Fine Phase Mixtures as Minimizers of Energy

Dedicated to James Serrin

J. M. BALL & R. D. JAMES

Table of Contents

1. Introduction

Solid-solid phase transformations often lead to certain characteristic microstructural features involving fine mixtures of the phases. In martensitic transformations one such feature is a plane interface which separates one homogeneous phase, austenite, from a very fine mixture of twins of the other phase, martensite. In quartz crystals held in a temperature gradient near the α-β transformation temperature, the α-phase breaks up into triangular domains called Dauphiné twins which become finer and finer in the direction of increasing temperature. In this paper we explore a theoretical approach to these fine phase mixtures based on the minimization of free energy.

In simplified terms the idea is the following. Suppose that for energetic reasons, a body prefers to be deformed, say, in three states specified by three constant deformation gradients $\mathbf{1}$, $\boldsymbol{F}^-$ and $\boldsymbol{F}^+$. Assume that conditions of geometric compatibility are satisfied across an interface separating regions deformed with gradients F^+ and F^-, *i.e.* that there are vectors a and n such that

$$F^+ - F^- = a \otimes n,$$

but that compatibility cannot be maintained across an interface separating $\mathbf{1}$ and F^+ or $\mathbf{1}$ and F^-, *i.e.*

$$F^\pm - \mathbf{1} \neq \text{a rank-one matrix}.$$

Thus, while it is possible to construct a continuous piecewise affine deformation consisting of layers having deformation gradients $F^+/F^-/F^+/F^-, \ldots$, it is not possible to construct a continuous piecewise affine deformation using all three matrices $\mathbf{1}$, F^- and F^+. However, we show that it is possible (for certain choices of F^+ and F^-) to arrange a very fine mixture of the layers $F^+/F^-/F^+/F^-, \ldots$, on one side of an appropriately oriented interface so that the "average" deformation gradient of these layers does approximately satisfy conditions of compatibility with $\mathbf{1}$. The approximation gets better as the distribution of layers gets finer and finer. We argue that this is the essential reason for fineness in some martensitic transformations.

The energetic interpretation of these configurations is in terms of minimizing sequences rather than minimizers. In fact, each of the minimizing sequences we study converges weakly to a deformation which is not itself a minimizer of the total free energy. Thus, the total free energy functional is not lower semicontinuous with respect to weak convergence in the Sobolev space $W^{1,p}(\Omega, \mathbb{R}^3)$, $p \geqq 1$. Our calculations in Section 5 show that this failure of lower semicontinuity is a typical property of free energy functionals for solids which change phase and results from a failure of ellipticity of these functionals. From the point of view of comparison of theory with experiment, the detailed structure of minimizing sequences appears to be as important in these problems as the minimizers.

In fact, it is well known from the pioneering work of L. C. YOUNG [52] that, in the absence of ellipticity conditions, integrals of the calculus of variations do not attain a minimum among ordinary functions, but can be thought of as attaining a minimum in a space of "generalized curves". Such generalized curves are the limits of minimizing sequences that necessarily oscillate more and more finely. The finely twinned configurations of martensite described above can be viewed as approximations of generalized curves. Another example from elasticity is the "infinitesimal wrinkling" of membranes studied by PIPKIN [38].

Our calculations are related to those involved in what is known as the *crystallographic theory of martensite* in the metallurgical literature and to emerging methods of *homogenization theory* in the mathematical literature. Treatments of the crystallographic theory of martensitic transformations are found in the books by CHRISTIAN [15], NISHIYAMA [35], and WAYMAN [49]. The theory was first put forth by BOWLES & MACKENZIE [11] and WECHSLER, LIEBERMAN & READ [50]. Our calculations of Section 5 are similar to those of the crystallographic theory. However, by developing the theory on the basis of a free energy minimization, we achieve several advantages. First, by looking at minimizers and minimizing sequences, we predict both the twinned martensite interface, with the observed twin planes, and the austenite/finely twinned martensite interface. Along the way, we clarify the role of fineness in energetic terms. Also, since our free energy accounts for general three-dimensional changes of shape, it can be used in conjunction with various loading potentials to study the effect of multiaxial loads on transformation temperatures.

There is some similarity between the observed geometrical configurations of martensite and the arrangement of constituent materials used to achieve optimal bounds and designs in homogenization theory*. For a striking example of this similarity, compare Figure 1 of a recent paper by MILTON [32] with the photographs of a "mishandled" crystal of InTl shown by BASINSKI & CHRISTIAN [10, plate III]. Our problem is different in that the material itself makes the "composite". Another major difference is that compatibility does not play an essential role in homogenization theory, while the fineness in our configurations is a consequence of the material striving to achieve compatibility.

We conclude the paper with some different examples of fineness in energy minimizers. In Section 7a we give an example which, while not applying directly to martensitic transformations because it does not satisfy the appropriate invariance requirements, suggests strongly that fine twinning can be initiated by a temperature or concentration gradient, this leading to a minimization problem which when interfacial energy is neglected only has a minimizer in the sense of generalized curves. In Section 7b we give an example of a strongly elliptic material which has potential wells and which supports configurations with very fine boundary wrinkles. In the Section 7c we return to the observations of Dauphiné twinning in quartz. We give an example of a configuration involving five deformation gradients in which compatibility is achieved by a self-similar mixing of smaller and smaller triangles on one side of an interface. We believe that this configuration is related to fine triangular domains in quartz observed by VAN TENDELOO, VAN LANDUYT & AMELINCKX [48], but our example is based on a simplified free energy function and therefore we are not able to make a quantitative comparison.

2. Internally Twinned Martensite

Our approach in Section 2 through Section 5 is suggested by observations of internally twinned martensite. One of the most studied of the alloys which form internally twinned martensite, because of its accessibility to low power optical microscopy and its simple crystal structure, is Indium–Thallium. The alloy consists of a substitutional solid solution of Tl in a crystalline matrix of In which at high temperature is a face-centered cubic.

If a single crystal of InTl is cooled to its transformation temperature (105 °C for In-18.5% Tl, 25 °C for In-23% Tl), it undergoes a diffusionless reversible change from a face-centered cubic to a face-centered tetragonal structure. The transformation is made evident by the movement of one or more interfaces across the

* See, for example, KLOSOWICZ & LURIE [27], KOHN & STRANG [28], LAVROV, LURIE & CHERKAEV [29], LURIE, CHERKAEV & FEDEROV [30], MILTON [32], MURAT & TARTAR [33], RAITUM [40] and TARTAR [45, 46].

specimen. A typical observation at the transformation temperature, redrawn from the photomicrograph of BASINSKI & CHRISTIAN [10, Figure 5], is shown in Figure 1. The cubic austenitic phase (to the right in Figure 1) is stable above the transformation temperature while the tetragonal martensite is stable below the transformation temperature. The twin spacing in the martensite is on the order of 20 μm.

Sometimes the transformation produces a more complicated arrangement of the phases; a fairly common observation is the X-shaped interface in Figure 2a (BASINSKI & CHRISTIAN [10, Figures 11–15], BURKART & READ [13, Figure 3]).

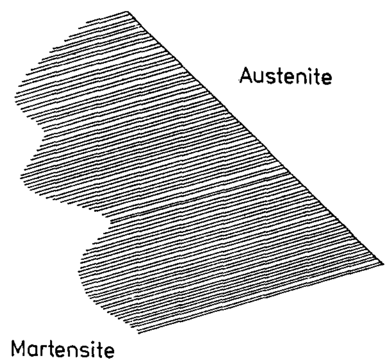

Fig. 1. Single interface transformation in InTl

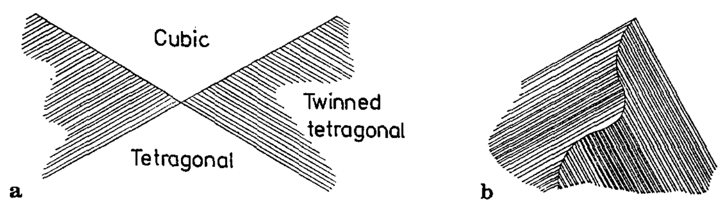

Fig. 2a. Transformation by an X-interface; b Curved martensite/martensite interface

At the bottom of Figure 2a is a single crystal of martensite. The single and X-interfaces are planar, but intriguing curved interfaces which separate twinned martensite from twinned martensite also are seen (BURKART & READ [13, Figure 3 and Figure 2b]). If the temperature is lowered further the phase boundaries move so as to eliminate all the austenite, often leaving a twinned crystal of martensite. Austenite and martensite co-exist in a crystal over a temperature range of about $3\frac{1}{2}$°C so actually there is not a single transformation temperature.*

In all cases the twin planes arise from the $\{1\ 1\ 0\}$ family of planes in the austenite. That is, if we adopt a reference configuration interpreted as the undistorted

* We return to this observation in Section 6.

austenite just above the transformation temperature, the six planes in the reference configuration with normals (1 1 0), (1 0 1), (0 1 1), (1 −1 0), (−1 0 1), (0 1 −1) relative to a basis parallel to the cubic axes are deformed into the twin planes by the transformation. Within each twin band the material is a tetragonal single crystal, but neighboring bands are oriented differently. Also, all the interfaces separating austenite from finely twinned martensite are observed to be very nearly {1 1 0} planes. Where the austenite/martensite interface meets the boundary of the body, this boundary bends sharply through a small angle.

Below the transformation temperature, the body is a single or partly twinned crystal of martensite. A partly twinned crystal is extremely flexible in that small applied loads easily change the spacing of the twinned layers. Below the transformation temperature, the general tendency of uniform loads on the faces of a crystal is to drive the twins out and leave a single crystal of martensite.

If an unloaded crystal of martensite, either twinned or not, is heated to the transformation temperature, the austenite/martensite interfaces reappear. Further heating to above the transformation temperature causes the crystal to return to a single untwinned crystal of austenite.

3. The Free Energy Functional and Minima

We now propose a free energy for materials which undergo reversible martensitic transformations and work out the details for InTl.

The change in crystal structure associated with the transformation in InTl is from face-centered cubic to face-centered tetragonal. Consider a regular reference configuration $\Omega \subset \mathbb{R}^3$ which is interpreted as the undistorted austenite at the transformation temperature $\theta_0 = \text{const}$. For later use we assume $\boldsymbol{o}$ belongs to the interior of Ω. The change of shape in going from fcc to fct can be described by a deformation $\boldsymbol{y} = \boldsymbol{U}_0\boldsymbol{x}$, $\boldsymbol{x} \in \Omega$, $\boldsymbol{U}_0$ being the constant positive-definite symmetric matrix given by

$$\boldsymbol{U}_0 = \eta_1 \mathbf{1} + (\eta_2 - \eta_1)\, \boldsymbol{e} \otimes \boldsymbol{e} \tag{3.1}$$

$|\boldsymbol{e}| = 1$, for some positive constants $\eta_1 \neq \eta_2$ (Burkart & Read [13]). For InTl $\eta_1 \doteq 1 - \varepsilon$, $\eta_2 \doteq 1 + 2\varepsilon$ with $\varepsilon \doteq .013$ for concentrations near 20% Tl. We refer to $\boldsymbol{U}_0$ as the *transformation strain*.

Since the transformation leads to a change of shape at a certain temperature, we are led to assume the existence of a free energy ϕ which depends on the change of shape, measured by the deformation gradient $\boldsymbol{F}$, and the temperature θ. Thus at a certain temperature θ let a deformation $\boldsymbol{y}\colon \Omega \to \mathbb{R}^3$ with gradient $\boldsymbol{F} = D\boldsymbol{y}(\boldsymbol{x})$ have a free energy per unit volume in Ω given by

$$\phi(\boldsymbol{F}, \theta). \tag{3.2}$$

Assume ϕ is defined and continuous for all $\boldsymbol{F} \in \mathcal{D} = \{\boldsymbol{F} \in M^{3\times 3} \mid \det \boldsymbol{F} > 0\}$ and for all θ in a neighborhood of the transformation temperature θ_0. Here $M^{m\times n}$ denotes the set of real $m \times n$ matrices.

We assume that the free energy is Galilean invariant: for all $F \in \mathscr{D}$, all θ near θ_0, and each orthogonal R with $\det R = 1$ (we call such R *rotations*),

$$\phi(RF, \theta) = \phi(F, \theta). \tag{3.3}$$

The restriction (3.3) implies that

$$\phi(F, \theta) = \phi(U, \theta)\big|_{U=(F^TF)^{\frac{1}{2}}}. \tag{3.4}$$

Let $\mathscr{D}^s$ be the subset of $\mathscr{D}$ consisting of positive definite symmetric matrices. Since we have interpreted Ω as undistorted austenite at the temperature θ_0, we assume that there is a finite group of rotations P^ν of order ν, representing the symmetry of austenite, such that

$$\phi(RUR^T, \theta) = \phi(U, \theta) \tag{3.5}$$

holds for all $U \in \mathscr{D}^s$, all $R \in P^\nu$ and all θ near θ_0. For the InTl alloy $\nu = 24$ and P^{24} consists of the 24 rotations which map a cube into itself. Also, one of the 4-fold rotations in P^{24} has the axis e of (3.1).

In an unloaded body the austenite is observed above and the martensite below the transformation temperature. We shall therefore presume an exchange of stability, in the sense that for all U in $\mathscr{D}^s$ which are unequal to $\mathbf{1}$, U_0 or any matrix of the form RU_0R^T, $R \in P^\nu$,

$$\phi(U, \theta_0) > \phi(\mathbf{1}, \theta_0) = \phi(U_0, \theta_0). \tag{3.6}$$

For $\theta > \theta_0$, we assume that some symmetric matrix $U_a(\theta)$ near $\mathbf{1}$ minimizes $\phi(\cdot, \theta)$, whereas for $\theta < \theta_0$, we assume that some symmetric matrix $U_m(\theta)$ near U_0 minimizes $\phi(\cdot, \theta)$. Here, nearness means that $|U_a - \mathbf{1}| \ll |U_0 - \mathbf{1}|$ and $|U_m - U_0| \ll |U_0 - \mathbf{1}|$. We use the notation $|A| = (\operatorname{tr} AA^T)^{\frac{1}{2}}$ for any $A \in M^{m\times n}$. See JAMES [23] for a fuller discussion of these kinds of energy functions. Note that this assumption means that the symmetry group of ϕ is smaller than that considered by ERICKSEN [16] and FONSECA [20], and, in particular, does not contain certain nontrivial shears.

In general, we have assumed that $\phi(\cdot, \theta_0)$ has (up to) $\nu + 1$ potential wells with minima at the matrices $\mathbf{1}, R_1U_0R_1^T, \ldots, R_\nu U_0 R_\nu^T$, in which $R_1, \ldots, R_\nu$ is an enumeration of the point group P^ν. Each distinct potential well with minimum of the form $R_iU_0R_i^T$ is associated with a *variant* of the martensite. For U_0 of the form (3.1) with $\eta_1 \neq \eta_2$, there are only three variants because some of the matrices of the form $R_iU_0R_i^T$ coincide.

The condition $U_0 = R_iU_0R_i^T$ in InTl is

$$\eta_1 \mathbf{1} + (\eta_2 - \eta_1) R_ie \otimes R_ie = \eta_1 \mathbf{1} + (\eta_2 - \eta_1)\, e \otimes e \tag{3.7}$$

or simply

$$R_ie = \pm e \tag{3.8}$$

which is satisfied by eight members of P^{24} (three rotations with axis e, four 180° rotations with axes perpendicular to e and the identity). Since P^{24} is a group, there are precisely 24/8 = 3 distinct matrices of the form $R_iU_0R_i^T$, and it is easily seen that these are $\eta_1 \mathbf{1} + (\eta_2 - \eta_1)\, \hat{e} \otimes \hat{e}$, $\hat{e}$ being a 4-fold axis of P^{24}.

To make a connection with calculations of the effect of stress on transformation temperatures in [25], we observe that **1** can be written as the convex combination

$$\tfrac{1}{3}\,\boldsymbol{U}_0 + \tfrac{1}{3}\,\boldsymbol{R}\boldsymbol{U}_0\boldsymbol{R}^T + \tfrac{1}{3}\,\boldsymbol{R}^2\boldsymbol{U}_0\boldsymbol{R}^{2T}, \tag{3.9}$$

$\boldsymbol{R}$ being any 3-fold rotation in P^{24}. Thus, in the notation of [25], $\mathbf{1} \in \delta\mathscr{H}_0$.

We define the total energy functional by

$$\mathscr{I}[y] = \int_\Omega \phi(Dy(\boldsymbol{x}), \theta_0)\, d\boldsymbol{x}. \tag{3.10}$$

(We shall only be concerned with stable configurations at the transformation temperature.) A deformation will be termed *stable* if it minimizes the total free energy. In precise terms the deformation

$$\tilde{\boldsymbol{y}} \in \mathscr{A} = \{\boldsymbol{y} \in W^{1,1}(\Omega, \mathbb{R}^3) \mid D\boldsymbol{y} \in \mathscr{D} \text{ a.e.}\}$$

is stable if

$$\mathscr{I}[\tilde{y}] \leqq \mathscr{I}[y] \quad \forall \boldsymbol{y} \in \mathscr{A}. \tag{3.11}$$

This stability criterion is appropriate for an unloaded body at the transformation temperature. Here and below $W^{1,p}(\Omega, \mathbb{R}^m)$ denotes the Sobolev space of mappings $\boldsymbol{y} : \Omega \to \mathbb{R}^m$ such that $\|\boldsymbol{y}\|_{1,p} < \infty$, where

$$\|\boldsymbol{y}\|_{1,p} = \begin{cases} \left(\int_\Omega (|\boldsymbol{y}|^p + |D\boldsymbol{y}|^p)\, d\boldsymbol{x}\right)^{1/p}, & 1 \leqq p < \infty, \\ \operatorname*{ess\,sup}_{\boldsymbol{x} \in \Omega} (|\boldsymbol{y}(\boldsymbol{x})| + |D\boldsymbol{y}(\boldsymbol{x})|), & p = \infty. \end{cases}$$

See ADAMS [1] for information on these spaces. We write $W^{1,p}(\Omega)$ for $W^{1,p}(\Omega, \mathbb{R}^1)$.

We now describe all stable deformations. A necessary and sufficient condition that $\tilde{\boldsymbol{y}}$ be stable is that $D\tilde{\boldsymbol{y}}(\boldsymbol{x})$ minimize the integrand $\phi(\cdot, \theta_0)$ for almost every $\boldsymbol{x}$. Because of the property (3.6), $D\tilde{\boldsymbol{y}}(\boldsymbol{x})$ minimizes the integrand if and only if for almost every $\boldsymbol{x}$ in Ω, the function defined by

$$\tilde{\boldsymbol{U}}(\boldsymbol{x}) = (D\tilde{\boldsymbol{y}}(\boldsymbol{x})^T\, D\tilde{\boldsymbol{y}}(\boldsymbol{x}))^{\frac{1}{2}} \tag{3.12}$$

takes on one of the values

$$\mathbf{1},\ \boldsymbol{R}_1\boldsymbol{U}_0\boldsymbol{R}_1^T, \ldots, \boldsymbol{R}_\nu\boldsymbol{U}_0\boldsymbol{R}_\nu^T. \tag{3.13}$$

A geometric characterization of all $\tilde{\boldsymbol{y}} \in \mathscr{A}$ satisfying the preceding condition does not appear to be available in the literature. Thus, we focus first on smooth interfaces separating austenite from itself, martensite from itself or martensite from austenite.

Let $\tilde{\boldsymbol{y}}$ be a stable, continuous and piecewise differentiable deformation and suppose a smooth interface separates two regions on which $D\tilde{\boldsymbol{y}}$ has constant values $\boldsymbol{F}^+$ and $\boldsymbol{F}^-$. It is well known that if $\boldsymbol{F}^+ \neq \boldsymbol{F}^-$ then the interface is a plane (with say reference normal $\boldsymbol{n}$, $|\boldsymbol{n}| = 1$) and that for some nonzero vector $\boldsymbol{a}$

$$\boldsymbol{F}^+ - \boldsymbol{F}^- = \boldsymbol{a} \otimes \boldsymbol{n}. \tag{3.14}$$

Let $F^+ = R^+U^+$ and $F^- = R^-U^-$ be the polar decompositions of F^+ and F^-. Since $\tilde{y}$ is stable, U^+ and U^- must each take on one of the values given in (3.13). We work out the possible interfaces below for U_0 given by (3.1).

(i) *Austenite/Austenite Interfaces.* These are governed by the condition

$$R^+ - R^- = a \otimes n, \quad a \neq 0, \tag{3.15}$$

which implies that

$$R = 1 + a' \otimes n, \tag{3.16}$$

where $R = R^{-T}R^+$ and $a' = R^{-T}a$. Equation (3.16) shows that R has two linearly independent axes ($\perp$ to n) which in turn implies that $R = 1$ and $R^+ = R^-$. Thus, there are no austenite/austenite interfaces. According to a theorem of RESHETNYAK [41, Corollary of Lemma 3], if $y \in W^{1,\infty}(\Omega, \mathbb{R}^3)$ is such that $Dy(x)$ is a rotation for almost all $x \in \Omega$ then Dy is necessarily a constant rotation. This implies the stronger result that the body cannot be inhomogeneously deformed in the austenite phase. Obviously, this result is independent of the choice of point group or transformation strain.

(ii) *Martensite/Martensite Interfaces.* These are governed by the equation

$$\hat{R}^+U_0R_i - \hat{R}^-U_0R_j = \hat{a} \otimes \hat{n}, \tag{3.17}$$

with $\hat{a} \neq 0$ and $|\hat{n}| = 1$.

Premultiply (3.17) by $\hat{R}^{-T}$ and postmultiply by R_j^T. Then, (3.17) becomes

$$RU_0\bar{R} - U_0 = a \otimes n, \tag{3.18}$$

where $\bar{R} = R_iR_j^T \in P^{24}$, $R = \hat{R}^{-T}\hat{R}^+$, $a = \hat{R}^{-T}\hat{a} \neq 0$, and $n = R_j\hat{n}$. Let $\{e, e_1, e_2\}$ be an orthonormal basis with e_1 and e_2 also four-fold axes of rotation in P^{24}. $\bar{R}e$ equals one of the vectors $\pm e_1$, $\pm e_2$ or $\pm e$ because of the structure of P^{24}. The case $\bar{R}e = \pm e$ yields no solutions of (3.18) by a quick calculation using the fact that $\bar{R}^{-T}U_0\bar{R} = U_0$. (This would correspond to an interface between one variant of martensite and itself.) There are various strategies for completing the calculation. The methods of ERICKSEN [18] and GURTIN [22] can be applied to the remaining cases $\bar{R}e = \pm e_1$ or $\pm e_2$. Alternately, our Proposition 4 of Section 5 can be used with $C = U_0^{-1}\bar{R}^TU_0^2\bar{R}U_0^{-1}$. The results are:

Twins

$$n = \frac{1}{\sqrt{2}}(\bar{e} + e)$$

$$a = \frac{\sqrt{2}(\eta_2^2 - \eta_1^2)}{\eta_2^2 + \eta_1^2} q_1$$

$$\bar{R} = -1 + 2n \otimes n$$

$$R = -1 + \frac{2}{|q_2|^2} q_2 \otimes q_2 \tag{3.19}$$

in which

$$q_1 = \eta_1 \bar{e} - \eta_2 e,$$ (3.20)

$$q_2 = \eta_2 \bar{e} + \eta_1 e,$$

and

$$\bar{e} \in \{e_1, -e_1, e_2, -e_2\}.$$ (3.21)

The equations (3.19) through (3.21) give all solutions of the equation (3.18) with U_0 given, $|n| = 1$ and $\bar{R} \in P^{24}$ in the sense that a and n are given by $(3.19)_{1,2}$ up to the replacement $a \to -a$ and $n \to -n$. The expressions for R and $\bar{R}$ are not unique. The various values of R and $\bar{R}$, which give rise to the a and n of (3.19) are associated with Type I and Type II twins (normal and parallel twins in the terminology of GURTIN [22]). The twins given by (3.19) are so-called compound twins, which means that they can be represented as both type I and type II twins. The expressions of R and $\bar{R}$ in (3.19) correspond to the type I description. All solutions of the original equation (3.17) are obtained by reversing the argument which leads from (3.17) to (3.18). The twin planes are all of the $\{1\ 1\ 0\}$ family and agree exactly with the observed twin planes described in Section 2. Solutions* of (3.18) for general U_0 andf or a more general family of groups than point groups are given by ERICKSEN [17].

(iii) *Austenite/Martensite Interfaces.* These are governed by the equation

$$RU_0R_k - \mathbf{1} = a \otimes n,$$ (3.22)

with $a \neq \mathbf{0}$, $|n| = 1$, which implies that

$$R_k^T U_0^2 R_k = (\mathbf{1} + n \otimes a)(\mathbf{1} + a \otimes n),$$ (3.23)

which in turn implies that a vector perpendicular to both a and n is an eigenvector of $R_k^T U_0^2 R_k$ with eigenvalue equal to 1. But this is impossible unless one of the eigenvalues η_1 or η_2 of U_0 equals 1. In particular, there are no austenite/martensite interfaces.

The latter conclusion does not agree with observations like those shown in Figure 1, which clearly show *some kind* of austenite/martensite interface.

4. Compatibility, Almost Compatibility, and Minimizing Sequences

In this section we explore the idea that the austenite/finely twinned martensite interface is modelled by certain minimizing sequences for the total free energy (3.10). The corresponding deformations are essentially piecewise affine, but a small correction is necessary close to the interface so as to render the deformation gradients compatible.

---

* It would be necessary to consider the more general groups to describe martensitic transformations involving slip, which does not occur under small loads in the internally twinned martensites.

We begin by giving a version of the Hadamard jump condition for deformations whose gradients take only two values, there being no assumption on the structure of the sets where these values are taken.

Proposition 1. *Let* $\Omega \in \mathbb{R}^n$ *be open and connected. Let* $\boldsymbol{y} \in W^{1,\infty}(\Omega, \mathbb{R}^m)$ *satisfy*

$$\begin{aligned} D\boldsymbol{y}(\boldsymbol{x}) &= \boldsymbol{A}, \quad \text{a.e. } \boldsymbol{x} \in \Omega_A, \\ D\boldsymbol{y}(\boldsymbol{x}) &= \boldsymbol{B}, \quad \text{a.e. } \boldsymbol{x} \in \Omega_B, \end{aligned} \tag{4.1}$$

where $\boldsymbol{A}, \boldsymbol{B} \in M^{m \times n}$ *and* Ω_A, Ω_B *are disjoint measurable sets with* $\Omega = \Omega_A \cup \Omega_B$, meas $\Omega_A > 0$, meas $\Omega_B > 0$. *Then*

$$\boldsymbol{A} - \boldsymbol{B} = \boldsymbol{c} \otimes \boldsymbol{n} \tag{4.2}$$

for some $\boldsymbol{c} \in \mathbb{R}^m$, $\boldsymbol{n} \in \mathbb{R}^n$, $|\boldsymbol{n}| = 1$, *and*

$$\boldsymbol{y}(\boldsymbol{x}) = \boldsymbol{y}_0 + \boldsymbol{B}\boldsymbol{x} + \theta(\boldsymbol{x})\, \boldsymbol{c}, \qquad \boldsymbol{x} \in \Omega, \tag{4.3}$$

where $\boldsymbol{y}_0 \in \mathbb{R}^m$, $\boldsymbol{y}_0 \cdot \boldsymbol{c} = 0$, $\theta \in W^{1,\infty}(\Omega)$ *satisfies* $D\theta(\boldsymbol{x}) = \chi_A(\boldsymbol{x})\, \boldsymbol{n}$ a.e., *and* χ_A *denotes the characteristic function of* Ω_A.

Proof. Let $\boldsymbol{z}(\boldsymbol{x}) = \boldsymbol{y}(\boldsymbol{x}) - \boldsymbol{B}\boldsymbol{x}$, $\boldsymbol{C} = \boldsymbol{A} - \boldsymbol{B}$, so that $D\boldsymbol{z} = \chi_A \boldsymbol{C}$. Since χ_A is not constant, there exists $\varrho \in C_0^\infty(\Omega)$ such that

$$\boldsymbol{n} \overset{\text{def}}{=} \int_{\Omega_A} D\varrho \, d\boldsymbol{x} = \int_{\Omega} \chi_A \, D\varrho \, d\boldsymbol{x}$$

is nonzero, and clearly we may suppose that $|\boldsymbol{n}| = 1$. But

$$\begin{aligned} 0 &= \int_{\Omega} (z^i_{,\alpha}\varrho_{,\beta} - z^i_{,\beta}\varrho_{,\alpha})\, d\boldsymbol{x} \\ &= C^i_\alpha n_\beta - C^i_\beta n_\alpha, \end{aligned}$$

and hence (4.2) holds with $\boldsymbol{c} = \boldsymbol{C}\boldsymbol{n}$. To obtain (4.3) we note that if $\boldsymbol{b} \cdot \boldsymbol{c} = 0$ then $D(\boldsymbol{z}(\boldsymbol{x}) \cdot \boldsymbol{b}) = \boldsymbol{0}$ a.e., so that $\boldsymbol{z}(\boldsymbol{x}) - \boldsymbol{z}(\boldsymbol{x}_0)$ is parallel to $\boldsymbol{c}$ for a.e. $\boldsymbol{x} \in \Omega$, where $\boldsymbol{x}_0 \in \Omega$ is fixed. Assuming without loss of generality that $\boldsymbol{c} \neq \boldsymbol{0}$, it follows from (4.3) that $D\theta(\boldsymbol{x}) = \chi_A(\boldsymbol{x})\, \boldsymbol{n}$ a.e., completing the proof. □

From (4.3) we see that, on any convex subset E of Ω, $\boldsymbol{y}$ has the form

$$\boldsymbol{y}(\boldsymbol{x}) = \boldsymbol{y}_0 + \boldsymbol{B}\boldsymbol{x} + f_E(\boldsymbol{x} \cdot \boldsymbol{n})\, \boldsymbol{c}, \tag{4.4}$$

where f_E is Lipschitz with derivative 0 or 1 a.e.. Thus $\Omega_A \cap E$ and $\Omega_B \cap E$ consist of parallel layers normal to $\boldsymbol{n}$. However, if Ω is not convex, there may be no representation (4.4) with f_E independent of E; for example, Ω_A could have the form of the shaded set in Figure 3.

We next consider deformations whose gradients to a good approximation take only two values.

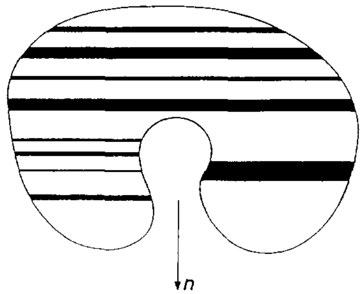

Fig. 3. Distribution of the sets Ω_A and Ω_B consistent with the hypotheses of Proposition 1

Proposition 2. *Let* $\Omega \subset \mathbb{R}^n$ *be bounded, open and connected. Let* $p > 2$, *and let* $A, B \in M^{m\times n}$ *be distinct. Let* $y^{(j)} \rightharpoonup y$ *in* $W^{1,p}(\Omega, \mathbb{R}^m)$ *and suppose that for every* $\varepsilon > 0$

$$\lim_{j\to\infty} \operatorname{meas}\{x \in \Omega: |Dy^{(j)}(x) - A| > \varepsilon \quad \text{and} \quad |Dy^{(j)}(x) - B| > \varepsilon\} = 0. \tag{4.5}$$

Then

$$Dy(x) = \lambda(x)\, A + (1 - \lambda(x))\, B, \quad \text{a.e. } x \in \Omega, \tag{4.6}$$

for some measurable function λ *satisfying* $0 \leqq \lambda(x) \leqq 1$ a.e., *and one of the following possibilities holds*:

(i) $\lambda(x) = 1$ a.e. *and* $Dy^{(j)}(x) \to A$ *in measure*,

(ii) $\lambda(x) = 0$ a.e. *and* $Dy^{(j)}(x) \to B$ *in measure*,

(iii) λ *equals neither* 0 a.e. *nor* 1 a.e. *and*

$$A - B = c \otimes n \tag{4.7}$$

for some $c \in \mathbb{R}^m$, $n \in \mathbb{R}^n$, $|n| = 1$.

Proof. Let $\Omega_A^{j,\varepsilon} = \{x \in \Omega: |Dy^{(j)}(x) - A| \leqq \varepsilon\}$, $\Omega_B^{j,\varepsilon} = \{x \in \Omega: |Dy^{(j)}(x) - B| \leqq \varepsilon\}$, and let $\chi_A^{j,\varepsilon}, \chi_B^{j,\varepsilon}$ denote the characteristic functions of $\Omega_A^{j,\varepsilon}, \Omega_B^{j,\varepsilon}$ respectively. Then for ε sufficiently small that $\Omega_A^{j,\varepsilon}, \Omega_B^{j,\varepsilon}$ are disjoint,

$$\begin{aligned} Dy^{(j)}(x) &= \chi_A^{j,\varepsilon}(x)\,(A + \theta^{j,\varepsilon}(x)) + \chi_B^{j,\varepsilon}(x)\,(B + \psi^{j,\varepsilon}(x)) \\ &\quad + (1 - \chi_A^{j,\varepsilon}(x) - \chi_B^{j,\varepsilon}(x))\, Dy^{(j)}(x), \quad \text{a.e. } x \in \Omega, \end{aligned} \tag{4.8}$$

where $\theta^{j,\varepsilon}(x) \overset{\text{def}}{=} Dy^{(j)}(x) - A$, $\psi^{j,\varepsilon}(x) \overset{\text{def}}{=} Dy^{(j)}(x) - B$. Since $\chi_A^{j,\varepsilon}$, $\chi_B^{j,\varepsilon}$ are uniformly bounded there exists a subsequence, again denoted $y^{(j)}$, such that $\chi_A^{j,\varepsilon} \overset{*}{\rightharpoonup} \lambda_A^\varepsilon$, $\chi_B^{j,\varepsilon} \overset{*}{\rightharpoonup} \lambda_B^\varepsilon$ in $L^\infty(\Omega)$. Since, by (4.5), $\lim_{j\to\infty} \int_\Omega (1 - \chi_A^{j,\varepsilon} - \chi_B^{j,\varepsilon})\, dx = 0$, it follows that

$$0 \leqq \lambda_A^\varepsilon(x) = 1 - \lambda_B^\varepsilon(x) \leqq 1, \quad \text{a.e. } x \in \Omega. \tag{4.9}$$

Also

$$\int_\Omega (1 - \chi_A^{j,\varepsilon} - \chi_B^{j,\varepsilon})\,|Dy^{(j)}(x)|\,dx$$

$$\leqq \left(\int_\Omega (1 - \chi_A^{j,\varepsilon} - \chi_B^{j,\varepsilon})\,dx\right)^{1/p'} \left(\int_\Omega |Dy^{(j)}(x)|^p\,dx\right)^{1/p},$$

where $1/p + 1/p' = 1$, so that the last term in (4.8) tends to zero as $j \to \infty$ strongly in L^1. Passing to the limit $j \to \infty$ in (4.8) we obtain

$$Dy(x) = \lambda_A^\varepsilon(x)\,A + (1 - \lambda_A^\varepsilon(x))\,B + H^\varepsilon(x), \quad \text{a.e. } x \in \Omega, \tag{4.10}$$

where $|H^\varepsilon(x)| \leqq \varepsilon$ a.e.. From (4.9) there exists a subsequence $\varepsilon_k \to 0$ such that $\lambda_A^{\varepsilon_k} \overset{*}{\rightharpoonup} \lambda(\cdot)$ in $L^\infty(\Omega)$, where $0 \leqq \lambda(x) \leqq 1$ a.e.. Passing to the limit in (4.10) we obtain (4.6).

Suppose that $\lambda(x) = 1$ a.e.. We claim that $Dy^{(j)} \to A$ in measure. If not, there would exist $\varepsilon > 0$ and a subsequence $y^{(\mu)}$ such that meas $\Omega_B^{\mu,\varepsilon} = \int_\Omega \chi_B^{\mu,\varepsilon}\,dx \geqq \delta > 0$. Applying the preceding argument to $y^{(\mu)}$ gives

$$Dy(x) = \bar\lambda(x)\,A + (1 - \bar\lambda(x))\,B, \quad \text{a.e. } x \in \Omega,$$

where $\int_\Omega (1 - \bar\lambda(x))\,dx \geqq \delta$, contradicting (4.6). A similar argument shows that if $\lambda(x) = 0$ a.e., then $Dy^{(j)} \to B$ in measure.

Suppose that λ equals neither 0 a.e. nor 1 a.e.. If λ takes only the values 0 and 1 a.e., then (4.7) follows from Proposition 1 applied to y. Hence we need only consider the case when $0 < \lambda(x) < 1$ on a set $\bar\Omega$ of positive measure. Since (4.7) says nothing if $m = 1$ or $n = 1$, we suppose also that $m \geqq 2$, $n \geqq 2$. If $M \in M^{m \times n}$ we denote by $J(M)$ some 2×2 minor of M. Since $p > 2$, we have that

$$J(Dy^{(j)} - A) \rightharpoonup J(Dy - A) \quad \text{in } L^{p/2}(\Omega) \tag{4.11}$$

(see RESHETNYAK [41], BALL [4], BALL, CURRIE & OLVER [5]). But, since $\chi_A^{j,\varepsilon}$ and $\chi_B^{j,\varepsilon}$ are characteristic functions of disjoint sets,

$$\begin{aligned} J(Dy^{(j)}(x) - A) &= \chi_A^{j,\varepsilon}(x)\,J(\theta^{j,\varepsilon}(x)) + \chi_B^{j,\varepsilon}(x)\,J(B - A + \psi^{j,\varepsilon}(x)) \\ &\quad + (1 - \chi_A^{j,\varepsilon}(x) - \chi_B^{j,\varepsilon}(x))\,J(Dy^{(j)}(x) - A), \quad \text{a.e. } x \in \Omega. \end{aligned} \tag{4.12}$$

Using a similar argument as for (4.8), we deduce from (4.11) and (4.12) that

$$J(Dy - A) = (1 - \lambda(x))\,J(B - A), \quad \text{a.e. } x \in \Omega,$$

and hence from (4.6) that

$$\lambda(x)\,(1 - \lambda(x))\,J(B - A) = 0 \quad \text{a.e. } x \in \Omega.$$

Since $\bar\Omega$ has positive measure it follows that $J(B - A) = 0$. Since a matrix all of whose 2×2 minors vanish is of rank one, this completes the proof. □

Remarks:

1. The argument using the minors J is only needed to handle the case when $\lambda(\boldsymbol{x}) \in (0, 1)$ is constant. If $\lambda(\boldsymbol{x})$ is not constant a.e. then (4.7) follows from (4.6) using an argument similar to that in Proposition 1.
2. In case (i) (respectively (ii)) it is easily shown that $\boldsymbol{y}^{(j)} \to \boldsymbol{y}_0 + \boldsymbol{A}\boldsymbol{x}$ (respectively $\boldsymbol{y}^{(j)} \to \boldsymbol{y}_0 + \boldsymbol{B}\boldsymbol{x}$) strongly in $W^{1,q}(\Omega, \mathbb{R}^m)$ for $1 \leqq q < p$, where $\boldsymbol{y}_0 \in \mathbb{R}^m$.
3. We do not know if Proposition 2 holds for $1 \leqq p \leqq 2$.
4. Suppose for simplicity that $\boldsymbol{y}^{(j)} \overset{*}{\rightharpoonup} \boldsymbol{y}$ in $W^{1,\infty}(\Omega; \mathbb{R}^m)$. Then the proof of Proposition 2 can be adapted to show that $\boldsymbol{y}^{(j)}$ converges in the sense of generalized curves (*cf.* TARTAR [46, Section 4]), so that for any continuous function $f\colon M^{m \times n} \to \mathbb{R}$
$$f(D\boldsymbol{y}^{(j)}) \overset{*}{\rightharpoonup} \langle \nu_{\boldsymbol{x}}, f \rangle \quad \text{in } L^\infty(\Omega),$$
and that the Young measure $\nu_{\boldsymbol{x}}$ is given by
$$\nu_{\boldsymbol{x}} = \lambda(\boldsymbol{x})\, \delta_{\boldsymbol{A}} + (1 - \lambda(\boldsymbol{x}))\, \delta_{\boldsymbol{B}} \quad \text{a.e. } \boldsymbol{x} \in \Omega,$$
where $\delta_{\boldsymbol{A}}, \delta_{\boldsymbol{B}}$ denote Dirac masses at $\boldsymbol{A}, \boldsymbol{B}$ respectively. See CHIPOT & KINDERLEHRER [14] and KINDERLEHRER [26] for further remarks in this direction.
5. By applying Proposition 2 we can strengthen the statement made in Section 3 concerning the nonexistence of austenite/martensite interfaces to the assertion that there is no sequence of deformations $\boldsymbol{y}^{(j)} \rightharpoonup \boldsymbol{y}$ in $W^{1,p}(\Omega; \mathbb{R}^3)$, $p > 2$, which, in the sense of Proposition 2, have gradients taking to a good approximation only the two values $\boldsymbol{R}$ and $\bar{\boldsymbol{R}} \boldsymbol{U}_0 \boldsymbol{R}_k$, where $\boldsymbol{R}$, $\bar{\boldsymbol{R}}$ are rotations and $\boldsymbol{R}_k \in P'$, unless $D\boldsymbol{y}^{(j)} \to \boldsymbol{R}$ in measure or $D\boldsymbol{y}^{(j)} \to \bar{\boldsymbol{R}} \boldsymbol{U}_0 \boldsymbol{R}_k$ in measure.

We can now address the case of deformations whose gradients to a good approximation take only two values on one side of an interface $\mathcal{S}$, and a third value on the opposite side. Here one goal is to understand why the austenite/finely twinned martensite interface is flat.

Theorem 3. *Let* $m \geqq 2$, $n \geqq 2$, $p > 2$. *Let* $\Omega \subset \mathbb{R}^n$ *be bounded, open and connected, and suppose that* Ω *can be written in the form* $\Omega = \Omega_{A,B} \cup \Omega_C \cup \mathcal{S}$, *where* $\Omega_{A,B}$ *and* Ω_C *are disjoint, open and connected, and where* $\mathcal{S} = \partial\Omega_{A,B} \cap \Omega = \partial\Omega_C \cap \Omega$. *Assume either that* meas $\mathcal{S} = 0$ *or that* $p > n$. *Let* $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \in M^{m \times n}$ *be distinct, and suppose that neither* $\boldsymbol{C} - \boldsymbol{A}$ *nor* $\boldsymbol{C} - \boldsymbol{B}$ *is of rank one. Let* $\boldsymbol{y}^{(j)} \rightharpoonup \boldsymbol{y}$ *in* $W^{1,p}(\Omega, \mathbb{R}^m)$ *satisfy for every* $\varepsilon > 0$

$$\lim_{j \to \infty} \text{meas}\, \{\boldsymbol{x} \in \Omega_{A,B} \colon |D\boldsymbol{y}^{(j)}(\boldsymbol{x}) - \boldsymbol{A}| > \varepsilon \ \text{ and } \ |D\boldsymbol{y}^{(j)}(\boldsymbol{x}) - \boldsymbol{B}| > \varepsilon\} = 0, \tag{4.13}$$

and

$$\lim_{j \to \infty} \text{meas}\, \{\boldsymbol{x} \in \Omega_C \colon |D\boldsymbol{y}^{(j)}(\boldsymbol{x}) - \boldsymbol{C}| > \varepsilon\} = 0. \tag{4.14}$$

Then the interface $\mathcal{S}$ *is necessarily planar*, i.e., *there exists a unit vector* $\boldsymbol{m} \in \mathbb{R}^n$ *and* $k \in \mathbb{R}$ *with*

$$\mathcal{S} \subset \{\boldsymbol{x} \in \mathbb{R}^n \colon \boldsymbol{x} \cdot \boldsymbol{m} = k\}, \tag{4.15}$$

and

$$A - B = c \otimes n, \tag{4.16}$$

$$C - B = -b \otimes m + \lambda c \otimes n, \tag{4.17}$$

where $n \in \mathbb{R}^n$ *is a unit vector that is not parallel to* m, *where* $b, c \in \mathbb{R}^m$ *are not parallel, and where* $0 < \lambda < 1$. *Furthermore, each point* $x_0 \in \mathscr{S}$ *has an open neighborhood* $N(x_0)$ *such that*

$$Dy(x) = B + \lambda c \otimes n, \quad x \in \Omega_{A,B} \cap N(x_0), \tag{4.18}$$

$$Dy(x) = C, \quad x \in \Omega_C \cap N(x_0). \tag{4.19}$$

Conversely, suppose that $\mathscr{S}$, A, B *and* C *have the forms* (4.15) *through* (4.17). *Then there exist sequences* $y^{(j)}$ *converging weak * in* $W^{1,\infty}(\Omega, \mathbb{R}^m)$ *to some* y *and satisfying* (4.13) *and* (4.14). *If* $m = n$, $\det A > 0$, $\det B > 0$ *and* $\det C > 0$, *then* $y^{(j)}$ *can be chosen so that* $\det Dy^{(j)}(x) \geq \delta > 0$ a.e. $x \in \Omega$, *for some constant* $\delta > 0$.

Proof. Applying Proposition 2 to $\Omega_{A,B}$, we see that one of the following three possibilities holds a.e. in $\Omega_{A,B}$:

(i) $y(x) = y_0 + Ax$ for some $y_0 \in \mathbb{R}^m$,

(ii) $y(x) = y_0 + Bx$ for some $y_0 \in \mathbb{R}^m$,

(iii) $A - B = c \otimes n$ for some $c \in \mathbb{R}^m$, $n \in \mathbb{R}^n$, $|n| = 1$,

and

$$Dy(x) = B + \lambda(x)\, c \otimes n, \tag{4.20}$$

where $0 \leq \lambda(x) \leq 1$ and $\lambda(x)$ equals neither 1 a.e. nor 0 a.e.; in this case, we have, using the argument at the end of the proof of Proposition 1, that

$$y(x) = y_0 + Bx + \theta(x)\, c, \tag{4.21}$$

where $y_0 \in \mathbb{R}^m$ and $\theta \in W^{1,\infty}(\Omega_{A,B})$, $D\theta(x) = \lambda(x)\, n$.

Similarly, we have

$$y(x) = y_1 + Cx, \quad \text{a.e. } x \in \Omega_C, \tag{4.22}$$

for some $y_1 \in \mathbb{R}^m$.

If $p > n$ then an appropriate choice of representative y is continuous on Ω, and the same holds if $\operatorname{meas} \mathscr{S} = 0$ since then $y \in W^{1,\infty}(\Omega, \mathbb{R}^m)$. Pick x_0 and suppose for contradiction that $(x_0 + r_i) \in \mathscr{S}$ for n linearly independent vectors r_i. We suppose (4.21) holds in $\Omega_{A,B}$; the cases (i) and (ii) are handled similarly. Since $c \neq o$, by (4.21) θ has a continuous extension, again denoted θ, to $\partial\Omega_{A,B} \cap \Omega$. By the continuity of y we therefore have

$$(C - B)\, x_0 = y_0 - y_1 + \theta(x_0)\, c,$$

$$(C - B)(x_0 + r_i) = y_0 - y_1 + \theta(x_0 + r_i)\, c, \quad i = 1, \ldots, n.$$

Subtracting, we find that $(C - B)\, r_i$ is parallel to c for $i = 1, \ldots, n$, contradicting our assumption that $C - B$ is not of rank one. Hence (4.15) holds for some

$\boldsymbol{m}$ and k. In the cases (i) (ii), (4.15) and Proposition 1 imply that either $\boldsymbol{C} - \boldsymbol{A}$ or $\boldsymbol{C} - \boldsymbol{B}$ is of rank one; thus these cases are impossible and (4.21) holds in $\Omega_{A,B}$.

Given $\boldsymbol{x}_0 \in \mathscr{S}$, choose $r > 0$ sufficiently small that the open ball $B(\boldsymbol{x}_0, r)$ with center $\boldsymbol{x}_0$ and radius r is contained in Ω. Let

$$B^+(\boldsymbol{x}_0, r) = \{\boldsymbol{x} \in B(\boldsymbol{x}_0, r) : \boldsymbol{x} \cdot \boldsymbol{m} > k\},$$

$$B^-(\boldsymbol{x}_0, r) = \{\boldsymbol{x} \in B(\boldsymbol{x}_0, r) : \boldsymbol{x} \cdot \boldsymbol{m} < k\}.$$

Then $\Omega_{A,B} \cap B(\boldsymbol{x}_0, r) = B^+(\boldsymbol{x}_0, r)$, $\Omega_C \cap B(\boldsymbol{x}_0, r) = B^-(\boldsymbol{x}_0, r)$ or *vice versa*, and $\mathscr{S} \cap B(\boldsymbol{x}_0, r) = \{\boldsymbol{x} \in B(\boldsymbol{x}_0, r) : \boldsymbol{x} \cdot \boldsymbol{m} = k\}$. Since $D\theta(\boldsymbol{x}) = \lambda(\boldsymbol{x})\, \boldsymbol{n}$ in $\Omega_{A,B}$ it follows that

$$\theta(\boldsymbol{x}) = f(\boldsymbol{x} \cdot \boldsymbol{n}) \quad \text{for all } \boldsymbol{x} \in \overline{\Omega}_{A,B} \cap B(\boldsymbol{x}_0, r), \tag{4.23}$$

for some $f \in W^{1,\infty}(\mathbb{R})$. For $\boldsymbol{x} \in B(\boldsymbol{x}_0, r)$ denote by $\boldsymbol{x}'$ the orthogonal projection of $\boldsymbol{x}$ onto $\mathscr{S}$. Then $\boldsymbol{x}' = \boldsymbol{x} + (k - (\boldsymbol{x} \cdot \boldsymbol{m}))\, \boldsymbol{m}$ and

$$(\boldsymbol{C} - \boldsymbol{B})\, \boldsymbol{x}' = \boldsymbol{y}_0 - \boldsymbol{y}_1 + f(\boldsymbol{x}' \cdot \boldsymbol{n})\, \boldsymbol{c},$$

so that

$$(\boldsymbol{C} - \boldsymbol{B})\, \boldsymbol{x} = \boldsymbol{y}_0 - \boldsymbol{y}_1 + f(\boldsymbol{x}' \cdot \boldsymbol{n})\, \boldsymbol{c} + ((\boldsymbol{x} \cdot \boldsymbol{m}) - k)\, \boldsymbol{d}, \tag{4.24}$$

where $\boldsymbol{d} = (\boldsymbol{C} - \boldsymbol{B})\, \boldsymbol{m}$. Taking the derivative of (4.23) with respect to $\boldsymbol{x}$, we find, using the chain rule for Lipschitz maps (*cf.* MARCUS & MIZEL [31, Lemma 2.1]) that

$$\boldsymbol{C} - \boldsymbol{B} = \boldsymbol{d} \otimes \boldsymbol{m} + f'(\boldsymbol{x}' \cdot \boldsymbol{n})\, \boldsymbol{c} \otimes [\boldsymbol{n} - (\boldsymbol{m} \cdot \boldsymbol{n})\, \boldsymbol{m}], \tag{4.25}$$

a.e. in $B(\boldsymbol{x}_0, r)$. Since $\boldsymbol{C} - \boldsymbol{B}$ is not of rank one, $\boldsymbol{m}$ is not parallel to $\boldsymbol{n}$, $(\boldsymbol{m} \cdot \boldsymbol{n})^2 \neq 1$, and $\boldsymbol{c} \neq \boldsymbol{o}$. Taking the inner product of (4.25) with $\boldsymbol{n}$, we thus deduce that $f'(\boldsymbol{x}' \cdot \boldsymbol{n}) = \lambda =$ constant a.e. in $B(\boldsymbol{x}_0, r)$. Since $\boldsymbol{m}$ is not parallel to $\boldsymbol{n}$, it follows that $f'(t) = \lambda$ for t in a neighborhood of $\boldsymbol{x}_0 \cdot \boldsymbol{n}$ and hence from (4.20), (4.21) and (4.23) that $\lambda(\boldsymbol{x}) = \lambda$, $0 < \lambda < 1$, for $\boldsymbol{x} \in \Omega_{A,B} \cap N(\boldsymbol{x}_0)$, where $N(\boldsymbol{x}_0)$ is some open neighborhood of $\boldsymbol{x}_0$, possibly smaller than $B(\boldsymbol{x}_0, r)$. The relation (4.17) follows from (4.25), $\boldsymbol{b}$ not being parallel to $\boldsymbol{c}$ since $\boldsymbol{C} - \boldsymbol{B}$ is not of rank one.

Conversely, suppose that $\mathscr{S}$, $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ have the forms (4.15) through (4.17). We suppose without loss of generality that $k = 0$. Let $\theta : \mathbb{R} \to \mathbb{R}$ be the 1-periodic function satisfying

$$\theta(t) = \begin{cases} (1 - \lambda)(t + \lambda) & \text{for } -\lambda \leqq t < 0, \\ -\lambda(t - 1 + \lambda) & \text{for } 0 \leqq t < 1 - \lambda. \end{cases} \tag{4.26}$$

Let $\mu > 0$ and

$$\boldsymbol{D}^{\pm} = \boldsymbol{C} + \left(\boldsymbol{b} \pm \frac{\lambda(1 - \lambda)}{\mu} \boldsymbol{c}\right) \otimes \boldsymbol{m}. \tag{4.27}$$

For $x \in \mathbb{R}^n$ define

$$z(x) = \begin{cases} (\lambda A + (1-\lambda) B)\, x + \theta(x \cdot n)\, c & \text{for } |x \cdot m| > \dfrac{\mu}{\lambda(1-\lambda)}\theta(x \cdot n), \\ D^+ x & \text{for } 0 \leqq x \cdot m \leqq \dfrac{\mu}{\lambda(1-\lambda)}\theta(x \cdot n), \\ D^- x & \text{for } -\dfrac{\mu}{\lambda(1-\lambda)}\theta(x \cdot n) \leqq x \cdot m \leqq 0. \end{cases} \tag{4.28}$$

It is easily verified that $z \in W^{1,\infty}_{\text{loc}}(\mathbb{R}^n, \mathbb{R}^m)$, that Dz takes almost everywhere only the four values A, B, D^+, D^- and that $z(x) = Cx$ for $x \cdot m = 0$. For $j = 1, 2, \dots$, define for $x \in \Omega$

$$y^{(j)}(x) = \begin{cases} j^{-1} z(jx) & \text{for } x \in \Omega_{A,B}, \\ Cx & \text{otherwise}, \end{cases} \tag{4.29}$$

(see Figure 4b). Since $y^{(j)}$ is continuous, $y^{(j)}(0) = 0$ and $Dy^{(j)}$ takes almost everywhere only the five values A, B, C and $D^{\pm}$, it follows that $y^{(j)}$ is a bounded sequence in $W^{1,\infty}(\Omega, \mathbb{R}^m)$. Since also

$$Dy^{(j)}(x) = \lambda A + (1-\lambda) B + \theta'(jx \cdot n)(A - B)$$

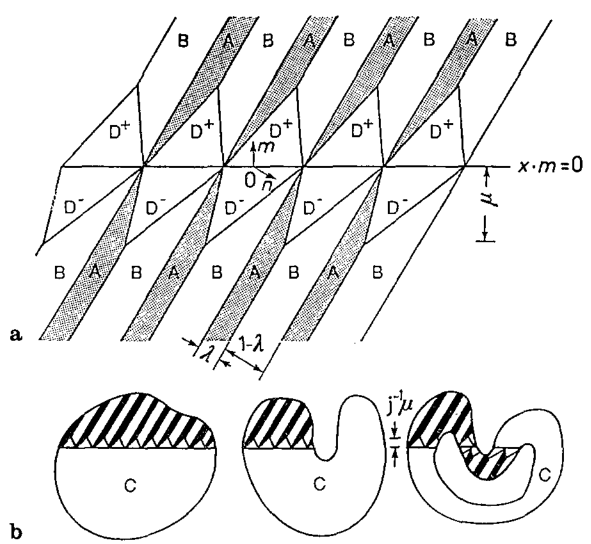

Fig. 4a. The orthogonal projection of $\mathbb{R}^n$ onto the (m, n) plane, showing the values taken in various regions by Dz when z is given by (4.28); b Possible two-dimensional cross sections of Ω parallel to the (m, n) plane, showing the division of Ω into regions where $Dy^{(j)}$ takes different values when $y^{(j)}$ is given by (4.29)

for a.e. $x \in \Omega_{A,B}$ satisfying $|x \cdot m| > j^{-1}\mu$, it is easily shown that $y^{(j)} \overset{*}{\rightharpoonup} y$ in $W^{1,\infty}(\Omega, \mathbb{R}^m)$, where

$$y(x) = \begin{cases} (\lambda A + (1-\lambda) B)\, x & \text{for } \; x \in \Omega_{A,B}, \\ Cx & \text{otherwise}, \end{cases} \tag{4.30}$$

and that (4.13) and (4.14) hold. Finally, if $m = n$ and $\det A > 0$, $\det B = 0$, $\det C > 0$, then by (4.16)

$$\det(\lambda A + (1-\lambda) B) = \lambda \det A + (1-\lambda) \det B.$$

Since

$$\det D^{\pm} = \det\left(\lambda A + (1-\lambda) B \pm \frac{\lambda(1-\lambda)}{\mu} c \otimes m\right),$$

it follows that $\det D^{\pm} > 0$ if μ is chosen sufficiently large, and in this case $\det Dy^{(j)}(x) \geqq \min\{\det A, \det B, \det C, \det D\} > 0$ a.e. for all j. □

Remarks:

1. Let $m = n$, $\det A > 0$, $\det B > 0$, $\det C > 0$ and let μ be chosen sufficiently large that $\det D^{\pm} > 0$. Then for domains Ω such as those in Figure 4b (ii) and (iii), the deformation y given by (4.30) need not be invertible, and hence $y^{(j)}$ given by (4.29) need not be invertible. Sufficient conditions for $y^{(j)}$ to be invertible are that $\Omega_{A,B} \subset \{x \in \mathbb{R}^n : x \cdot m > k\}$ and $\Omega_C \subset \{x \in \mathbb{R}^n : x \cdot m < k\}$.
2. It follows from the proof of Proposition 2 that

$$\lim_{\varepsilon \to 0} \lim_{j \to \infty} \frac{\operatorname{meas}\{x \in N(x_0) : |Dy^{(j)}(x) - A| < \varepsilon\}}{\operatorname{meas}(N(x_0) \cap \Omega_{A,B})} = \lambda,$$

$$\lim_{\varepsilon \to 0} \lim_{j \to \infty} \frac{\operatorname{meas}\{x \in N(x_0) : |Dy^{(j)}(x) - B| < \varepsilon\}}{\operatorname{meas}(N(x_0) \cap \Omega_{A,B})} = 1 - \lambda,$$

so that λ, $1-\lambda$ denote the asymptotic proportions of A and B respectively in $N(x_0) \cap \Omega_{A,B}$.

We now return to the minimization problem (3.11) with the free energy function described in Section 3, not necessarily specialized to InTl. Assume that F^+ and F^- satisfy for some rotations R^+ and R^-

$$F^{\pm} = R^{\pm} U_0. \tag{4.31}$$

Recall that ϕ is a Galilean invariant, that U_0 is at the minimum of a potential well and that $\phi(\mathbf{1}, \theta_0) = \phi(U_0, \theta_0) = 0$. We apply Theorem 3 with $m = n = 3$, $A = F^+$, $B = F^-$ and $C = \mathbf{1}$. Let $y^{(j)}$ be a minimizing sequence for $\mathscr{I}$ in $\mathscr{A}$, so that

$$\lim_{j \to \infty} \mathscr{I}(y^{(j)}) = 0, \tag{4.32}$$

and suppose that the hypotheses of the necessity part of Theorem 3 holds. (The condition that a minimizing sequence is weakly convergent in $W^{1,p}(\Omega; \mathbb{R}^3)$ is satisfied for some subsequence of that sequence provided ϕ satisfies an estimate of the form

$$\phi(F) \geqq K_0 |F|^p + K_1 \quad \text{for all } F \in \mathscr{D} \tag{4.33}$$

for some constants $K_0 > 0, K_1$.) From Theorem 3 we deduce that the interface $\mathscr{S}$ is *flat* and, by (4.16) and (4.17), that

$$\begin{aligned} F^+ &= \mathbf{1} + (1 - \lambda)\, c \otimes n + b \otimes m, \\ F^- &= \mathbf{1} - \lambda c \otimes n + b \otimes m, \end{aligned} \tag{4.34}$$

for nonparallel vectors $b, c \in \mathbb{R}^3$, nonparallel unit vectors $m, n \in \mathbb{R}^3$ and some $\lambda \in (0, 1)$. Conversely, if the conditions (4.34) hold and $\mathscr{S}$ has the form (4.15) then the $y^{(j)}$ constructed in the theorem and satisfying (4.13) and (4.14) will be a minimizing sequence for $\mathscr{I}$ in $\mathscr{A}$, since ϕ is continuous, $Dy^{(j)}$ bounded in L^∞ and $\det Dy^{(j)}(x) \geqq \delta > 0$ a.e. in Ω. In fact, since F^+ and F^- are at the minima of potential wells,

$$\int_\Omega \phi(Dy^{(j)}, \theta_0)\, dx = \int_{\{x|Dy^{(j)}(x)=D^+\}} \phi(D^+, \theta_0)\, dx + \int_{\{x|Dy^{(j)}(x)=D^-\}} \phi(D^-, \theta_0)\, dx$$
$$\to 0 \quad \text{as } j \to \infty. \tag{4.35}$$

That is, the only contribution to the total energy of $y^{(j)}$ is from the layer of triangular prisms pictured in Figure 4b whose total volume tends to zero as $j \to \infty$. The existence of rotations $R^\pm$ such that $F^\pm$ given by (4.31) satisfy (4.34) will be established for InTl in Section 5.

Simple examples show that if we drop the hypothesis in Theorem 3 that the open set $\Omega_{A,B}$ is connected, then $\mathscr{S}$ need not be contained in a plane. The X-interface in Figure 2a provides an example where the finely twinned region is a disconnected open set and the austenite/martensite interface $\mathscr{S}$ is not contained in a plane. Note, however, that the deformation gradient in this configuration takes to a good approximation four values. The X-interface is easily understood by patching together two deformations of the type given by Theorem 3. The curved martensite/martensite interface shown in Figure 2b is obviously not covered by Theorem 3. In principle, it should be possible to relate the orientation of this interface to the local twin concentrations $\lambda(x)$ and $\lambda'(x)$ on each side of the interface.

Of course the weak limit y of a minimizing sequence $y^{(j)}$ is not in general a minimizer; for example, with $y^{(j)} \rightharpoonup y$ as above, (4.18) becomes

$$Dy(x) = \lambda F^+ + (1 - \lambda)\, F^-, \tag{4.36}$$

and $\lambda F^+ + (1 - \lambda)\, F^- = \mathbf{1} + b \otimes m$ does not in general, or in particular for InTl, yield one of the stretch matrices $\mathbf{1}, R_i U_0 R_i^T, i = 1, \ldots, \nu$. Hence $\mathscr{I}$ is generally not sequentially weakly lower semicontinuous (swlsc) in $W^{1,q}(\Omega, \mathbb{R}^3)$ for any $q \geqq 1$. This typical feature of multidimensional phase change problems results from the failure of strong ellipticity of ϕ (*cf.* ERICKSEN [16], BALL [4]) and contrasts with certain models of rubber-like materials for which $\mathscr{I}$ *is* swlsc and for which the direct method of the calculus of variations can consequently be applied to establish the existence of minimizers (BALL [3]). When ϕ is not $W^{1,1}$-quasiconvex, a condition closely related to strong ellipticity, an argument of BALL & MURAT [8, Theorem 5.1] in fact shows that for appropriate boundary conditions and body forces the total free energy does *not* attain a minimum (see also the examples in Section 7).

5. Materials Which Can Form Internally Twinned Martensite

Given a transformation strain and a family of energy minimizing twins, we now consider the algebraic problem of whether the equations $(4.34)_{1,2}$ can be satisfied. If so, we can construct minimizing sequences by the methods given in Section 4. The critical physical question is whether the austenite/martensite planes work out correctly. These calculations are closely related to those of the crystallographic theory of martensite, although necessary and sufficient conditions for the existence of solutions of $(4.34)_{1,2}$ in the case of a general transformation strain and point group appear to be absent from the literature. We restrict attention to the case $m = n = 3$ throughout this section.

According to the analysis of Section 3, all classical interfaces between minimizing deformation gradients are martensite/martensite twins. The deformation gradients associated with these twins, $\boldsymbol{F}^+$ and $\boldsymbol{F}^-$, have the forms

$$\begin{aligned} \boldsymbol{F}^+ &= \hat{\boldsymbol{R}}^+ \boldsymbol{U}_0 \boldsymbol{R}_i, \\ \boldsymbol{F}^- &= \hat{\boldsymbol{R}}^- \boldsymbol{U}_0 \boldsymbol{R}_j, \end{aligned} \tag{5.1}$$

with $\boldsymbol{R}_i$ and $\boldsymbol{R}_j$ in P^ν. It is sufficient* for our purposes to take $\boldsymbol{R}_j = \boldsymbol{1}$ and to rewrite (5.1) in the form

$$\begin{aligned} \boldsymbol{F}^+ &= \hat{\boldsymbol{R}} \boldsymbol{R} \boldsymbol{U}_0 \bar{\boldsymbol{R}}, \\ \boldsymbol{F}^- &= \hat{\boldsymbol{R}} \boldsymbol{U}_0, \end{aligned} \tag{5.2}$$

where

$$\boldsymbol{F}^+ - \boldsymbol{F}^- = \boldsymbol{c} \otimes \boldsymbol{n}, \quad \boldsymbol{c} = \hat{\boldsymbol{R}} \boldsymbol{a}. \tag{5.3}$$

For InTl, the values of $\boldsymbol{R}$, $\bar{\boldsymbol{R}}$, $\boldsymbol{a}$ and $\boldsymbol{n}$ satisfying the equations (5.2) and (5.3) are obtained from (3.19) with $\eta_1 = 1 - \varepsilon$ and $\eta_2 = 1 + 2\varepsilon$, $\varepsilon \doteq .013$. In this section we allow $\boldsymbol{U}_0$ to be an arbitrary positive-definite symmetric matrix and the twins in (5.2) to be general in that they are subject only to (3.18), that is

$$\boldsymbol{R} \boldsymbol{U}_0 \bar{\boldsymbol{R}} = \boldsymbol{U}_0 + \boldsymbol{a} \otimes \boldsymbol{n}, \tag{5.4}$$

where $\boldsymbol{R}$, $\bar{\boldsymbol{R}}$ are rotations, $\boldsymbol{a} \neq \boldsymbol{0}$ and $|\boldsymbol{n}| = 1$. We note the relations

$$\boldsymbol{U}_0^{-1} \boldsymbol{a} \cdot \boldsymbol{n} = 0, \tag{5.5}$$

$$2 \boldsymbol{U}_0 \boldsymbol{a} \cdot \boldsymbol{n} + |\boldsymbol{a}|^2 = 0, \tag{5.6}$$

$$2 \boldsymbol{U}_0^{-2} \boldsymbol{a} \cdot \boldsymbol{U}_0^{-1} \boldsymbol{n} - |\boldsymbol{U}_0^{-1} \boldsymbol{a}|^2 \, |\boldsymbol{U}_0^{-1} \boldsymbol{n}|^2 = 0, \tag{5.7}$$

---

* The twins represented by (5.1) can be obtained from those represented by (5.2) and (5.3) by replacing $\boldsymbol{F}^+$ and $\boldsymbol{F}^-$ in (5.1) by $\boldsymbol{R}_j^T \boldsymbol{F}^+ \boldsymbol{R}_j$ and $\boldsymbol{R}_j^T \boldsymbol{F}^- \boldsymbol{R}_j$. Note that these replacements do not alter the forms of $(4.34)_{1,2}$.

which follow from (5.4) by taking determinants and by calculating tr $(\boldsymbol{R}\boldsymbol{U}_0^2\boldsymbol{R}^T)$ and tr $(\boldsymbol{R}\boldsymbol{U}_0^{-2}\boldsymbol{R}^T)$.

To decide whether the twins (5.2) can participate in an austenite/finely twinned martensite interface, we consider the algebraic problem of whether $\boldsymbol{F}^+$ and $\boldsymbol{F}^-$ given by (5.2) can assume the forms $(4.34)_{1,2}$ for some choice of $\hat{\boldsymbol{R}}$. We view $\boldsymbol{U}_0$, $\boldsymbol{a}$ and $\boldsymbol{n}$ as given, consistent with (3.18). Since $(4.34)_1$ follows immediately from $(4.34)_2$ and (5.3), we only need to consider $(4.34)_2$ which becomes

$$\boldsymbol{F}^- = \hat{\boldsymbol{R}}\boldsymbol{U}_0 = \boldsymbol{1} - \lambda\hat{\boldsymbol{R}}\boldsymbol{a} \otimes \boldsymbol{n} + \boldsymbol{b} \otimes \boldsymbol{m}, \tag{5.8}$$

or equivalently,

$$\boldsymbol{U}_0 + \lambda\boldsymbol{a} \otimes \boldsymbol{n} = \hat{\boldsymbol{R}}^T(\boldsymbol{1} + \boldsymbol{b} \otimes \boldsymbol{m}), \tag{5.9}$$

which is to be solved for $\lambda \in (0, 1)$, $\boldsymbol{b}$, the unit vector $\boldsymbol{m}$ and the rotation $\hat{\boldsymbol{R}}$.

Let $\boldsymbol{C}_0(\lambda)$ be defined by

$$\boldsymbol{C}_0(\lambda) \stackrel{\text{def}}{=} (\boldsymbol{U}_0 + \lambda\boldsymbol{n} \otimes \boldsymbol{a})(\boldsymbol{U}_0 + \lambda\boldsymbol{a} \otimes \boldsymbol{n}). \tag{5.10}$$

According to the polar decomposition theorem, the basic equation (5.9) with $\det(\boldsymbol{U}_0 + \lambda\boldsymbol{a} \otimes \boldsymbol{n}) > 0$ is equivalent to

$$\boldsymbol{C}_0(\lambda) = (\boldsymbol{1} + \boldsymbol{m} \otimes \boldsymbol{b})(\boldsymbol{1} + \boldsymbol{b} \otimes \boldsymbol{m}), \tag{5.11}$$

together with the restriction that $\det(\boldsymbol{1} + \boldsymbol{b} \otimes \boldsymbol{m}) = 1 + \boldsymbol{b} \cdot \boldsymbol{m} > 0$. Hence, we begin with a characterization of $\boldsymbol{b}$ and $\boldsymbol{m}$ satisfying (5.11).

Proposition 4. *Necessary and sufficient conditions for a symmetric* 3×3 *matrix* $\boldsymbol{C}$ *with eigenvalues* $\lambda_1 \leqq \lambda_2 \leqq \lambda_3$ *to be expressible in the form*

$$\boldsymbol{C} = (\boldsymbol{1} + \boldsymbol{m} \otimes \boldsymbol{b})(\boldsymbol{1} + \boldsymbol{b} \otimes \boldsymbol{m}) \tag{5.12}$$

for nonzero $\boldsymbol{b}$ *and* $\boldsymbol{m}$ *are that* $\lambda_1 \geqq 0$ (i.e., $\boldsymbol{C} \geqq 0$) *and* $\lambda_2 = 1$.

The solutions are given by:

a) $\boldsymbol{C} \neq \boldsymbol{1}$,

$$\boldsymbol{b} = \varrho\left(\sqrt{\frac{\lambda_3(1-\lambda_1)}{\lambda_3-\lambda_1}}\,\boldsymbol{e}_1 + \varkappa\sqrt{\frac{\lambda_1(\lambda_3-1)}{\lambda_3-\lambda_1}}\,\boldsymbol{e}_3\right), \tag{5.13}$$

$$\boldsymbol{m} = \varrho^{-1}\left(\frac{\bar{\varkappa}\sqrt{\lambda_3}-\sqrt{\lambda_1}}{\sqrt{\lambda_3-\lambda_1}}\right)\left(-\bar{\varkappa}\sqrt{1-\lambda_1}\,\boldsymbol{e}_1 + \varkappa\sqrt{\lambda_3-1}\,\boldsymbol{e}_3\right),$$

where $\varrho \neq 0$ *is a constant, and* $\boldsymbol{e}_1$, $\boldsymbol{e}_3$ *are normalized eigenvectors of* $\boldsymbol{C}$ *corresponding to* λ_1, λ_3 *respectively, and where each of* $\varkappa$, $\bar{\varkappa}$ *can take the values* ± 1. *For these solutions* $1 + \boldsymbol{b} \cdot \boldsymbol{m} = \bar{\varkappa}\sqrt{\lambda_1\lambda_3}$.

b) $\boldsymbol{C} = \boldsymbol{1}$,

$$\boldsymbol{b} = \varrho\boldsymbol{e}, \tag{5.14}$$

$$\boldsymbol{m} = -2\varrho^{-1}\boldsymbol{e},$$

where $\varrho \neq 0$ *is a constant and* $|\boldsymbol{e}| = 1$. *For these solutions* $1 + \boldsymbol{b} \cdot \boldsymbol{m} = -1$.

Proof. Necessity. Let $\boldsymbol{p}$ be perpendicular to $\boldsymbol{b}$ and $\boldsymbol{m}$. Then $\boldsymbol{C}\boldsymbol{p} = \boldsymbol{p}$ so that one eigenvalue of $\boldsymbol{C}$ equals 1. If $\boldsymbol{b}$ and $\boldsymbol{m}$ are linearly dependent, then the eigenvalue 1 has a multiplicity of at least two, so $\lambda_2 = 1$. Suppose that $\boldsymbol{b}$ and $\boldsymbol{m}$ are linearly independent. Consider

$$\begin{aligned} \boldsymbol{x} \cdot \boldsymbol{C}\boldsymbol{x} - |\boldsymbol{x}|^2 &= 2(\boldsymbol{x} \cdot \boldsymbol{b})(\boldsymbol{x} \cdot \boldsymbol{m}) + (\boldsymbol{x} \cdot \boldsymbol{m})^2 |\boldsymbol{b}|^2 \\ &= (\boldsymbol{x} \cdot \boldsymbol{m})[2(\boldsymbol{x} \cdot \boldsymbol{b}) + (\boldsymbol{x} \cdot \boldsymbol{m})|\boldsymbol{b}|^2]. \end{aligned} \tag{5.15}$$

If we choose $\boldsymbol{x} \cdot \boldsymbol{b} > 0$ and $\boldsymbol{x} \cdot \boldsymbol{m} > 0$, the expression (5.15) is positive. If we choose $\boldsymbol{x} \cdot \boldsymbol{b} = -1$ and $\boldsymbol{x} \cdot \boldsymbol{m}$ small and positive, the expression (5.15) is negative. Hence $\lambda_1 < 1 < \lambda_3$. Finally, $\lambda_1 \lambda_3 = \det \boldsymbol{C} = (1 + \boldsymbol{b} \cdot \boldsymbol{m})^2 \geqq 0$, so that $\lambda_1 \geqq 0$.

Sufficiency. First suppose that $\boldsymbol{C} \neq \boldsymbol{1}$, so that $\boldsymbol{b} \neq \boldsymbol{0}$, and suppose that $\lambda_1 > 0$. If $\boldsymbol{b}$ and $\boldsymbol{m}$ satisfy (5.12), we have

$$\begin{aligned} \boldsymbol{C}\boldsymbol{b} &= (\boldsymbol{1} + \boldsymbol{m} \otimes \boldsymbol{b})\, \boldsymbol{b}(1 + \boldsymbol{b} \cdot \boldsymbol{m}), \\ &= \pm(\det \boldsymbol{C})^{1/2} (\boldsymbol{b} + |\boldsymbol{b}|^2 \boldsymbol{m}). \end{aligned} \tag{5.16}$$

Hence,

$$\boldsymbol{m} = [\pm (\det \boldsymbol{C})^{-1/2} \boldsymbol{C} - \boldsymbol{1}] \left(\frac{\boldsymbol{b}}{|\boldsymbol{b}|^2} \right). \tag{5.17}$$

In view of (5.17) a necessary and sufficient condition that nonzero vectors $\boldsymbol{b}$ and $\boldsymbol{m}$ satisfy (5.12) is that $\boldsymbol{b} \neq \boldsymbol{0}$ satisfies

$$\boldsymbol{C} = \boldsymbol{1} - \frac{\boldsymbol{b} \otimes \boldsymbol{b}}{|\boldsymbol{b}|^2} + \frac{\boldsymbol{C}\boldsymbol{b} \otimes \boldsymbol{C}\boldsymbol{b}}{|\boldsymbol{b}|^2 \det \boldsymbol{C}}. \tag{5.18}$$

In the orthonormal basis of eigenvectors $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$, we have say $\boldsymbol{b} = (b_1, b_2, b_3)$ and $\boldsymbol{C}\boldsymbol{b} = (\lambda_1 b_1, b_2, \lambda_3 b_3)$. Then (5.18) is equivalent to

$$\begin{bmatrix} \lambda_1 - 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \lambda_3 - 1 \end{bmatrix}$$

$$= \frac{1}{|\boldsymbol{b}|^2} \begin{bmatrix} b_1^2 \left(-1 + \frac{\lambda_1}{\lambda_3}\right) & b_1 b_2 \left(-1 + \frac{1}{\lambda_3}\right) & 0 \\ b_1 b_2 \left(-1 + \frac{1}{\lambda_3}\right) & b_2^2 \left(-1 + \frac{1}{\lambda_1 \lambda_3}\right) & b_2 b_3 \left(-1 + \frac{1}{\lambda_1}\right) \\ 0 & b_2 b_3 \left(-1 + \frac{1}{\lambda_1}\right) & b_3^2 \left(-1 + \frac{\lambda_3}{\lambda_1}\right) \end{bmatrix}. \tag{5.19}$$

Since by assumption λ_1 and λ_3 are not both 1, (5.19) holds if and only if $b_2 = 0$ and

$$\frac{b_1^2}{|\boldsymbol{b}|^2} = \frac{\lambda_3(1 - \lambda_1)}{\lambda_3 - \lambda_1}, \quad \frac{b_3^2}{|\boldsymbol{b}|^2} = \frac{\lambda_1(\lambda_3 - 1)}{\lambda_3 - \lambda_1}. \tag{5.20}$$

The equations (5.20) are consistent and so (5.18) holds if and only if b has the form $(5.13)_1$ with $\varrho \neq 0$. Then we get $(5.13)_2$ from (5.17) as required.

If $C = 1$ then (5.12) is satisfied by nonzero vectors b and m if and only if

$$b = \tau m \quad \text{with } 2\tau + \tau^2 |m|^2 = 0, \tag{5.21}$$

so that the nonzero solutions of (5.12) are given by (5.14).

If $\lambda_1 = 0$, then from (5.16), which holds without the restriction $\lambda_1 > 0$, we have $Cb = 0$, so $b = \varrho e_1$. If we write $m = m_1 e_1 + m_2 e_2 + m_3 e_3$, we can write (5.12) in the form

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \lambda_3 - 1 \end{pmatrix} = \varrho \begin{pmatrix} 2m_1 & m_2 & m_3 \\ m_2 & 0 & 0 \\ m_3 & 0 & 0 \end{pmatrix} + \varrho^2 \begin{pmatrix} m_1^2 & m_1 m_2 & m_1 m_3 \\ m_2 m_1 & m_2^2 & m_2 m_3 \\ m_3 m_1 & m_3 m_2 & m_3^2 \end{pmatrix}, \tag{5.22}$$

that is,

$$\begin{aligned} (1 + \varrho m_1)^2 &= 0, \\ \varrho m_2 (1 + \varrho m_1) &= 0, \\ \varrho m_3 (1 + \varrho m_1) &= 0, \\ \varrho^2 m_2^2 &= 0, \\ \varrho^2 m_2 m_3 &= 0, \\ \varrho^2 m_3^2 &= \lambda_3 - 1, \end{aligned} \tag{5.23}$$

with solutions

$$m_1 = -\frac{1}{\varrho}, \quad m_2 = 0, \quad m_3 = \frac{\pm\sqrt{\lambda_3 - 1}}{\varrho}, \tag{5.24}$$

which are already covered by (5.13). □

Remarks.

1. Consider solutions of (5.12) with $1 + b \cdot m \geqq 0$. If $\lambda_1 < \lambda_2 = 1 < \lambda_3$, then there are two essentially distinct such solutions $b^- \otimes m^-$ and $b^+ \otimes m^+$, consistent with the analysis of [24, Appendix 1]. These are related by a rotation $\tilde{R}$ in the sense that $1 + b^+ \otimes m^+ = \tilde{R}(1 + b^- \otimes m^-)$; this follows from the polar decomposition theorem in the case $\lambda_1 > 0$ and by an explicit calculation if $\lambda_1 = 0$. If λ_1 or λ_3 equals 1, there is only one solution. If λ_1 and λ_3 both equal 1, there is no solution.
2. Note that if $m^\perp$ is perpendicular to m then

$$m^\perp \cdot C m^\perp = |m^\perp|^2, \tag{5.25}$$

so that m is normal to an "undistorted" plane for C. Similarly, if $b^\perp$ is perpendicular to b, then

$$b^\perp \cdot C^{-1} b^\perp = |b^\perp|^2 \tag{5.26}$$

so that b is normal to an "undistorted" plane for C^{-1}.

3. The formula for $\boldsymbol{b}$ can be written

$$\boldsymbol{b} = \frac{\varrho}{\sqrt{\lambda_1^{-1} - \lambda_3^{-1}}} \left(\sqrt{\lambda_1^{-1} - 1}\, \boldsymbol{e}_1 + \varkappa \sqrt{1 - \lambda_3^{-1}}\, \boldsymbol{e}_3\right) \tag{5.27}$$

which has the same form as the formula for $\boldsymbol{m}$.

Proposition 4 makes it clear in particular that in order to solve the basic equation (5.11) we must show that $C_0(\lambda)$ has an eigenvalue equal to 1 for some λ, so we consider the following proposition:

Proposition 5. *Let the* 3×3 *nonsingular matrix* $\boldsymbol{U}_0 = \boldsymbol{U}_0^T$, *the vector* $\boldsymbol{a}$, *the unit vector* $\boldsymbol{n}$ *and rotations* $\boldsymbol{R}$ *and* $\bar{\boldsymbol{R}}$ *be given subject to the twinning relations* (5.4). *Let* $C_0(\lambda) = (\boldsymbol{U}_0 + \lambda \boldsymbol{n} \otimes \boldsymbol{a})(\boldsymbol{U}_0 + \lambda \boldsymbol{a} \otimes \boldsymbol{n})$ *and let*

$$g(\lambda) = \det (C_0(\lambda) - \mathbf{1}). \tag{5.28}$$

Then $g(\lambda)$ *is a quadratic function of* λ *which satisfies* $g(\lambda) = g(1 - \lambda)$.

Proof. To show that $g(\lambda)$, which appears to be a sixth order polynomial, is in fact only quadratic, note that by (5.5)

$$\det (\boldsymbol{U}_0 + \lambda \boldsymbol{a} \otimes \boldsymbol{n}) = \det \boldsymbol{U}_0 \neq 0. \tag{5.29}$$

Hence

$$\begin{aligned} g(\lambda) &= \det [(\boldsymbol{U}_0 + \lambda \boldsymbol{n} \otimes \boldsymbol{a})(\boldsymbol{U}_0 + \lambda \boldsymbol{a} \otimes \boldsymbol{n}) - \mathbf{1}] \\ &= \det \boldsymbol{U}_0 \det [(\boldsymbol{U}_0 + \lambda \boldsymbol{a} \otimes \boldsymbol{n}) - (\boldsymbol{U}_0 + \lambda \boldsymbol{n} \otimes \boldsymbol{a})^{-1}] \\ &= \det \boldsymbol{U}_0 \det [(\boldsymbol{U}_0 - \boldsymbol{U}_0^{-1}) + \lambda(\boldsymbol{a} \otimes \boldsymbol{n} + \boldsymbol{U}_0^{-1}\boldsymbol{n} \otimes \boldsymbol{U}_0^{-1}\boldsymbol{a})]. \end{aligned} \tag{5.30}$$

Since the matrix multiplying λ is singular, the right-hand side of $(5.30)_3$ is at most quadratic in λ.

Since

$$g(1) = \det (\bar{\boldsymbol{R}}^T(\boldsymbol{U}_0^2 - 1)\bar{\boldsymbol{R}}) = \det (\boldsymbol{U}_0^2 - 1) = g(0),$$

if follows that $g(\lambda) = g(1 - \lambda)$. □

From Proposition 5 we can see how to make one eigenvalue of $C_0(\lambda)$ equal to 1 at some λ so as to satisfy part of the conditions in Proposition 4. For the remaining part we need to show that the other two eigenvalues of $C_0(\lambda)$ bound 1 above and below, using the following proposition:

Proposition 6. *Suppose that for some* λ, $C_0(\lambda)$ *has the unordered triple of eigenvalues* $1, \lambda_1, \lambda_3$. *Then*

$$(1 - \lambda_1)(\lambda_3 - 1) = \operatorname{tr} \boldsymbol{U}_0^2 - \det \boldsymbol{U}_0^2 - 2 + (\lambda^2 - \lambda)\,|\boldsymbol{a}|^2. \tag{5.31}$$

Proof. Using (5.6) we obtain

$$\begin{aligned} 1 + \lambda_1 + \lambda_3 &= \operatorname{tr} C_0(\lambda) \\ &= \operatorname{tr} U_0^2 + 2\lambda(U_0 a \cdot n) + \lambda^2 |a|^2 \qquad (5.32) \\ &= \operatorname{tr} U_0^2 + (\lambda^2 - \lambda) |a|^2. \end{aligned}$$

Since also

$$\lambda_1 \lambda_3 = \det C_0(\lambda) = \det U_0^2, \tag{5.33}$$

the result follows. □

We now combine Propositions 4 through 6 to get an existence theorem for the original equation (5.9). For the purpose of Theorem 7 a triple $(\hat{R}, \lambda, b \otimes m)$ consisting of a rotation $\hat{R}$, a scalar $\lambda \in (0, 1)$ and a rank-one matrix $b \otimes m$ such that

$$U_0 + \lambda a \otimes n = \hat{R}^T(1 + b \otimes m) \tag{5.34}$$

will be termed a *solution* of (5.34).

Theorem 7. *Let the positive-definite symmetric matrix* U_0 *satisfy the twinning relation*

$$R U_0 \bar{R} = U_0 + a \otimes n \tag{5.35}$$

for some pair of rotations R *and* $\bar{R}$ *and for vectors* $a \neq 0$ *and* n, $|n| = 1$.

I. *Assume* U_0 *does not have an eigenvalue equal to* 1. *Necessary and sufficient conditions that* (5.34) *has a solution are that*

$$1 + \tfrac{1}{2}\delta^* \leqq 0 \tag{5.36}$$

and that

$$\operatorname{tr} U_0^2 - \det U_0^2 - 2 + \frac{1}{2\delta^*} |a|^2 \geqq 0, \tag{5.37}$$

where

$$\delta^* = a \cdot U_0 (U_0^2 - 1)^{-1} n. \tag{5.38}$$

If further

$$1 + \tfrac{1}{2}\delta^* < 0, \tag{5.39}$$

then strict inequality holds also in (5.37) *and there are exactly four distinct solutions of* (5.34), *these having the form*

$$\begin{aligned} &(\hat{R}_1, \lambda^*, b_1^+ \otimes m_1^+), \\ &(\hat{R}_2, \lambda^*, b_1^- \otimes m_1^-), \\ &(\hat{R}_3, 1 - \lambda^*, b_2^+ \otimes m_2^+), \\ &(\hat{R}_4, 1 - \lambda^*, b_2^- \otimes m_2^-), \end{aligned} \tag{5.40}$$

where

$$\lambda^* = \tfrac{1}{2}\left(1 - \sqrt{1 + \frac{2}{\delta^*}}\right), \tag{5.41}$$

so that $0 < \lambda^* < 1/2$. *If*

$$1 + \tfrac{1}{2}\delta^* = 0, \tag{5.42}$$

then all solutions have $\lambda = 1/2$; *if strict inequality holds in* (5.37) *then there are exactly two distinct solutions, while if equality holds in* (5.37) *there is just one solution.*

II. *Assume* U_0 *has an eigenvalue equal to* 1. *A necessary and sufficient condition that* (5.34) *has a solution is that*

$$\mu^* \stackrel{\text{def}}{=} \operatorname{tr} U_0^2 - \det U_0^2 - 2 > 0. \tag{5.43}$$

All solutions are given as follows:

If $\mu^* > \dfrac{|\boldsymbol{a}|^2}{4}$ *then for each* $\lambda \in (0, 1)$ *there are exactly two distinct solutions*

$$\begin{aligned} &(\hat{\boldsymbol{R}}_\lambda^+, \lambda, \boldsymbol{b}_\lambda^+ \otimes \boldsymbol{m}_\lambda^+),\\ &(\hat{\boldsymbol{R}}_\lambda^-, \lambda, \boldsymbol{b}_\lambda^- \otimes \boldsymbol{m}_\lambda^-). \end{aligned} \tag{5.44}$$

If $0 < \mu^* \leqq \dfrac{|\boldsymbol{a}|^2}{4}$ *and* $\bar{\lambda} \stackrel{\text{def}}{=} \tfrac{1}{2}\left(1 - \sqrt{1 - \dfrac{4\mu^*}{|\boldsymbol{a}|^2}}\right)$, *so that* $0 < \bar{\lambda} \leqq \tfrac{1}{2}$, *then for each* $\lambda \in (0, \bar{\lambda}) \cup (1 - \bar{\lambda}, 1)$ *there are exactly two distinct solutions of the form* (5.44), *while if* $\lambda = \bar{\lambda}$ *or* $1 - \bar{\lambda}$ *with either* $\bar{\lambda} \neq \tfrac{1}{2}$ $\left(\text{i.e., } \mu^* < \dfrac{|\boldsymbol{a}|^2}{4}\right)$ *or* $\bar{\lambda} = \tfrac{1}{2}$ *and* $\det U_0 \neq 1$, *then there is one solution* $(\hat{\boldsymbol{R}}_\lambda, \lambda, \boldsymbol{b}_\lambda \otimes \boldsymbol{m}_\lambda)$.

In all the cases above, formulas for $\boldsymbol{b} \otimes \boldsymbol{m}$ *associated with a solution* $(\hat{\boldsymbol{R}}, \lambda, \boldsymbol{b} \otimes \boldsymbol{m})$ *are given by* (5.13) *evaluated at the ordered eigenvalues of* $C_0(\lambda)$.

Proof. By Proposition 4 and the polar decomposition theorem, necessary and sufficient conditions that the basic equation (5.34) has a solution with $\lambda = \lambda^*$ are that the eigenvalues $0 < \lambda_1 \leqq \lambda_2 \leqq \lambda_3$ of $C_0(\lambda^*)$ satisfy $\lambda_2 = 1$ and $(\lambda_1 - 1)^2 + (\lambda_3 - 1)^2 \neq 0$. (The condition $\lambda_1 > 0$ is automatically satisfied because $\det(U_0 + \lambda \boldsymbol{a} \otimes \boldsymbol{n}) = \det U_0 > 0$ for all λ.)

Part I. By Proposition 5, g can be written in the form

$$g(\lambda) = a(\lambda - \tfrac{1}{2})^2 + b, \tag{5.45}$$

for some constants a and b with $b = g(\tfrac{1}{2})$ and

$$g(0) = \tfrac{1}{4}a + b, \quad g'(0) = -a. \tag{5.46}$$

The condition that $C_0(\lambda^*)$ has an eigenvalue equal to 1 for some $\lambda^* \in (0, 1)$ is that $g(\lambda^*) = \det(C_0(\lambda^*) - \mathbf{1}) = 0$ for some $\lambda^* \in (0, 1)$, which holds if and only if $g(0)\, g(\tfrac{1}{2}) \leqq 0$ with strict inequality if $\lambda^* \neq \tfrac{1}{2}$.

By direct calculation,

$$g(0) = \det(U_0^2 - \mathbf{1}),$$
$$g'(0) = 2\boldsymbol{a} \cdot U_0 \operatorname{adj}(U_0^2 - \mathbf{1})\,\boldsymbol{n}, \tag{5.47}$$

so that by (5.46) the inequality $g(0)\,g(\frac{1}{2}) \leqq 0$ is equivalent to (5.36). If $g(\lambda^*) = 0$, $\lambda^* \in (0, 1)$, then by (5.45) through (5.47)

$$\lambda^{*2} - \lambda^* = \frac{1}{2\delta^*}, \tag{5.48}$$

with δ^* defined by (5.38). Hence, by Proposition 6, the eigenvalues of $C_0(\lambda^*)$ satisfy $\lambda_1 \leqq \lambda_2 = 1 \leqq \lambda_3$ if and only if (5.36) and (5.37) hold. The eigenvalues of $C_0(\lambda^*)$ are not all 1, since $C_0(\lambda^*) = \mathbf{1}$ implies that $C_0(\lambda^*)\,\boldsymbol{e} = U_0^2\boldsymbol{e} = \boldsymbol{e}$ for $\boldsymbol{e}$ perpendicular to $\boldsymbol{n}$ and $U_0\boldsymbol{a}$, and U_0^2 does not have an eigenvalue 1.

If (5.39) holds then $g(0)\,g(\frac{1}{2}) < 0$, so that g has the two roots λ^*, $1 - \lambda^*$ with λ^* given by (5.41). By Proposition 6 equality holds in (5.37) if and only if 1 is a double eigenvalue of $C_0(\lambda^*)$. But this is impossible, since there would then exist a corresponding eigenvector $\boldsymbol{e}$ with $\boldsymbol{e} \cdot \boldsymbol{n} = 0$, so that

$$U_0^2\boldsymbol{e} + \lambda^*(U_0\boldsymbol{a} \cdot \boldsymbol{e})\,\boldsymbol{n} = \boldsymbol{e}, \tag{5.49}$$

which with (5.47) implies that

$$g(0)\,(U_0\boldsymbol{a} \cdot \boldsymbol{e}) = -\tfrac{1}{2}\,\lambda^* g'(0)\,(U_0\boldsymbol{a} \cdot \boldsymbol{e}). \tag{5.50}$$

The quantity $U_0\boldsymbol{a} \cdot \boldsymbol{e}$ does not vanish because of (5.49) and the assumption that U_0 does not have an eigenvalue equal to 1. Hence (5.50) implies that

$$g(0) = -\tfrac{1}{2}\,\lambda^* g'(0), \tag{5.51}$$

which immediately gives $\lambda^* = \frac{1}{2}$, a contradiction. Hence, by Proposition 4 there are four distinct solutions of (5.34) as claimed.

If (5.42) holds then by (5.48) $\lambda^* = \frac{1}{2}$ and by Proposition 4 there are two solutions if strict inequality holds in (5.38) and only one otherwise.

Part II. If U_0 has an eigenvalue equal to 1, then $g(0) = g(1) = \det(U_0^2 - \mathbf{1}) = 0$. Hence, if there is a solution of (5.34) with $\lambda = \lambda^*$ (which means that $\lambda^* \in (0, 1)$), then $g(\lambda) = 0$ for all λ. Hence, one eigenvalue of $C_0(\lambda)$ equals 1 for each $\lambda \in (0, 1)$ and it remains to examine the other two eigenvalues. Let

$$\theta(\lambda) = \operatorname{tr} U_0^2 - \det U_0^2 - 2 + (\lambda^2 - \lambda)\,|\boldsymbol{a}|^2. \tag{5.52}$$

A necessary and sufficient condition that $\theta(\lambda) \geqq 0$ for some $\lambda \in (0, 1)$ is that $\mu^* > 0$. If $\theta(\lambda) > 0$, $\lambda \in (0, 1)$, by Proposition 6 the eigenvalues of $C_0(\lambda)$ satisfy $\lambda_1 < \lambda_2 = 1 < \lambda_3$ and hence there are two distinct solutions of the form (5.44); this case occurs if $\mu^* > |\boldsymbol{a}|^2/4$ or if $0 < \mu^* \leqq |\boldsymbol{a}|^2/4$ and $\lambda \in (0, \bar{\lambda}) \cup (1 - \bar{\lambda}, 1)$. If $\theta(\lambda) = 0$, that is if $\lambda = \bar{\lambda}$ or $1 - \bar{\lambda}$, there is a single solution $(\hat{\boldsymbol{R}}_\lambda, \lambda, \boldsymbol{b}_\lambda \otimes \boldsymbol{m}_\lambda)$ if and only if $C_0(\lambda) \neq \mathbf{1}$. But $C_0(\lambda) = \mathbf{1}$ implies by (5.5), (5.6) that

$$0 = U_0^{-1}\boldsymbol{a} \cdot C_0(\lambda)\,\boldsymbol{n} = (\lambda - \tfrac{1}{2})\,|\boldsymbol{a}|^2 \tag{5.53}$$

and hence that $\lambda = \frac{1}{2}$. Hence there is a single solution if $\bar{\lambda} \neq \frac{1}{2}$. If $\bar{\lambda} = \frac{1}{2}$ then $C_0(\bar{\lambda}) = \mathbf{1}$ implies $\det U_0 = (\det C_0(\bar{\lambda}))^{1/2} = 1$, while $\det U_0 = 1$ implies by (5.32) and $\theta(\bar{\lambda}) = 0$ that $\operatorname{tr} C_0(\bar{\lambda}) = 3$, which together with the fact that $C_0(\bar{\lambda})$ has two eigenvalues equal to 1 gives $C_0(\bar{\lambda}) = \mathbf{1}$. Hence there is a solution if and only if $\det U_0$ is not equal to one. □

Remarks:

1. An alternative method to that in Proposition 5 for evaluating $g(\lambda)$ is to write

$$\begin{aligned} g(\lambda) &= \det(C_0(\lambda) - \mathbf{1}) \\ &= \det C_0(\lambda) - \operatorname{tr} \operatorname{adj} C_0(\lambda) + \operatorname{tr} C_0(\lambda) - 1 \end{aligned} \tag{5.54}$$

and note that

$$\begin{aligned} \operatorname{tr} \operatorname{adj} C_0(\lambda) &= \det C_0(\lambda) \operatorname{tr} C_0^{-1}(\lambda) \\ &= \det U_0^2 \operatorname{tr} (\mathbf{1} - \lambda U_0^{-1} \boldsymbol{a} \otimes \boldsymbol{n}) U_0^{-2} (\mathbf{1} - \lambda \boldsymbol{n} \otimes U_0^{-1} \boldsymbol{a}) \\ &= \det U_0^2 (\operatorname{tr} U_0^{-2} + (\lambda^2 - \lambda) \mid U_0^{-1} \boldsymbol{a} \mid^2 \mid U_0^{-1} \boldsymbol{n} \mid^2), \end{aligned} \tag{5.55}$$

where we have used (5.7). Thus, from (5.32),

$$g(\lambda) = \det(U_0^2 - \mathbf{1}) + (\lambda^2 - \lambda)(\mid \boldsymbol{a} \mid^2 - (\det U_0^2) \mid U_0^{-1} \boldsymbol{a} \mid^2 \mid U_0^{-1} \boldsymbol{n} \mid^2), \tag{5.56}$$

and hence

$$\delta^* = \frac{\det U_0^2 \mid U_0^{-1} \boldsymbol{a} \mid^2 \mid U_0^{-1} \boldsymbol{n} \mid^2 - \mid \boldsymbol{a} \mid^2}{2 \det(U_0^2 - 1)}. \tag{5.57}$$

2. A different way of writing the necessary and sufficient conditions in Theorem 7 that there be a solution of (5.34) with $\lambda \neq 1/2$ can be obtained by noting that if $C_0(\lambda)$ has the unordered triple of eigenvalues $1, \lambda_1, \lambda_3$, and if $\boldsymbol{e}$ is any eigenvector of $C_0(\lambda)$ corresponding to the eigenvalue 1, then

$$\begin{aligned} &(1 - \lambda_1)(\lambda_3 - 1)(\boldsymbol{e} \cdot \boldsymbol{n})^2 \\ &\qquad = \operatorname{tr} C_0(\lambda) - \boldsymbol{n} \cdot C_0(\lambda) \boldsymbol{n} - \boldsymbol{n} \cdot \operatorname{adj} C_0(\lambda) \boldsymbol{n} - 1. \end{aligned} \tag{5.58}$$

Since $\operatorname{tr} C_0(\lambda) - \boldsymbol{n} \cdot C_0(\lambda) \boldsymbol{n} = \operatorname{tr} U_0^2 - \boldsymbol{n} \cdot U_0^2 \boldsymbol{n}$ and $\boldsymbol{n} \cdot \operatorname{adj} C_0(\lambda) \boldsymbol{n} = \boldsymbol{n} \cdot \operatorname{adj} U_0^2 \boldsymbol{n}$, we have that

$$(1 - \lambda_1)(\lambda_3 - 1)(\boldsymbol{e} \cdot \boldsymbol{n})^2 = \operatorname{tr} U_0^2 - \boldsymbol{n} \cdot U_0^2 \boldsymbol{n} - \boldsymbol{n} \cdot \operatorname{adj} U_0^2 \boldsymbol{n} - 1. \tag{5.59}$$

Since, as is shown in the proof of Theorem 7, $\boldsymbol{e} \cdot \boldsymbol{n}$ can vanish only if $\lambda = 1/2$, it follows that necessary and sufficient conditions that (5.34) has a solution with $\lambda \neq 1/2$ are that (5.39) holds and

$$\operatorname{tr} U_0^2 - \boldsymbol{n} \cdot U_0^2 \boldsymbol{n} - \boldsymbol{n} \cdot \operatorname{adj} U_0^2 \boldsymbol{n} - 1 > 0. \tag{5.60}$$

From the physical point of view, it is rare to have a material whose measured transformation strain has an eigenvalue equal to 1 within experimental error. We have not been able to find any such examples* in the literature after an extensive search, although even in some relatively common alloys the transforma-

* See Notes added in proof.

tion strain has not been measured. Nevertheless, the transformation strain is generally a continuous function of composition and while the composition of an alloy does not change during the martensitic transformation, it can be adjusted when the crystal is originally grown. Thus, it seems possible that for very special compositions, alloys could be made that have an eigenvalue of $\boldsymbol{U}_0$ equal to 1 in addition to the property (5.43). Such alloys would be interesting because of the great variety of austenite/martensite interfaces possible, *i.e.*, those given by (5.44). However, austenite/martensite interfaces in a cubic-to-tetragonal transformation are only possible when $\boldsymbol{U}_0$ does not have an eigenvalue equal to 1, as is evident from (5.43) applied to (3.1) with either $\eta_1 = 1$ or $\eta_2 = 1$. The possibility mentioned above can occur in cubic-to-orthorhombic transformations, for example.

We now specialize the calculations to the cubic-to-tetragonal transformation and then to the specific case of InTl. Let $\boldsymbol{U}_0$ have the form (3.1) with $\eta_2 \neq 1$ and $\eta_1 \neq 1$ and let $\boldsymbol{a}$ and $\boldsymbol{n}$ be given by the twinning formulas $(3.19)_{1,2}$. Since $\boldsymbol{U}_0$ in this case does not have an eigenvalue equal to 1, we turn to Part I of Theorem 7. There is a solution of the basic equation (5.34) with $\lambda^* \neq 1/2$ if and only if (5.39) and (5.37) are satisfied with strict inequality. This pair of inequalities is equivalent to the two conditions

$$\begin{aligned} &\eta_1 < 1 < \eta_2 \quad \text{and} \quad \frac{1}{\eta_1^2} + \frac{1}{\eta_2^2} < 2, \quad \text{or} \\ &\eta_2 < 1 < \eta_1 \quad \text{and} \quad \eta_1^2 + \eta_2^2 < 2. \end{aligned} \tag{5.61}$$

The inequalities (5.61) delineate the hatched region of Figure 5. There are also solutions for which $\lambda = 1/2$ that occur when the conditions (5.37) and (5.42)

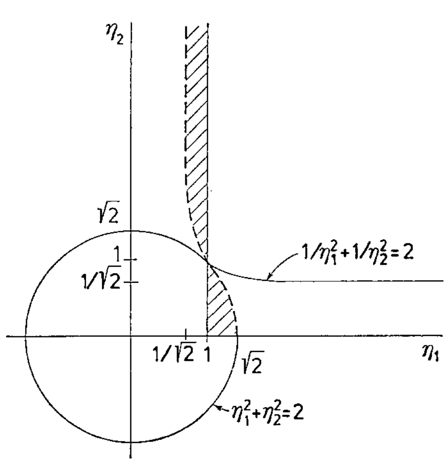

Fig. 5. Values of $\eta_1 > 0$ and $\eta_2 > 0$ for which a solution of (5.34) exists with $\lambda \neq 1/2$ are contained in the hatched region, not including its boundary. Solutions of (5.34) with $\lambda = 1/2$ exist for η_1 and η_2 on the dashed curve, not including the point (1, 1).

are satisfied. These conditions are

$$\begin{aligned} &\eta_1 < 1 \quad \text{and} \quad \frac{1}{\eta_1^2} + \frac{1}{\eta_2^2} = 2, \quad \text{or} \\ &\eta_1 > 1 \quad \text{and} \quad \eta_1^2 + \eta_2^2 = 2, \end{aligned} \tag{5.62}$$

which are satisfied on the dashed curves of Figure 5. No solutions are possible when either $\eta_1 = 1$ or $\eta_2 = 1$. We now give formulas for the solutions under the conditions (5.61). The value of λ^* is given by (5.41), which becomes

$$\lambda^* = \tfrac{1}{2}\left[1 - \left(2(\eta_2^2 - 1)(\eta_1^2 - 1)(\eta_1^2 + \eta_2^2)(\eta_2^2 - \eta_1^2)^{-2} + 1\right)^{1/2}\right]. \tag{5.63}$$

To complete the calculation we need to find the eigenvalues and eigenvectors of $C_0(\lambda^*)$, according to the last statement of Theorem 7. Once these are found, the vectors $\boldsymbol{b}_{\mathrm{I}}^{\pm}$ and $\boldsymbol{m}_{\mathrm{I}}^{\pm}$ of (5.40) are determined by (5.13). To find the eigenvalues $\lambda_1 \leqq \lambda_2 = 1 \leqq \lambda_3$ of $C_0(\lambda^*)$, it is helpful to note that by (5.32) and (5.33)

$$\begin{aligned} 1 + \lambda_1 + \lambda_3 &= 1 + \eta_1^2 + \eta_1^2\eta_2^2, \\ \lambda_1\lambda_3 &= \eta_1^4\eta_2^2. \end{aligned} \tag{5.64}$$

Hence $\{\lambda_1, \lambda_3\} = \{\eta_1^2, \eta_1^2\eta_2^2\}$. Let $\{\boldsymbol{e}, \bar{\boldsymbol{e}}, \bar{\bar{\boldsymbol{e}}}\}$ be the orthonormal basis introduced in (3.1) and (3.19) to (3.21). To find the eigenvectors of $C_0(\lambda^*)$, notice that by a direct calculation $\bar{\bar{\boldsymbol{e}}}$ is an eigenvector of $C_0(\lambda^*)$ corresponding to the eigenvalue η_1^2. The remaining two eigenvectors are found by a brute force computation which is simplified by the identity

$$(2\eta_1^2\eta_2^2 - \eta_1^2 - \eta_2^2)(\eta_1^2 + \eta_2^2 - 2) = 2(\eta_1^2 - 1)(\eta_2^2 - 1)(\eta_1^2 + \eta_2^2) + (\eta_2^2 - \eta_1^2)^2. \tag{5.65}$$

It is also helpful to change to the variables δ and τ defined by

$$\begin{aligned} \delta &= [(\eta_2^2 + \eta_1^2 - 2)(1 - \eta_1^2)^{-1}]^{1/2}, \\ \tau &= [(2\eta_1^2\eta_2^2 - \eta_1^2 - \eta_2^2)(1 - \eta_1^2)^{-1}]^{1/2}. \end{aligned} \tag{5.66}$$

Then, in the basis $\{\boldsymbol{e}, \bar{\boldsymbol{e}}, \bar{\bar{\boldsymbol{e}}}\}$, the other two eigenvectors are

$$\frac{1}{(2(\tau^2 + \delta^2))^{1/2}}(\tau \pm \delta, \tau \mp \delta, 0). \tag{5.67}$$

Therefore, by Proposition 4 with ϱ chosen to make the third component of $\boldsymbol{m}_{\mathrm{I}}^{\pm}$ equal to 1 for simplicity, we get the formulas

$$\begin{aligned} \boldsymbol{b}_{\mathrm{I}}^{\pm} &= (\mp\tfrac{1}{2}\zeta(\delta + \tau), \pm\tfrac{1}{2}\zeta(\delta - \tau), \beta), \\ \boldsymbol{m}_{\mathrm{I}}^{\pm} &= (\mp\tfrac{1}{2}(\delta + \tau), \pm\tfrac{1}{2}(\delta - \tau), 1) \end{aligned} \tag{5.68}$$

where δ and τ are given by (5.66) and

$$\begin{aligned} \zeta &= (1 - \eta_1^2)(1 + \eta_2)^{-1}, \\ \beta &= \eta_2(\eta_1^2 - 1)(1 - \eta_2)^{-1}. \end{aligned} \tag{5.69}$$

The signs are taken in parallel (either all upper or all lower). The expressions for $b_2^{\pm}$ and $m_2^{\pm}$ associated with $(1 - \lambda^*)$ are obtained by changing the signs just before τ on the right-hand sides of $(5.69)_{1,2}$.

The expressions (5.68) are equivalent to the formulas given by WECHSLER, LIEBERMAN & READ [50] and summarized elsewhere as the crystallographic theory of martensite, which is not based on energy considerations as here. The formulas (5.63) and (5.68) simplify considerably in the case tr $U_0^2 = 3$ and have been given by ERICKSEN [19]; this case arises naturally from his theory of constrained crystals. With $\eta_1 = 1 - \varepsilon$ and $\eta_2 = 1 + 2\varepsilon$, all quantities become functions of ε and as $\varepsilon \to 0$, $\lambda^* \to 1/3$ and $m_1^{\pm} \to (\mp 1, 0, 1)$. For InTl, with $\varepsilon = .013$, we get

$$\begin{aligned} \lambda^* &\doteq .338, \\ m_1^{\pm} &\doteq (\mp .993, \pm .0265, 1). \end{aligned} \tag{5.70}$$

This expression for each $m_1^{\pm}$ differs by about 1° from a member of the $\{1\ 1\ 0\}$ family of planes. Note that the acute angle between n and $m_1^{\pm}$ is very nearly 60°, which can be seen from the pictures of BASINSKI & CHRISTIAN [10]. Finally, if we go back and use the full set of symmetry related twins represented by (3.17), we get $24(= 6 \times 4)$ distinct austenite/finely twinned martensite interfaces.

6. Surface Energy and Scaling

The analysis of the minimizing sequences yields the observed austenite/martensite interfaces but otherwise has an obvious flaw. That is, the sequences suggest infinite fineness whereas the spacing of the observed twins is small but nonzero. In this section we explore the idea that interfacial energies, which have been excluded from our total free energy, can account for limited fineness without nullifying the overall conclusions we have reached.

Twin boundaries contribute a small free energy, not accounted for by our ϕ, which is most simply introduced as an energy per unit area assigned on twin boundaries (GIBBS [21, p. 314–328]). For simplicity, we assume that the interfacial free energy per unit area is a constant σ. This assumption is open to question near places where the twin boundaries meet the boundary of the body or the austenite/martensite interface. Whether or not an interfacial energy should be assigned to the austenite/martensite interface is debatable. (Recall that our calculation of Section 4 does assign a bulk energy to this interface of $O(j^{-1})$, *cf.* equation (4.35).) An electron micrograph of an austenite/martensite interface by NAKANISHI [34] does not suggest a surface of discontinuity of the deformation gradient, so we omit the interfacial energy there.

Since the twin boundaries are essentially plane and parallel, we do not expect that the introduction of interfacial energy will change our expressions for the bulk deformation gradients, stresses or energies. Hence, we assume F^+ and F^- are given by the equations $(4.34)_{1,2}$, and we repeat the construction in Theorem 3, except that now we assume that the total free energy is

$$\bar{\mathscr{I}}[y^{(j)}] = \int_{\Omega} \phi(Dy^{(j)}, \theta_0)\, dx + \sigma A(j), \tag{6.1}$$

$A(j)$ being the total area of twin boundaries in $y^{(j)}(\Omega)$. Here we allow j to take any nonnegative real value, defining $y^{(j)}$ by (4.29) for $j > 0$ and $y^{(0)} = y$ given by (4.30). $A(j)$ implicitly depends on Ω and the deformation $y^{(j)}$. In any case,

$$A(j) \begin{cases} = \text{const.} & \text{at } j = 0, \\ \to \infty & \text{as } j \to \infty. \end{cases} \tag{6.2}$$

With a suitable choice of Ω, $A(j)$ is continuous. The first term in (6.1) tends to zero as $j \to \infty$ and is a continuous function of j. Thus $\bar{\mathscr{I}}[y^{(j)}]$ attains its minmum at some $j_0 < \infty$. A quantitative calculation of j_0 based on linear elasticity is given by BURKART & READ [13].

In this calculation we have assumed a certain interpolation of the austenite/martensite interface involving the deformation gradients D^+ and D^- (see Figure 4). This choice was made so as to simplify the task of showing that each $y^{(j)}(x)$ is invertible. It appears that this particular interpolation does not have any special physical significance. Other interpolations might lead to a uniformly lower value of the bulk energy and therefore a higher value of j_0. It appears that with any fixed interpolation and realistic free energy functions we would always get limited fineness by this calculation.

It is interesting to speculate on the reasons why some materials that are very similar to InTl do not form internally twinned martensite. The high temperature A-15 superconductors (see ERICKSEN [16]) undergo a reversible cubic to tetragonal transformation,* the martensite is found twinned on the $\{1\ 1\ 0\}$ planes, but the austenite and martensite do *not* co-exist at equilibrium as in InTl. In this connection we note that the deformations found in this section are at best metastable with regard to the total energy given by (6.1) since any of the linear deformations $y = F^+x$, $y = F^-x$, $y = x$ have less energy than twinned deformations. Whether the twinned deformations can really be some kind of relative minima of a total energy, which includes interfacial energy in a general way, appears to be a delicate matter.

Another possibly significant fact is that any single crystal of an alloy in solid solution, which is grown from the melt, inevitably contains a slight concentration gradient due to segregation during growth. Also, it is impossible to eliminate completely temperature gradients in a heat bath. Both concentration and temperature gradients can be modelled by explicit dependence of the free energy on x. We explore the consequences of this in Section 7(a), where we show that it is a possible mechanism for the initiation of fine twinning.

The reasoning summarized by the total energy (6.1) suggests that places in Ω where the cross-section

$$\{x \in \Omega_{F^+,F} : x \cdot n = \text{const.}\} \tag{6.3}$$

has a small area should contain a larger twin density than in regions where this area is large. To explore this suggestion via a heuristic calculation, we consider a specimen with uniform length and width but with a variable thickness and suppose an austenite/martensite interface divides the specimen perpendicular to the thickness direction. Assume $m \wedge n$ is parallel to the width. We use the

* However, it is not clear that this transformation is really of 1st order ($U_0 \neq 1$).

calculation of bulk energy given in Section 4 directly and therefore ignore the fact that with limited fineness the boundary of the body will not quite be free of traction. The bulk energy per unit length is given approximately by (const. j^{-1}) while the surface energy per unit length is given approximately by (const. jA), A being the area of the set given in (6.3) and the constants being positive. The total energy per unit length is therefore

$$\frac{\text{const.}}{j} + \text{const.}\ Aj \tag{6.4}$$

which is minimized among all $j > 0$ at

$$j = \text{const.}\ A^{-1/2}. \tag{6.5}$$

Since j^{-1} is proportional to the twin spacing (see Figure 4), equation (6.5) gives an inverse square root relation between the fineness and the cross-sectional area. Apparently, experiments to test this relation are not available.

OTSUKA & SHIMIZU [37] observe that "the reason for the absence of internal twins in 'small' martensites is not known presently". In the spirit of the calculation just above, consider a cube of side L divided by an austenite/martensite interface which is parallel to a pair of faces. Again by use of expressions from Section 4 and the total energy (6.1), the bulk energy in this situation is approximately (const. $L^2 j^{-1}$) while the surface energy is approximately (const. $L^3 j$), the constants being positive and independent of L. The total energy is minimized as a function of $j > 0$ when

$$j = \text{const.}\ L^{-1/2}. \tag{6.6}$$

The twin spacing is therefore proportional to $L^{1/2}$. However, the cube is of side L. Hence, if L is sufficiently small the twin spacing will be larger than a side of the cube, suggesting that small crystals containing both austenite and martensite will not be stable, as is observed.

7. Other Similar Phenomena

a. Fine twins in a problem with no absolute minimizer

The minimization problem we have studied so far has linear absolute minima in addition to the minimizing sequences described in Section 4. By allowing the free energy ϕ to depend explicitly on x, we now construct a similar example in which the total free energy does not have an absolute minimizer in $W^{1,1}(\Omega, \mathbb{R}^3)$.

One way to think of doing this is to put the body in a temperature gradient so as to introduce dependence on x through the composition $\phi(F, \theta(x))$. Within the context of a thermodynamic theory based on the Planck inequality and a Fourier Law of heat conduction, BALL & KNOWLES [9] justify the criterion of stability

$$\min_y \int_\Omega \phi(Dy(x), \theta(x))\, dx \tag{7.1}$$

as appropriate for an unloaded body having a steady temperature distribution $\theta(x)$.

To avoid technicalities, we ignore the Galilean invariance of the free energy and consider a linear temperature distribution. Specifically, let a smooth function $\tilde{\phi}: \mathscr{D} \times \mathbb{R} \to \mathbb{R}$ satisfy for each $\theta < 0$

$$\tilde{\phi}(\boldsymbol{F}, \theta) > \tilde{\phi}(\boldsymbol{F}^+, \theta) = \tilde{\phi}(\boldsymbol{F}^-, \theta) \quad \text{for } \boldsymbol{F} \neq \boldsymbol{F}^\pm, \tag{7.2}$$

while for $\theta > 0$, let

$$\tilde{\phi}(\boldsymbol{F}, \theta) > \tilde{\phi}(\mathbf{1}, \theta) \quad \text{for } \boldsymbol{F} \neq \mathbf{1}. \tag{7.3}$$

Assume $\boldsymbol{F}^\pm$ are given by (4.34) for appropriate choices of the vectors. As done there, assume $\boldsymbol{F}^\pm - \mathbf{1}$ is not a rank-one matrix. Suppose $\boldsymbol{o} \in \Omega$. Let

$$\tilde{\phi}(\boldsymbol{F}, \boldsymbol{x}) = \tilde{\phi}(\boldsymbol{F}, \boldsymbol{x} \cdot \boldsymbol{m}) \tag{7.4}$$

and consider the total free energy

$$\mathscr{I}[y] = \int_\Omega \tilde{\phi}(Dy(\boldsymbol{x}), \boldsymbol{x})\, d\boldsymbol{x}. \tag{7.5}$$

As before let $\mathscr{A} = \{y \in W^{1,1}(\Omega, \mathbb{R}^3): Dy \in \mathscr{D} \text{ a.e.}\}$. Then

$$\inf_{y \in \mathscr{A}} \mathscr{I}[y] = \int_{\Omega \cap \{\boldsymbol{x} \cdot \boldsymbol{m} < 0\}} \tilde{\phi}(\boldsymbol{F}^\pm, \boldsymbol{x} \cdot \boldsymbol{m})\, dx + \int_{\Omega \cap \{\boldsymbol{x} \cdot \boldsymbol{m} > 0\}} \tilde{\phi}(\mathbf{1}, \boldsymbol{x} \cdot \boldsymbol{m})\, d\boldsymbol{x},$$

and examples of minimizing sequences are given by the family of functions $\boldsymbol{y}^{(j)}(\boldsymbol{x})$ constructed in Theorem 3.

We claim that the absolute minimum of $\mathscr{I}$ in $\mathscr{A}$ is not attained. This in fact follows from Theorem 3. That is, any absolute minimizer $\boldsymbol{y} \in \mathscr{A}$ must have the property

$$Dy = \boldsymbol{F}^+ \quad \text{or} \quad \boldsymbol{F}^- \text{ a.e. on } \Omega \cap \{\boldsymbol{x} \cdot \boldsymbol{m} < 0\} \tag{7.6}$$

and the property

$$Dy = \mathbf{1} \text{ a.e. on } \Omega \cap \{\boldsymbol{x} \cdot \boldsymbol{m} \geqq 0\}. \tag{7.7}$$

Any such $\boldsymbol{y}$ belongs to $W^{1,\infty}(\Omega, \mathbb{R}^3)$ and is therefore continuous on Ω. Let B be an open ball with center $\boldsymbol{o}$ and contained in Ω. On $\Omega \cap \{\boldsymbol{x} \cdot \boldsymbol{m} \geqq 0\}$ any such $\boldsymbol{y}$ satisfies

$$\boldsymbol{y}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{c}_1, \tag{7.8}$$

for some $\boldsymbol{c}_1 = \text{const.}$, while on $E = \Omega \cap \{\boldsymbol{x} \cdot \boldsymbol{m} \leqq 0\}$ $\boldsymbol{y}$ is given by the expression

$$\boldsymbol{y}(\boldsymbol{x}) = \boldsymbol{c}_0 + \boldsymbol{F}^- \boldsymbol{x} + f_E(\boldsymbol{x} \cdot \boldsymbol{n})\, \boldsymbol{c} \tag{7.9}$$

according to (4.4), where f_E is Lipschitz with derivative 0 or 1 a.e. and $\boldsymbol{c}_0 = \text{const.}$ However, it is easily seen using (4.34) that no continuous $\boldsymbol{y}$ satisfies both (7.9) and (7.8) on $\Omega \cap \{\boldsymbol{x} \cdot \boldsymbol{m} = 0\}$. Hence there are no absolute minimizers of $\mathscr{I}$ in $\mathscr{A}$.

Similar arguments show that for appropriate boundary conditions of place, say $\boldsymbol{y}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{b}(\boldsymbol{x} \cdot \boldsymbol{m})$ for $\boldsymbol{x} \in \partial\Omega$, $\boldsymbol{x} \cdot \boldsymbol{m} < 0$, a minimizer of the original problem (3.11) fails to exist.

b. Strongly elliptic energies with minimizers having fine boundary wrinkles

We now consider free energy functions for isotropic, n-dimensional elastic materials of the type analyzed by BALL [6, Section 6.4] and BALL & MARSDEN [7]. Let $1 < \alpha < n$, $0 < \lambda < \mu < \infty$. Let $\phi_1: (0, \infty) \to (0, \infty)$ be a smooth function satisfying

$$\phi_1' > 0, \quad \phi_1'' > 0, \tag{7.10}$$

and

$$\phi_1(\gamma) = \gamma^\alpha, \quad \lambda \leqq \gamma \leqq \mu. \tag{7.11}$$

Now choose a smooth function $\phi_2: (0, \infty) \to \mathbb{R}$ with the properties

$$\begin{aligned} &\phi_2'' > 0, \\ &\phi_2(\tau) = -n\tau^{\alpha/n}, \qquad \tau \in [\lambda^n, \mu^n], \\ &\phi_2(\tau) > -n\phi_1(\tau^{1/n}), \quad \tau \notin [\lambda^n, \mu^n]. \end{aligned} \tag{7.12}$$

There are functions ϕ_2 satisfying (7.12) because $(\tau^{\alpha/n})'' < 0$ for $\tau \in [\lambda^n, \mu^n]$. Define

$$\phi_3(\gamma_1, \ldots, \gamma_n) = \sum_{i=1}^n \phi_1(\gamma_i) + \phi_2\left(\prod_{i=1}^n \gamma_i\right) \tag{7.13}$$

and

$$\phi(F) = \phi_3(\gamma_1, \ldots, \gamma_n), \tag{7.14}$$

where $\gamma_i = \gamma_i(F)$, $i = 1, \ldots, n$, are the eigenvalues of $(F^TF)^{1/}$.

From (7.10) it follows that ϕ_1 is strictly convex in $\log \gamma$, and hence that the minimum of $\sum_{i=1}^n \phi_1(\gamma_i)$ subject to $\gamma_i > 0$, $\prod_{i=1}^n \gamma_i = \tau$, is attained exactly when $\gamma_1 = \gamma_2 = \ldots = \gamma_n = \tau^{1/n}$. Since by construction the nonnegative function $n\phi_1(\tau^{1/n}) + \phi_2(\tau)$ is zero only when $\tau \in [\lambda^n, \mu^n]$, we have shown that the absolute minima of ϕ subject to $\det F > 0$ are given precisely by F with

$$\gamma_1 = \gamma_2 = \ldots = \gamma_n = \gamma \in [\lambda, \mu]. \tag{7.15}$$

We have thus constructed a free energy function ϕ for an isotropic nonlinear elastic material which has a continuous line of absolute minimizers at dilatations. Surprisingly, ϕ is also strictly polyconvex (see BALL [6]) and strongly elliptic, unlike our energies for crystals. Furthermore, by suitably choosing ϕ_1 and ϕ_2, $\phi(F)$ can be chosen to grow as fast as desired as $|F| \to \infty$.

Consider the problem appropriate to an unloaded body,

$$\min_{y \in \mathscr{A}} \int_\Omega \phi(Dy(x))\, dx, \tag{7.16}$$

where $\Omega \subset \mathbb{R}^n$ is bounded and open and $\mathscr{A} = \{y \in W^{1,1}(\Omega, \mathbb{R}^n): \det Dy(x) > 0$ a.e. $x \in \Omega\}$. The absolute minimizers for this problem are those $y \in \mathscr{A}$ having the property that

$$(Dy(x))^T Dy(x) = \gamma(x)^2\, \mathbf{1} \quad \text{a.e. } x \in \Omega, \tag{7.17}$$

where $\gamma(\boldsymbol{x}) \in [\lambda, \mu]$ a.e.. Equivalently,

$$Dy(\boldsymbol{x}) = \gamma(\boldsymbol{x})\, \boldsymbol{R}(\boldsymbol{x}) \quad \text{a.e. } \boldsymbol{x} \in \Omega, \tag{7.18}$$

for some rotation-valued measurable function $\boldsymbol{R}(\boldsymbol{x})$. The condition (7.18) says that $\boldsymbol{y}$ is *conformal.*

We first consider the case $n = 2$. The conformal mappings are representable by analytic functions $w = f(z)$ with the correspondence $\gamma(x_1, x_2) = |f'(x_1 + ix_2)|$. Thus any function analytic in Ω and such that $\lambda \leqq |f'(z)| \leqq \mu$ generates an absolute minimizer of the problem (7.16). We first take the example

$$f(z) = \gamma z + \varepsilon\varrho e^{iz/\varepsilon}, \tag{7.19}$$

with $\Omega = (0, 1)^2$ and $\gamma \in (\lambda, \mu)$. We pick $\varrho > 0$ sufficiently small that $\gamma \in [\lambda + \varrho, \mu - \varrho]$, and take $\varepsilon > 0$. Since $|e^{iz/\varepsilon}| = |e^{-x_2/\varepsilon} e^{ix_1/\varepsilon}| \leqq 1$ for $x_2 \geqq 0$ we have that $\lambda \leqq |f'(z)| \leqq \mu$ in Ω. Note that f is invertible for $\operatorname{Re} z \geqq 0$, since $|e^{iz} - e^{iw}| \leqq |z - w|$ for $\operatorname{Re} z$, $\operatorname{Re} w \geqq 0$ and $\gamma > \varrho$. The mapping $\boldsymbol{y}^\varepsilon$ corresponding to (7.19) is given by

$$\begin{aligned} y_1^\varepsilon &= \gamma x_1 + \varepsilon\varrho e^{-x_2/\varepsilon} \cos(x_1/\varepsilon), \\ y_2^\varepsilon &= \gamma x_2 + \varepsilon\varrho e^{-x_2/\varepsilon} \sin(x_1/\varepsilon). \end{aligned} \tag{7.20}$$

As $\varepsilon \to 0$, $\boldsymbol{y}^\varepsilon$ has finer and finer oscillations near $x_2 = 0$. Note that $\boldsymbol{y}^\varepsilon \to \gamma\boldsymbol{x}$ as $\varepsilon \to 0$ strongly in $W^{1,p}(\Omega, \mathbb{R}^2)$ for $1 \leqq p < \infty$, but that $\boldsymbol{y}^\varepsilon|_{x_2=0}$ only converges weakly to $\gamma\boldsymbol{x}$ in $W^{1,p}((0,1), \mathbb{R}^2)$ (weak * if $p = \infty$). As a second example, we let

$$f(z) = z^{1+i}, \tag{7.21}$$

taking the principal value, with Ω the unit disc centered at $z = i$. Thus with suitable choices of λ, μ we obtain a minimizer $\boldsymbol{y} \in W^{1,\infty}(\Omega, \mathbb{R}^2)$ which is smooth except at $\boldsymbol{o} \in \Omega$ where there is a spiral singularity with Dy discontinuous. It is easily shown that $\boldsymbol{y}$ is invertible on $\bar{\Omega}$. It is probably significant that the equilibrium equations, when linearized about the deformation $\boldsymbol{y} = \gamma\boldsymbol{x}$, $\gamma \in [\lambda, \mu]$, fail to satisfy the complementing condition of Agmon, Douglis & Nirenberg [2] with respect to boundary conditions of null traction. Simpson & Spector [43] discuss in detail exactly this linearized problem within the context of elasticity theory. See also Simpson & Spector [44] for a discussion of the complementing condition in the context of nonlinear elasticity.

We turn to the case $n \geqq 3$. The conformal transformations are now characterized by Liouville's theorem as products of inversions. Under our regularity assumptions (*i.e.*, $\boldsymbol{y} \in W^{1,\infty}(\Omega, \mathbb{R}^3)$ by (7.18)) an appropriate version of Liouville's theorem has been proved by Reshetnyak [42]. For n odd an example is given by

$$\boldsymbol{y}(\boldsymbol{x}) = -\frac{\boldsymbol{x}}{|\boldsymbol{x}|^2}. \tag{7.22}$$

If $\boldsymbol{o} \in \bar{\Omega}$, then $\boldsymbol{y}$ satisfies (7.18) with $\gamma(\boldsymbol{x}) = |\boldsymbol{x}|^{-2}$. Note that when Ω is convex this furnishes an example of a nontrivial deformation which is an absolute minimizer of the total energy for a strictly polyconvex isotropic unloaded material, and thus bears on a conjecture of NOLL [36] (see also TRUESDELL [47]) to the effect that for rubber-like materials, the absolute minimum is homogeneous and unique up to rigid-body translation and rotation.

c. *Minimizers of energy having a finer and finer mixture of phases as an interface is approached from one side*

VAN TENDELOO, VAN LANDUYT & AMELINCKX [48, Figure 8] show arrays of triangular Dauphiné twins in quartz which get finer and finer in the direction of increasing temperature. This suggests another phenomenon whereby compatibility at an interface is achieved by mixing different deformations in triangles which themselves get finer and finer as the interface is approached from one side.

We choose a simplified free energy function which is not intended to model the behavior of quartz. Also, we ignore the temperature gradient. A free energy function which accounts for the α-β transformation and Dauphiné twinning has been given by JAMES [23].

Let $\{e_1, e_2, e_3\}$ be an orthonormal basis and consider a Galilean invariant free energy function $\phi(F)$ whose point group contains a 180° rotation about e_1. That is, for all F with $\det F > 0$, assume

$$\phi(F) = \phi(U)|_{U=(F^TF)^{1/2}} = \phi(RUR^T) \tag{7.23}$$

with $R = -1 + 2e_1 \otimes e_1$. Let $\varepsilon \in (0, 1)$ and $\gamma \neq 0$ be given constants. Assume that $\phi(U)$ has absolute minima at the three positive symmetric matrices U_1, U_2, U_3 where

$$\begin{aligned} U_1^2 &= (\tfrac{1}{4}\varepsilon^2)\, e_1 \otimes e_1 + (\tfrac{1}{2}\varepsilon)(e_1 \otimes e_2 + e_2 \otimes e_1) + 1, \\ U_2^2 &= (\tfrac{1}{4}\varepsilon^2)\, e_1 \otimes e_1 + (\tfrac{1}{2}\varepsilon)(1-\varepsilon)(e_1 \otimes e_2 + e_2 \otimes e_1) + \varepsilon(\varepsilon - 2)\, e_2 \otimes e_2 + 1, \\ U_3^2 &= \gamma(e_1 \otimes e_2 + e_2 \otimes e_1) + \gamma^2 e_2 \otimes e_2 + 1. \end{aligned} \tag{7.24}$$

Referring to Figure 6(a), consider a deformation $\tilde{y}(x)$ defined on the indicated reference configuration Ω by

$$\tilde{y}(x) = \begin{cases} x & \text{at circled nodes,} \\ x + 2^{-(i+1)}\varepsilon e_2 & \text{at uncircled nodes } x \text{ satisfying} \\ & \quad x \cdot e_2 = (1 - 2^{-i}),\ i = 0, 1, 2, \ldots, \\ x + \gamma e_1[(x \cdot e_2) - 1] & \text{for } x \cdot e_2 > 1, \end{cases} \tag{7.25}$$

and which is linear in each triangle. Assume also $D\tilde{y}e_3 = e_3$. Constructed in this way $D\tilde{y}$ has a.e. only the five distinct values

$$\begin{aligned} F_1 &= 1 + \gamma e_1 \otimes e_2, \\ F_2 &= 1 + \tfrac{1}{2}\varepsilon e_2 \otimes e_1, \\ F_3 &= 1 - \tfrac{1}{2}\varepsilon e_2 \otimes e_1, \\ F_4 &= 1 + \tfrac{1}{2}\varepsilon e_2 \otimes e_1 - \varepsilon e_2 \otimes e_2, \\ F_5 &= 1 - \tfrac{1}{2}\varepsilon e_2 \otimes e_1 - \varepsilon e_2 \otimes e_2. \end{aligned} \tag{7.26}$$

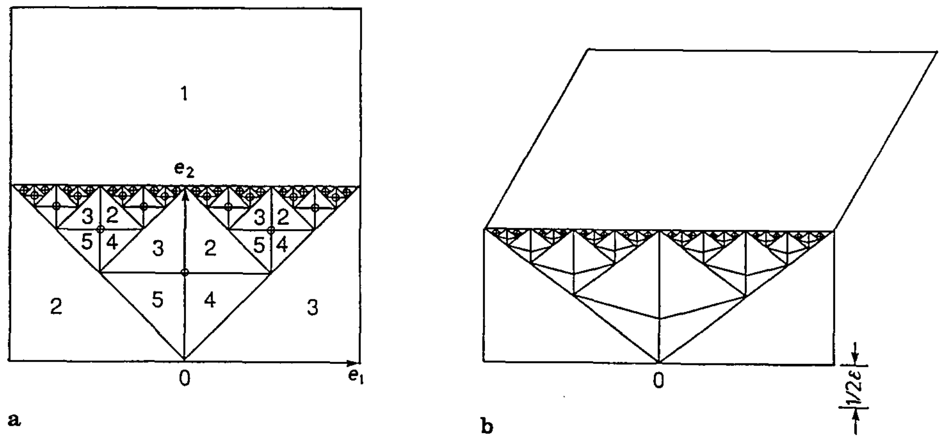

Fig. 6a and b. Compatibility at an interface achieved by a fine mixture of five deformation gradients. a Reference configuration; b Deformed configuration

The subscripts on the deformation gradients in (7.26) correspond to the numbered regions in Figure 6a. Clearly $\tilde{y}(x)$ is in $W^{1,\infty}(\Omega, \mathbb{R}^3)$ and is globally invertible.

The function $\tilde{y}$ we have constructed is an absolute minimizer of the total free energy

$$\int_\Omega \phi(Dy)\,dx \tag{7.27}$$

in the class $W^{1,1}(\Omega, \mathbb{R}^3)$ because

$$\begin{aligned} F_1^T F_1 &= U_3^2, \\ F_2^T F_2 &= R^T F_3^T F_3 R = U_1^2, \\ F_4^T F_4 &= R^T F_5^T F_5 R = U_2^2. \end{aligned} \tag{7.28}$$

However, it fails to satisfy the classical conditions of compatibility because

$$F_1 - F_i \neq \text{a rank-one matrix}, \quad i = 2, \dots, 5. \tag{7.29}$$

As in the example of finely twinned martensite, here compatibility is achieved by mixing a fine distribution of "phases" near an interface. However, here the minimum is achieved in $W^{1,\infty}(\Omega, \mathbb{R}^3)$ by the fine phase mixture rather than merely approached by a minimizing sequence.

The example raises the question of what are the conditions of compatibility at an "interface". In this regard we note that both the example presented in this section and the one involving fine twins have the property that

$$D\tilde{y}(x_2) - D\tilde{y}(x_1) = \text{a rank-two matrix} \tag{7.30}$$

for almost every x_1 and x_2 on opposite sides of what one would think of as the interface. We also note that in the example of this section

$$\left(\tfrac{1}{4} \sum_{i=2}^{5} F_i\right) - F_1 = \text{a rank-one matrix}, \tag{7.31}$$

an analog of the equation

$$(\lambda F^+ + (1-\lambda) F^-) - \mathbf{1} = \boldsymbol{b} \otimes \boldsymbol{m} \tag{7.32}$$

found in Section 4.

The example also raises the question of what are the fewest number of deformation gradients of a function in $W^{1,\infty}(\Omega, \mathbb{R}^3)$ such that at least one of them does not differ from any of the others by a rank-one matrix. We conjecture that the answer is four and we are pursuing this and related questions in a further study of fine phase mixtures.

Acknowledgements. We thank J. W. CAHN for telling us about reference [48].

The research of J.M.B. was supported by an SERC Senior Fellowship and by the Institute of Mathematics and its Applications, University of Minnesota. The research of R.D.J. was supported by an SERC Visiting Fellowship and by the National Science Foundation under Grant No. MSM-8612420.

Note 1 added in proof. Professor J. W. CHRISTIAN has shown us an example of a cubic to orthorhombic transformation in titanium-tantalum alloys which exhibits an exact interface between cubic and orthorhombic phases (K. A. BYWATER & J. W. CHRISTIAN, Martensitic transformations in titanium-tantulum alloys, Phil. Mag. **25** (1972), p. 1249-1272). To get the exact interfaces, these authors adjusted the concentration of tantulum during preparation of the alloy in order to make one eigenvalue of the transformation strain equal to 1, while maintaining the condition that the other two eigenvalues are greater and less than 1.

Note 2 added in proof. By Proposition 4 the conditions in Part II of Theorem 1 that U_0 have an eigenvalue equal to 1 and that $\mu^* > 0$ are also sufficient that there be an exact austenite/martensite interface.

References

1. R. A. ADAMS, *Sobolev Spaces*. New York: Academic Press, 1975.
2. S. AGMON, A. DOUGLIS & L. NIRENBERG, Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions, II. Comm. Pure Appl. Math. **17** (1964), 35–92.
3. J. M. BALL, Convexity conditions and existence theorems in nonlinear elasticity. Arch. Rational Mech. Anal. **63** (1977), 337–403.
4. J. M. BALL, Strict convexity, strong ellipticity, and regularity in the calculus of variations. Math. Proc. Camb. Phil. Soc. **87** (1980), 501–513.
5. J. M. BALL, J. C. CURRIE & P. J. OLVER, Null Lagrangians, weak continuity and variational problems of arbitrary order. J. Functional Anal. **41** (1981), 135–174.
6. J. M. BALL, Discontinuous equilibrium solutions and cavitation in nonlinear elasticity. Phil. Trans. Roy. Soc. London A **306** (1982), 557–611.
7. J. M. BALL & J. E. MARSDEN, Quasiconvexity, positivity of the second variation, and elastic stability. Arch. Rational Mech. Anal. **86** (1984), 251–277.
8. J. M. BALL & F. MURAT, $W^{1,p}$-quasiconvexity and variational problems for multiple integrals. J. Funct. Anal. **58** (1984), 225–253.
9. J. M. BALL & G. KNOWLES, Liapunov functions for thermomechanics with spatially varying boundary temperatures. Arch. Rational Mech. Anal. **92** (1986), 193–204.

10. Z. S. BASINSKI & J. W. CHRISTIAN, Crystallography of deformation by twin boundary movements in indium-thallium alloys. Acta Met. **2** (1954), 101–116; also, Experiments on the martensitic transformation in single crystals of indium-thallium alloys. Acta Met. **2** (1954), 148–166.
11. J. S. BOWLES & J. K. MACKENZIE, The crystallography of martensitic transformations I and II. Acta Met. **2** (1954), 129–137, 138–147.
12. D. A. G. BRUGGEMAN, Berechnung verschiedener physikalischer Konstanten, von Heterogenen Substanzen. Ann. Phys. **5** (1935), 636–664.
13. M. W. BURKART & T. A. READ, Diffusionless phase change in the indium-thallium system. Trans. AIME J. Metals **197** (1953), 1516–1524.
14. M. CHIPOT & D. KINDERLEHRER, Equilibrium configurations of crystals, to appear.
15. J. W. CHRISTIAN, *The Theory of Transformations in Metals and Alloys*. Pergamon Press, 1975.
16. J. L. ERICKSEN, Some phase transitions in crystals. Arch. Rational Mech. Anal. **73** (1980), 99–124.
17. J. L. ERICKSEN, Constitutive theory for some constrained elastic crystals. IMA Preprint # 123, Institute for Mathematics and its Applications, University of Minnesota.
18. J. L. ERICKSEN, Some surface defects in unstressed thermoelastic solids. Arch. Rational Mech. Anal. **88** (1985), 337–345.
19. J. L. ERICKSEN, Some Constrained Elastic Crystals, in *Material Instabilities in Continuum Mechanics*, (ed. J. M. BALL). Oxford University Press, to appear.
20. I. FONSECA, Variational methods for elastic crystals. Arch. Rational Mech. Anal., **97** (1987), 189–220.
21. J. W. GIBBS, On the equilibrium of heterogeneous substances, in *The Scientific Papers of J. Willard Gibbs*, Vol. 1. Dover Publications, New York, 1961.
22. M. E. GURTIN, Two-phase deformations of elastic solids. Arch. Rational Mech. Anal. **84** (1983), 1–29.
23. R. D. JAMES, Finite deformation by mechanical twinning. Arch. Rational Mech. Anal. **77** (1981), 143–176.
24. R. D. JAMES, The stability and metastability of quartz, in *Metastability and Incompletely Posed Problems*, IMA, Vol. 3 (ed. S. ANTMAN, J. L. ERICKSEN, D. KINDERLEHRER & I. MÜLLER) Springer-Verlag, 1987, 147-176.
25. R. D. JAMES, Displacive phase transformations in solids. J. Mech. Phys. Solids **34** (1986), 359–394.
26. D. KINDERLEHRER, Remarks about equilibrium configurations of crystals, in *Material Instabilities in Continuum Mechanics*, (ed. J. M. BALL). Oxford University Press, to appear.
27. B. KLOSOWICZ & K. A. LURIE, On the optimal nonhomogeneity of a torsional elastic bar. Arch. of Mech. **24** (1971), 239–249.
28. R. V. KOHN & G. STRANG, Explicit relaxation of a variational problem in optimal design. Bull. Amer. Math. Soc. **9** (1983), 211–214.
29. N. A. LAVROV, K. A. LURIE & A. V. CHERKAEV, Nonuniform rod of extremal torsional stiffness. Mech. of Sol. **15** (1980), 74–80.
30. K. A. LURIE, A. V. CHERKAEV & A. V. FEDOROV, Regularization of optimal design problems for bars and plates I, II. J. Opt. Th. Appl. **37** (1982), 499–522 and 523–543; also, On the existence of solutions to some problems of optimal design for bars and plates. J. Opt. Th. Appl. **42** (1984), 247–281.
31. M. MARCUS & V. J. MIZEL, Absolute continuity on tracks and mappings of Sobolev spaces. Arch. Rational Mech. Anal. **45** (1972), 294–320.
32. G. W. MILTON, Modelling the properties of composites by laminates, in *Proc.*

Workshop on Homogenization and Effective Moduli of Materials and Media, 1984, to appear.
33. F. MURAT & L. TARTAR, Calcul des variations et homogénéisation, in *Ecole d'été d'homogénéisation*, Bureau sans Nappe, Eyrolles, Paris, July 1983, to appear.
34. N. NAKANISHI, Characteristics of stress-strain behaviour associated with thermoelastic martensitic transformation. Arch. Mech. **35** (1983), 37–62.
35. Z. NISHIYAMA, *Martensitic Transformation*. Academic Press, 1978.
36. W. NOLL, A general framework for problems in the statics of finite elasticity, in *Contemporary Developments in Continuum Mechanics and Partial Differential Equations* (ed. G. M. DE LA PENHA & L. A. MEDEIROS), North-Holland, 1978.
37. K. OTSUKA & K. SHIMIZU, Morphology and crystallography of thermoelastic γ' Cu–Al–Ni martensite. Japanese J. Appl. Phys. **8** (1969), 1196–1204.
38. A. C. PIPKIN, Some examples of crinkles, in *Homogenization and Effective Moduli of Materials and Media*. IMA, Vol. 1 (ed. J. L. ERICKSEN, D. KINDERLEHRER, R. KOHN & J. L. LIONS), Springer-Verlag, 1986.
39. U. E. RAITUM, The extension of extremal problems connected with a linear elliptic equation. Soviet Math. Dokl. **19** (1978), 1342–1345.
40. U. E. RAITUM, On optimal control problems for linear elliptic equations. Soviet Math. Dokl. **20** (1979), 129–132.
41. YU. G. RESHETNYAK, On the stability of conformal mappings in multidimensional spaces. Siberian Math. J. **8** (1967), 69–85.
42. YU. G. RESHETNYAK, Liouville's theorem on conformal mappings under minimal regularity assumptions. Siberian Math. J. **8** (1967), 631–653.
43. H. C. SIMPSON & S. J. SPECTOR, On failure of the complementing condition and nonuniqueness in linear elastostatics. J. Elasticity **15** (1985), 229–231.
44. H. C. SIMPSON & S. J. SPECTOR, On the sign of the second variation in finite elasticity, Arch. Rational Mech. Anal. **98** (1987), 1–30.
45. L. TARTAR, The compensated compactness method applied to systems of conservation laws, in *Systems of nonlinear partial differential equations* (ed. J. M. BALL). Reidel, 1983.
46. L. TARTAR, Étude des oscillations dans les équations aux dérivées partielles nonlinéaires, in *Lecture Notes in Physics 195*. Springer-Verlag, 1984, 384–412.
47. C. TRUESDELL, Some challenges offered to analysis by rational thermomechanics in *Contemporary Developments in Continuum Mechanics and Partial Differential Equations* (ed. G. M. DE LA PENHA & L. A. J. MEDEIROS), North-Holland, 1978.
48. G. VAN TENDELOO, J. VAN LANDUYT & S. AMELINCKX, The α-β phase transition in quartz and $AlPO_4$ as studied by electron microscopy and diffraction. Phys. Stat. Sol. a**33** (1976), 723–735.
49. C. M. WAYMAN, *Introduction to the Crystallography of Martensitic Transformations*. MacMillan, 1964.
50. M. S. WECHSLER, D. S. LIEBERMAN & T. A. READ, On the theory of the formation of martensite. Trans. AIME J. Metals, **197** (1953), 1503–1515.
51. H. WEYL, *The Classical Groups*. Princeton, 1946.
52. L. C. YOUNG, *Lectures on the Calculus of Variations and Optimal Control Theory*. Chelsea, 1980.

Heriot-Watt University
Edinburgh
University of Minnesota
Minneapolis

(Received April 20, 1987)