xi^+) = 0 \,, \quad \lambda_1 \geq 0 \,, \tag{C.39}$$

$$\lambda_2\Delta\xi^- = 0 \,, \quad \lambda_2 \geq 0 \,, \tag{C.40}$$

$$\lambda_3(\Delta\xi^+ + \Delta\xi^- + \xi^{(n)} - 1) = 0 \,, \quad \lambda_3 \geq 0 \,, \tag{C.41}$$

$$\lambda_4(-\Delta\xi^+ - \Delta\xi^- - \xi^{(n)}) = 0 \,, \quad \lambda_4 \geq 0 \,. \tag{C.42}$$

The variation of $\mathcal{L}$ with respect to $\Delta\mathbf{u}$ leads to

$$\mathrm{Div}\boldsymbol{\sigma}^{(n+1)} = 0 \,, \tag{C.43}$$

where $\boldsymbol{\sigma}^{(n+1)} = \mathcal{C} : (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\mathrm{t}(n)} - \Delta\boldsymbol{\varepsilon}^{\mathrm{t}})$ .

The variation of $\mathcal{L}$ with respect to $\boldsymbol{\Lambda}^{\mathrm{f}}$ leads to

$$-\boldsymbol{\sigma}^{(n+1)}\Delta\xi^{+} + 2\lambda_5\boldsymbol{\Lambda}^{\mathrm{f}} - \lambda_6\mathbf{I} = 0 \,. \tag{C.44}$$

Taking the trace of the equation above, we can see that $\lambda_6 = \frac{1}{3}\mathrm{trace}(\boldsymbol{\sigma}^{(n+1)})\Delta\xi^{+}$, which implies that $\boldsymbol{\Lambda}^{\mathrm{f}}$ is collinear with the deviatoric part of $\boldsymbol{\sigma}^{(n+1)}$. Therefore,

$$\boldsymbol{\Lambda}^{\mathrm{f}} = \sqrt{\frac{3}{2}}\bar{\varepsilon}^{\mathrm{t}}\frac{\boldsymbol{\sigma}_0^{(n+1)}}{\|\boldsymbol{\sigma}_0^{(n+1)}\|} \,. \tag{C.45}$$

The variation of $\mathcal{L}$ with respect to $\Delta\xi^{+}$ leads to

$$\begin{aligned} 0 = &-\boldsymbol{\sigma}^{(n+1)} : \boldsymbol{\Lambda}^{\mathrm{f}} - \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) - S_0\ell_{\mathrm{e}}^2\nabla^2\xi^{(n+1)} + Y\left[\frac{\Delta\xi}{d^{\mathrm{t}}} - \frac{\partial}{\partial x_J}\left(\frac{\ell_{\mathrm{d}}^2\Delta\xi_{,J}}{d^{\mathrm{t}}}\right)\right] \\ &- \lambda_1 + \lambda_3 - \lambda_4 \,. \end{aligned} \tag{C.46}$$

As a result, Eqn (C.39) leads to

$$\boldsymbol{\sigma}^{(n+1)} : \boldsymbol{\Lambda}^{\mathrm{f}} + \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) + S_0\ell_{\mathrm{e}}^2\nabla^2\xi^{(n+1)} - Y\left[\frac{\Delta\xi}{d^{\mathrm{t}}} - \frac{\partial}{\partial x_J}\left(\frac{\ell_{\mathrm{d}}^2\Delta\xi_{,J}}{d^{\mathrm{t}}}\right)\right] - (\lambda_3 - \lambda_4) \leq 0 \,, \tag{C.47}$$

$$\left\{\boldsymbol{\sigma}^{(n+1)} : \boldsymbol{\Lambda}^{\mathrm{f}} + \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) + S_0\ell_{\mathrm{e}}^2\nabla^2\xi^{(n+1)} - Y\left[\frac{\Delta\xi}{d^{\mathrm{t}}} - \frac{\partial}{\partial x_J}\left(\frac{\ell_{\mathrm{d}}^2\Delta\xi_{,J}}{d^{\mathrm{t}}}\right)\right] - (\lambda_3 - \lambda_4)\right\}\Delta\xi^{+} = 0 \,. \tag{C.48}$$

If the dissipative length scale is zero ($\ell_{\mathrm{d}} = 0$), $\lim_{\Delta\xi\to 0^{+}} \frac{\Delta\xi}{d^{\mathrm{t}}} - \frac{\partial}{\partial x_J}\left(\frac{\ell_{\mathrm{d}}^2\Delta\xi_{,J}}{d^{\mathrm{t}}}\right) = 1$ . Eqn (C.47) can be viewed as a nonlocal yield condition for forward transformation. The

forward transformation is possible if $\xi^{(n)} < 1$ [2] and

$$\sqrt{\frac{3}{2}}\bar{\varepsilon}^{\mathrm{t}}\|\boldsymbol{\sigma}_0^{\mathrm{pre}}\| + \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) + S_0\ell_{\mathrm{e}}^2\nabla^2\xi^{(n)} - Y \geq 0\,. \tag{C.49}$$

This is also the yield condition that has been used by Thamburaja in [112]. If the dissipative length scale is not zero, $\lim_{\Delta\xi\to 0^+} \frac{\Delta\xi}{d^{\mathrm{t}}} - \frac{\partial}{\partial x_J}(\frac{\ell_{\mathrm{d}}^2\Delta\xi_{,J}}{d^{\mathrm{t}}})$ is unknown at the beginning of this load increment, and Eqn (C.47) cannot be viewed as a yield condition.

The variation of $\mathcal{L}$ with respect to $\Delta\xi^-$ leads to

$$0 = -\boldsymbol{\sigma}^{(n+1)} : \boldsymbol{\Lambda}^{\mathrm{r}} - \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) - S_0\ell_{\mathrm{e}}^2\nabla^2\xi^{(n+1)} + Y\left[\frac{\Delta\xi}{d^{\mathrm{t}}} - \frac{\partial}{\partial x_J}(\frac{\ell_{\mathrm{d}}^2\Delta\xi_{,J}}{d^{\mathrm{t}}})\right] + \lambda_2 + \lambda_3 - \lambda_4\,. \tag{C.50}$$

As a result, Eqn (C.40) leads to

$$\boldsymbol{\sigma}^{(n+1)} : \boldsymbol{\Lambda}^{\mathrm{r}} + \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) + S_0\ell_{\mathrm{e}}^2\nabla^2\xi^{(n+1)} - Y\left[\frac{\Delta\xi}{d^{\mathrm{t}}} - \frac{\partial}{\partial x_J}(\frac{\ell_{\mathrm{d}}^2\Delta\xi_{,J}}{d^{\mathrm{t}}})\right] - (\lambda_3 - \lambda_4) \geq 0\,, \tag{C.51}$$

$$\left\{\boldsymbol{\sigma}^{(n+1)} : \boldsymbol{\Lambda}^{\mathrm{r}} + \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) + S_0\ell_{\mathrm{e}}^2\nabla^2\xi^{(n+1)} - Y\left[\frac{\Delta\xi}{d^{\mathrm{t}}} - \frac{\partial}{\partial x_J}(\frac{\ell_{\mathrm{d}}^2\Delta\xi_{,J}}{d^{\mathrm{t}}})\right] - (\lambda_3 - \lambda_4)\right\}\Delta\xi^- = 0\,. \tag{C.52}$$

If the dissipative length scale is zero, $\lim_{\Delta\xi\to 0^-} \frac{\Delta\xi}{d^{\mathrm{t}}} - \frac{\partial}{\partial x_J}(\frac{\ell_{\mathrm{d}}^2\Delta\xi_{,J}}{d^{\mathrm{t}}}) = -1$ . Eqn (C.51) can be viewed as a nonlocal yield condition for reverse transformation. The reverse transformation is possible if $\xi^{(n)} > 0$ [3] and

$$\sqrt{\frac{3}{2}}\bar{\varepsilon}^{\mathrm{t}}\boldsymbol{\sigma}^{\mathrm{pre}} : \frac{\boldsymbol{\varepsilon}^{\mathrm{t}(n)}}{\|\boldsymbol{\varepsilon}^{\mathrm{t}(n)}\|} + \Delta s_{\mathrm{eq}}(T - T_{\mathrm{eq}}) + S_0\ell_{\mathrm{e}}^2\nabla^2\xi^{(n)} + Y \leq 0\,. \tag{C.53}$$

If the dissipative length scale is not zero, $\lim_{\Delta\xi\to 0^-} \frac{\Delta\xi}{d^{\mathrm{t}}} - \frac{\partial}{\partial x_J}(\frac{\ell_{\mathrm{d}}^2\Delta\xi_{,J}}{d^{\mathrm{t}}})$ is unknown at the beginning of this load increment, and Eqn (C.51) cannot be viewed as a yield condition.

[2] Condition $\xi^{(n)} < 1$ results in $\lim_{\Delta\xi\to 0^+} \lambda_3 = \lim_{\Delta\xi\to 0^+} \lambda_4 = 0$.

[3] Condition $\xi^{(n)} > 0$ results in $\lim_{\Delta\xi\to 0^-} \lambda_3 = \lim_{\Delta\xi\to 0^-} \lambda_4 = 0$.

# Bibliography


[1] R. Abeyaratne, C. Chu, and R. D. James. Kinetics of materials with wiggly energies: Theory and application to the evolution of twinning microstructures in a Cu-Al-Ni shape memory alloy. *Philosophical Magazine A*, 73(2):457–497, 1996.

[2] R. Abeyaratne and J.K. Knowles. A continuum model of a thermoelastic solid capable of undergoing phase transitions. *Journal of the Mechanics and Physics of Solids*, 41:541–571, 1993.

[3] R. Abeyaratne and J.K. Knowles. *Evolution of phase transitions: a continuum theory.* Cambridge University Press, 2006.

[4] L. Anand and M.E. Gurtin. Thermal effects in the superelasticity crystalline shape-memmory materials. *Journal of the Mechanics and Physics of Solids*, 51:1015–1058, 2003.

[5] L. Anand, M.E. Gurtin, S. P. Lele, and C. Gething. A one-dimensional theory of strain-gradient plasticity: Formulation, analysis, numerical results. *Journal of the Mechanics and Physics of Solids*, 53:1789–1826, 2005.

[6] S. Aubry, M. Fago, and M. Ortiz. A constrained sequential-lamination algorithm for the simulation of sub-grid microstructure in martensitic materials. *Computer Methods in Applied Mechanics and Engineering*, 192(26-27):2823–2843, 2003.

[7] F. Auricchio, R.L. Taylor, and J. Lubliner. Shape-memory alloys: macromodelling and numerical simulations of the superelastic behavior. *Computer Methods in Applied Mechanics and Engineering*, 146:281–312, 1997.

[8] J.M. Ball and R.D. James. Fine phase mixtures as minimizers of energy. *Archive for Rational Mechanics Analysis*, 100(1):13–52, 1987.

[9] Dimitri P. Bertsekas. *Convex Analysis and Variational Problems.* Athena Scientific, 1999.

[10] K. Bhattacharya. *Microstructure of Martensite Why it forms and how it gives rise to the shape-memory effect?* Oxford University Press, 2003.

[11] K Bhattacharya, S Conti, G Zanzotto, and J Zimmer. Crystal symmetry and the reversibility of martensitic transformations. *Nature*, 428(6978):55–59, 2004.

[12] Zhonghe Bo and Dimitris C. Lagoudas. Thermomechanical modeling of polycrystalline SMAs under cyclic loading, Part III: evolution of plastic strains and two-way shape memory effect. *Int. J. Eng. Sci.*, 37(9):1175–1203, 1999.

[13] J.G. Boyd and D.C. Lagoudas. A thermodynamical constitutive model for shape memory materials. part i. the monolithic shape memory alloy. *International Journal of Plasticity*, 12:805–842, 1996.

[14] L.C. Brinson. One-dimensional constitutive behavior of shape memory alloys: Thermomechanical derivation with non-constant material functions and redefined martensite internal variable. *Journal of Intelligent Material Systems and Structures*, 4(2):229–242, 1993.

[15] L.C. Brinson and R. Lammering. Finite element analysis of the behavior of shape memory alloys and their applications. *International Journal of Solids and Structures*, 30(23):3261–3280, 1993.

[16] W. J. Buehler, J. V. Gilfrich, and R. C. Wiley. Effect of lowtemperature phase changes on the mechanical properties of alloys near composition tini. *Journal of Applied Physics*, 34(5):1475–1477, 1963.

[17] C. B. Carter and M. G. Norton. *Ceramic Materials Science and Engineering.* Springer, New York, 2006.

[18] Y. Chen and C. A. Schuh. Size effects in shape memory alloy microwires. *Acta Materialia*, 59:537 – 553, 2011.

[19] D. Christ and S. Reese. A finite element model for shape memory alloys considering thermomechanical couplings at large strains. *International Journal of Solids and Structures*, 46:3694–3709, 2009.

[20] C.L. Chu, S.K. Wu, and Y.C. Yen. Oxidation behavior of equiatomic TiNi alloy in high temperature air environment. *Mater. Sci. Eng. A-Struct*, 216(1-2):193–200, 1996.

[21] Bernard Dacorogna. *Direct Methods in the Calculus of Variations.* Springer New York, 2007.

[22] A. Duval, M. Haboussi, and T.B. Zineb. Nonlocal modelling of superelastic behavior of shape memory alloys. *8th European Symposium on Martensitic Transformation*, 2009.

[23] P. Entel, R. Meyer, and K. Kadau. Molecular dynamics simulations of martensitic transitions. *Philosophical Magazine Part B*, 80(2):183–194, 2000.

[24] Lawrence C. Evans. *Partial differential equations.* American Mathematical Society, 1998.

[25] F. Falk. Model free energy, mechanics, and thermodynamics of shape memory alloys. *Acta Metallurgica*, 28(12):1773–1780, 1980.

[26] F. Falk. Ginzburg-Landau theory of static domain walls in shape-memory alloys. *Z. Physik B Condensed Matter*, 51:177–185, 1983.

[27] F. Falk and P. Konopka. Three-dimensional Landau theory describing the martensitic phase transformation of shape-memory alloys. *Journal of Physics: Condensed Matter*, 2(1):61, 1990.

[28] N.A. Fleck and J.W. Hutchinson. Strain gradient effects in plasticity. *Journal of the Mechanics and Physics of Solids*, 41:1825–1857, 1993.

[29] N.A. Fleck, G.M. Muller, M.F. Ashby, and J.W. Hutchinson. Strain gradient plasticity: Theory and Experiment. *Acta Materialia*, 42:475–487, 1994.

[30] M. Frémond. *Non-smooth thermomechanics*. Springer, 2001.

[31] C. P. Frick, B. G. Clark, S. Orso, P. Sonnweber-Ribic, and E. Arzt. Orientation-independent pseudoelasticity in small-scale NiTi compression pillars. *Scripta Mater*, 59(1):7–10, 2008.

[32] C. P. Frick, S. Orso, and E. Arzt. Loss of pseudoelasticity in nickel-titanium sub-micron compression pillars. *Acta Mater*, 55(11):3845–3855, 2007.

[33] Y.Q. Fu, Sam Zhang, M.J. Wu, W.M. Huang, H.J. Du, J.K. Luo, A.J. Flewitt, and W.I. Milne. On the lower thickness boundary of sputtered TiNi films for shape memory application. *Thin Solid Films*, 515(1):80–86, 2006.

[34] H. Gao, Y. Huang, W.D. Nix, and J.W. Hutchinson. Mechanism-based strain gradient plasticity i. theory. *Journal of the Mechanics and Physics of Solids*, 47:1239–1263, 1999.

[35] I. M. Gelfand and S. V. Fomin. *Calculus of Variations*. Dover Publications, 2000.

[36] Gene H. Golub and Charles F. van Van Loan. *Matrix Computations*. The Johns Hopkins University Press, 1996.

[37] S. Govindjee and C. Miehe. A multi-variant martensitic phase transformation model: formulation and numerical implementation. *Computer Methods in Applied Mechanics and Engineering*, 191:215–238, 2001.

[38] Julia R. Greer and Jeff Th.M. De Hosson. Plasticity in small-sized metallic systems: Intrinsic versus extrinsic size effect. *Prog. Mater. Sci.*, 56(6):654–724, 2011.

[39] Karthik Guda Vishnu and Alejandro Strachan. Size effects in niti from density functional theory calculations. *Phys. Rev. B*, 85(1):014114, 2012.

[40] P. Gudmundson. A unified treatment of strain gradient plasticity. *Journal of the Mechanics and Physics of Solids*, 52:1379–1406, 2004.

[41] M.E. Gurtin and L. Anand. A theory of strain-gradient plasticity for isotropic, plastically irrotational materials. Part I: Small deformations. *Journal of the Mechanics and Physics of Solids*, 53:1624–1649, 2005.

[42] M.E. Gurtin and L. Anand. A theory of strain-gradient plasticity for isotropic, plastically irrotational materials. Part II: Finite deformations. *International Journal of Plasticity*, 21:2297–2318, 2005.

[43] Weimin Han and B. Daya Reddy. *Plasticity: mathematical theory and numerical analysis*. Springer, 2013.

[44] K.F. Hane and T.W. Shield. Microstructure in the cubic to monoclinic transition in titanium-nickel shape memory alloys. *Acta Materialia*, 47(9):2602–2617, 1999.

[45] Felix E. Hildebrand and Rohan Abeyaratne. An atomistic investigation of the kinetics of detwinning. *Journal of the Mechanics and Physics of Solids*, 56(4):1296–1319, 2008.

[46] Y. Huang, H. Gao, W.D. Nix, and J.W. Hutchinson. Mechanism-based strain gradient plasticity i. analysis. *Journal of the Mechanics and Physics of Solids*, 48:99–128, 2000.

[47] Thomas J.R. Hughes. *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*. Dover Publications, 2000.

[48] A Ishida and M Sato. Thickness effect on shape memory behavior of Ti-50.0at.%Ni thin film. *Acta Mater*, 51(18):5571–5578, 2003.

[49] R.D. James and K.F. Hane. Martensitic transformations and shape memory alloys. *Acta Materialia*, 48:197–222, 2000.

[50] J.C. Simo and T.J.R. Hughes. *Computational Inelasticity*. Springer, Berlin, 1998.

[51] Y. Jung, P. Papadopoulos, and R.O. Ritchie. Constitutive modelling and numerical simulation of multivariant phase transformation in superelastic shape-memory alloys. *International Journal for Numerical Methods in Engineering*, 60:429–460, 2004.

[52] Oliver Kastner. *First Principles Modelling of Shape Memory Alloys*. Springer, 2012.

[53] D. C. Lagoudas, P. B. Entchev, P. Popov, L. C. Brinson E. Patoor, and Xiujie Gao. Shape memory alloys, part ii: Modeling of polycrystals. *Mechanics of Materials*, 38:430–462, 2006.

[54] D.C. Lagoudas, editor. *Shape memory alloys: modeling and engineering applications*. Springer, 2008.

[55] Dimitris C. Lagoudas and Pavlin B. Entchev. Modeling of transformation-induced plasticity and its effect on the behavior of porous shape memory alloys. Part I: constitutive model for fully dense SMAs. *Mech. Mater.*, 36(9):865–892, 2004.

[56] S. P. Lele and L. Anand. A small-deformation strain-gradient theory for isotropic viscoplastic materials. *Philosophical Magazine*, 88(30-32):3655–3689, 2008.

[57] Suvrat P. Lele and Lallit Anand. A large-deformation strain-gradient theory for isotropic viscoplastic materials. *International Journal of Plasticity*, 25(3):420 – 453, 2009.

[58] V.I. Levitas and I.B. Ozsoy. Micromechanical modeling of stress-induced phase transformations. part 1. thermodynamics and kinetics of coupled interface propagation and reorientation. *International Journal of Plasticity*, 25:239–280, 2009.

[59] V.I. Levitas and I.B. Ozsoy. Micromechanical modeling of stress-induced phase transformations. part 2. computational algorithms and examples. *International Journal of Plasticity*, 25:546–583, 2009.

[60] V.I. Levitas and D.L. Preston. Three-dimensional landau theory for multivariant stress-induced martensitic phase transformations. i. austenite↔ martensite. *Physical Review B*, 66:134206, 2002.

[61] B. Li, X. M. Zhang, P.C. Clapp, and J. A. Rifkin. Molecular dynamics simulations of the effects of defects on martensite nucleation. *Journal of Applied Physics*, 95(4):1698–1705, 2004.

[62] Zhiping Li. Computations of needle-like microstructures. *Applied Numerical Mathematics*, 39(1):1–15, 2001.

[63] Zhiping Li. Mesh transformation and regularization in numerical simulation of austenitic- martensitic phase transition. *Computational Materials Science*, 21(3):418–428, 2001.

[64] C. Liang and C.A. Rogers. A multi-dimensional constitutive model for shape memory alloys. *Journal of Engineering Mathematics*, 26(3):429–443, 1992.

[65] V. A. L'vov, C. Picornell, J. Pons, and E. Cesari. Statistical description of mechanical stabilization of cu-al-ni martensite. *Materials Transactions*, 46:983 – 989, 2005.

[66] V. A. L'vov, A. A. Rudenko, V. A. Chernenko, E. Cesari, J. Pons, and T. Kanomata. Stress-induced martensitic transformation and superelastcity of alloys: experiment and theory. *Materials Transactions*, 46:790 – 797, 2005.

[67] MATLAB. *version 7.14.0 (R2012a).* The MathWorks Inc., Natick, Massachusetts, 2012.

[68] C. Miehe, F. Welschinger, and M. Hofacker. Thermodynamically consistent phase-field models of fracture: Variational principles and multi-field FE implementations. *International Journal for Numerical Methods in Engineering*, 83:1273–1311, 2010.

[69] Christian Miehe. A multi-field incremental variational framework for gradient-extended standard dissipative solids. *Journal of the Mechanics and Physics of Solids*, 59(4):898–923, 2011.

[70] A. Mielke and T. Roubíček. A rate-independent model for inelastic behavior of shape-memory alloys. *Multiscale Modeling and Simulation*, 1:571–597, 2003.

[71] S. Montecinos, A. Cuniberti, and A. Sepúlveda. Grain size and pseudoelastic behaviour of a Cu-Al-Be alloy. *Materials Characterization*, 59(2):117–123, 2008.

[72] D. Mutter and P. Nielaba. Simulation of the thermally induced austenitic phase transition in niti nanoparticles. *The European Physical Journal B*, 84(1):109–113, 2011.

[73] C. Ni, H. Ding, and X.J. Jin. Grain size dependence of the martensitic transformation in a nano-grained feni polycrystal a molecular dynamics study. *Journal of Alloys and Compounds*, 546(0):1–6, 2013.

[74] D. R. Oakley and N. F. Knight. Adaptive dynamic relaxation algorithm for non-linear hyperelastic structures part i. formulation. *Computer Methods in Applied Mechanics and Engineering*, 126(1-2):67–89, 1995.

[75] D. R. Oakley and N. F. Knight. Adaptive dynamic relaxation algorithm for non-linear hyperelastic structures part ii. single-processor implementation. *Computer Methods in Applied Mechanics and Engineering*, 126(1-2):91–109, 1995.

[76] D. R. Oakley, N. F. Knight, and D. D. Warner. Adaptive dynamic relaxation algorithm for non-linear hyperelastic structures part iii. parallel implementation. *Computer Methods in Applied Mechanics and Engineering*, 126(1-2):111–129, 1995.

[77] M. Ortiz, R. A. Radovitzky, and E. A. Repetto. The computation of the exponential and logarithmic mappings and their first and second linearizations. *International Journal for Numerical Methods in Engineering*, 52(12):1431–1441, 2001.

[78] M. Ortiz and L. Stainier. The variational formulation of viscoplastic updates. *Computer Methods in Applied Mechanics and Engineering*, 171:419–444, 1999.

[79] K. Otsuka and X. Ren. Physical metallurgy of TiNi-based shape memory alloys. *Progress in Materials Science*, 50(5):511–678, 2005.

[80] K. Otsuka and C.M. Wayman, editors. *Shape memory materials.* Cambridge University Press, 1998.

[81] Joseph R.H. Otter, Alfred C. Cassell, and Roger E. Hobbs. Dynamic relaxation. *ICE Proceeding*, 35(4):633–656, 1966.

[82] S. Ozgen and O. Adiguzel. Investigation of the thermoelastic phase transformation in a nial alloy by molecular dynamics simulation. *Journal of Physics and Chemistry of Solids*, 65(5):861–865, 2004.

[83] M. Papadrakakis. A method for the automatic evaluation of the dynamic relaxation parameters. *Computer Methods in Applied Mechanics and Engineering*, 25(1):35–48, 1981.

[84] Harold S. Park. Stress-induced martensitic phase transformation in intermetallic nickel aluminum nanowires. *Nano Letters*, 6(5):958–962, 2006.

[85] E. Patoor, D.C. Lagoudas, P.B. Entchev, L.C. Brison, and X. Gao. Shape memory alloys, part i: General properties and modeling of single crystals. *Mechanics of Materials*, 38:391–429, 2006.

[86] H. Petryk and S. Stupkiewicz. Interfacial energy and dissipation in martensitic phase transformations. Part I: Theory. *J. Mech. Phys. Solids*, 58(3):390–408, 2010.

[87] H. Petryk, S. Stupkiewicz, and G. Maciejewski. Interfacial energy and dissipation in martensitic phase transformations. Part II: Size effects in pseudoelasticity. *Journal of the Mechanics and Physics of Solids*, 58:373–389, 2010.

[88] F.R. Phillips, D. Fang, H. Zheng, and D.C. Lagoudas. Phase transformation in free-standing SMA nanowires. *Acta Materialia*, 59:1871–1880, 2011.

[89] D.A. Porter, K.E. Easterling, and M.Y. Sherif. *Phase tranformations in metals and alloys.* CRC Press, 2009.

[90] L. Qiao, J.J. Rimoli, Y. Chen, C. A. Schuh, and R. Radovitzky. Nonlocal superelastic model of size-dependent hardening and dissipation in single crystal Cu-Al-Ni shape memory alloys. *Phys. Rev. Lett.*, 106:085504, 2011.

[91] M. A. Qidwai and D. C. Lagoudas. Numerical implementation of a shape memory alloy thermomechanical constitutive model using return mapping algorithms. *International Journal for Numerical Methods in Engineering*, 47(6):1123–1168, 2000.

[92] R. Radovitzky and M. Ortiz. Error estimation and adaptive meshing in strongly nonlinear dynamic problems. *Computer Methods in Applied Mechanics and Engineering*, 172:203–240, 1999.

[93] V. Recarte, R. B. Pérez-Sáez, E. H. Bocanegra, M. L. Nó, and J. San Juan. Influence of al and ni concentration on the martensitic transformation in cu-al-ni shape-memory alloys. *Metallurgical and Materials Transactions A*, 33A:2581 – 2591, 2002.

[94] S. Reese and D. Christ. Finite deformation pseudo-elasticity of shape memory alloys-constitutive modelling and finite element implementation. *International Journal of Plasticity*, 24:455–482, 2008.

[95] J. San Juan, M.L. Nó, and C.A. Schuh. Superelasticity and shape memory in micro- and nanometer-scale pillars. *Advanced Materials*, 20:272–278, 2008.

[96] J. San Juan, M.L. Nó, and C.A. Schuh. Nanoscale shape-memory alloys for ultrahigh mechanical damping. *Nature Nanotechnology*, 4:415–419, 2009.

[97] Jose San Juan, Maria L. Nó, and Christopher A. Schuh. Thermomechanical behavior at the nanoscale and size effects in shape memory alloys. *J. Mater. Res.*, 26:2461–2469, 2011.

[98] Olaf Schenk and Klaus Grtner. Solving unsymmetric sparse systems of linear equations with PARDISO. *Future Generation Computer Systems*, 20(3):475 – 487, 2004.

[99] H. Sehitoglu, I. Karaman, R. Anderson, X. Zhang, K. Gall, H. J. Maier, and Y. Chumlyakov. Compressive response of NiTi single crystals. *Acta Mater*, 48(13):3311–3326, 2000.

[100] A. Sengupta, P. Papadopoulos, and R.L. Taylor. Multiscale finite element modeling of superelasticity in nitinol polycrystals. *Computational Mechanics*, 43:573–584, 2009.

[101] N. Siredey, E. Patoor, M. Berveiller, and A. Eberhardt. Constitutive equations for polycrystalline thermoelastic shape memory alloys. part i. intragranular interactions and behavior of the grain. *International Journal of Solids and Structures*, 36:4289–4315, 1999.

[102] M. Somerday, R.J. Comstock Jr., and J.A. Wert. Effect of grain size on the observed pseudoelastic behavior of a cu-zn-al shape memory alloy. *Metallurgical and Materials Transactions A*, 28(11):2335–2341, 1997.

[103] H. Soul, A. Isalgue, A. Yawny, V. Torra, and F.C. Lovey. Pseudoelastic fatigue of niti wires: frequency and size effects on damping capacity. *Smart Materials and Structures*, 19(8):085006, 2010.

[104] N.A. Stelmashenko, M.G. Walls, L.M. Brown, and Yu.V. Milman. Microindentations on w and mo oriented single crystals: an stm study. *Acta Metallurgica et Materialia*, 41:2855–2865, 1993.

[105] J.S. Stolken and A.G. Evans. A microbend test method for measuring the plasticity length scale. *Acta Materialia*, 46:5109–5115, 1998.

[106] S. Stupkiewicz and H. Petryk. Grain-size effect in micromechanical modelling of hysteresis in shape memory alloys. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 90(10-11):783–795, 2010.

[107] Q.P. Sun and Y.J. He. A multiscale continuum model of the grain-size dependence of the stress hysteresis in shape memory alloy polycrystals. *International Journal of Solids and Structures*, 45:3868–3896, 2008.

[108] Y. Sutou, T. Omori, R. Kainuma, and K. Ishida. Grain size dependence of pseudoelasticity in polycrystalline CuAlMn-based shape memory sheets. *Acta Materialia*, 61(10):3842–3850, 2013.

[109] Y. Sutou, T. Omori, K. Yamauchi, N. Ono, R. Kainuma, and K. Ishida. Effect of grain size and texture on pseudoelasticity in Cu-Al-Mn-based shape memory wire. *Acta Materialia*, 53(15):4121–4133, 2005.

[110] K. Tanaka, S. Kobayashi, and Y. Sato. Thermomechanics of transformation pseudoelasticity and shape memory effect in alloys. *International Journal of Plasticity*, 2:59–72, 1986.

[111] Weijia Tang. Thermodynamic study of the low-temperature phase B19' and the martensitic transformation in near-equiatomic Ti-Ni shape memory alloys. *Metall. Mater. Trans. A*, 28:537–544, 1997.

[112] P. Thamburaja. A finite-deformation-based phenomenological theory for shape-memory alloys. *International Journal of Plasticity*, 26:1195–1219, 2010.

[113] P. Thamburaja and L. Anand. Polycrystalline shape-memory materials: effect of crystallographic texture. *Journal of the Mechanics and Physics of Solids*, 49:709–737, 2001.

[114] P. Thamburaja and N. Nikabdullah. A macroscopic constitutive model for shape-memory alloys: Theory and finite-element simulations. *Computer Methods in Applied Mechanics and Engineering*, 198:1074–1086, 2009.

[115] T Uehara, C Asai, and N Ohno. Molecular dynamics simulation of shape memory behaviour using a multi-grain model. *Modelling and Simulation in Materials Science and Engineering*, 17:035011, 2009.

[116] S. M. Ueland, Y. Chen, and C. A. Schuh. Oligocrystalline shape memory alloys. *Adv. Funct. Mater.*, 22(10):2094–2099, 2012.

[117] S. M. Ueland and C. A. Schuh. Superelasticity and fatigue in oligocrystalline shape memory alloy microwires. *Acta Mater*, 60(1):282–292, 2012.

[118] P. Underwood. Dynamic relaxation. *Computational Mehtods for Transient Analysis*, pages 245–265, 1983.

[119] Srikanth Vedantam and Rohan Abeyaratne. A Helmholtz free-energy function for a CuAlNi shape memory alloy. *International Journal of NonLinear Mechanics*, 40(23):177–193, 2005.

[120] R. Vishnoi and D. Kaur. Size dependence of martensite transformation temperature in nanostructured Ni-Mn-Sn ferromagnetic shape memory alloy thin films. *Surface and Coatings Technology*, 204(23):3773–3782, 2010.

[121] T. Waitz, T. Antretter, F. D. Fischer, N. K. Simha, and H. P. Karnthaler. Size effects on the martensitic phase transformation of NiTi nanograins. *J. Mech. Phys. Solids*, 55(2):419–444, 2007.

[122] T. Waitz, T. Antretter, F.D. Fischer, and H.P. Karnthaler. Size effects on martensitic phase transformations in nanocrystalline NiTi shape memory alloys. *Materials Science and Technology*, 24:937–940, 2008.

[123] T. Waitz, K. Tsuchiya, T. Antretter, and F.D. Fischer. Phase transformations of nanocrystalline martensitic materials. *MRS Bull.*, 34:814–821, 2009.

[124] X. Wang, M. Rein, and J. J. Vlassak. Crystallization kinetics of amorphous equiatomic NiTi thin films: Effect of film thickness. *J. Appl. Phys.*, 103:023501, 2008.

[125] Y. Wang and A. G. Khachaturyan. Three-dimensional field model and computer modeling of martensitic transformations. *Acta Materialia*, 45:759 – 773, 1997.

[126] Y. Xiang and J.J. Vlassak. Bauschinger and size effects in thin-film plasticity. *Acta Materialia*, 54:5449–5460, 2006.

[127] Kazuya Yamauchi, Ichizo Ohkata, Koichi Tsuchiya, and Shuichi Miyazaki, editors. *Shape Memory and Superelastic Alloys: Technologies and Applications*. Woodhead Publishing, 2011.

[128] Wenyi Yan, Chun Hui Wang, Xin Ping Zhang, and Yiu-Wing Mai. Theoretical modelling of the effect of plasticity on reverse transformation in superelastic shape memory alloys. *Mater. Sci. Eng. A-Struct*, 354(1-2):146–157, 2003.

[129] J. Ye, R.K. Mishra, A.R. Pelton, and A.M. Minor. Direct observation of the NiTi martensitic phase transformation in nanoscale volumes. *Acta Materialia*, 58:490–498, 2010.

[130] Chao Yu, Guozheng Kang, Di Song, and Qianhua Kan. Micromechanical constitutive model considering plasticity for super-elastic NiTi shape memory alloy. *Comp. Mater. Sci.*, 56(0):1–5, 2012.

[131] X.Y. Zhang, L.C. Brinson, and Q.P. Sun. The variant selection criteria in single-crystal CuAlNi shape memory alloys. *Smart Materials and Structures*, 9:571–581, 2000.

[132] Z. Zhao, S. Kuchnicki, R. Radovitzky, and A. Cuitio. Influence of in-grain mesh resolution on the prediction of deformation textures in fcc polycrystals by crystal plasticity FEM. *Acta Materialia*, 55(7):2361–2373, 2007.

[133] Yuan Zhong, Ken Gall, and Ting Zhu. Atomistic study of nanotwins in NiTi shape memory alloys. *Journal of Applied Physics*, 110:033532, 2011.

[134] Yuan Zhong, Ken Gall, and Ting Zhu. Atomistic characterization of pseudoelasticity and shape memory in NiTi nanopillars. *Acta Materialia*, 60(18):6301–6311, 2012.