Materials Science and Engineering A 378 (2004) 61–69

www.elsevier.com/locate/msea

Mathematical models of martensitic microstructure

J.M. Ball*

Mathematical Institute, University of Oxford, 24-29 St. Giles', Oxford OX1 3LB, UK

Received 15 June 2003; received in revised form 10 November 2003

Abstract

Martensitic microstructures are studied using variational models based on nonlinear elasticity. Some relevant mathematical tools from nonlinear analysis are described, and applications given to austenite–martensite interfaces and related topics.

Keywords: Martensitic transformations; Nonlinear elasticity; Microstructure

1. Introduction

This article reviews some mathematical contributions to the study of martensitic microstructures via continuum models based on nonlinear elasticity. These contributions have exploited techniques from branches of mathematics such as nonlinear analysis, the calculus of variations, partial differential equations and geometry. As often happens, however, when different disciplines come into contact, their interaction is a two-way process. This is particularly striking for continuum models of martensitic microstructure, which have suggested quite new mathematical questions, for example of multi-dimensional calculus, and for which the relation of experimental observation to fundamental unresolved mathematical issues is strikingly immediate.

This is not intended to be, and space prevents it from being, a comprehensive review of the field. Rather it emphasises some issues which have concerned the author and collaborators over recent years, together with other related research. For a broader view, the reader is referred to the books Bhattacharya [19], Dolzmann [28], Müller [45], Pitteri and Zanzotto [49] and Ball and James [10]. Some specific issues not considered here (together with representative references), to which similar mathematical techniques have provided insight, include the mechanical behaviour of polycrystals (see Bhattacharya and Kohn [21], Kohn and Niethammer [40], Paroni [47]), magnetoelastic materials (see James and Kinderlehrer [34], DeSimone and James [27]), hysteresis induced by incompatibility between parent and product phases (Ball et al. [9,8], Ball and James [11], Forclaz [32]), thin films (see Bhattacharya and James [20], Luskin [42]), and generalizations of the Hadamard jump condition (Ball and Carstensen [5], Iwaniec et al. [33]).

2. Nonlinear elasticity model

We describe the elements of the nonlinear elasticity model used by Ball and James [12,13]. Consider a single crystal occupying in a reference configuration the bounded region Ω of three-dimensional Euclidean space $\mathbf{R}^3$. In a deformed configuration, the material point at $x \in \Omega$ in the reference configuration is displaced to the point $y(x) \in \mathbf{R}^3$. Here, we are considering statics, so there is no dependence on the time t. Thus, the deformation of the crystal is described by a mapping $y : \Omega \to \mathbf{R}^3$.

We denote the set of real 3×3 matrices by $M^{3\times3}$, and set $|A| = (\operatorname{tr} A^{\mathrm{T}} A)^{1/2}$, $M^{3\times3}_{+} = \{A \in M^{3\times3} : \det A > 0\}$. Here, $\operatorname{tr} E = \sum_{i=1}^{3} E_{ii}$ denotes the trace of $E \in M^{3\times3}$. Wherever it is defined, the deformation gradient $\nabla y(x) = (\partial y_i/\partial x_j)$ is required to satisfy $\det \nabla y(x) > 0$, so that $\nabla y(x) \in M^{3\times3}_{+}$. The relation of this condition to the invertibility of y, and thus to non-interpenetration of matter, is somewhat subtle and is discussed in Ball [2], Ciarlet and Necas [25], Šverák [52].

Because we want to include deformations y whose deformation gradients can jump across a smooth surface, such as a twin plane, it is important to specify precisely what such singularities in y will be allowed. This *choice of function space* is part of the mathematical model, and affects its pre-

* *E-mail address:* ball@maths.ox.ac.uk (J.M. Ball).

0921-5093/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.msea.2003.11.055

dictions. For example, we could decide to allow deformations y which are discontinuous and allow the crystal to fracture. Because fracture is not the main issue for martensitic transformations, however, we will not allow such singularities. Instead, we will suppose that y belongs to the *Sobolev space* $W^{1,1} = W^{1,1}(\Omega; \mathbf{R}^3)$, which roughly speaking is the set of mappings $y : \Omega \to \mathbf{R}^3$ such that

$$\int_\Omega (|y(x)| + |\nabla y(x)|)\,\mathrm{d}x < \infty. \tag{1}$$

Here, $\nabla y(x)$ is the *weak derivative* of y, which is defined except possibly for a set of points x having zero volume. The precise definition and properties of weak derivatives, the integral in (1), and the Sobolev spaces $W^{1,p}$, $1 \le p \le \infty$, can be found in numerous standard texts, for example, Adams and Fournier [1]. For the purposes of understanding the main points of this article, it is enough to think of the allowed deformations as including piecewise continuously differentiable mappings.

We consider the problem of minimizing the total free energy of the crystal

$$I_\theta(y) = \int_\Omega \psi(\nabla y(x), \theta)\,\mathrm{d}x \tag{2}$$

among $y \in W^{1,1}$ satisfying suitable boundary conditions, for example that y is specified on a portion $\partial\Omega_1$ of the boundary $\partial\Omega$ of Ω, so that $y(x) = \bar{y}(x)$ for all $x \in \partial\Omega_1$, where $\bar{y}$ is a given mapping. The meaning to be ascribed to this boundary condition for $y \in W^{1,1}$ can be found in the previously cited texts. The fact that no boundary condition is specified on the remainder $\partial\Omega_2$ of the boundary corresponds to the surface tractions being zero there, this formally being a 'natural boundary condition' for the variational problem.

In (2), $\psi = \psi(A, \theta)$ is the free-energy density of the crystal, and θ is the temperature, regarded as a constant parameter (that is, independent of x). We assume that ψ is *frame-indifferent*, that is

$$\psi(RA, \theta) = \psi(A, \theta) \quad \text{for all } R \in \mathrm{SO}(3), \tag{3}$$

where SO(3) denotes the set of proper orthogonal matrices (rotations), and that ψ satisfies the *material symmetry* condition

$$\psi(AQ, \theta) = \psi(A, \theta) \quad \text{for all } Q \in \mathcal{S}, \tag{4}$$

where $\mathcal{S}$ is the point group of the crystal, which is a subgroup of SO(3). A well-known argument using the polar decomposition theorem for matrices implies that ψ satisfies (3) if and only if it has the representation $\psi = \psi(U, \theta)$, where $U = (A^{\mathrm{T}}A)^{1/2}$, and then (4) reduces to the requirement that

$$\psi(QUQ^T, \theta) = \psi(U, \theta) \quad \text{for all } Q \in \mathcal{S}. \tag{5}$$

An important role is played by the set

$$K(\theta) = \{A \in M^{3\times 3}_+ : A \text{ minimizes } \psi(A, \theta)\} \tag{6}$$

of energy-minimizing deformation gradients. By adding to ψ a suitable function of θ we may and do assume that the minimum value of $\psi(A, \theta)$ is zero for all θ, so that $K(\theta) = \{A \in M^{3\times 3} : \psi(A, \theta) = 0\}$. From (5), it follows that if $U \in K(\theta)$ then $QUQ^T \in K(\theta)$ for all $Q \in \mathcal{S}$. For a martensitic transformation with transformation temperature θ_c, and transformation strain $U(\theta)$, $\theta < \theta_c$, the N distinct matrices $QU(\theta)Q^T$ for $Q \in \mathcal{S}$ describe the N variants of martensite with strains

$$U_1(\theta), \ldots, U_N(\theta).$$

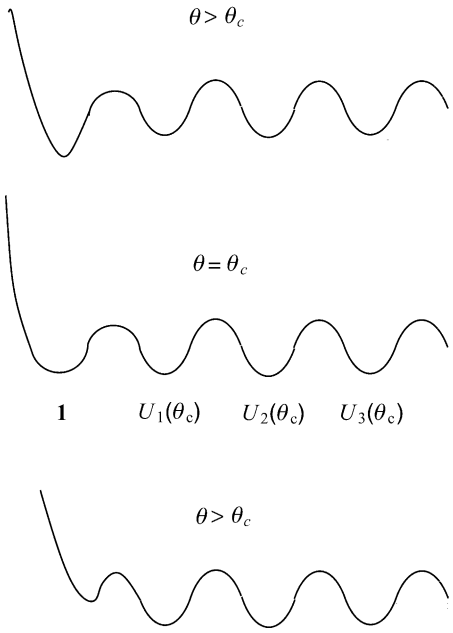

Fig. 1. Schematic of the assumed energy-well structure for the free-energy density ψ showing the exchange of stability as the temperature θ passes through θ_c.

Taking the reference configuration Ω to be undistorted austenite at the temperature θ_c, we then assume (see Fig. 1) that $K(\theta)$ has the form

$$K(\theta) = \begin{cases} \alpha(\theta)\,\mathrm{SO}(3), & \theta > \theta_c \\ \mathrm{SO}(3) \cup \bigcup_{i=1}^N \mathrm{SO}(3)U_i(\theta_c), & \theta = \theta_c \\ \bigcup_{i=1}^N \mathrm{SO}(3)U_i(\theta), & \theta < \theta_c \end{cases} \tag{7}$$

where the thermal expansion coefficient α satisfies $\alpha(\theta_c) = 1$. Thus, for $\theta < \theta_c$ the ith variant of martensite is associated with an energy-well that locally rises above its zero set $\mathrm{SO}(3)U_i$.

For cubic austenite we have that $\mathcal{S} = \mathcal{P}^{24}$, the group consisting of the 24 rotations of a cube into itself. For a cubic to tetragonal transformation, we find that $N = 3$ and that

$$U_1(\theta) = \mathrm{diag}(\eta_2, \eta_1, \eta_1), \quad U_2(\theta) = \mathrm{diag}(\eta_1, \eta_2, \eta_1),$$
$$U_3(\theta) = \mathrm{diag}(\eta_1, \eta_1, \eta_2),$$

where $\eta_1 = \eta_1(\theta) > 0, \eta_2 = \eta_2(\theta) > 0$ are the deformation parameters. Other transformations give rise to different values of N; for example, $N = 6$ for cubic to orthorhombic and $N = 12$ for cubic to monoclinic transformations.

Why is it preferable to use nonlinear elasticity, rather than linear elasticity, as a model for martensitic transformations? To understand this, one should remember that linear elasticity is not a special case of nonlinear elasticity, but a linearization of it about a natural state. Not only the stress–strain behaviour, but also rotations, are linearized, so that the frame-indifference condition (3) no longer holds. In fact, a rigid rotation of a body in a stress-free state, which according to nonlinear elasticity correctly results in another stress-free state, produces non-physical 'phantom' stresses according to linear elasticity. In martensitic transformations, rotations of several degrees are commonplace, so this becomes an important issue. It is possible to construct a 'linearized' theory of martensitic transformations, in which the austenite and each variant of martensite are assigned their own linearized stress–strain law, as was done by Khachaturyan [35,36], Khachaturyan and Shatalov [37], and Roitburd [50,51]. Since the regions occupied by each phase are unknowns, this theory is still nonlinear, but in certain cases is more tractable. It is not easy to give this linearized theory a satisfactory status with respect to the nonlinear one, though an attempt to do this was made in Ball and James (Section 9 in [12]), and of course in situations where there are significant rotations it will not give good results (for a comparison of the theories see Bhattacharya [18,19]).

The energy functional (2) ignores, among other things, the interfacial energy associated with surfaces of discontinuity of ∇y, such as twin boundaries. As a consequence, the theory based on (2) predicts infinitely fine microstructures, whereas including small interfacial energy will typically set a length-scale for these microstructures resulting from the balance between bulk and interfacial energy. Expressed differently, the theory based on (2) has no preferred length-scale; if, for example, the deformation $y : \mathbf{R}^3 \to \mathbf{R}^3$ describes a deformation with $\nabla y(x) \in K(\theta)$ for all x (excepting, perhaps, a set of points of zero volume), the deformation

$$y^\varepsilon(x) = \varepsilon y\left(\frac{x}{\varepsilon}\right) \tag{8}$$

defines for arbitarily small $\varepsilon > 0$ another deformation with the same properties, since $\nabla y^\varepsilon(x) = \nabla y(\varepsilon x)$. Nevertheless, the zero interfacial energy theory based on (2) is of interest because its simplified, though singular, mathematical structure leads to explicit predictions of certain features of observed microstructures that can be compared, often successfully, to experiment. The situation is analogous to the relationship of the Navier-Stokes equations for flow of a viscous fluid, with those for inviscid flow and gas dynamics obtained by setting the viscosity to be zero; these equations likewise have a singular structure, with solutions representing vortices and shock-waves, for example, that can give a good description of, and much insight concerning, the flow of fluids with small viscosity.

The theory based on (2) also, through the assumption (7), ignores the symmetries of the free-energy density ψ corresponding to lattice invariant shears associated to slip and plastic flow (cf. Ericksen [29]). Including these symmetries in a naive way leads to the material behaving like an elastic fluid (see Fonseca [30]). It is not clear how to give the theory described here, with the smaller point group as symmetry group, a clear status with respect to that with the full symmetry group, or to developing theories of elasto-plasticity (cf. Ortiz [46]). Likewise, reconstructive transformations are not covered.

Consider now a smooth surface S containing the point $z \in \mathbf{R}^3$ and having unit normal N there. If ∇y is continuous on either side of S with limits $\nabla^+ y(z) = A$, $\nabla^- y(z) = B$ from above and below S, respectively, then equating the tangential derivatives at z leads to the *Hadamard jump condition*

$$A - B = a \otimes N. \tag{9}$$

for some $a \in \mathbf{R}^3$, where the right-hand side of (9) is the 3×3 matrix of rank one (provided $a \neq 0$) with entries $(a \otimes N)_{ij} = a_i N_j$. An important special case is when S is a plane $\{x \cdot N = k\}$. In this case, it is worth making the following remark. Suppose that $y : \mathbf{R}^3 \to \mathbf{R}^3$ has bounded gradient, so that $|\nabla y(x)| \leq M < \infty$ for all x, and that

$$\lim_{x \cdot N \to +\infty} \nabla y(x) = A, \qquad \lim_{x \cdot N \to -\infty} \nabla y(x) = B. \tag{10}$$

Then (9) still holds. In fact, using the elasticity scaling (8) we obtain a sequence of deformations $y^\varepsilon : \mathbf{R}^3 \to \mathbf{R}^3$ with bounded gradient converging to a deformation y satisfying $\nabla y(x) = A$ if $x \cdot N > 0$, $\nabla y(x) = B$ if $x \cdot N < 0$, so that (10) follows.

Because of the Hadamard jump condition, zero-energy interfaces between variants are in one-to-one correspondence with rank-one connections between the sets $\mathrm{SO}(3)U_i$. More generally, given $U = U^T > 0$, $V = V^T > 0$, we seek rank-one connections between $\mathrm{SO}(3)U$ and $\mathrm{SO}(3)V$. That is, we ask when there are rotations R_1, R_2 and vectors a, N such that

$$R_1 U = R_2(V + a \otimes N). \tag{11}$$

Theorem 2.1. *Let $D = U^2 - V^2$ have eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$. Then $\mathrm{SO}(3)U$ and $\mathrm{SO}(3)V$ are rank-one connected if and only if $\lambda_2 = 0$. There are exactly two solutions up to rigid rotation provided $\lambda_1 < \lambda_2 = 0 < \lambda_3$, and the corresponding normals N_1, N_2 are orthogonal if and only if $\mathrm{tr}\, U^2 = \mathrm{tr}\, V^2$, that is $\lambda_3 = -\lambda_1$.*

Theorem 2.1 is taken from Ball and James [12] (see also Khachaturyan [36]). However, the simple observation concerning the orthogonality of the normals seems to be new. In the case of martensitic variants we have $U = U_i$, $V = U_j$, and since $U_j = QU_iQ^T$ for some rotation Q the requirement $\mathrm{tr}\, U_i^2 = \mathrm{tr}\, U_j^2$ is automatically satisfied, and the condition $\lambda_1 < \lambda_2 = 0 < \lambda_3$ holds if and only if $\det(U_i^2 - U_j^2) = 0$. If this condition holds then up to rigid rotation there are exactly two such rank-one connections (twins) and the corresponding twin planes are orthogonal. For example, for the

case of a cubic to tetragonal transformation, taking $U = U_2$, $V = U_1$ and $R_2 = \mathbf{1}$, the two twins are given by

$$a = \sqrt{2}\frac{\eta_2^2 - \eta_1^2}{\eta_1^2 + \eta_2^2}(-\eta_2, \kappa\eta_1, 0), \quad N = \frac{1}{\sqrt{2}}(1, \kappa, 0),$$

where $\kappa = \pm 1$. An equivalent condition for the existence of twins, due to Forclaz [31], is that $\det(U_i - U_j) = 0$.

3. Mathematical tools

We briefly describe some mathematical tools that prove useful for describing and analyzing martensitic microstructures at the continuum level. The first is *weak convergence*, or convergence in the sense of averages. Consider a sequence of deformation gradients $\nabla y^{(j)}$, which for simplicity we assume to be uniformly bounded independently of j, i.e. $|\nabla y^{(j)}(x)| \leq M < \infty$ for all j and $x \in \Omega$ (again with the possible exception of a set of points of zero volume). We say that $\nabla y^{(j)}$ *converges weakly* to the deformation gradient ∇y, written $\nabla y^{(j)} \rightharpoonup \nabla y$, if

$$\int_E \nabla y^{(j)}(x)\,\mathrm{d}x \to \int_E \nabla y(x)\,\mathrm{d}x$$

for all open subregions E of Ω (equivalently, for all balls, or all cubes, contained in Ω). For example consider the simple laminate shown in Fig. 2 formed from gradients A, B satisfying $A - B = c \otimes N$ with separating interfaces with normal N, the A layers having thickness λ/j and the B layers thickness $(1-\lambda)$/j, where $0 < \lambda < 1$. Then $\nabla y^{(j)}$ has weak limit $\nabla y(x) = \lambda A + (1-\lambda)B$. Note that $\nabla y^{(j)}(x)$ does not converge to $\nabla y(x)$ in the usual sense for any x.

Next, we describe the *Young measure* corresponding to a sequence of gradients $\nabla y^{(j)}$. This concept was introduced by Young (see [56]), while Tartar [54] drew attention to its importance for nonlinear partial differential equations and for carrying information from microscales to macroscales.

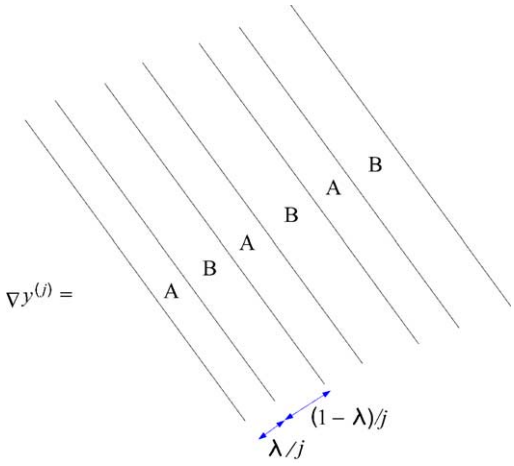

Fig. 2. Simple laminate formed from layering compatible gradients A, B with volume fractions λ, $1-\lambda$ respectively.

To define it, we act like a microscopist, fixing j and a point $x \in \Omega$, and looking at the deformation gradients in a ball $B(x,\delta)$ with centre x and small radius $\delta > 0$. We pick points z distributed uniformly at random from $B(x,\delta)$, and look at the corresponding probability distribution of the matrices $\nabla y^{(j)}(z)$. The probability that $\nabla y^{(j)}(z)$ belongs to a subset $G \subset M^{3\times 3}$ is given by

$$\nu_{x,j,\delta}(G) = \frac{\text{volume}\{z \in B(x,\delta) \text{ with } \nabla y^{(j)}(z) \in G\}}{\text{volume } B(x,\delta)}. \quad (12)$$

Now, we let $j \to \infty$, to obtain the limiting value of this probability, and finally $\delta \to 0$, to localize the probability to the point x. Thus, we expect to obtain a probability distribution on 3×3 matrices given by

$$\nu_x(G) = \lim_{\delta\to 0}\lim_{j\to\infty} \nu_{x,j,\delta}(G). \quad (13)$$

In fact, it is a theorem (see for example Ball [3]) that such a limiting probability exists for any sequence of gradients satisfying a suitable bound such as $|\nabla y^{(j)}(x)| \leq M < \infty$, provided we extract a suitable subsequence of the $\nabla y^{(j)}$. We call $(\nu_x)_{x\in\Omega}$ the (gradient) *Young measure* corresponding to $\nabla y^{(j)}$. The Young measure contains exactly the information needed to determine the weak limit of $h(\nabla y^{(j)})$ for any continuous function h. In fact, this weak limit is given by the expectation of h with respect to the Young measure, that is

$$h(\nabla y^{(j)}) \rightharpoonup \langle \nu_x, h\rangle = \int_{M^{3\times 3}} h(A)\,\mathrm{d}\nu_x(A). \quad (14)$$

In particular, taking $h(A) = A$ for all $A \in M^{3\times 3}$ we have that $\nabla y^{(j)} \rightharpoonup \nabla y(x)$, where $\nabla y(x) = \bar{\nu}_x$ and the centre of mass $\bar{\nu}_x$ is defined by

$$\bar{\nu}_x = \int_{M^{3\times 3}} A\,\mathrm{d}\nu_x(A). \quad (15)$$

From the definition, we see immediately that the Young measure corresponding to the simple laminate in Fig. 2 corresponds to a limiting probability λ of finding the matrix A, $1-\lambda$ of finding the matrix B, and zero of finding any other matrix. That is

$$\nu_x = \lambda\delta_A + (1-\lambda)\delta_B, \quad (16)$$

where for any matrix C the *Dirac mass* δ_C is defined for any $G \subset M^{3\times 3}$ by

$$\delta_C(G) = \begin{cases} 1, & \text{if } C \in G \\ 0, & \text{if } C \notin G \end{cases}$$

Finally, we describe the deeper idea of *quasiconvexity* due to Morrey [43,44]. An integrand $f = f(A)$ is *quasiconvex* if

$$\int_\Omega f(\nabla z(x))\,\mathrm{d}x \geq \int_\Omega f(A)\,\mathrm{d}x = (\text{volume } \Omega)\, f(A) \quad (17)$$

whenever $z : \Omega \to \mathbf{R}^3$ is smooth with $z(x) = Ax$ for all $x \in \partial\Omega$. Despite appearances, the condition does not depend

on Ω. Quasiconvexity of f has important implications for whether or not the functional

$$I(y) = \int_\Omega f(\nabla y)\,\mathrm{d}x$$

attains a minimum subject to given boundary conditions. If the minimum is not attained then although there are admissible y for which $I(y)$ becomes arbitrarily close to its greatest lower bound, there is no y which realizes this lower bound (just as there is no real number $t > 0$ that realizes the lower bound zero of the function $1/t$). In fact, under suitable technical conditions quasiconvexity implies that I attains a minimum, while lack of quasiconvexity is suggestive of nonattainment. If f is quasiconvex then f is *rank-one convex*, that is the function $g(t) = f(A + tc \otimes N)$ is convex (i.e. the tangents to the graph of g never lie above the graph) for any A, c and N. However, the converse is false (Šverák [53]).

For martensitic transformations, the existence of twins implies immediately that $\psi(\cdot, \theta)$ is not rank-one convex, since in the corresponding rank-one directions the function g is minimized at exactly two points. Hence $\psi(\cdot, \theta)$ is not quasiconvex, suggesting that the minimum of I_θ is not in general attained. In this case, we expect the deformation gradients $\nabla y^{(j)}$ of a minimizing sequence for I_θ to have a nontrivial Young measure corresponding to an infinitely fine microstructure, thus explaining in the context of the elasticity model why fine microstructures are formed.

Even though $\psi(\cdot, \theta)$ is not quasiconvex, quasiconvexity plays an important role in analyzing I_θ. Let ψ^{qc} be the *quasiconvexification* of ψ, that is the greatest quasiconvex function less than or equal to $\psi(\cdot, \theta)$, defined by

$$\psi^{\mathrm{qc}}(A, \theta) = \sup\{f(A) : f \text{ quasiconvex}, f(B) \le \psi(B, \theta) \text{ for all } B\}. \tag{18}$$

Then ψ^{qc} can be interpreted as being the *macroscopic* free-energy function corresponding to ψ. (This interpretation follows from the relaxation theorem of Dacorogna [26], which, however, does not strictly speaking apply to elasticity; for a discussion see Ball [4].) In particular we can consider the quasiconvexification $K(\theta)^{\mathrm{qc}}$ of $K(\theta)$, namely the set defined equivalently as

$$\begin{aligned} K(\theta)^{\mathrm{qc}} &= \{A : A \text{ minimizes } \psi^{\mathrm{qc}}(A, \theta)\} \\ &= \{\bar\nu : \nu \text{ an } x\text{-independent gradient Young measure with } \operatorname{supp}\nu \subset K(\theta)\}. \end{aligned}$$

More generally, gradient Young measures are characterized (see Kinderlehrer and Pedregal [38,39]) by the fact that $\bar\nu_x$ is a gradient and

$$\int_{M^{3\times 3}} f(A)\,\mathrm{d}\nu_x(A) \ge f(\bar\nu_x) \tag{19}$$

for all quasiconvex f. Now, the minors (subdeterminants) of A are up to linear combinations the only functions f for which equality holds in (17). Thus, from (19) applied to $f + \pm J$ we deduce the *minors relations*

$$\int_{M^{3\times 3}} J(A)\,\mathrm{d}\nu(A) = J(\bar\nu) \tag{20}$$

for any minor $J = J(A)$, that is for $J(A) = A_{ij}$, $J(A) = (\operatorname{cof} A)_{ij}$, or $J(A) = \det A$. These useful and nontrivial relations can be verified explicitly for the special case of the Young measure (16) of a simple laminate.

For martensitic transformations with $\theta < \theta_c$, $K(\theta)^{\mathrm{qc}}$ can be interpreted as the set of stress-free macroscopic deformation gradients corresponding to zero-energy microstructures.

Unfortunately, there is no known tractable characterization of quasiconvexity, and it has been shown by Kristensen [41] that there is no such local characterization. Thus, for example, there can be no set of inequalities on f and its derivatives at an arbitrary matrix A which is necessary and sufficient for f to be quasiconvex. Thus, we are in the awkward position that the key mathematical concept for the analysis of microstructure is shrouded in mystery.

4. Some successes of the theory

4.1. The crystallographic theory of martensite

As was shown by Ball and James [12], the nonlinear elasticity model incorporates the crystallographic theory of martensite, due to Wechsler et al. [55] and Bowles and Mackenzie [24]. A (classical) austenite–martensite interface is described by a minimizing sequence $y^{(j)}$ for I_{θ_c} in which a simple laminate comprised of two twin-related martensitic variants with gradients $A = R_1U_r$, $B = R_2U_s$, where $R_1, R_2 \in \mathrm{SO}(3)$, meets undistorted austenite, represented by the constant deformation gradient $\nabla y^{(j)} = \mathbf{1}$ at a habit plane with normal m. Because neither variant is compatible with the austenite (that is there is in general no rank-one connection between $\mathrm{SO}(3)$ and $\mathrm{SO}(3)U_i$), a boundary layer is required to interpolate between the laminate and the austenite, the volume of this layer tending to zero as $j \to \infty$ (see Fig. 3).

The construction of such a minimizing sequence with $I_{\theta_c}(y^{(j)}) \to 0$ is possible if and only if $\lambda A + (1-\lambda)B = \mathbf{1} + b \otimes m$ for some λ, $0 < \lambda < 1$, and vectors b, m. Solving for R_1, R_2, λ, b and m leads to the formulae of the crystallographic theory. For example, in the case of a cubic to tetragonal transformation there are 24 possible habit planes with $\lambda = \lambda^*$ or $1 - \lambda^*$, where

$$\lambda^* = \frac{1}{2}\left(1 - \sqrt{\frac{2(\eta_2^2 - 1)(\eta_1^2 - 1)(\eta_1^2 + \eta_2^2)}{(\eta_2^2 - \eta_1^2)^2} + 1}\right),$$

and

$$\begin{aligned} m &= \left(\frac{1}{2}\chi(\delta + \nu\tau), \frac{1}{2}\chi\kappa(\nu\tau - \delta), 1\right), \\ b &= \left(\frac{1}{2}\chi\zeta(\delta + \nu\tau), \frac{1}{2}\chi\zeta\kappa(\nu\tau - \delta), \beta\right), \end{aligned} \tag{21}$$

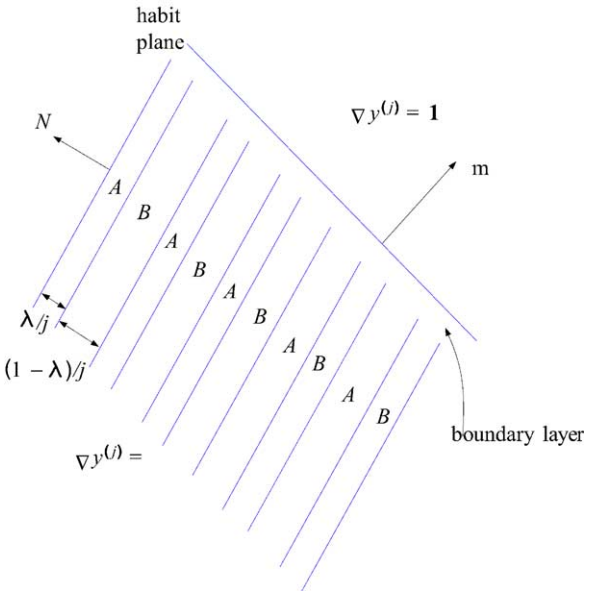

Fig. 3. Mathematical description of an austenite–martensite interface.

where $\nu = 1$ for $\lambda = \lambda^*$, $\nu = -1$ for $\lambda = 1 - \lambda^*$. Here, $\delta = [(\eta_2^2+\eta_1^2-2)(1-\eta_1^2)^{-1}]^{1/2}$, $\tau = [(2\eta_1^2\eta_2^2-\eta_1^2-\eta_2^2)(1-\eta_1^2)^{-1}]^{1/2}$, $\zeta = (1-\eta_1^2)/(1+\eta_2)$, $\beta = \eta_2(\eta_1^2-1)/(1+\eta_2)$, $\chi = \pm 1$. These solutions exist provided the inequalities $\eta_1^2 + \eta_2^2 < 2$ if $\eta_1 > 1$, $\eta_1^{-2} + \eta_2^{-2} < 2$ if $\eta_1 < 1$ hold (see Fig. 4).

4.2. The two well problem

The general question of characterizing $K(\theta)^{qc}$ for $\theta \leq \theta_c$ is open. However, in the case when there are just two martensitic variants, such as for orthorhombic to monoclinic transformatios, $K(\theta)^{qc}$ can be calculated using the minors relations (19). In this case, we can take

$$K(\theta) = \mathrm{SO}(3)U_1 \cup \mathrm{SO}(3)U_2, \tag{22}$$

where $U_1 = \mathrm{diag}\,(\eta_2, \eta_1, \eta_1)$, $U_2 = \mathrm{diag}\,(\eta_1, \eta_2, \eta_1)$ are the first two variants for a cubic to tetragonal transformation.

Theorem 4.1 (Ball and James [13,10]). *For $K(\theta)$ given by (22) $K(\theta)^{\mathrm{qc}}$ consists of those $A \in M^{3\times 3}_+$ such that*

$$A^{\mathrm{T}}A = \begin{pmatrix} a & c & 0 \\ c & b & 0 \\ 0 & 0 & \eta_2^2 \end{pmatrix},$$

where $ab - c^2 = \eta_1^2\eta_2^2$, $a + b + 2|c| \leq \eta_1^2 + \eta_2^2$.

If y is invertible and $\nabla y(x) \in K(\theta)^{\mathrm{qc}}$ for all x then y is a plane strain.

Since a plane strain cannot coincide on the boundary of a three-dimensional region Ω with a linear mapping unless it is itself linear, this provides a case when we can rigorously prove that the minimum of the energy is not attained.

Corollary 4.2 (Ball and Carstensen [7]). *Let $A \in K(\theta)^{\mathrm{qc}}$ with $A \notin K(\theta)$. Then the minimum of $I_\theta(y)$ subject to $y|_{\partial\Omega} = Ax$ is not attained.*

4.3. The role of special deformation parameters

Certain microstructures are only geometrically possible if the deformation parameters satisfy special relations. A case in point is provided by the analysis by Bhattacharya [16] of the *wedge microstructure* observed in certain shape-memory alloys, in which austenite surrounds a wedge-shaped region of martensite consisting of two simply-laminated plates meeting along a midrib plane. Bhattacharya analyzed this microstructure for cubic to tetragonal and for cubic to orthorhombic transformations and showed that it was only possible at zero energy if certain special relations held, and that these relations were indeed nearly satisfied for alloys for which wedges are observed. In the case of a cubic to tetragonal transformation there is a single relation

$$\eta_1^2 = \frac{(1-\eta_2^2)^2 + 4\eta_2^2(1+\eta_2^2)}{(1-\eta_2^2)^2 + 8\eta_2^4}, \tag{23}$$

and the wedge must involve all three variants, with the microtwin planes in each plate meeting at the midrib at 120°. For other results in the same spirit see Bhattacharya [17], Pitteri and Zanzotto [48].

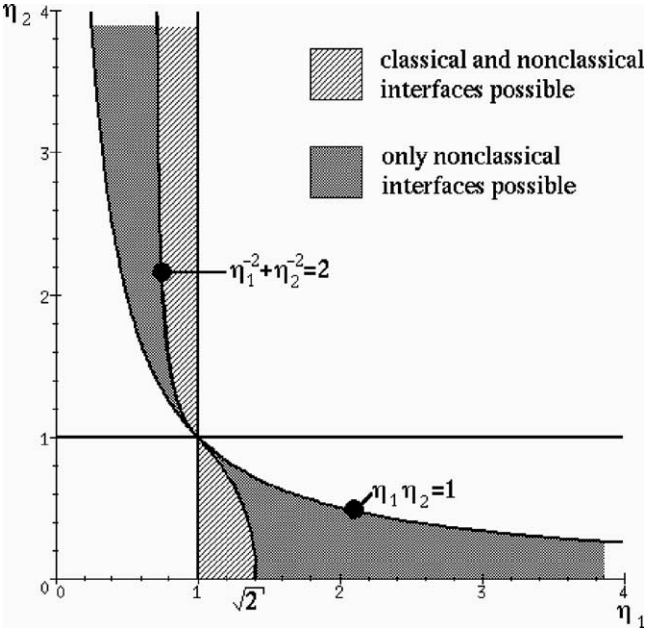

Fig. 4. Deformation parameters η_1, η_2 allowing classical and nonclassical austenite–martensite interfaces.

5. Towards a more predictive theory

Although much new and useful information can be obtained from the theory described above, it is not truly predictive of the microstructure morphology that arises in a given experimental situation. For example, in the analyses of Sections 4.1 and 4.3 the basic microstructure geometry is assumed (guided by experimental observation) and this geometry is then shown to be consistent with a zero-energy microstructure provided various quantitative relations hold. The theory does not tell us why the assumed microstructure geometry is preferred to others.

Of course, the problem of predicting microstructure geometry is a *pattern formation problem*, which can only be satisfactorily treated using appropriate dynamic equations. Such a dynamic theory should tell us what features of microstructure morphology are predictable, that is which should arise in repetitions of the same experiment, and which are not. In principle, one might expect the concepts of *dynamical systems theory*, such as invariant manifolds and attractors, to be important in such an analysis. Unfortunately, to carry out such a programme at present seems out of reach, firstly because it is not clear what dynamic equations to use (a key issue being the laws governing the motion of interfaces—see Bhattacharya et al. [22]), and secondly because the mathematical analysis of such systems is poorly developed. For example, the author is not aware of a satisfactory three-dimensional treatment of motion of a single austenite martensite interface.

Nevertheless, there are examples for which the static theory can be made somewhat more predictive, or give some limited insight into the dynamic formation of patterns of microstructure, and we discuss two of these.

5.1. Nonclassical austenite–martensite interfaces

Is there a reason why simple laminates of martensite are preferred in austenite–martensite interfaces, or could a more complicated geometry arise, for example a double laminate (layers within layers structure) or a fractal-like microstructure which refines in a self-similar way as the habit plane or surface is approached? From a mathematical perspective, one difference between the double laminate case and the fractal one is that the former can be represented by a gradient Young measure $(\nu_x)_{x\in\Omega}$ that is independent of x in the martensite, whereas the latter corresponds to ν_x varying as the habit plane or surface is approached. It turns out that for cubic to tetragonal transformations the former case can be analyzed. The analysis, which does not make any extra assumptions about the microstructure morphology in the martensite, reduces to determining the possible rank-one connections between SO(3) and the quasiconvexification $K(\theta)^{\mathrm{qc}}$, where $K(\theta) = \cup_{i=1}^{3}\mathrm{SO}(3)U_i$. Even though $K(\theta)^{\mathrm{qc}}$ is unknown in this case, it turns out to be possible to determine the values of the deformation parameters η_1, η_2 for which such rank-one connections are possible. In fact (see Ball and Carstensen [6,7]), $\mathbf{1} + b \otimes m \in K(\theta)^{\mathrm{qc}}$ for some b, m if and only if $\eta_{\min} \le \eta_{\mathrm{mid}}^{-1} \le \eta_{\max}$, where $\eta_{\min} \le \eta_{\mathrm{mid}} \le \eta_{\max}$ are the numbers η_1, η_1, η_2 in nondecreasing order (so that $\eta_{\mathrm{mid}} = \eta_1$). This region is shown in Fig. 4. Note that in the neighbourhood of the point $\eta_1 = \eta_2 = 1$ the region for which only nonclassical interfaces are possible is cusped, suggesting that one is more likely to observe a classical interface with simply-laminated martensite for such deformation parameters. However, the situation is not quite as simple as this, since even in the region corresponding to classical interfaces planar nonclassical interfaces are possible with a much larger set of normals m than the 24 allowed for classical interfaces. Thus, a nonclassical interface might be preferred if the conditions of the experiment (such as the orientation of a temperature gradient) favoured one of these extra normals. Whether these more complicated austenite-martensite interfaces will arise in suitable experiments, or whether they are disfavoured for reasons outside the model, such as interfacial energy or dynamic effects, remains to be determined.

5.2. Macrotwin formation by coalescence of martensitic plates

Boullay and Schryvers [23] made HRTEM observations of macrotwins in quenched $Ni_{65}Al_{35}$ polycrystals (Fig. 5).

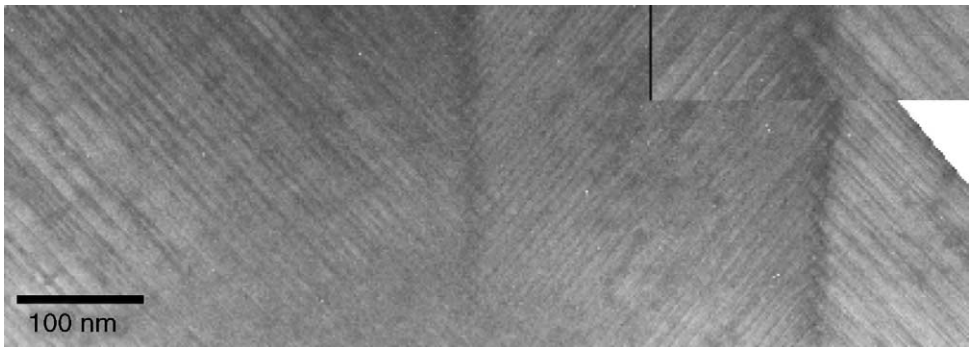

Fig. 5. Low magnification image of crossing-type macrotwins. The insert shows details of the crossing at twice the magnification. Bands of different grey levels correspond to different variants U_1 and U_2.

Table 1
Rotations Q_1 and Q_2 that bring Plate II into compatibility with Plate I ($\kappa_1 = \chi_1 = \nu_1 = 1$) and the corresponding macrotwin normals N_1 and N_2

Parameter values			Q_1			Q_2		
κ_2	χ_2	ν_2	Axis	Angle (°)	N_1	Axis	Angle (°)	N_2
−1	1	1	(0.70, 0, −0.71)	1.64	(0, 1, 0)	(0.75, 0, 0.66)	1.75	(1, 0, 0)
−1	−1	1	(0, 0.99, 0.16)	7.99	(1, 0, 0)	(0, 0.99, −0.14)	7.99	(0, 1, 0)
−1	1	−1	(0.65, 0.48, −0.59)	6.76	(0.59, −0.81, 0)	(0.68, 0.50, 0.54)	6.91	(−0.81, −0.59, 0)
−1	−1	−1	(−0.48, 0.65, 0.59)	6.76	(−0.81, −0.59, 0)	(−0.50, 0.68, −0.54)	6.91	(0.59, −0.81, 0)
1	1	−1	(−0.54, 0.54, 0.64)	5.87	$\frac{1}{\sqrt{2}}(1, 1, 0)$	(−0.57, 0.57, −0.59)	6.08	$\frac{1}{\sqrt{2}}(1, -1, 0)$
1	−1	−1	(0.60, 0.60, −0.52)	7.37	$\frac{1}{\sqrt{2}}(1, -1, 0)$	(0.62, 0.62, 0.47)	7.47	$\frac{1}{\sqrt{2}}(1, 1, 0)$

The direction of rotation is that of a right-handed screw in the direction of the given axis. For the case $\kappa_2 = \nu_2 = 1$, $\chi_2 = -1$ see the text.

The macrotwin interfaces separate a pair of simple laminates of martensite. Although it is natural to believe that these macrotwins arise via coalescence of impinging martensitic plates, such a time evolution is too fast to observe. Does the static theory nevertheless provide any supporting evidence for this scenario, or other insights? This alloy undergoes a cubic (bcc) to tetragonal (bct) transformation. For macrotwins involving just two of the three tetragonal variants in the two contingent laminates, the corresponding microtwin planes are close to orthogonal. Now, the rank-one connections required for the wedge microstructure are exactly the same as those needed for two martensitic plates to be simultaneously compatible with the austenite and with each other across the macrotwin plane. Thus, although the special relationship (23) holds approximately for $Ni_{65}Al_{35}$, and macrotwins involving all three variants with microtwins meeting at nearly 120° are seen, Bhattacharya's analysis means that when only two variants are involved the corresponding martensitic plates are never compatible at zero energy. Note that the precise structure at the macrotwin interface does not affect this conclusion since the jump condition (9) holds under the condition (10), provided that the deformation is describable by elasticity. However, it turns out that the plates can be nearly compatible. Without loss of generality we can assume that the first plate, Plate I, is given by the choice of parameters $\kappa_1 = \nu_1 = \chi_1 = 1$ in (21). Then any other distinct plate, Plate II, with parameters κ_2, ν_2, χ_2, is compatible with Plate I if subjected to a small prior rigid rotation Q, i.e.

$$\mathbf{1} + b_1 \otimes m_1 = Q(\mathbf{1} + b_2 \otimes m_2) + c \otimes N \tag{24}$$

for vectors c, N, $|N| = 1$. In the case $\kappa_2 = \nu_2 = 1$, $\chi_2 = -1$ the two plates have macroscopic gradients which are rigid rotations of each other, so that there this only one solution Q to (24) with $c = 0$. In all other cases there are exactly two such rotations Q_1, Q_2 with corresponding vectors c_1, N_1 and c_2, N_2 respectively. The rotations Q_i and normals N_i are shown in Table 1 taken from Ball and Schryvers [14].

Note the orthogonality of N_1, N_2, a consequence of Theorem 2.1. From the table, we see that the case when the plates are the most compatible, i.e. for which the angle of rotation is the least, is when $\kappa_2 = -1$, $\chi_2 = 1$, $\nu_2 = 1$, which is thus expected to be preferred on energetic grounds. In fact, the macrotwin normals observed for 'crossing-type' macrotwins correspond exactly to those in the table for this case. Further, the larger angles of rotation for the case $\nu_2 = -1$ of reversed volume fractions suggests why this situation was not seen by Boullay and Schryvers. In order to achieve compatibility at the macrotwin plane, the microtwins bend slightly as they approach this plane. The corresponding angles and directions of rotations agree well qualitatively, and fairly well quantitatively, with those observed, lending further confirmation to the coalescence scenario. For more details the reader is referred to Boullay et al. [23], and Ball and Schryvers [15,14]. Much remains to be done to complete our understanding of these macrotwins and their genesis.

Acknowledgements

This research was supported by the EC TMR Contract FMRX-CT98-0229 on 'Phase Transitions in Crystalline Solids'.

References

[1] R. Adams, J.J.F. Fournier, Sobolev Spaces, Academic Press, second edition, 2003.
[2] J.M. Ball, Global invertibility of Sobolev functions and the interpenetration of matter, Proc. R. Soc. Edinburgh 88A (1981) 315–328.
[3] J.M. Ball, A version of the fundamental theorem for Young measures, in: M. Rascle, D. Serre, M. Slemrod (Eds.), Proceedings of the Conference on 'Partial Differential Equations and Continuum Models of Phase Transitions', Springer Lecture Notes in Physics, No. 359, 1989, pp. 3–16.
[4] J.M. Ball, Some open problems in elasticity, in: Geometry, Mechanics, and Dynamics, Springer, New York, 2002, pp. 3–59.
[5] J.M. Ball, C. Carstensen, Hadamard's compatibility condition for microstructures, in preparation.
[6] J.M. Ball, C. Carstensen, Nonclassical austenite–martensite interfaces, J. Phys. IV (France) 7 (C5) (1997) 35–40.
[7] J.M. Ball, C. Carstensen, Compatibility conditions for microstructures and the austenite–martensite transition, Mater. Sci. Eng. A 273–275 (1999) 231–236.
[8] J.M. Ball, C. Chu, R.D. James, Metastability and martensite, in preparation.
[9] J.M. Ball, C. Chu, R.D. James, Hysteresis during stress-induced variant rearrangement, J. Phys. IV C8 (1995) 245–251.

[10] J.M. Ball, R.D. James, From Microscales to Macroscales in Materials, in preparation.
[11] J.M. Ball, R.D. James, Incompatible sets of gradients and metastability, in preparation.
[12] J.M. Ball, R.D. James, Fine phase mixtures as minimizers of energy, Arch. Rational Mech. Anal. 100 (1987) 13–52.
[13] J.M. Ball, R.D. James, Proposed experimental tests of a theory of fine microstructure, and the two-well problem, Phil. Trans. R. Soc. Lond. A 338 (1992) 389–450.
[14] J.M. Ball, D. Schryvers, The analysis of macrotwins in NiAl martensite, J. Phys. IV 112 (2003) 159–162.
[15] J.M. Ball, D. Schryvers, The formation of macrotwins in NiAl martensite, in: Q.-P. Sun (Ed.), Proceedings of IUTAM Symposium on Mechanics of Martensitic Phase Transformation in Solids, Hong Kong, 2001, Kluwer, 2002, pp. 27–36.
[16] K. Bhattacharya, Wedge-like microstructure in martensites, Acta Metall. Mater. 39 (1991) 2431–2444.
[17] K. Bhattacharya, Self-accommodation in martensite, Arch. Rational Mech. Anal. 120 (1992) 201–244.
[18] K. Bhattacharya, Comparison of geometrically nonlinear and linear theories of martensitic transformation, Continuum Mech. Thermodyn. 5 (1993) 205–242.
[19] K. Bhattacharya, Microstructure of Martensite, Oxford University Press, 2003.
[20] K. Bhattacharya, R.D. James, A theory of thin films of martensitic materials with applications to microactuators, J. Mech. Phys. Solids 47 (1999) 531–576.
[21] K. Bhattacharya, R.V. Kohn, Symmetry, texture and the recoverable strain of shape-memory polycrystals, Acta Mater. 44 (1996) 529–542.
[22] K. Bhattacharya, P. Purohit, B. Craciun, Mobility of twin and phase boundaries, J. Phys. IV 112 (2003) 163–166.
[23] Ph. Boullay, D. Schryvers, J.M. Ball, Nano-structures at martensite macrotwin interfaces in bulk and splat-cooled $Ni_{65}Al_{35}$, Acta Mater. 51 (2003) 1421–1436.
[24] J.S. Bowles, J.K. MacKenzie, The crystallography of martensitic transformations i and ii, Acta Met. 2 (1954) 129–147.
[25] P.G. Ciarlet, J. Nečas, Unilateral problems in nonlinear three-dimensional elasticity, Arch. Rational Mech. Anal. 87 (1985) 319–338.
[26] B. Dacorogna, Quasiconvexity and relaxation of non convex variational problems, J. Funct. Anal. 46 (1982) 102–118.
[27] A. DeSimone, R.D. James, A constrained theory of magnetoelasticity, J. Mech. Phys. Solids 50 (2002) 283–320.
[28] G. Dolzmann, Variational Methods for Crystalline Microstructure—Analysis and Computation, volume 1803 of Lecture Notes in Mathematics, Springer-Verlag.
[29] J.L. Ericksen, Special topics in elastostatics, in: C.-S. Yih (Ed.), Advances in Applied Mechanics, vol. 17, Academic Press, 1977, pp. 189–244.
[30] I. Fonseca, The lower quasiconvex envelope of the stored energy function of an elastic crystal, J. Math. Pures Appl. 67 (1988) 175–195.
[31] A. Forclaz, A simple criterion for the existence of rank-one connections between martensitic wells, J. Elasticity 57 (1999) 281–305.
[32] A. Forclaz, Variational methods in materials science, PhD thesis, University of Oxford, 2002.
[33] T. Iwaniec, G.C. Verchota, A.L. Vogel, The failure of rank-one connections, Arch. Ration. Mech. Anal. 163 (2002) 125–169.
[34] R.D. James, D. Kinderlehrer, Theory of magnetostriction with applications to $Tb_xDy_{1-x}Fe_2$, Philos. Mag. B 68 (1993) 237–274.
[35] A.G. Khachaturyan, Some questions concerning the theory of phase transformations in solids, Soviet Phys. Solid State 8 (1967) 2163–2168.
[36] A.G. Khachaturyan, Theory of Structural Transformations in Solids, John Wiley, 1983.
[37] A.G. Khachaturyan, G.A. Shatalov, Theory of macroscopic periodicity for a phase transition in the solid state, Soviet Phys. JETP 29 (1969) 557–561.
[38] D. Kinderlehrer, P. Pedregal, Characterizations of Young measures generated by gradients, Arch. Rational Mech. Anal. 115 (1991) 329–365.
[39] D. Kinderlehrer, P. Pedregal, Gradient Young measures generated by sequences in Sobolev spaces, J. Geom. Anal. 4 (1994) 59–90.
[40] R.V. Kohn, B. Niethammer, Geometrically nonlinear shape-memory polycrystals made from a two-variant material, M2AN Math. Model. Numer. Anal. 34 (2000) 377–398.
[41] J. Kristensen, On the non-locality of quasiconvexity, Ann. Inst. H. Poincaré, Anal. Non Linéaire 16 (1999) 1–13.
[42] M. Luskin, Computational modeling of microstructure, in: Proceedings of the International Congress of Mathematicians, vol. III, Beijing, Higher Ed. Press, 2002, pp. 707–716.
[43] C.B. Morrey, Quasi-convexity and the lower semicontinuity of multiple integrals, Pacific J. Math. 2 (1952) 25–53.
[44] C.B. Morrey, Multiple Integrals in the Calculus of Variations, Springer, 1966.
[45] S. Müller, Variational methods for microstructure and phase transitions, in: Calculus of Variations and Geometric Evolution Problems, volume 1713 of Lecture Notes in Mathematics, Springer, Berlin, 1999, pp. 85–210.
[46] M. Ortiz, Variational problems in mechanics and the link between microstructure and microscopic behavior, in: Proceedings of ICIAM 2003, Sydney.
[47] R. Paroni, Homogenization of polycrystalline aggregates, Arch. Ration. Mech. Anal. 151 (2000) 311–337.
[48] M. Pitteri, G. Zanzotto, Non-generic twinning concentrations for a class of active alloys, in: Applied and Industrial Mathematics, Venice, 1998, Kluwer Academic Publishers, Dordrecht, 2000, pp. 245–257.
[49] M. Pitteri, G. Zanzotto, Continuum Models for Phase Transitions and Twinning in Crystals, Chapman & Hall/CRC Press, 2003.
[50] A.L. Roitburd, Kristallografiya, 1967, p. 567 ff. (in Russian).
[51] A.L. Roitburd, Martensitic transformation as a typical phase transformation in solids, Solid State Phys. 33 (1978) 317–390.
[52] V. Šverák, Regularity properties of deformations with finite energy, Arch. Rational Mech. Anal. 100 (1988) 105–127.
[53] V. Šverák, Rank-one convexity does not imply quasiconvexity, Proc. R. Soc. Edinburgh 120A (1992) 185–189.
[54] L. Tartar, Compensated compactness and applications to partial differential equations, in: R.J. Knops (Ed.), Nonlinear Analysis and Mechanics, Heriot–Watt Symposium, vol. IV, Pitman Research Notes in Mathematics, 1979, pp. 136–192.
[55] M.S. Wechsler, D.S. Lieberman, T.A. Read, On the theory of the formation of martensite, Trans. AIME J. Met. 197 (1953) 1503–1515.
[56] L.C. Young, Lectures on the Calculus of Variations and Optimal Control Theory, Saunders, 1969. Reprinted by A.M.S. Chelsea, 1969.