Noname manuscript No.
(will be inserted by the editor)

Incompatible sets of gradients and metastability

J.M. Ball · R.D. James

July 23, 2014

Abstract We give a mathematical analysis of a concept of metastability induced by incompatibility. The physical setting is a single parent phase, just about to undergo transformation to a product phase of lower energy density. Under certain conditions of incompatibility of the energy wells of this energy density, we show that the parent phase is metastable in a strong sense, namely it is a local minimizer of the free energy in an L^1 neighbourhood of its deformation. The reason behind this result is that, due to the incompatibility of the energy wells, a small nucleus of the product phase is necessarily accompanied by a stressed transition layer whose energetic cost exceeds the energy lowering capacity of the nucleus. We define and characterize incompatible sets of matrices, in terms of which the transition layer estimate at the heart of the proof of metastability is expressed. Finally we discuss connections with experiment and place this concept of metastability in the wider context of recent theoretical and experimental research on metastability and hysteresis.

1 Introduction

Materials that undergo first order phase transformations without diffusion typically exhibit hysteresis loops, that is, loops in a plot of a measured property vs. temperature as the temperature is cycled back and forth through the transformation temperature. It is the rule rather than the exception that the area within these loops does not tend to zero as the temperature is cycled more and more slowly. Thus, while there is an issue of the time-scale of such experiments, hysteresis is apparently not entirely due to viscosity or other thermally activated mechanisms. An alternative explanation is metastability,

Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, U.K. E-mail: ball@maths.ox.ac.uk · Department of Aerospace Engineering and Mechanics, University of Minnesota, Minneapolis, MN 55455, USA E-mail: james@umn.edu

as quantified by the presence of local minimizers in a continuum level elastic energy. This paper is a mathematical analysis of this possibility appropriate to cases in which the two phases are geometrically incompatible in a certain precise sense.

To illustrate our analysis in a simple case, let $\Omega \subset \mathbb{R}^n$ be a bounded domain with sufficiently smooth boundary $\partial\Omega$, and consider the energy functional

$$I(y) = \int_\Omega W(Dy(x))\,dx, \tag{1.1}$$

defined for mappings $y : \Omega \to \mathbb{R}^m$, where $Dy(x) = \left(\frac{\partial y_i}{\partial x_\alpha}(x)\right)$ denotes the gradient of y, so that $Dy(x)$ belongs to the set $M^{m\times n}$ of real $m \times n$ matrices for each x. Suppose that $W : M^{m\times n} \to \mathbb{R}$ is a continuous function satisfying $W(A) \geq C(1 + |A|^p)$ for constants $C > 0$, $p > 1$, and having exactly two local minimizers at matrices A_1, A_2 with $W(A_1) > W(A_2)$. Thus, imposing no boundary conditions on $\partial\Omega$, the global minimizers of I are given by affine mappings $y_{\min}(x) = a_2 + A_2x, a_2 \in \mathbb{R}^m$ having constant gradient A_2. Under suitable structural conditions on W, we prove that if A_1, A_2 are incompatible in the sense that $\operatorname{rank}(A_1 - A_2) > 1$, and if $W(A_1) - W(A_2)$ is sufficiently small, then $y^*(x) = a_1 + A_1x, a_1 \in \mathbb{R}^m$ is a local minimizer of I in L^1, i.e. there exists $\sigma > 0$ such that $I(y) \geq I(y^*)$ if $\|y - y^*\|_1 < \sigma$.

Notice that if $\|y - y^*\|_1 < \sigma$ then it can happen that $Dy(x)$ belongs to a small neighbourhood of A_2 on a set $E \subset \Omega$ of positive measure, so that $W(Dy(x)) < W(A_1)$ for $x \in E$. The basic idea underlying the analysis is that, if a nucleus E of the product phase of arbitrary form is introduced in this way so as to lower the energy, then, due to the incompatibility between the two phases, this nucleus is necessarily accompanied by a transition layer that interpolates between the nucleus and the parent phase A_1. This transition layer costs more energy than the lowering of energy due to the presence of the new phase. The analysis is delicate because the energy (1.1) contains no contribution from interfacial energy that would dominate at small scales. Thus, for example, scaling down of the nucleus and transition layer using geometric similarity preserves the ratio of transition layer and nucleus energies.

The above result is a special case of the considerably more general metastability theorem (Theorem 21) proved in this paper, in which the parent and product phases are represented by disjoint compact sets of matrices K_1 and K_2 respectively. Since the multiwell elastic energies we consider can exhibit nonattainment of the minimum of I, we formulate the problem more generally in terms of gradient Young measures, so that the metastability theorem applies to microstructures. We assume that K_1, K_2 are incompatible in the sense that if an L^∞ gradient Young measure $\nu = (\nu_x)_{x\in\Omega}$ is such that $\operatorname{supp}\nu_x \subset K_1 \cup K_2$ for a.e. $x \in \Omega$, then either $\operatorname{supp}\nu_x \subset K_1$ for a.e. $x \in \Omega$, or $\operatorname{supp}\nu_x \subset K_2$ for a.e. $x \in \Omega$. We can then estimate the energy of a transition layer that must be present if a gradient Young measure has nontrivial support near both K_1 and K_2. The delicate case is when the support of the gradient Young measure near either K_1 or K_2 is vanishingly small; to handle this, we find a way of moving and rescaling suitable convex subsets of Ω so as to get half of the support of the gradient Young measure in the subset near K_1, and half near K_2, which enables us to use a version of the Vitali covering lemma to obtain the desired estimate. This method of varying the volume fractions of a gradient Young measure has other applications and will be developed in a forthcoming paper [12]. Using the estimate for the energy of the transition layer, we show that a gradient Young measure supported on K_1 is a local minimizer w,ith respect to the L^1 norm of the difference between the underlying deformations, for energy densities that have a well at K_1 and a slightly lower well at K_2.

The shape of the domain Ω matters for our analysis. It is possible to defeat metastability as discussed here using the "rooms and passages" domain of Fraenkel [35], which consists of a bounded domain formed from an infinite sequence of rooms of vanishingly small diameters, each connected to the two adjacent rooms by passages of even smaller diameter. For such a domain the parent phase is not an L^1 local minimizer, because one can reduce the energy through deformations that are arbitrarily close in L^1, whose gradients lie entirely in the parent phase except for a nucleus of the product phase occupying a single room, together with transition layers in the two adjacent passages. To quantify the effect of domain shape on metastability we introduce a concept of a domain connected with respect to rigid-body motions of a convex set C (see Section 2), for which the method outlined in the previous paragraph can be applied, the constants in the transition layer estimate depending on C. This shape dependence is expected to have physical implications regarding the size of the hysteresis, for example in more conventional domains with sharply outward pointing corners. This phenomenon is therefore different from the well-known lowering of hysteresis that occurs in magnetism due to sharp *inward* pointing corners, and which is one explanation of the coercivity paradox.

In applications, K_2 usually grows with a parameter, either stress or temperature (Section 6). As discussed by C. Chu and the authors [10], one can derive upper bounds to the size of the hysteresis by considering test functions. The easiest upper bound is found when the stress, say, reaches a point where K_2 has grown sufficiently that there are matrices $A \in K_1$ and $B \in K_2$ such that $\operatorname{rank}(B-A) = 1$. This upper bound is directly related to the Schmid Law [62], though the conventional reasoning behind this law is completely different than the one offered here (see Section 6.1). In fact, for the problem of variant rearrangement discussed in [10] and Section 6.1 there is a more complicated test function that implies a loss of metastability earlier than the simple rank-one connection between A and B [10]. Curiously, these more complicated test functions require $\partial\Omega$ to have a sharp corner. A more careful analysis of Forclaz [34] seems to suggest that this is necessary.

Our differential constraint implying compatibility conditions is $\operatorname{curl} F = 0$, where F is a gradient. Our framework applies to other constraints in the theory of compensated compactness, except possibly that, in the case of compact sets K_1 and K_2, we use Zhang's lemma (see [77] and Lemma 1) to show that the definition of incompatibility is independent of the Sobolev exponent p. The interesting question of what are the incompatible sets for other important differential constraints seems not to have been explicitly investigated.

The first metastability result of the type given here is due to Kohn & Sternberg [47] who used Γ-convergence to prove under quasiconvexity assumptions the existence of local minimizers for (1.1) with gradient near A_1 (see also [46] for an improved version in particular showing that y^* is a local minimizer). Our work is also related to the important results of Grabovsky & Mengesha [36]. They prove, under assumptions of quasiconvexity, quasiconvexity at the boundary, and nonegativity of the second variation, all imposed locally at the gradients of a C^1 solution of the Euler-Lagrange equations, that this solution is an L^∞-local minimizer. Our approach differs from theirs in that we assume a multiwell structure of the energy, but make much weaker assumptions on the eventual local minimizer, which in our case is allowed to be a gradient Young measure. The idea behind the concept of metastability that we discuss here was first introduced without proof by C. Chu and the authors in [10,14].

The plan of the paper is as follows. In Section 2 we give some necessary technical background, especially concerning gradient Young measures, quasiconvexity and quasiconvexifications, and define and discuss C-connected domains. In Section 3 we define incompatible sets, and characterize them in terms of quasiconvexity, analyzing various examples. The fundamental transition layer estimate is proved in Section 4, and applied to prove metastability in Section 5. Finally, in Section 6 we give various applications of the metastability theorem. The first application is to the experiments of Chu & James on variant rearrangement in CuAlNi single crystals under biaxial dead loads, which originally motivated this paper. Then we discuss purely dilatational phase transformations, and the interesting case of Terephthalic acid. Finally, in Section 7 we give a perspective on metastability and hysteresis, discussing in particular other concepts of metastability [38,78,44,28,20,45,79] that have recently appeared in the literature, as well as experiments that show a dramatic dependence of the size of the hysteresis on conditions of compatibility [25,78,75,63]. These observations answer some questions and raise others.

2 Technical preliminaries

2.1 Gradient Young measures and quasiconvexity

Let $m \geq 1, n \geq 1$. We denote by $M^{m\times n}$ the set of real $m\times n$ matrices. Lebesgue measure in $\mathbb{R}^n$ is denoted by $\mathcal{L}^n$. Let $\Omega \subset \mathbb{R}^n$ be a bounded domain. Fix p with $1 \leq p \leq \infty$. We consider $\mathbb{R}^m-$valued distributions y in Ω whose gradients Dy belong to $L^p(\Omega; M^{m\times n})$. Without further hypotheses on Ω such distributions need not in general belong to $W^{1,p}(\Omega;\mathbb{R}^m)$, but it is proved in Maz'ya [51, p. 21] that they do so if Ω satisfies *the cone condition* with respect to a fixed cone $C^* = \{x \in \mathbb{R}^n : |x| \leq \rho,\ x\cdot e_1 \geq |x|\cos\alpha\}$, where $\rho > 0,\ 0 < \alpha < \frac{\pi}{2}$; that is, any point $x \in \Omega$ is the vertex of a cone congruent to C^* and contained in Ω, so that $x + QC^* \subset \Omega$ for some $Q \in SO(n)$.

Given a sequence $y^{(j)}$ such that $Dy^{(j)}$ is weakly convergent in $L^p(\Omega; M^{m\times n})$ (weak* if $p=\infty$) there exist (see, for example, [7]) a subsequence $y^{(\mu)}$ and a family of probability measures $(\nu_x)_{x\in\Omega}$ on $M^{m\times n}$, depending measurably on $x\in\Omega$, such that for any continuous function $f:\mathbb{R}^m\to\mathbb{R}$ and measurable $E\subset\Omega$

$$f(Dy^{(\mu)}) \rightharpoonup \langle \nu_x, f\rangle \text{ in } L^1(E)$$

whenever this weak limit exists. We call the family $\nu=(\nu_x)_{x\in\Omega}$ the L^p *gradient Young measure* generated by the sequence $Dy^{(\mu)}$ (alternative names in common use are $W^{1,p}$ gradient Young measure,or p-gradient Young measure). If $\nu_x=\nu$ is independent of x we say that the gradient Young measure is *homogeneous.* If $1<p\leq\infty$ then the weak relative compactness condition is equivalent to boundedness of $Dy^{(j)}$ in $L^p(\Omega; M^{m\times n})$, whereas if $p=1$ it is equivalent to equi-integrability of $Dy^{(j)}$. If $K\subset M^{m\times n}$ closed and $Dy^{(\mu)}\to K$ in measure, that is

$$\lim_{j\to\infty}\mathcal{L}^n(\{x\in\Omega : \operatorname{dist}(Dy^{(\mu)}(x),K)>\varepsilon\})=0 \text{ for all } \varepsilon>0,$$

then $\operatorname{supp}\nu_x\subset K$ for a.e. $x\in\Omega$.

Definition 1 *A continuous function $\varphi: M^{m\times n}\to\mathbb{R}$ is quasiconvex if*

$$⨍_G \varphi(A+D\theta(x))\,dx \geq \varphi(A) \tag{2.1}$$

for any bounded open set $G\subset\mathbb{R}^n$, all $A\in M^{m\times n}$ and any $\theta\in W_0^{1,\infty}(G;\mathbb{R}^m)$.

As is well known (see, for example, [26, p. 172]) this definition does not depend on G.

We recall the characterization of L^p gradient Young measures in terms of quasiconvexity due to Kinderlehrer & Pedregal. In the following statement we combine together various of their results.

Theorem 1 (Kinderlehrer & Pedregal [39, 40]) *Let $1\leq p\leq\infty$. A family $\nu=(\nu_x)_{x\in\Omega}$ of probability measures on $M^{m\times n}$, depending measurably on x, is an L^p gradient Young measure if and only if*
(i) $\bar\nu_x := \int_{M^{m\times n}} A\,d\nu_x(A) = Dy(x)$ for a.e. $x\in\Omega$ and some $\mathbb{R}^m-$valued distribution y with $Dy\in L^p(\Omega;M^{m\times n})$
(ii) for any quasiconvex $\varphi: M^{m\times n}\to\mathbb{R}$ satisfying $|\varphi(A)|\leq C(1+|A|^p)$ for all $A\in M^{m\times n}$, where $C>0$ is constant, (no growth condition required if $p=\infty$) we have

$$\langle\nu_x,\varphi\rangle := \int_{M^{m\times n}}\varphi(A)\,d\nu_x(A)\geq\varphi(\bar\nu_x) \text{ for a.e. } x\in\Omega,$$

(iii) if $1\leq p<\infty$ then $\int_\Omega\int_{M^{m\times n}}|A|^p\,d\nu_x(A)\,dx<\infty$; if $p=\infty$ then $\operatorname{supp}\nu_x\subset G$ *for some compact $G\subset M^{m\times n}$.*
Furthermore, if $1\leq p<\infty$ any L^p gradient Young measure $(\nu_x)_{x\in\Omega}$ is generated by some sequence of gradients $Dz^{(j)}$ (possibly different from the generating sequence $Dy^{(j)}$ in the definition) such that $|Dz^{(j)}|^p$ converges weakly in $L^1(\Omega)$ to some $g\in L^1(\Omega)$.

Remark 1 In [40] no assumption is stated concerning the bounded domain Ω, but the proof uses the Sobolev embedding theorem for Ω and thus implicitly makes some assumption. However, the proof can be easily modified by, in Lemma 5.1, writing Ω as a disjoint union of scaled copies of a cube, rather than of scaled copies of Ω. For an alternative approach to L^p gradient Young measures see Sychev [68].

We will make frequent use of the following version of Zhang's lemma that is a consequence of Müller [55, Corollary 3]. The original version is due to Zhang [77, Lemma 3.1].

Lemma 1 *Let $K \subset M^{m\times n}$ be compact, and suppose $\nu = (\nu_x)_{x\in\Omega}$ is an L^1 gradient Young measure with* $\operatorname{supp}\nu_x \subset K$ *for a.e. $x \in \Omega$. Then ν is an L^∞ gradient Young measure; that is it can be generated by a sequence $z^{(j)}$ whose gradients $Dz^{(j)}$ are bounded in $L^\infty(\Omega, M^{m\times n})$.*

2.2 Quasiconvex functions taking the value $+\infty$

Some care needs to be taken when defining quasiconvexity for functions which take the value $+\infty$. For example, as pointed out in [18, Example 3.5], the function φ defined by $\varphi(0) = \varphi(a \otimes b) = 0, \varphi(A) = +\infty$ otherwise, where $a \in \mathbb{R}^m, b \in \mathbb{R}^n$ are nonzero vectors, satisfies (2.1) for any bounded open set $G \subset \mathbb{R}^n$, all $A \in M^{m\times n}$ and any $\theta \in W_0^{1,\infty}(G;\mathbb{R}^m)$, even though φ is not rank-one convex and $I(y) = \int_\Omega \varphi(Dy)\,dx$ is not sequentially weak* lower semicontinuous in $W^{1,\infty}(\Omega;\mathbb{R}^m)$. See [9, p. 9] for further discussion, and another example related to Example 6. In this paper we will define quasiconvexity for functions which take the value $+\infty$ differently from in [18], as follows.

Definition 2 *A function $\varphi : M^{m\times n} \to \mathbb{R}\cup\{\infty\}$ is quasiconvex if there exists a nondecreasing sequence $\varphi^{(j)} : M^{m\times n} \to \mathbb{R}$ of continuous quasiconvex functions with*

$$\varphi(A) = \lim_{j\to\infty} \varphi^{(j)}(A) \quad \textit{for all} \;\; A \in M^{m\times n}.$$

Remark 2 Suppose $\varphi : M^{m\times n} \to \mathbb{R}$ is quasiconvex according to the above definition. Let G be a bounded domain, $A \in M^{m\times n}$, $\theta \in W_0^{1,\infty}(G;\mathbb{R}^m)$. Then for each j we have

$$\fint_G \varphi(A + D\theta(x))\,dx \geq \fint_G \varphi^{(j)}(A + D\theta(x))\,dx \geq \varphi^{(j)}(A),$$

so that passing to the limit $j \to \infty$ we deduce that (2.1) holds. It follows by the argument in Müller [56, Lemma 4.3] that φ is rank-one convex, and hence φ is continuous. Thus φ is quasiconvex in the sense of Definition 1.

Let $\varphi : M^{m\times n} \to \mathbb{R} \cup \{\infty\}$ be quasiconvex. Then φ is lower semicontinuous because it is the supremum of continuous functions. Let $(\nu_x)_{x\in\Omega}$ be an L^∞

gradient Young measure corresponding to a sequence $y^{(k)}$ with $Dy^{(k)} \overset{*}{\rightharpoonup} Dy$ in $L^\infty(\Omega;\mathbb{R}^m)$. For each j we have

$$\int_\Omega \varphi^{(j)}(Dy^{(k)})\,dx \leq \int_\Omega \varphi(Dy^{(k)})\,dx.$$

Since $\varphi^{(j)}$ is quasiconvex, letting $k \to \infty$ we obtain, using the lower semicontinuity of $\int_\Omega \varphi^{(j)}(Dz)\,dx$ with respect to weak* convergence in $W^{1,\infty}(\Omega;\mathbb{R}^m)$ (see, for example, [26, p369]),

$$\int_\Omega \varphi^{(j)}(Dy)\,dx \leq \int_\Omega \langle \nu_x, \varphi^{(j)}\rangle\,dx \leq \liminf_{k\to\infty} \int_\Omega \varphi(Dy^{(k)})\,dx.$$

(In order to apply the lower semicontinuity when we just have $Dy^{(k)} \overset{*}{\rightharpoonup} Dy$ in L^∞, we can, for example, write Ω as a disjoint union of cubes. In each cube we can fix $y^{(k)}$ to be zero at the centre of the cube, from which weak* convergence in $W^{1,\infty}$ follows. Thus we have the desired lower semicontinuity on each cube, from which that on Ω follows.) Letting $j \to \infty$, noting that $\varphi^{(j)}(Dy) \geq \varphi^{(1)}(Dy)$, and using monotone convergence, it follows that

$$-\infty < \int_\Omega \varphi(Dy)\,dx \leq \int_\Omega \langle \nu_x, \varphi\rangle\,dx \leq \liminf_{k\to\infty} \int_\Omega \varphi(Dy^{(k)})\,dx.$$

Thus the functional

$$I(y) = \int_\Omega \varphi(Dy)\,dx$$

is sequentially lower semicontinuous with respect to weak* convergence of the gradient in L^∞. Also, if $\nu_x = \nu$ is homogeneous then we obtain

$$\varphi(\bar{\nu}) \leq \langle \nu, \varphi\rangle.$$

Lemma 2 *Assume that $\varphi : M^{m\times n} \to \mathbb{R}\cup\{+\infty\}$ is such that $\operatorname{dom}\varphi = \{A \in M^{m\times n} : \varphi(A) < \infty\}$ is bounded. Then φ is quasiconvex if and only if φ is lower semicontinuous and $\langle \mu, \varphi\rangle \geq \varphi(\bar{\mu})$ for all homogeneous L^∞ gradient Young measures μ.*

Proof The necessity of the conditions has already been proved without the extra condition on φ. Conversely, suppose that φ is lower semicontinuous and that $\langle \mu, \varphi\rangle \geq \varphi(\bar{\mu})$ for all homogeneous gradient Young measures μ. Since $\operatorname{dom}\varphi$ is bounded and φ lower semicontinuous, φ is bounded below. Also, the lower semicontinuity implies (for example by [52, Theorem 3.8, p. 76]) that there is a nondecreasing sequence of continuous functions $\psi^{(j)}$ such that $\lim_{j\to\infty} \psi^{(j)}(A) = \varphi(A)$ for all $A \in M^{m\times n}$. Since $\operatorname{dom}\varphi$ is bounded we may also assume that $\psi^{(j)}(A) \geq C|A|^p - C_1$ for all $A \in M^{m\times n}$, where $C > 0$ and C_1 are constants and $p > 1$. Let $\varphi^{(j)} = (\psi^{(j)})^{\mathrm{qc}}$ be the quasiconvexification of $\psi^{(j)}$, that is the supremum of all continuous real-valued quasiconvex functions less than or equal to $\psi^{(j)}$. Then $\varphi^{(j)}$ is continuous and quasiconvex [26, p. 271], and it suffices to show that $\lim_{j\to\infty}\varphi^{(j)}(A)=\varphi(A)$ for all A. Suppose this is not the case, that there exists $A\in M^{m\times n}$ with $\varphi^{(j)}(A)\leq M<\infty$, where $M<\varphi(A)$. By the characterization [26, p. 271] of quasiconvexifications,

$$\varphi^{(j)}(A)=\inf_{\theta\in W_0^{1,\infty}(Q;\mathbb{R}^m)}\fint_Q \psi^{(j)}(A+D\theta)\,dx,$$

where $Q=(0,1)^n$. Hence there exist $\varepsilon>0$ and a sequence $\theta^{(j)}\in W_0^{1,\infty}(Q;\mathbb{R}^m)$ such that

$$\fint_Q \psi^{(j)}(A+D\theta^{(j)})\,dx\leq M+\varepsilon<\varphi(A).$$

Thus for any $j\geq k$ we have

$$\fint_Q \tilde{\psi}^{(k)}(A+D\theta^{(j)})\,dx\leq\fint_Q \psi^{(k)}(A+D\theta^{(j)})\,dx\leq M+\varepsilon,$$

where $\tilde{\psi}^{(k)}=\min(k,\psi^{(k)})$. From the growth condition on $\psi^{(j)}$, a subsequence (not relabelled) of $A+D\theta^{(j)}$ generates an L^p gradient Young measure $(\nu_x)_{x\in\Omega}$. Passing to the limit $j\to\infty$, noting that $\tilde{\psi}^{(k)}$ is bounded, we deduce that

$$\fint_Q\langle\nu_x,\tilde{\psi}^{(k)}\rangle\,dx\leq M+\varepsilon,$$

and then letting $k\to\infty$ we obtain by monotone convergence that

$$\fint_Q\langle\nu_x,\varphi\rangle\,dx\leq M+\varepsilon.$$

But then $\langle\mu,\varphi\rangle\leq M+\varepsilon$, where $\mu=\fint_Q\nu_x\,dx$, which by [40, Theorem 3.1] is a homogeneous L^p gradient Young measure with centre of mass $\bar{\mu}=A$. Since $\varphi(A)=\infty$ for $A\notin\operatorname{dom}\varphi$ we deduce that $\operatorname{supp}\mu\subset\overline{\operatorname{dom}\varphi}$. Since $\overline{\operatorname{dom}\varphi}$ is compact, it follows from Lemma 1 that μ is an L^∞ gradient Young measure. Hence by our assumption we have that $\varphi(A)\leq M+\varepsilon<\varphi(A)$, a contradiction.

Remark 3 Lemma 2 is a $p=\infty$ version of a result of Kristensen [48], who showed using a similar argument that if $\varphi:M^{m\times n}\to\mathbb{R}\cup\{+\infty\}$ satisfies the growth condition

$$\varphi(A)\geq C|A|^p-C_1 \text{ for all } A\in M^{m\times n}, \tag{2.2}$$

for some $C>0, C_1, p>1$, then φ is the supremum of a nondecreasing sequence of continuous quasiconvex functions $\varphi^{(j)}:M^{m\times n}\to\mathbb{R}$ satisfying $M\leq\varphi^{(j)}(A)\leq\alpha_j|A|^p+\beta_j$ for constants $\alpha_j>0,\beta_j,M$ (so that in particular φ is quasiconvex according to Definition 2) if and only if φ is lower semicontinuous and

$$\langle\mu,\varphi\rangle\geq\varphi(\bar{\mu}) \tag{2.3}$$

for any homogeneous L^p gradient Young measure μ (i.e. φ is closed $W^{1,p}$ quasiconvex in the sense of Pedregal [59]).

Note, however, that (2.3) is not in general a necessary condition for such φ to be quasiconvex, as can be seen by taking φ to be a finite quasiconvex function satisfying (2.2) that is not $W^{1,p}$ quasiconvex (see [18] with, for example, $m = n = 3, p = 2$).

Remark 4 The same proof shows that if $\varphi : M^{m\times n} \to \mathbb{R} \cup \{+\infty\}$ is a lower semicontinuous function with $\operatorname{dom}\varphi$ bounded, and if $F \in M^{m\times n}$, then there exists a nondecreasing sequence of continuous quasiconvex functions $\varphi^{(j)} : M^{m\times n} \to \mathbb{R}$ with $\varphi^{(j)}(F) \to \varphi(F)$ if and only if φ

$$\langle \mu, \varphi \rangle \geq \varphi(F) \tag{2.4}$$

for all homogeneous gradient Young measures μ with $\bar{\mu} = F$.

2.3 Quasiconvexification of sets

A closed set $E \subset M^{m\times n}$ is *quasiconvex* if $E = \varphi^{-1}(0)$ for some nonnegative finite quasiconvex function φ. Given $A \subset M^{m\times n}$ we can thus define the *quasiconvexification* A^{qc} of A by

$$A^{\mathrm{qc}} = \bigcap \{E \supset A : E \text{ quasiconvex}\}.$$

We recall the following equivalent characterizations of K^{qc} for compact $K \subset M^{m\times n}$:

Proposition 2 *If $K \subset M^{m\times n}$ is compact then*

$$\begin{aligned} K^{\mathrm{qc}} &= \{\bar{\nu} : \nu \text{ a homogeneous gradient Young measure with } \operatorname{supp}\nu \subset K\} \\ &= \{A \in M^{m\times n} : \varphi(A) \leq \max_{B\in K} \varphi(B) \text{ for all finite quasiconvex } \varphi\} \\ &= (\mathrm{dist}_{\mathrm{K}}^{\mathrm{qc}})^{-1}(0), \end{aligned}$$

where $\mathrm{dist}_{\mathrm{K}}$ *is the distance function to the set K.*

Proof The equality of the three sets in the proposition is proved in [56, Theorem 4.10, p. 54]. Since $(\mathrm{dist}_{\mathrm{K}}^{\mathrm{qc}})^{-1}(0)$ is quasiconvex and $\mathrm{dist}_{\mathrm{K}}^{\mathrm{qc}}(\mathrm{A}) = 0$ for all $A \in K$, we have that $K^{\mathrm{qc}} \subset (\mathrm{dist}_{\mathrm{K}}^{\mathrm{qc}})^{-1}(0)$. But if $\varphi(A) \leq \max_{B\in K} \varphi(B)$ for all finite quasiconvex φ then A belongs to any quasiconvex set $E \supset K$. Hence $K^{\mathrm{qc}} \subset \{A \in M^{m\times n} : \varphi(A) \leq \max_{B\in K} \varphi(B) \text{ for all finite quasiconvex } \varphi\} \subset K^{\mathrm{qc}}$, so that all three sets in the proposition equal K^{qc}.

Theorem 3 *Let $K_1, \ldots, K_N$ be compact subsets of $M^{m\times n}$ whose quasiconvexifications K_r^{qc} are disjoint. Let $\nu = (\nu_x)_{x\in\Omega}$ be an L^∞ gradient Young measure such that* $\operatorname{supp}\nu_x \subset \bigcup_{r=1}^N K_r^{\mathrm{qc}}$ *for a.e. $x \in \Omega$. Then there is an L^∞ gradient Young measure $\nu^* = (\nu_x^*)_{x\in\Omega}$ such that* $\operatorname{supp}\nu_x^* \subset \bigcup_{r=1}^N K_r$*, $\bar{\nu}_x^* = \bar{\nu}_x$ and $\nu_x^*(K_r) = \nu_x(K_r^{\mathrm{qc}}), r = 1, \ldots, N$, for a.e. $x \in \Omega$. If ν is homogeneous then ν^* can be chosen to be homogeneous.*

In order to prove Theorem 3 we will need two technical lemmas. Let $\mathcal{P}(M^{m\times n})$ denote the set of probability measures on $M^{m\times n}$. Given a compact set $K \subset M^{m\times n}$ we denote by $\mathrm{GYM}(K)$ the set of homogeneous (L^∞) gradient Young measures μ with $\operatorname{supp}\mu \subset K$.

Lemma 3 *Let $K \subset M^{m\times n}$ be compact. For $A \in K^{\mathrm{qc}}$ define*

$$F(A) = \{\mu \in \mathrm{GYM}(K) : \bar{\mu} = A\}.$$

Then $F(A)$ is a nonempty, sequentially weak closed subset of $\mathcal{P}(M^{m\times n})$.*

Proof Let $\mu_j \in F(A)$ with $\mu_j \overset{*}{\rightharpoonup} \mu$ (that is $\langle \mu_j, f\rangle \to \langle \mu, f\rangle$ for all $f \in C_0(M^{m\times n})$, where $C_0(M^{m\times n})$ denotes the space of all continuous functions $f : M^{m\times n} \to \mathbb{R}$ such that $\lim_{|A|\to\infty} f(A) = 0$). If $\psi \in C_0(M^{m\times n})$ with $\psi = 0$ on K, then $\langle \mu, \psi\rangle = \lim_{j\to\infty}\langle \mu_j, \psi\rangle = 0$, and so $\operatorname{supp}\mu \subset K$. Then, choosing $f \in C_0(M^{m\times n})$ with $f = 1$ on K, and noting that $\langle \mu_j, f\rangle = 1$ we have that $\langle \mu, f\rangle = \mu(K) = 1$, and so $\mu \in \mathcal{P}(M^{m\times n})$. Let $h \in C_0(M^{m\times n})$ with $h(B) = B$ for all $B \in K$. Then $A = \bar{\mu}_j = \langle \mu_j, h\rangle$, so that $\lim_{j\to\infty}\langle \mu_j, h\rangle = \langle \mu, h\rangle = \bar{\mu} = A$. If g is finite and quasiconvex, we have by Theorem 1 that $\langle \mu_j, g\rangle \geq g(A)$ for all j, so that passing to the limit (using $\operatorname{supp}\mu_j \subset K$) we obtain $\langle \mu, g\rangle \geq g(A)$, so that, again using Theorem 1, we have $\mu \in \mathrm{GYM}(K)$ as required.

Lemma 4 *There is a Borel measurable map $A \mapsto \mu_A$ from K^{qc} to the set $\mathcal{P}(K)$ of probability measures on K endowed with the weak* topology, such that $\mu_A \in F(A)$ for all $A \in K^{\mathrm{qc}}$.*

Proof By Parthasarathy [58, Theorems 6.3 6.4, 6.5 pp. 44-46] $\mathcal{P}(K)$ endowed with the weak* topology is a Polish space, i.e. separable and completely metrizable. We first claim that the multivalued map $F : K^{\mathrm{qc}} \to \mathcal{P}(K)$ is upper semicontinuous, i.e. for every closed $E \subset \mathcal{P}(K)$ the set $\{A \in K^{qc} : F(A) \cap E \neq \emptyset\}$ is closed in $M^{m\times n}$. Indeed if $A_j \in K^{qc}$ with $\mu_{A_j} \in F(A_j) \cap E$ and $A_j \to A$ then we may assume that $\mu_{A_j} \overset{*}{\rightharpoonup} \mu$ (since μ_{A_j} is bounded in the dual space of $C_0(M^{m\times n})$, namely the space of measures). By a similar argument to that of the proof of Lemma 3 we deduce that $\mu \in F(A) \cap E$ as required.

We now apply the measurable selection theorem of Kuratowski & Ryll-Nardzewski [49], which in the statement by Wagner [70, Theorem 4.1] implies that a Borel measurable selection $\mu_A \in F(A)$ exists whenever $F(A)$ is closed for all $A \in K^{\mathrm{qc}}$ and $A \mapsto F(A)$ is weakly measurable. In our case weak measurability means that $\{A \in K^{qc} : F(A) \cap U \neq \emptyset\}$ is Borel measurable, and it is shown in [70, Theorem 4.2] that this is implied by upper semicontinuity, giving the required result since $F(A)$ is closed by Lemma 3.

Proof of Theorem 3 Let $\mathcal{K} = \cup_{r=1}^N K_r$. We apply Lemma 4 to each compact set K_r, and denote the corresponding Borel measurable selection μ^r_A, so that for each $r = 1, \ldots, N$ and $A \in K_r^{\mathrm{qc}}$ we have $\mu^r_A \in \mathrm{GYM}(K_r)$ with $\bar{\mu}^r_A = A$. We then define the required gradient Young measure $\nu^* = (\nu^*_x)_{x\in\Omega}$ by the action of ν_x on functions $f \in C(\mathcal{K})$ through the formula

$$\langle \nu^*_x, f\rangle = \sum_{r=1}^N \langle \nu_x, \langle \mu^r_A, f\rangle\rangle, \tag{2.5}$$

that is

$$\langle \nu_x^*, f\rangle = \sum_{r=1}^{N} \int_{K_r^{\mathrm{qc}}} \int_{K_r} f(B)\, d\mu_A^r(B)\, d\nu_x(A). \tag{2.6}$$

(Note that $\langle \nu_x^*, f\rangle$ is well defined because we can extend f outside the compact set $\mathcal{K}$ to a function $f \in C_0(M^{m\times n})$ and $\operatorname{supp}\mu_A^r \subset K_r$.) Since $\langle \nu_x^*, f\rangle \geq 0$ for $f \geq 0$, ν_x^* is a positive measure. Choosing $f = 1$ we see that $\int_{M^{m\times n}} d\nu_x(A) = \int_{M^{m\times n}} d\nu_x^*(A) = 1$, so that $\nu_x^* \in \mathcal{P}(\mathcal{K})$. Similarly, choosing $f(A) = A$ we deduce that $\bar{\nu}_x^* = \sum_{r=1}^{N} \int_{K_r^{\mathrm{qc}}} A\, d\nu_x(A) = \bar{\nu}_x$. In particular $\bar{\nu}_x^* = Dy(x)$ for some $Dy \in L^\infty(\Omega; M^{m\times n})$. If φ is finite and quasiconvex, then

$$\begin{aligned}
\langle \nu_x^*, \varphi\rangle &\geq \sum_{r=1}^{N} \int_{K_r^{\mathrm{qc}}} \varphi(\bar{\mu}_A^r)\, d\nu_x(A) \\
&= \sum_{r=1}^{N} \int_{K_r^{\mathrm{qc}}} \varphi(A)\, d\nu_x(A) \\
&= \int_{M^{m\times n}} \varphi(A)\, d\nu_x(A) \geq \varphi(\bar{\nu}_x),
\end{aligned}$$

where we have used the necessity of condition (ii) of Theorem 1 twice. By construction $\operatorname{supp}\nu_x^* \subset \mathcal{K}$. Hence, by the sufficiency part of Theorem 1, ν^* is an L^∞ gradient Young measure, which is homogeneous if ν is homogeneous. Finally, choosing f to be the characteristic function of K_s we see that $\nu_x^*(K_s) = \nu_x(K_s^{\mathrm{qc}})$ as required.

2.4 Domains connected with respect to rigid motion of a convex set

Let $n > 1$. We recall that two subsets E, F of $\mathbb{R}^n$ are *directly congruent* if

$$E = \xi + QF \quad \text{for some } \xi \in \mathbb{R}^n, Q \in \mathrm{SO}(n). \tag{2.7}$$

Let $\Omega \subset \mathbb{R}^n$ be a bounded domain, and let $C \subset \mathbb{R}^n$ be bounded, open and convex. We suppose without loss of generality that $0 \in C$; this implies in particular that $\lambda C \subset C$ for any $\lambda \in [0, 1]$.

We define the *outer radius* $R(C)$ by

$$R(C) = \inf\{\rho > 0 : B(a, \rho) \supset C \text{ for some } a \in \mathbb{R}^n\}, \tag{2.8}$$

the *inner radius* $r(C)$ by

$$r(C) = \sup\{\rho > 0 : B(a, \rho) \subset C \text{ for some } a \in \mathbb{R}^n\}, \tag{2.9}$$

and the *eccentricity* $E(C)$ by

$$E(C) = \sqrt{1 - \frac{r(C)^2}{R(C)^2}}. \tag{2.10}$$

Note that there exists a unique minimal ball $B(a(C), R(C))$ containing C, but that there may be infinitely many maximal balls $B(b(C), r(C))$ contained in C.

Definition 3 *Ω is C-filled if any $x \in \Omega$ belongs to a subset of Ω that is directly congruent to C.*

Thus Ω is C-filled if and only if

$$\Omega = \bigcup \{E \subset \Omega : E \text{ directly congruent to } C\}. \tag{2.11}$$

Definition 4 *Let C_1, C_2 be subsets of Ω directly congruent to C. We say that C_1, C_2 are congruently connected, written $C_1 \sim C_2$, if C_1 can be moved continuously to C_2 as a rigid body while remaining in Ω, i.e. there exist continuous maps $\xi : [0,1] \to \mathbb{R}^n$, $Q : [0,1] \to SO(n)$, such that $C_1 = \xi(0) + Q(0)C$, $C_2 = \xi(1) + Q(1)C$ and $\xi(t) + Q(t)C \subset \Omega$ for all $t \in [0,1]$.*

Clearly $\sim$ is an equivalence relation on the family $\mathcal{K}(C)$ of subsets of Ω that are directly congruent to C.

Definition 5 *Ω is C-connected if there is an equivalence class of $\mathcal{K}(C)$ with respect to $\sim$ that covers Ω. Ω is strongly C-connected if it is C-filled and every pair of subsets of Ω directly congruent to C are congruently connected.*

Thus Ω is C-connected if Ω is covered by a collection of directly congruent copies of C any pair of which can be moved from one to the other as a rigid body while remaining in Ω, while Ω is strongly C-connected if in addition there is a single equivalence class with respect to $\sim$. Example 2 below shows that C-connectedness does not imply strong C-connectedness.

Proposition 4 *Let $0 < \lambda \leq 1$, and let Ω be convex. Then the subsets of Ω of the form $a + \lambda\Omega$, $a \in \mathbb{R}^n$, cover Ω and are pairwise congruently connected. In particular Ω is strongly $\lambda\Omega$-connected.*

Proof Let $x \in \Omega$. Since Ω is convex, $\lambda(\Omega - x) \subset \Omega - x$, and hence $x \in (1-\lambda)x + \lambda\Omega \subset \Omega$. Thus the subsets of Ω of the form $a + \lambda\Omega$ cover Ω.

If $a_1 + \lambda\Omega$ and $a_2 + \lambda\Omega$ are two such subsets then $t \mapsto (1-t)a_1 + ta_2 + \lambda\Omega$, $t \in [0,1]$, defines a suitable continuous path of directly congruent subsets of Ω joining them.

If Ω is C-connected then obviously Ω is C-filled. The following example shows that if Ω is C-filled then it need not be C-connected.

Example 1 For $0 < \alpha < 1$ define $\Omega^\alpha \subset \mathbb{R}^n$ by

$$\Omega^\alpha = B(0,1) \cup B((2-\alpha)e_1, 1).$$

Then Ω^α is $B(0,1)$-filled but is only $B(0,r)$-connected for $0 < r \leq r_\alpha = \sqrt{\alpha - \frac{1}{4}\alpha^2}$, since the diameter of the opening joining the two balls comprising Ω^α is $2r_\alpha$.

Proposition 5 *If Ω is C-filled, it is λC-connected for all sufficiently small $\lambda > 0$.*

To prove Proposition 5 we need the following definition and lemma.

Definition 6 *If $\delta > 0$, a δ-tube joining $x_1, x_2 \in \Omega$ is a continuous path $\xi : [0,1] \to \Omega$ with $\xi(0) = x_1, \xi(1) = x_2$ such that $\xi(t) + \overline{B(0,\delta)} \subset \Omega$ for all $t \in [0,1]$.*

Lemma 6 *Let Ω be a bounded domain and let $\varepsilon > 0$ be sufficiently small. Then there exists $\delta = \delta(\varepsilon) > 0$ such that any pair of points $x_1, x_2 \in \Omega$ with $\operatorname{dist}(x_i, \partial\Omega) \geq \varepsilon$ are joined by a δ-tube.*

Proof Fix $\bar{x} \in \Omega$ with $\operatorname{dist}(\bar{x}, \partial\Omega) \geq \varepsilon$. For $\delta > 0$ let $E_\delta = \{x \in \Omega : \text{there exists a } \delta\text{-tube joining } \bar{x} \text{ and } x\}$. We claim that $E_\delta \supset \{x \in \Omega : \operatorname{dist}(x, \partial\Omega) \geq \varepsilon\}$ for δ sufficiently small. If not there would exist $x^{(j)} \in \Omega$ with $\operatorname{dist}(x^{(j)}, \partial\Omega) \geq \varepsilon$ such that there is no $\frac{1}{j}$-tube joining $\bar{x}$ to $x^{(j)}, j = 1, 2, \ldots$. But we may assume that $x^{(j)} \to x$ with $\operatorname{dist}(x, \partial\Omega) \geq \varepsilon$. Since Ω is connected there is a δ-tube joining $\bar{x}$ to x for some $\delta > 0$, so that this path followed by the straight line from x to $x^{(j)}$ defines a $\frac{1}{j}$-tube for large j, a contradiction. Hence for δ sufficiently small any points $x_1, x_2 \in \Omega$ with $\operatorname{dist}(x_i, \partial\Omega) \geq \varepsilon$ are joined to $\bar{x}$, and hence to each other, by a δ-tube.

Proof of Proposition 5 Let $\varepsilon > 0$ be such that $B(0, \varepsilon) \subset C$, and let $\delta = \delta(\varepsilon)$ be as in Lemma 6. Pick $\lambda > 0$ sufficiently small so that $\lambda C \subset B(0, \delta)$.

Let $\mathcal{E}_\lambda(C) = \{b + \lambda QC : b \in \mathbb{R}^n, Q \in SO(n), b + \lambda QC \subset a + QC \subset \Omega$ for some $a \in \mathbb{R}^n\}$. Since Ω is C-filled, $\mathcal{K}(C)$ covers Ω, and by Proposition 4 applied to $a + QC$, so does $\mathcal{E}_\lambda(C)$.

Suppose that $b_i + \lambda Q_i C \in \mathcal{E}_\lambda(C),\ i = 1, 2$. Then by Proposition 4, $b_i + \lambda Q_i C$ is congruently connected to $a_i + \lambda Q_i C$, where $a_i + Q_i C \subset \Omega,\ i = 1, 2$. But $a_i + \lambda Q_i C \subset B(a_i, \delta)$ and $\operatorname{dist}(a_i, \partial\Omega) \geq \operatorname{dist}(a_i, \partial(a_i + Q_i C)) \geq \varepsilon$. Hence by Lemma 6 there exists a δ-tube $\xi : [0,1] \to \Omega$ joining a_1 and a_2. Let $Q : [0,1] \to SO(n)$ be continuous with $Q(0) = Q_1, Q(1) = Q_2$. Then $\xi(t) + \lambda Q(t) C \subset \Omega$ for all $t \in [0,1]$, and so $a_1 + \lambda Q_1 C$, $a_2 + \lambda Q_2 C$ are congruently connected. Hence $b_1 + \lambda Q_1 C$, $b_2 + \lambda Q_2 C$ are congruently connected. Hence Ω is λC-connected.

The following example shows that Proposition 5 does not hold for strong C-connectedness. That is, a bounded domain may be C-filled but not strongly λC-connected for all sufficiently small $\lambda > 0$.

Example 2 Let $C \subset \mathbb{R}^2$ be the interior of the equilateral triangle of side 1 with vertices at $(0,0)$, $(\frac{\sqrt{3}}{2}, \pm\frac{1}{2})$. Let Ω consist of a large ball $B(0,R)$ from which the origin $(0,0)$ and the points $A_i = (\frac{2^{-i}}{\sqrt{3}}, \frac{2^{-i}}{3}),\ B_i = (\frac{2^{-i}}{\sqrt{3}}, -\frac{2^{-i}}{3}),\ i = 0, 1, 2, ..,$ are removed. The points A_i, B_i lie on the half-lines L_A and L_B defined by $\{\sqrt{3}x_2 - x_1 = 0, x_1 \geq 0\}$ and $\{\sqrt{3}x_2 + x_1 = 0, x_1 \geq 0\}$ respectively, which meet at the origin at an angle of $60°$. Then Ω is C-filled. Indeed Ω consists of C together with points lying outside C which are clearly inside congruent copies of C lying in Ω (for example, for the points on L_A, L_B we can use an equilateral triangle of side 1 which lies outside C except for a small region near one of its vertices).

Now consider the open equilateral triangle Δ of side 1 with vertices at $(\frac{2}{\sqrt{3}}, 0)$ and $(\frac{1}{2\sqrt{3}}, \pm\frac{1}{2})$, and the corresponding scaled equilateral triangles $\Delta_i = 2^{-i}\Delta$ of side 2^{-i}. Note that $\Delta_i \subset \Omega$, and that the edges of Δ_i intersect L_A and L_B in the points A_i, A_{i+1} and B_i, B_{i+1} respectively. We claim that Δ_i cannot be continuously moved to a position far from the origin while remaining in Ω. This is even true for a slightly smaller equilateral triangle contained in Δ_i. A rigorous proof can be constructed by noting that the width of Δ_i, that is the minimal distance between parallel lines that enclose Δ_i, is $2^{-(i+1)}\sqrt{3}$, which is greater than any of the distances of the openings through which it would have to pass, namely $|A_iA_{i+1}| = |B_iB_{i+1}| = \frac{2^{-i}}{3}$ and $|A_iB_i| = \frac{2^{1-i}}{3}$ (see Strang [65]). Hence Ω is not strongly λC-connected for sufficiently small $\lambda > 0$.

Proposition 7 *The bounded domain Ω is C-connected for some bounded open convex C if and only if Ω satisfies the cone condition with respect to some cone C^*.*

Proof Let Ω satisfy the cone condition with respect to C^*. If $x \in \Omega$ with $x + QC^* \subset \Omega$ then $x \in x + Q(C^* - \varepsilon e_1) \subset \Omega$ for $\varepsilon > 0$ sufficiently small. Hence Ω is $(\operatorname{int} C^*)$-filled, and hence, by Proposition 5, $\lambda(\operatorname{int} C^*)$-connected for sufficiently small $\lambda > 0$.

Conversely, let Ω be C-connected for some C. Since C is convex it is Lipschitz (see Morrey[54, p. 72]) and hence satisfies the cone condition with respect to some C^*. Since Ω is C-filled it follows immediately that Ω also satisfies the cone condition with respect to C^*.

Despite this result, the concept of C-connectedness is of interest since we will show that the constants in the transition layer estimate of Theorem 13 can be chosen to depend on Ω through C.

2.5 The Vitali Covering Lemma

The following simpler version [64] of the Vitali covering lemma is used in an important way in the transition layer estimate.

Lemma 8 *Let E be a measurable subset of $\mathbb{R}^n$ which is covered by the union of a family of balls $\{B_i\}$ of bounded diameter. From this family we can select a countable or finite disjoint subsequence $B_{i(k)}, k = 1, 2, \dots$ such that*

$$\sum_k \mathcal{L}^n(B_{i(k)}) \geq c_n \mathcal{L}^n(E).$$

Here, $c_n > 0$ depends only on the dimension n. The choice $c_n = 5^{-n}$ suffices.

3 Incompatible sets

Let $\Omega \subset \mathbb{R}^n$ be a bounded domain. Fix p with $1 \leq p \leq \infty$.

Definition 7 *The closed subsets $K_1, \ldots, K_N$ of $M^{m\times n}$ are L^p incompatible if they are disjoint, and if whenever $\nu = (\nu_x)_{x\in\Omega}$ is an L^p gradient Young measure satisfying*

$$\operatorname{supp}\nu_x \subset \bigcup_{r=1}^{N} K_r \quad \textit{for a.e. } x \in \Omega,$$

then for some $i,\ 1 \le i \le N$,

$$\operatorname{supp}\nu_x \subset K_i \quad \textit{for a.e. } x \in \Omega.$$

Remark 5 a. It is easily seen that the sets $K_1, \ldots, K_N$ are L^p incompatible if and only if for each $i = 1, \ldots, N$ the pair of sets K_i, $\bigcup_{r\neq i} K_r$ are L^p incompatible. The latter condition is obviously necessary, and it is sufficient since if $\operatorname{supp}\nu_x \subset \bigcup_{r=1}^{N} K_r$ for a.e. $x \in \Omega$ then we have for each i either

$$\operatorname{supp}\nu_x \subset K_i \quad \text{for a.e. } x \in \Omega$$

or

$$\operatorname{supp}\nu_x \subset \bigcup_{r\neq i} K_r,$$

and $\bigcap_{i=1}^{N} \bigcup_{r\neq i} K_r$ is empty. For this reason we can often restrict attention to the case $N = 2$.

b. The definition does not depend on Ω. By the above remark we may assume that $N = 2$. So let K_1, K_2 be L^p incompatible with respect to Ω and let $\tilde{\Omega} \subset \mathbb{R}^n$ be another bounded domain. Let $D\tilde{y}^{(j)}$ be a sequence of gradients that is relatively weakly compact in $L^p(\tilde{\Omega}; M^{m\times n})$ with corresponding gradient Young measure $(\tilde{\nu}_x)_{x\in\tilde{\Omega}}$ satisfying $\operatorname{supp}\tilde{\nu}_x \subset K_1 \cup K_2$ for a.e. $x \in \tilde{\Omega}$. Let $E_1 = \{x \in \tilde{\Omega} : \operatorname{supp}\tilde{\nu}_x \cap K_1 \neq \emptyset\}$, $E_2 = \{x \in \tilde{\Omega} : \operatorname{supp}\tilde{\nu}_x \cap K_2 \neq \emptyset\}$ and suppose for contradiction that $\mathcal{L}^n(E_1) > 0, \mathcal{L}^n(E_2) > 0$. By hypothesis we have that

$$\mathcal{L}^n(\tilde{\Omega}\backslash(E_1 \cup E_2)) = 0. \tag{3.1}$$

Let x_1, x_2 be Lebesgue points of E_1, E_2 respectively. Since $\tilde{\Omega}$ is connected there is a continuous arc $x(t)$, $t \in [0,1]$, with $x(0) = x_1$, $x(1) = x_2$ and $x(t) \in \tilde{\Omega}$ for all $t \in [0,1]$. Then there exists $\varepsilon_1 > 0$ such that $x(t) + \varepsilon\Omega \subset \tilde{\Omega}$ for all $t \in [0,1]$, $0 < \varepsilon \le \varepsilon_1$. Fix $0 < \varepsilon \le \varepsilon_1$ sufficiently small so that $\mathcal{L}^n((x_1 + \varepsilon\Omega) \cap E_1) > 0$ and $\mathcal{L}^n((x_2 + \varepsilon\Omega) \cap E_2) > 0$, which is possible since x_1, x_2 are Lebesgue points. Define for $i = 1, 2$

$$f_i(t) = \frac{\mathcal{L}^n((x(t) + \varepsilon\Omega) \cap E_i)}{\varepsilon^n \mathcal{L}^n(\Omega)}.$$

Then each f_i is continuous in t, and by construction $f_1(0) > 0, f_2(1) > 0$. But from (3.1)

$$f_1(t) + f_2(t) \ge 1,$$

from which it follows easily that there exists $t_0 \in [0,1]$ with $0 < f_i(t_0) \leq 1$ for $i = 1, 2$, i.e.

$$\mathcal{L}^n((x(t_0) + \varepsilon\Omega) \cap E_1) > 0, \;\; \mathcal{L}^n((x(t_0) + \varepsilon\Omega) \cap E_2) > 0. \tag{3.2}$$

Now let $y^{(j)}(x) = \varepsilon^{-1}\tilde{y}^{(j)}(x(t_0) + \varepsilon x)$, which is well defined because $\tilde{y}^{(j)} \in L^1_{\mathrm{loc}}(\tilde{\Omega}; \mathbb{R}^m)$. Then $Dy^{(j)}(x) = D\tilde{y}^{(j)}(x(t_0) + \varepsilon x)$ and so $Dy^{(j)}$ is relatively weakly compact in $L^p(\Omega; M^{m\times n})$ and has Young measure

$$\nu_x = \tilde{\nu}_{x(t_0)+\varepsilon x}, \;\; x \in \Omega. \tag{3.3}$$

Furthermore $\operatorname{supp}\nu_x \subset K_1 \cup K_2$ for a.e. $x \in \Omega$, and so either $\operatorname{supp}\nu_x \subset K_1$ for a.e. $x \in \Omega$ or $\operatorname{supp}\nu_x \subset K_2$ for a.e. $x \in \Omega$. This implies that $\operatorname{supp}\tilde{\nu}_x \subset K_1$ for a.e. $x \in x(t_0) + \varepsilon\Omega$ or $\operatorname{supp}\tilde{\nu}_x \subset K_2$ for a.e. $x \in x(t_0) + \varepsilon\Omega$, contradicting (3.2).

c. If the sets $K_1, \ldots, K_N$ are compact then the definition is independent of p. Consequently in this case we say simply that $K_1, \ldots, K_n$ are *incompatible.* In fact suppose that $K_1, \ldots, K_N$ are compact and L^∞ incompatible. Let $1 \leq p < \infty$ and let $Dy^{(j)}$ be weakly relatively compact in L^p and have Young measure $(\nu_x)_{x\in\Omega}$ with $\operatorname{supp}\nu_x \subset \bigcup_{r=1}^N K_r$ for a.e. $x \in \Omega$. Then by Lemma 1 there is a sequence of gradients $Dz^{(j)}$ which is bounded in L^∞ and has the same Young measure, so that $K_1, \ldots, K_n$ are L^p incompatible.

d. *The case* $p = 1$. An alternative definition of L^1 incompatible sets would have been to replace the weak relative compactness of $Dy^{(j)}$ by boundedness of $Dy^{(j)}$ in $L^1(\Omega; M^{m\times n})$. But with such a modification no family of disjoint closed subsets of $M^{m\times n}$ would be L^1 incompatible. In fact if K_1, K_2 were a pair of L^1 incompatible sets in this sense, we could let $\Omega = [-1,1]^n, A \in K_1, \; B \in K_2$, and define

$$y^{(j)}(x) = \begin{cases} Ax & \text{if } x_1 \leq 0, \\ jx_1 Bx + (1 - jx_1)Ax & \text{if } 0 < x_1 < \frac{1}{j}, \\ Bx & \text{if } x_1 \geq \frac{1}{j}. \end{cases}$$

Then

$$Dy^{(j)}(x) = jx_1 B + (1 - jx_1)A + j(B - A)x \otimes e_1$$

for $0 < x_1 < \frac{1}{j}$, so that

$$\int_{[-1,1]^n} \mid Dy^{(j)} \mid dx \leq C < \infty.$$

But the corresponding Young measure $(\nu_x)_{x\in\Omega}$ is given by

$$\nu_x = \begin{cases} \delta_A \text{ if } x_1 < 0, \\ \delta_B \text{ if } x_1 > 0. \end{cases}$$

Definition 8 *The closed subsets $K_1, \ldots, K_n$ of $M^{m\times n}$ are homogeneously L^p incompatible if they are disjoint, and if whenever ν is a homogeneous L^p gradient Young measure generated by a sequence satisfying*

$$\operatorname{supp}\nu \subset \bigcup_{r=1}^{N} K_r,$$

then for some i, $1 \leq i \leq N$,

$$\operatorname{supp}\nu \subset K_i.$$

The same arguments as in Remark 5 show that this definition too is independent of Ω and, in the case when the K_r are compact, also of p with $1 \leq p \leq \infty$ (in which case we say that the K_r are *homogeneously incompatible*).

Definition 9 *The closed subsets $K_1, \ldots, K_N$ of $M^{m\times n}$ are L^p gradient incompatible if they are disjoint, and if whenever $Dy \in L^p(\Omega; M^{m\times n})$ with*

$$Dy(x) \in \bigcup_{r=1}^{N} K_r \quad \text{for a.e. } x \in \Omega$$

then

$$Dy(x) \in K_i \quad \text{for a.e. } x \in \Omega$$

for some i.

Again the definition is independent of Ω and, in the case when the K_r are compact, also of p with $1 \leq p \leq \infty$ (in which case we say that the K_r are *gradient incompatible*).

Note that if $n = 1$ or $m = 1$ then no pair of disjoint nonempty closed sets K_1, K_2 can be homogeneously L^p incompatible, since if $A_1 \in K_1$, $A_2 \in K_2$ then $\operatorname{rank}(A_1 - A_2) = 1$, so that $\frac{1}{2}(\delta_{A_1} + \delta_{A_2})$ is a homogeneous L^∞ gradient Young measure supported nontrivially on $K_1 \cup K_2$; similarly K_1 and K_2 are not L^∞ gradient incompatible. Thus most of the results of this paper are only relevant for $n \geq 2$ and $m \geq 2$.

Of course if $K_1, \ldots, K_N$ are L^p incompatible they are also L^p gradient incompatible. However the converse is false (for other examples see Examples 5, 6).

Example 3 Let $m = n = 2$, $\{e_1, e_2\}$ be an orthonormal basis of $\mathbb{R}^2$, $K_1 = \{\mathbf{1}\}, K_2 = \{\mathbf{0}, 2e_2 \otimes e_2\}$. Then K_1, K_2 are not incompatible. To see this note that $\mathbf{1} = e_1 \otimes e_1 + e_2 \otimes e_2$, so that

$$\begin{aligned} \mathbf{1} - e_2 \otimes e_2 &= e_1 \otimes e_1 \\ e_2 \otimes e_2 &= \frac{1}{2}(\mathbf{0} + 2e_2 \otimes e_2). \end{aligned} \tag{3.4}$$

Thus a double laminate can be constructed having homogeneous gradient Young measure

$$\nu = \frac{1}{2}\delta_{\mathbf{1}} + \frac{1}{4}\delta_{\mathbf{0}} + \frac{1}{4}\delta_{2e_2\otimes e_2}.$$

However, K_1, K_2 are gradient incompatible. In fact if $Dy(x) \in K_1 \cup K_2$ a.e. in $\Omega = (0,1)^2$, we have

$$Dy(x) = \lambda(x)\mathbf{1} + 2\mu(x)e_2 \otimes e_2,$$

where $\lambda(x)\mu(x) = 0$, $\lambda(x) \in \{0,1\}$ and $\mu(x) \in \{0,1\}$ a.e.. Hence $y_{,1} = \lambda e_1, y_{,2} = (\lambda + 2\mu)e_2$ and so

$$\lambda_{,2} = (\lambda + 2\mu)_{,1} = 0$$

in the sense of distributions. Hence $\lambda = \lambda(x_1), \lambda + 2\mu = f(x_2)$ from which it follows easily that either $\lambda = 0$ a.e. or $\lambda = 1$ a.e. as required.

3.1 Characterization of incompatible sets

Clearly if $K_1, \dots, K_N$ are L^p incompatible they are homogeneously L^p incompatible. We do not know if the converse holds, even if the K_r are compact (but see Remark 6 for the case $m = n = 2$). It is possible to characterize homogeneously incompatible sets in terms of quasiconvex functions. We first prove some preliminary results relating incompatibility of the sets K_r to that of the sets K_r^{qc}.

Lemma 5 *If $K_1, \dots, K_N$ are homogeneously L^∞ incompatible then $(\bigcup_{r=1}^N K_r)^{qc}$ is the disjoint union of the sets K_r^{qc}.*

Proof We first show that $K_r^{\mathrm{qc}} \bigcap K_s^{\mathrm{qc}}$ is empty if $r \neq s$. Suppose the contrary, that there exists an $A \in K_r^{\mathrm{qc}} \bigcap K_s^{\mathrm{qc}}$. Then there exist homogeneous L^∞ Young measures ν^r and ν^s with $\operatorname{supp} \nu^r \subset K_r$, $\operatorname{supp} \nu^s \subset K_s$ and $\bar{\nu}^r = \bar{\nu}^s = A$. But the set of homogeneous L^∞ Young measures with a given centre of mass A is convex (Kinderlehrer & Pedregal [39]), and thus $\nu = \frac{1}{2}(\nu^r + \nu^s)$ is a homogeneous L^∞ Young measure with $\operatorname{supp} \nu \subset K_r \cup K_s$ and both $\operatorname{supp} \nu \cap K_r$ and $\operatorname{supp} \nu \cap K_s$ nonempty. Thus $K_1, \dots, K_N$ are not homogeneously L^∞ incompatible.

Next, let $A \in (\bigcup_{r=1}^n K_r)^{\mathrm{qc}}$. Then $A = \bar{\nu}$ for some homogeneous L^∞ Young measure ν with $\operatorname{supp} \nu \subset \bigcup_{r=1}^n K_r$, and by hypothesis $\operatorname{supp} \nu \subset K_i$ for some i. Hence $A \in K_i^{\mathrm{qc}}$, completing the proof.

Proposition 9 *The compact sets $K_1, \dots, K_N$ are incompatible (resp. homogeneously incompatible) if and only if $K_1^{qc}, \dots, K_N^{qc}$ are incompatible (resp. homogeneously incompatible).*

Proof Suppose that $K_1, \dots, K_N$ are incompatible. By Lemma 5 the K_r^{qc} are disjoint. Let $\nu = (\nu_x)_{x \in \Omega}$ be an L^∞ gradient Young measure with $\operatorname{supp} \nu_x \subset \bigcup_{r=1}^n K_r^{\mathrm{qc}}$ for a.e. $x \in \Omega$. Then by Theorem 3 there is an L^∞ gradient Young measure $\nu^* = (\nu_x^*)_{x \in \Omega}$ with $\operatorname{supp} \nu_x^* \subset \bigcup_{r=1}^\infty K_r$ and $\nu_x^*(K_r) = \nu_x(K_r^{\mathrm{qc}})$ for all r and a.e. $x \in \Omega$. Since the K_r are incompatible, we have that $\nu_x^*(K_i) = 1$ for some i and a.e. $x \in \Omega$. Hence $\nu_x(K_i^{\mathrm{qc}}) = 1$ for a.e. $x \in \Omega$ and thus $K_1^{\mathrm{qc}}, \dots, K_N^{\mathrm{qc}}$ are incompatible. The same argument shows that if the K_r are homogeneously incompatible then so are the K_r^{qc}. The converse direction is obvious.

Theorem 10 *The compact sets $K_1,\dots,K_N$ are homogeneously incompatible if and only if*
(i) the sets K_r^{qc}, $r=1,\dots,N$, are disjoint,
(ii) for each $i=1,\dots,N$ the function $\varphi_i : M^{m\times n} \longrightarrow [0,\infty]$ defined by

$$\varphi_i(A) = \begin{cases} 1 & \text{if } A \in K_i^{qc}, \\ 0 & \text{if } A \in \bigcup_{r\neq i} K_r^{qc}, \\ +\infty & \text{otherwise}, \end{cases}$$

is quasiconvex.

Proof Let $K_1,\dots,K_N$ be homogeneously incompatible. Then (i) holds by Lemma 5. To prove (ii), by Lemma 2 with $K=\bigcup_{r=1}^N K_r$ it suffices to show that

$$\langle \mu, \varphi_i\rangle \ge \varphi_i(\bar\mu) \tag{3.5}$$

for any homogeneous L^∞ gradient Young measure μ. Since (3.5) obviously holds if $\langle\mu,\varphi_i\rangle=\infty$, we may assume that $\operatorname{supp}\nu\subset\bigcup_{r=1}^N K_r^{\rm qc}$. Then it follows from Proposition 9 that $\operatorname{supp}\mu\subset K_j^{\rm qc}$ for some j, so that also $\bar\mu\in K_j^{\rm qc}$. Thus if $j\neq i$ both sides of (3.5) are zero, while if $j=i$ then both sides are one.

Conversely, suppose that (i) and (ii) hold, and let ν be a homogeneous L^∞ gradient Young measure with supp $\nu\subset\bigcup_{r=1}^N K_r$. Then $\nu=\sum_{r=1}^N\lambda_r\nu^r$, where $\lambda_r\ge 0, \sum_{r=1}^N\lambda_r=1$ and ν^r is a probability measure with supp $\nu^r\subset K_r$. For any k we have (since φ_k is quasiconvex)

$$\varphi_k(\bar\nu)\le\langle\nu,\varphi_k\rangle=\lambda_k.$$

In particular $\varphi_k(\bar\nu)<\infty$ and so $\bar\nu\in K_i^{\rm qc}$ for some i. Choosing $k=i$ we obtain $\lambda_i\ge 1$ and so $\nu=\nu^i$ and $\operatorname{supp}\nu\subset K_i$. Hence $K_1,\dots,K_N$ are homogeneously incompatible.

Theorem 11 *The compact sets $K_1,...,K_N$ are incompatible if and only if*
(i) The sets $K_1^{qc},...,K_N^{qc}$ are gradient incompatible,
(ii) for each $i=1,\dots,N$ the function $\varphi_i : M^{m\times n} \longrightarrow [0,\infty]$ defined by

$$\varphi_i(A) = \begin{cases} 1 & \text{if } A \in K_i^{qc}, \\ 0 & \text{if } A \in \bigcup_{r\neq i} K_r^{qc}, \\ +\infty & \text{otherwise}, \end{cases}$$

is quasiconvex.

Proof Let $K=\bigcup_{r=1}^N K_r$. Suppose that $K_1,\dots,K_N$ are incompatible. Then $K_1,\dots,K_N$ are homogeneously incompatible, so that by Lemma 5 and Theorem 10 the sets $K_r^{\rm qc}$ are disjoint, $K^{\rm qc}=\bigcup_{r=1}^N K_r^{\rm qc}$ and (ii) holds. To show that the $K_r^{\rm qc}$ are gradient incompatible, suppose that $Dy\in L^\infty(\Omega;M^{m\times n})$ satisfies $Dy(x)\in K^{\rm qc}$ a.e.. It follows from Theorem 3 applied to the gradient Young measure $\nu=(\delta_{Dy(x)})_{x\in\Omega}$ that there exists a gradient Young measure $(\nu_x^*)_{x\in\Omega}$ with $\operatorname{supp}\nu_x^*\subset K$ and $\bar\nu_x^*=Dy(x)$ a.e.. But then by hypothesis $\operatorname{supp}\nu_x^*\subset K_s$ a.e. for some s and so $Dy(x)\in K_s^{\rm qc}$ a.e..

Conversely, let (i) and (ii) hold, and let $(\nu_x)_{x\in\Omega}$ be an L^∞ gradient Young measure with $\operatorname{supp}\nu_x \subset \bigcup_{r=1}^N K_r$ a.e.. Then, for a.e. $x\in\Omega$, ν_x is a homogeneous L^∞ gradient Young measure, and so by Theorem 10 $\operatorname{supp}\nu_x \subset K_{r(x)}$ for some $r(x)$, and hence $\bar\nu_x \in K^{\rm qc}_{r(x)}$. Thus $Dy(x)=\bar\nu_x\in\bigcup_{r=1}^N K_r^{\rm qc}$ a.e., and so $Dy(x)\in K_s^{\rm qc}$ a.e. for some s. Since the $K_r^{\rm qc}$ are disjoint, $r(x)=s$ a.e. and hence $\operatorname{supp}\nu_x\subset K_s$ a.e..

Corollary 12 *The compact sets $K_1,\dots,K_N$ are incompatible if and only if $K_1,\dots,K_N$ are homogeneously incompatible and $K_1^{qc},\dots,K_N^{qc}$ are gradient incompatible.*

Proof This follows immediately from Theorems 10, 11.

Remark 6 When $m=n=2$, Kirchheim & Székelyhidi [43], using results from Faraco & Székelyhidi [32], show that two disjoint compact sets K_1,K_2 are incompatible if and only if $(K_1\cup K_2)^{\rm rc}$ is the disjoint union of $K_1^{\rm rc}$ and $K_2^{\rm rc}$, where $K^{\rm rc}$ denotes the rank-one convex envelope of a compact set K. They also show that K_1,K_2 are incompatible if and only if they are homogeneously incompatible. For compact sets $K_1,K_2\subset M^{2\times2}$ that are left invariant under $SO(2)$ and consist of matrices with positive determinant, necessary and sufficient conditions for K_1,K_2 to be incompatible are given by Heinz [37].

3.2 Examples

A necessary condition that $K_1,\dots,K_N$ be homogeneously L^∞ incompatible is that there are no rank-one connections between any of the K_r. This follows from Lemma 5 and the fact that quasiconvex sets are rank-one convex. However the absence of such rank-one connections is not sufficient (see the well-known Example 6 below).

Example 4 (Two matrices) If $K_1=\{A\},K_2=\{B\}$, where $A,B\in M^{m\times n}$ with $\operatorname{rank}(A-B)>1$, then K_1,K_2 are L^p incompatible for any $p>1$. We give two proofs of this fact.
First proof Let $(\nu_x)_{x\in\Omega}$ be an L^p gradient Young measure with $\operatorname{supp}\nu_x\subset\{A,B\}$ for a.e. $x\in\Omega$, i.e. $\nu_x=\lambda(x)\delta_A+(1-\lambda(x))\delta_B$ where $0\le\lambda(x)\le1$. In particular $\operatorname{supp}\nu_x$ is contained in a bounded set for a.e. x, and so $(\nu_x)_{x\in\Omega}$ is an L^∞ gradient Young measure by Lemma 1. Thus by the results in [13], based on the weak continuity of minors, $\nu_x=\delta_A$ for a.e. $x\in\Omega$ or $\nu_x=\delta_B$ for a.e. $x\in\Omega$ as required.
Second proof This is due to V. Šverák [66]. Without loss of generality we suppose that $A=0$ and define $h(E)=(\operatorname{dist}(E,L))^2$ for $E\in M^{m\times n}$, where $L=\{tB;t\in\mathbb{R}\}$. Thus

$$h(E)=|E|^2-\frac{(E\cdot B)^2}{|B|^2}.$$

h is quadratic and strongly elliptic, since tB is not rank-one for any t. If $Dy^{(j)}$ is bounded in $L^p(\Omega; M^{m\times n})$ with $\operatorname{supp}\nu_x \subset \{A, B\}$ then $Dh(Dy^{(j)}) \to 0$ in measure, and hence $Dh(Dy^{(j)}) \to 0$ strongly in $L^s(\Omega; M^{m\times n})$ if $1 < s < p$. So

$$\text{div}\ \ Dh(Dy^{(j)}) = \text{div}\ \ f^{(j)},\ \ x \in \Omega,$$

where $f^{(j)} \to 0$ strongly in $L^s(\Omega; M^{m\times n})$. By elliptic regularity theory this implies that $Dy^{(j)}$ is relatively compact in $L^s_{\text{loc}}(\Omega; M^{m\times n})$, so that $\nu_x = \delta_{Dy(x)}$ a.e. for some y with $Dy(x) \in \{A, B\}$ a.e.. But elliptic regularity implies that Dy is smooth, so that $\nu_x = \delta_A$ a.e. or $\nu_x = \delta_B$ a.e., as required.

Example 5 (3 matrices) Let $K_1 = \{A_1\}, K_2 = \{A_2\}, K_3 = \{A_3\}$, where $A_r \in M^{m\times n}$ with $\text{rank}\,(A_r - A_s) > 1$ for $r \neq s$. Then K_1, K_2, K_3 are incompatible. This is a consequence of a deep result of Šverák [66,67], which uses in particular the result of Zhang [76] that K_1, K_2, K_3 are gradient incompatible.

Example 6 (4 matrices) Let $K_r = \{A_r\}, 1 \le r \le 4$, with $\text{rank}\,(A_r - A_s) > 1$ for $r \neq s$. Then $K_1, \ldots, K_4$ are not in general incompatible. This follows from the construction of [19] that was motivated by the example of [2] and Tartar [69]. Chlebík & Kirchheim [22] showed that $K_1, \ldots, K_4$ are nevertheless gradient incompatible.

Example 7 (5 matrices) Let $K_r = \{A_r\}, 1 \le r \le 5$, with $\text{rank}\,(A_r - A_s) > 1$ for $r \neq s$. Then $K_1, \ldots, K_5$ are not in general gradient incompatible (Kirchheim & Preiss [41,42]).

Example 8 (Incompatible energy wells in $M^{2\times 2}$) Let $K_r = SO(2)U_r, 1 \le r \le N$, where $U_r = U_r^T > 0$ and there are no rank-one connections between the different K_r. Then $K_1, \ldots, K_N$ are incompatible. This follows from the result of Firoozye [19,33] and Šverák [67].

Example 9 (Incompatible energy wells in $M^{3\times 3}$) Let $K_1 = SO(3)U_1, K_2 = SO(3)U_2$, where $U_1 = U_1^T > O$, $U_2 = U_2^T > O$, and $\text{rank}\,(A_1 - A_2) > 1$ for all $A_1 \in K_1, A_2 \in K_2$. Then it is not known whether in general K_1, K_2 are incompatible. However under stronger conditions on U_1, U_2 incompatibility is proved by Dolzmann, Kirchheim, Müller & Šverák [30] (see also Matos [50] and Kohn, Lods & Haraux [46]). If K_1, K_2, K_3 are three such energy wells without rank-one connections then it is shown in [19] that K_1, K_2, K_3 need not be incompatible, using Example 6.

4 The transition layer estimate

In this section we suppose that $K_1, \ldots, K_N$ are disjoint compact subsets of $M^{m\times n}$. Given $y \in W^{1,p}(\Omega; \mathbb{R}^m)$ and $\varepsilon > 0$ we consider for $r = 1, \ldots, N$ the sets

$$\Omega_{r,\varepsilon}(y) := \{x \in \Omega : Dy(x) \in N_\varepsilon(K_r)\},$$

where

$$N_\varepsilon(K) := \{A \in M^{m\times n} : \mathrm{dist}\,(A, K) \le \varepsilon\},$$

and the corresponding 'transition layer'

$$T_\varepsilon(y) := \{x \in \Omega : Dy(x) \notin \bigcup_{r=1}^{N} N_\varepsilon(K_r)\}.$$

The main result is

Theorem 13 *Let $1 < p < \infty$ and let Ω be C-connected. Then $K_1, \ldots, K_N$ are incompatible if and only if there exist constants $\varepsilon_0 > 0$ and $\gamma > 0$ such that if $0 \le \varepsilon < \varepsilon_0$ and $y \in W^{1,p}(\Omega;\mathbb{R}^m)$ then*

$$\int_{T_\varepsilon(y)} (1 + |Dy|^p)\,dx \ge \gamma \max_{1\le r\le N} \min(\mathcal{L}^n(\Omega_{r,\varepsilon}(y)), \mathcal{L}^n(\bigcup_{s\neq r} \Omega_{s,\varepsilon}(y))). \quad (4.1)$$

The constant ε_0 can be chosen to depend only on the eccentricity $E(C)$, the sets $K_1, \ldots, K_N$ and p, while the constant γ can be chosen to depend only on these quantities and $\mathcal{L}^n(C)/\mathcal{L}^n(\Omega)$.

Remark 7 An alternative way of writing the right-hand side of (4.1) is

$$\gamma \min(\mathcal{L}^n(\Omega_{\bar r,\varepsilon}(y)), \sum_{r\neq \bar r} \mathcal{L}^n(\Omega_{r,\varepsilon}(y))),$$

where $\bar r = \bar r(\varepsilon, y)$ is such that

$$\mathcal{L}^n(\Omega_{\bar r,\varepsilon}(y)) = \max_{1\le r\le N} \mathcal{L}^n(\Omega_{r,\varepsilon}(y)).$$

To see this, fix ε and y and let $a_r = \mathcal{L}^n(\Omega_{r,\varepsilon}(y))$. Suppose without loss of generality that $a_N \ge a_{N-1} \ge \ldots \ge a_1$ and let $c = \sum_{r=1}^N a_r$. Then we have to show that

$$\max_{1\le r\le N} \min(a_r, c - a_r) = \min(a_N, c - a_N).$$

But this follows from the fact that $a_r \le c - a_N$ if $1 \le r < N$.

We state the case $N = 2$ of Theorem 13 separately.

Theorem 14 *Let $1 < p < \infty$ and let Ω be C-connected. Two disjoint compact sets K_1, K_2 are incompatible if and only if there exist constants $\varepsilon_0 > 0$ and $\gamma > 0$ such that if $0 \le \varepsilon < \varepsilon_0$ and $y \in W^{1,p}(\Omega;\mathbb{R}^m)$ then*

$$\int_{T_\varepsilon(y)} (1 + |Dy|^p)\,dx \ge \gamma \min(\mathcal{L}^n(\Omega_{1,\varepsilon}(y)), \mathcal{L}^n(\Omega_{2,\varepsilon}(y))). \quad (4.2)$$

The constant ε_0 can be chosen to depend only on $E(C), K_1, K_2$ and p, while the constant γ can be chosen to depend only on these quantities and $\mathcal{L}^n(C)/\mathcal{L}^n(\Omega)$.

Note that Theorem 13 follows from Theorem 14 by applying it to the pair of sets K_r and $\bigcup_{s\neq r} K_s$ for each r, remarking that the set $T_\varepsilon(y)$ is the same for each r, and applying Remark 5a. It therefore suffices to prove Theorem 14. We use the following lemma.

Lemma 15 *Let $0 \le E < 1$, and let K_1, K_2 be incompatible. Then there exist constants $\varepsilon_0 = \varepsilon_0(E, K_1, K_2, p) > 0$ and $\gamma_0 = \gamma_0(E, K_1, K_2, p) > 0$ such that if $\tilde{C} \subset \mathbb{R}^n$ is any bounded open convex set with $E(\tilde{C}) \le E$ and if $0 \le \varepsilon < \varepsilon_0$, $y \in W^{1,p}(\tilde{C}; \mathbb{R}^m)$, with for some $i = 1, 2$*

$$\frac{3}{4}\mathcal{L}^n(\tilde{C}) \ge \mathcal{L}^n(\{x \in \tilde{C} : Dy(x) \in N_\varepsilon(K_i)\}) \ge \frac{1}{4}\mathcal{L}^n(\tilde{C}),$$

then

$$\int_{T_{\varepsilon,\tilde{C}}(y)} (1 + |Dy|^p)\, dx \ge \gamma_0 \mathcal{L}^n(\tilde{C}),$$

where $T_{\varepsilon,\tilde{C}}(y) := \{x \in \tilde{C} : Dy(x) \notin N_\varepsilon(K_1) \cup N_\varepsilon(K_2)\}$.

Proof Suppose not. Then for $j = 1, 2, \ldots$ there exist $\varepsilon^{(j)} \le 1/j$, bounded open convex sets $C^{(j)} \subset \mathbb{R}^n$ with $E(C^{(j)}) \le E$ and mappings $y^{(j)} \in W^{1,p}(C^{(j)}; \mathbb{R}^m)$ with for some $i = 1, 2$ (independent of j)

$$\frac{3}{4}\mathcal{L}^n(C^{(j)}) \ge \mathcal{L}^n(\{x \in C^{(j)} : Dy^{(j)}(x) \in N_{\varepsilon^{(j)}}(K_i)\}) \ge \frac{1}{4}\mathcal{L}^n(C^{(j)}), \quad (4.3)$$

and

$$\int_{T_{\varepsilon^{(j)},C^{(j)}}(y^{(j)})} (1 + |Dy^{(j)}|^p)\, dx \le \frac{1}{j}\mathcal{L}^n(C^{(j)}).$$

For definiteness we suppose (4.3) holds for $i = 1$. Let $B(a^{(j)}, R_j)$ be the unique minimal ball containing $C^{(j)}$, so that $R_j = R(C^{(j)})$. We normalize $C^{(j)}$ by setting

$$\tilde{C}^{(j)} = \frac{1}{R_j}(C^{(j)} - a^{(j)}). \quad (4.4)$$

Thus $R(\tilde{C}^{(j)}) = 1$ and $B(0, 1)$ is the unique minimal ball containing $\tilde{C}^{(j)}$. Define $z^{(j)} \in W^{1,p}(\tilde{C}^{(j)}; \mathbb{R}^m)$ by

$$z^{(j)}(x) = \frac{1}{R_j} y^{(j)}(a^{(j)} + R_j x). \quad (4.5)$$

Then

$$Dz^{(j)}(x) = Dy^{(j)}(a^{(j)} + R_j x) \quad (4.6)$$

and we have that

$$\frac{3}{4}\mathcal{L}^n(\tilde{C}^{(j)}) \ge \mathcal{L}^n(\{x \in \tilde{C}^{(j)} : Dz^{(j)}(x) \in N_{\varepsilon^{(j)}}(K_1)\}) \ge \frac{1}{4}\mathcal{L}^n(\tilde{C}^{(j)}) \quad (4.7)$$

and

$$\int_{T_j} (1 + |Dz^{(j)}|^p)\, dx \leq \frac{1}{j} \mathcal{L}^n(\tilde{C}^{(j)}), \tag{4.8}$$

where $T_j := \{x \in \tilde{C}^{(j)} : Dz^{(j)}(x) \notin N_{\varepsilon^{(j)}}(K_1) \cup N_{\varepsilon^{(j)}}(K_2)\}$. Since the closures $D^{(j)}$ of $\tilde{C}^{(j)}$ lie in $\overline{B(0,1)}$, a subsequence (which we do not relabel) of the $D^{(j)}$ converges in the Hausdorff metric to a closed convex set $D \subset \overline{B(0,1)}$. Since $E(\tilde{C}^{(j)}) = E(C^{(j)}) \leq E$, there is a closed ball B_j contained in $D^{(j)}$ with radius at least $\sqrt{1-E^2}$. We can suppose that these balls also converge to a ball $B \subset D$ of radius at least $\sqrt{1-E^2} > 0$, and hence D has nonempty interior $\tilde{C}$. Note that $\mathcal{L}^n(\tilde{C}^{(j)}) \to \mathcal{L}^n(\tilde{C})$. Let G be open and convex with $\bar{G} \subset \tilde{C}$ and $\mathcal{L}^n(\tilde{C} \backslash G) < \frac{1}{8}\mathcal{L}^n(\tilde{C})$. Then for sufficiently large j we have $G \subset \tilde{C}^{(j)}$. Hence, for sufficiently large j,

$$\mathcal{L}^n(\tilde{C}^{(j)}) < \frac{8}{7}\mathcal{L}^n(G). \tag{4.9}$$

Also

$$\begin{aligned}
&\mathcal{L}^n(\{x \in \tilde{C}^{(j)} : Dz^{(j)}(x) \in N_{\varepsilon^{(j)}}(K_1)\}) \\
&\qquad \geq \mathcal{L}^n(\{x \in G : Dz^{(j)}(x) \in N_{\varepsilon^{(j)}}(K_1)\}) \\
&\qquad \geq \mathcal{L}^n(\{x \in \tilde{C}^{(j)} : Dz^{(j)}(x) \in N_{\varepsilon^{(j)}}(K_1)\}) - \mathcal{L}^n(\tilde{C}^{(j)} \backslash G) \\
&\qquad \geq \frac{1}{4}\mathcal{L}^n(\tilde{C}^{(j)}) - \mathcal{L}^n(\tilde{C}^{(j)} \backslash G) \\
&\qquad \geq \frac{1}{8}\mathcal{L}^n(G).
\end{aligned} \tag{4.10}$$

Hence, combining (4.9), (4.10) and the left-hand inequality in (4.7), we have

$$\frac{6}{7}\mathcal{L}^n(G) \geq \mathcal{L}^n(\{x \in G : Dz^{(j)}(x) \in N_{\varepsilon^{(j)}}(K_1)\}) \geq \frac{1}{8}\mathcal{L}^n(G). \tag{4.11}$$

Since K_1, K_2 are bounded, it follows in particular from (4.8) that $Dz^{(j)}$ is bounded in $L^p(G; M^{m\times n})$, and so we may assume that $Dz^{(j)}$ generates a Young measure $(\nu_x)_{x\in G}$. Let U_1, U_2 be open neighbourhoods of K_1, K_2 respectively. Since $\{x \in G : Dz^{(j)}(x) \notin U_1 \cup U_2\} \subset T_j$ for sufficiently large j, and $\mathcal{L}^n(T_j) \to 0$, we have that $Dz^{(j)}(x) \to K_1 \cup K_2$ in measure, and hence $\operatorname{supp}\nu_x \subset K_1 \cup K_2$ for a.e. $x \in G$. Since K_1, K_2 are incompatible we thus have either $\operatorname{supp}\nu_x \subset K_1$ a.e. or $\operatorname{supp}\nu_x \subset K_2$ a.e. in G. Now let $\varphi_i : M^{m\times n} \to [0,1]$, $i = 1,2$, be continuous functions such that $\varphi_i = 1$ on $N_{\delta/2}(K_i)$, $\varphi_i = 0$ outside $N_\delta(K_i)$, where $\delta > 0$ is sufficiently small so that $N_\delta(K_1) \cap N_\delta(K_2)$ is empty. Then from (4.11) we have that

$$\int_G \varphi_1(Dz^{(j)})\, dx \geq \frac{1}{8}\mathcal{L}^n(G) \tag{4.12}$$

for all sufficiently large j. Since for sufficiently large j

$$\begin{aligned}
&\mathcal{L}^n(\{x \in G : Dz^{(j)}(x) \in N_{\varepsilon^{(j)}}(K_2)\}) \\
&\qquad \geq \mathcal{L}^n(G) - \mathcal{L}^n(\{x \in G : Dz^{(j)}(x) \in N_{\varepsilon^{(j)}}(K_1)\}) - \mathcal{L}^n(T_j),
\end{aligned}$$

we have from (4.7), (4.8) that for sufficiently large j

$$\mathcal{L}^n(\{x \in G : Dz^{(j)}(x) \in N_{\delta/2}(K_2)\}) \geq \frac{1}{7}\mathcal{L}^n(G) - \frac{1}{j}\mathcal{L}^n(\tilde{C}^{(j)})$$

and thus

$$\int_G \varphi_2(Dz^{(j)}(x))\, dx \geq \frac{1}{8}\mathcal{L}^n(G). \tag{4.13}$$

But

$$\lim_{j\to\infty}\int_G \varphi_i(Dz^{(j)})\, dx = \int_G \langle \nu_x, \varphi_i\rangle\, dx$$

for $i = 1, 2$, and one of these integrals is zero, contradicting (4.12),(4.13).

Proof of Theorem 14
Sufficiency. Let $Dy^{(j)}$ be bounded in $L^\infty(\Omega; M^{m\times n})$ and have Young measure $(\nu_x)_{x\in\Omega}$ with $\operatorname{supp}\nu_x \subset K_1 \cup K_2$ a.e.. Choose $\varepsilon \in (0, \varepsilon_0)$ sufficiently small so that $N_\varepsilon(K_1), N_\varepsilon(K_2)$ are disjoint. Then since $Dy^{(j)} \to K_1 \cup K_2$ in measure we have $\lim_{j\to\infty}\mathcal{L}^n(T_\varepsilon(y^{(j)})) = 0$ and hence by (4.2)

$$\min(\mathcal{L}^n(\Omega_{1,\varepsilon}(y^{(j)})), \mathcal{L}^n(\Omega_{2,\varepsilon}(y^{(j)}))) \to 0. \tag{4.14}$$

Let $f : M^{m\times n} \to [0, 1]$ be continuous with $f = 1$ on K_1, $f = 0$ outside $N_\varepsilon(K_1)$. Then

$$\lim_{j\to\infty} \fint_\Omega f(Dy^{(j)})\, dx = \fint_\Omega \langle \nu_x, f\rangle\, dx = \fint_\Omega \nu_x(K_1)\, dx. \tag{4.15}$$

From (4.14) there exists a subsequence $y^{(j_k)}$ of $y^{(j)}$ such that either $\mathcal{L}^n(\Omega_{1,\varepsilon}(y^{(j_k)})) \to 0$ or $\mathcal{L}^n(\Omega_{2,\varepsilon}(y^{(j_k)})) \to 0$, and so from (4.15) we have that

$$\fint_\Omega \nu_x(K_1)\, dx = 0 \text{ or } 1,$$

implying either that $\operatorname{supp}\nu_x \subset K_1$ a.e. or that $\operatorname{supp}\nu_x \subset K_2$ a.e. as required.

Necessity. Fix ε with $0 \leq \varepsilon < \varepsilon_0$, where ε_0 is given by Lemma 15 with E being the eccentricity of C (so that in particular $N_\varepsilon(K_1)$ and $N_\varepsilon(K_2)$ are disjoint), and let $y \in W^{1,p}(\Omega; \mathbb{R}^m)$. First suppose that

$$\mathcal{L}^n(T_\varepsilon(y)) \geq \frac{1}{4}\mathcal{L}^n(C).$$

Then

$$\int_{T_\varepsilon(y)} (1 + |Dy|^p)\, dx \geq \frac{1}{4}\frac{\mathcal{L}^n(C)}{\mathcal{L}^n(\Omega)}\mathcal{L}^n(\Omega),$$

so that (4.2) holds with $\gamma = \frac{1}{4}\frac{\mathcal{L}^n(C)}{\mathcal{L}^n(\Omega)}$.

We thus assume that

$$\mathcal{L}^n(T_\varepsilon(y)) < \frac{1}{4}\mathcal{L}^n(C). \tag{4.16}$$

Since Ω is C-connected, there is an equivalence class $\mathcal{C}$ of $\mathcal{K}(C)$ with respect to $\sim$ that covers Ω. Suppose that there exist two sets $C_1, C_2 \in \mathcal{C}$ (in particular, both directly congruent to C) such that

$$\mathcal{L}^n(\{x \in C_i : Dy(x) \in N_\varepsilon(K_i)\}) \geq \frac{1}{4}\mathcal{L}^n(C) \tag{4.17}$$

for $i = 1, 2$. By the definition of $\sim$ there exist continuous functions $\xi : [0,1] \to \Omega, Q : [0,1] \to SO(n)$, such that $\xi(0) + Q(0)C = C_1, \xi(1) + Q(1)C = C_2$, and $C(t) := \xi(t) + Q(t)C \subset \Omega$ for all $t \in [0,1]$. For $i = 1, 2$ define

$$\theta_i(t) = \frac{\mathcal{L}^n(\{x \in C(t) : Dy(x) \in N_\varepsilon(K_i)\})}{\mathcal{L}^n(C)}.$$

Then by (4.16) $\theta_i : [0,1] \to [0,1]$ is continuous, $\theta_1(t) + \theta_2(t) \geq \frac{3}{4}$ for all $t \in [0,1]$, and by (4.17) $\theta_1(0) \geq \frac{1}{4}$, $\theta_2(1) \geq \frac{1}{4}$. Hence there exists $\tau \in [0,1]$ with $\theta_1(\tau) \geq \frac{1}{4}$, $\theta_2(\tau) \geq \frac{1}{4}$. By Lemma 15 applied to $\tilde{C} = C(\tau)$ we deduce that

$$\int_{T_\varepsilon(y)} (1 + |Dy|^p)\, dx \geq \gamma_0 \frac{\mathcal{L}^n(C)}{\mathcal{L}^n(\Omega)} \mathcal{L}^n(\Omega)$$

so that (4.2) holds with $\gamma = \gamma_0 \frac{\mathcal{L}^n(C)}{\mathcal{L}^n(\Omega)}$.

It therefore remains to consider the case when for some $i = 1, 2$

$$\mathcal{L}^n(\{x \in D : Dy(x) \in N_\varepsilon(K_i)\}) < \frac{1}{4}\mathcal{L}^n(C) \tag{4.18}$$

for *every* $D \in \mathcal{C}$.

Let $\tilde{x}$ be any Lebesgue point of $\Omega_{i,\varepsilon} = \Omega_{i,\varepsilon}(y)$. Since $\mathcal{C}$ covers Ω there exist $\xi(\tilde{x}) \in \mathbb{R}^n$, $\tilde{Q}(\tilde{x}) \in SO(n)$ such that $\tilde{C}(\tilde{x}) := \xi(\tilde{x}) + \tilde{Q}(\tilde{x})C$ belongs to $\mathcal{C}$ and $\tilde{x} \in \tilde{C}(\tilde{x})$. For $0 < r \leq 1$ let $\tilde{C}_r(\tilde{x}) = r\tilde{C}(\tilde{x}) + (1-r)\tilde{x}$. Note that $\tilde{x} \in \tilde{C}_r(\tilde{x}) \subset \tilde{C}(\tilde{x})$. Define

$$f(\tilde{x}, r) = \frac{\mathcal{L}^n(\tilde{C}_r(\tilde{x}) \cap \Omega_{i,\varepsilon})}{\mathcal{L}^n(\tilde{C}_r(\tilde{x}))}.$$

Then $f(\tilde{x}, r)$ is continuous in $r \in (0,1]$, and since $\tilde{x}$ is a Lebesgue point of $\Omega_{i,\varepsilon}$ we have

$$\lim_{r \to 0} f(\tilde{x}, r) = 1.$$

But by (4.18) applied to $\tilde{C}(\tilde{x})$, we have $f(\tilde{x}, 1) < \frac{1}{4}$, and so there exists $r(\tilde{x}) \in (0,1]$ such that

$$\mathcal{L}^n(\{x \in \tilde{C}_{r(\tilde{x})}(\tilde{x}) : Dy(x) \in N_\varepsilon(K_i)\}) = \frac{1}{2}\mathcal{L}^n(\tilde{C}_{r(\tilde{x})}(\tilde{x})).$$

Let $B(a(C), R(C))$ be the minimal ball containing C. Then the balls

$$B_{\tilde{x}} = B(r(\tilde{x})[\tilde{Q}(\tilde{x})a(C) + \xi(\tilde{x})] + (1 - r(\tilde{x}))\tilde{x}, r(\tilde{x})R(C))$$

are such that $C_{r(\tilde{x})}(\tilde{x}) \subset B_{\tilde{x}}$ and in particular they cover the set E_i of Lebesgue points of $\Omega_{i,\varepsilon}$. It follows from Lemma 8 that there exists a finite or countable disjoint subfamily $\{B_j\}$, where $B_j = B_{\tilde{x}_j}$, such that

$$\sum_j \mathcal{L}^n(B_j) \geq c_n \mathcal{L}^n(\Omega_{i,\varepsilon}).$$

Hence by Lemma 15, writing $\tilde{C}_j = \tilde{C}_{r(\tilde{x}_j)}(\tilde{x}_j)$,

$$\begin{aligned}
\int_{T_\varepsilon(y)} (1 + |Dy|^p)\, dx &\geq \sum_j \int_{T_\varepsilon(y)\cap\tilde{C}_j} (1 + |Dy|^p)\, dx \\
&\geq \gamma_0 \sum_j \mathcal{L}^n(\tilde{C}_j) \\
&= \gamma_0 \frac{\mathcal{L}^n(C)}{\mathcal{L}^n(B(0, R(C)))} \sum_j \mathcal{L}^n(B_j) \\
&\geq \gamma_0 c_n \frac{\mathcal{L}^n(C)}{\mathcal{L}^n(B(0, R(C)))} \mathcal{L}^n(\Omega_{i,\varepsilon}) \\
&\geq \gamma_0 c_n (1 - E^2)^{\frac{n}{2}} \mathcal{L}^n(\Omega_{i,\varepsilon}). \qquad (4.19)
\end{aligned}$$

Combining this with the previous cases we deduce that (4.12) holds with

$$\gamma = \min[\gamma_1 \frac{\mathcal{L}^n(C)}{\mathcal{L}^n(\Omega)}, \gamma_0 c_n (1 - E^2)^{\frac{n}{2}}], \qquad (4.20)$$

where $\gamma_1 = \min(\gamma_0, \frac{1}{4})$.

The transition layer estimate can be given an equivalent formulation in terms of gradient Young measures.

Theorem 16 *Let $1 < p < \infty$ and let Ω be C-connected. Then $K_1, \ldots, K_N$ are incompatible if and only if there exist constants $\varepsilon_0 > 0$ and $\gamma > 0$ such that if $0 \leq \varepsilon < \varepsilon_0$ and $(\nu_x)_{x\in\Omega}$ is an L^p gradient Young measure then*

$$\int_\Omega \int_{[\bigcup_{r=1}^N N_\varepsilon(K_r)]^c} (1 + |A|^p)\, d\nu_x(A)\, dx \geq$$
$$\gamma \max_{1\leq r\leq N} \min \left(\int_\Omega \nu_x(N_\varepsilon(K_r))\, dx, \int_\Omega \nu_x(\bigcup_{s\neq r} N_\varepsilon(K_s))\, dx \right). \qquad (4.21)$$

The constant ε_0 can be chosen to depend only on $E(C), K_1, \ldots, K_N$ and p, while the constant γ can be chosen to depend only on these quantities and $\mathcal{L}^n(C)/\mathcal{L}^n(\Omega)$.

Note that Theorem 13 corresponds to the special case when $\nu_x = \delta_{Dy(x)}$. Again we need only prove the case $N = 2$ of Theorem 16, namely

Theorem 17 *Let $1 < p < \infty$ and let Ω be C-connected. A pair of disjoint compact sets K_1, K_2 are incompatible if and only if there exist constants $\varepsilon_0 > 0$ and $\gamma > 0$ such that if $0 \le \varepsilon < \varepsilon_0$ and $(\nu_x)_{x\in\Omega}$ is an L^p gradient Young measure then*

$$\int_\Omega \int_{[N_\varepsilon(K_1)\cup N_\varepsilon(K_2)]^c} (1+|A|^p)\, d\nu_x(A)\, dx \ge \gamma \min\left(\int_\Omega \nu_x(N_\varepsilon(K_1))\, dx, \int_\Omega \nu_x\left(N_\varepsilon(K_2)\right)\, dx\right). \tag{4.22}$$

The constant ε_0 can be chosen to depend only on $E(C), K_1, K_2$ and p, while the constant γ can be chosen to depend only on these quantities and $\mathcal{L}^n(C)/\mathcal{L}^n(\Omega)$.

Proof of Theorem 17 Since Theorem 14 is a special case of Theorem 17 we need only show that if K_1, K_2 are incompatible then (4.22) holds. Let ε_0, γ be as in Theorem 14, and let $0 \le \varepsilon < \varepsilon' < \varepsilon_0$. Let $(\nu_x)_{x\in\Omega}$ be an L^p gradient Young measure. By Theorem 1, we may suppose that $(\nu_x)_{x\in\Omega}$ is generated by a sequence $Dy^{(j)}$ of gradients which is such that $|Dy^{(j)}|^p$ is weakly convergent in $L^1(\Omega)$. Also

$$\int_\Omega \int_{M^{m\times n}} |A|^p d\nu_x(A)\, dx < \infty. \tag{4.23}$$

For $k = 1, 2, \ldots$ let $\varphi_k : M^{m\times n} \to [0,1]$ be continuous and satisfy

$$\varphi_k(A) = \begin{cases} 1 \text{ if } A \in [N_{\varepsilon'}(K_1)\cup N_{\varepsilon'}(K_2)]^c \\ 0 \text{ if } A \in N_{\varepsilon'-\frac{1}{k}}(K_1)\cup N_{\varepsilon'-\frac{1}{k}}(K_2), \end{cases} \tag{4.24}$$

with φ_k nonincreasing in k. The existence of $\tilde\varphi_k$ satisfying all but the last condition follows from Urysohn's lemma, and then we may set $\varphi_k = \min_{j\le k}\tilde\varphi_j$. Clearly $\varphi_k \to \chi_{\varepsilon'}$ pointwise, where $\chi_{\varepsilon'}$ is the characteristic function of the closure of $[N_{\varepsilon'}(K_1)\cup N_{\varepsilon'}(K_2)]^c$. Similarly, for $l = 1, 2$ let $\varphi_k^l : M^{m\times n} \to [0,1]$ be continuous and satisfy

$$\varphi_k^l(A) = \begin{cases} 0 \text{ if } A \in N_{\varepsilon'}(K_l)^c \\ 1 \text{ if } A \in N_{\varepsilon'-\frac{1}{k}}(K_l), \end{cases} \tag{4.25}$$

with φ_k^l nondecreasing in k. Clearly $\varphi_k^l \to \chi(\mathrm{int}\, N_{\varepsilon'}(K_l))$ pointwise.

For each j, k we have by Theorem 14 that

$$\begin{aligned} \int_\Omega \varphi_k(Dy^{(j)})(1+|Dy^{(j)}|^p)\, dx &\ge \int_{T_{\varepsilon'}(y^{(j)})} (1+|Dy^{(j)}|^p)\, dx \\ &\ge \gamma \min\left(\mathcal{L}^n(\Omega_{1,\varepsilon'}(y^{(j)})), \mathcal{L}^n(\Omega_{2,\varepsilon'}(y^{(j)}))\right) \\ &\ge \gamma \min\left(\int_\Omega \varphi_k^1(Dy^{(j)})\, dx, \int_\Omega \varphi_k^2(Dy^{(j)})\, dx\right). \end{aligned}$$

Since $|Dy^{(j)}|^p$ is weakly convergent in $L^1(\Omega)$, it is equi-integrable, and hence so is $\varphi_k(Dy^{(j)})(1+|Dy^{(j)}|^p)$, which thus has an L^1 weakly convergent subsequence. Letting $j \to \infty$ in this subsequence we deduce from the fundamental properties of Young measures that

$$\int_\Omega \langle \nu_x, \varphi_k(A)(1+|A|^p)\rangle \, dx \geq \gamma \min\left(\int_\Omega \langle \nu_x, \varphi_k^1\rangle \, dx, \int_\Omega \langle \nu_x, \varphi_k^2\rangle \, dx\right) \tag{4.26}$$

Passing to the limit $k \to \infty$, using the everywhere monotone convergence of $\varphi_k, \varphi_k^1, \varphi_k^2$, we obtain

$$\begin{aligned}\int_\Omega \int_{M^{m\times n}} & \chi_{\varepsilon'}(A)(1+|A|^p)\, d\nu_x(A)\, dx \\ &\geq \gamma \min\left(\int_\Omega \nu_x(\operatorname{int} N_{\varepsilon'}(K_1))\, dx, \int_\Omega \nu_x(\operatorname{int} N_{\varepsilon'}(K_2))\, dx\right) \\ &\geq \gamma \min\left(\int_\Omega \nu_x(N_\varepsilon(K_1))\, dx, \int_\Omega \nu_x(N_\varepsilon(K_2))\, dx\right).\end{aligned}$$

Letting $\varepsilon' \to \varepsilon+$, and noting that $\chi_{\varepsilon'} \to \chi([N_\varepsilon(K_1) \cup N_\varepsilon(K_2)]^c)$ monotonically, we deduce by (4.23) and monotone convergence that (4.22) holds.

Corollary 18 *Let $K_1, \ldots K_N$ be incompatible. Then there exists $\varepsilon_0 > 0$ such that $N_\varepsilon(K_1), \ldots, N_\varepsilon(K_N)$ are incompatible for $0 \leq \varepsilon < \varepsilon_0$.*

Proof By Remark 5b we may assume that Ω is C-connected, while by Remark 5a we need only show that $N_\varepsilon(K_r)$ and $\bigcup_{s\neq r} N_\varepsilon(K_s)$ are incompatible. Let $\operatorname{supp}\nu_x \subset \bigcup_{r=1}^N N_\varepsilon(K_r)$ a.e.. Then the left-hand side of (4.21) is zero. Hence for each r either $\nu_x(N_\varepsilon(K_r)) = 0$ a.e. or $\nu_x(\bigcup_{s\neq r} N_\varepsilon(K_s)) = 0$ a.e., and hence either $\operatorname{supp}\nu_x \subset \bigcup_{s\neq r} N_\varepsilon(K_s)$ a.e. or $\operatorname{supp}\nu_x \subset N_\varepsilon(K_r)$ a.e., giving the result.

Applying the above Corollary 18 to the case when each K_r consists of a single matrix we immediately obtain

Corollary 19 *For any N the set of points $(A_1, \ldots, A_N) \in (M^{m\times n})^N$ with $\{A_1\}, \ldots, \{A_N\}$ incompatible is open.*

When $N = 2$ this already gives interesting information. Indeed it implies a special case of Šverák's three matrix theorem [66]. In fact if $A_1, A_2 \in M^{m\times n}$, with $\operatorname{rank}(A_1 - A_2) > 1$, we have that $\{A_1\}, \{A_2\}$ are incompatible, and so if A_3 is taken sufficiently close to A_2 with $\operatorname{rank}(A_2 - A_3) > 1$ we have that the sets $K_1 = \{A_1\}$ and $K_2 = \{A_2, A_3\}$ are incompatible. Thus if $(\nu_x)_{x\in\Omega}$ is a gradient Young measure with $\operatorname{supp}\nu_x \subset \{A_1, A_2, A_3\}$ a.e. then either $\nu_x = \delta_{A_1}$ a.e. or $\operatorname{supp}\nu_x \subset \{A_2, A_3\}$ a.e.. In the latter case, since $\{A_2\}, \{A_3\}$ are incompatible, we have that either $\nu_x = \delta_{A_2}$ a.e. or $\nu_x = \delta_{A_3}$ a.e.. Hence $\nu_x = \delta_{A_i}$ a.e. for some i, which is the statement of Šverák's theorem in this special case. As remarked to us by V. Šverák, this special case cannot be proved using the minors relations alone. For example, taking $m=n=2$, the probability measure

$$\nu = \frac{\varepsilon^2}{4-\varepsilon^2}\delta_0 + \frac{2-\varepsilon^2}{4-\varepsilon^2}(\delta_{\mathbf{1}} + \delta_{A_\varepsilon}),$$

where $A_\varepsilon = \begin{pmatrix} 1-\varepsilon & 0 \\ 0 & 1+\varepsilon \end{pmatrix}$ and $\varepsilon > 0$ is sufficiently small, satisfies the minors relation $\det \bar{\nu} = \langle \nu, \det \rangle$, but by the above $\{0\}, \{\mathbf{1}\}, \{A_\varepsilon\}$ are incompatible. By Theorem 10, Corollary 18 thus implies the existence of quasiconvex functions that are not polyconvex. In [12] we give a new proof of the three matrix theorem in the general case, using similar techniques as in the proof of Theorem 13 plus ingredients from the theory of quasiregular maps.

The following simple example shows that Theorems 13, 14, 16, 17 are not true if $1+|A|^p$ is replaced by $|A|^p$ in the integrals over the transition layer, even when the volume of the transition layer is arbitrarily small.

Example 10 Let $m=n=2$, $\Omega = (0,1)^2$ and let $A_1 = e_2 \otimes e_2$, $A_2 = (e_1+e_2)\otimes(e_1+e_2)$. Then $K_1 = \{A_1\}, K_2 = \{A_2\}$ are incompatible, but 0 is compatible with both A_1 and A_2. Define for small $\delta > 0$ and for $x \in \Omega$,

$$y_\delta(x) = \begin{cases} x_2 e_2 & \text{if } 0 < x_2 < 1-\delta, \\ (1-\delta)e_2 & \text{if } x_2 \geq 1-\delta,\ x_1+x_2 \leq 2-\delta, \\ (e_1+e_2)(x_1+x_2) + (\delta-2)e_1 - e_2 & \text{if } x_2 \geq 1-\delta,\ x_1+x_2 > 2-\delta. \end{cases}$$

Then

$$Dy_\delta(x) = \begin{cases} A_1 & \text{if } 0 < x_2 < 1-\delta, \\ 0 & \text{if } x_2 \geq 1-\delta,\ x_1+x_2 \leq 2-\delta, \\ A_2 & \text{if } x_2 \geq 1-\delta,\ x_1+x_2 > 2-\delta, \end{cases}$$

and we have for any $p > 1$

$$\int_{T_0(y_\delta)} |Dy_\delta|^p \, dx = 0, \quad \min\{\mathcal{L}^2(\Omega_{1,0}(y_\delta)), \mathcal{L}^2(\Omega_{2,0}(y_\delta))\} = \frac{1}{2}\delta^2.$$

We now show that Theorems 13, 14, 16, 17 do not hold for general bounded domains Ω. Since by Proposition 7 the hypothesis in these theorems that Ω be C-connected is equivalent to the cone condition, for a counterexample we need a domain not satisfying the cone condition.

Example 11 We take Ω to be the 'rooms and passages' domain of Fraenkel [35]. For simplicity we let $m=n=2$. This domain Ω consists of the union of a sequence of square rooms $Q_j = (a_j, 0) + h_j(-1,1)^2$, $j = 1, 2, \ldots$, of decreasing side $2h_j > 0$, centred at the points $(a_j, 0) \in \mathbb{R}^2$ on the x_1-axis, where $a_1 = 0, a_j > 0$, together with the rectangular connecting corridors $C_j = [a_j+h_j, a_{j+1}-h_{j+1}] \times (-d_j, d_j)$ of length $l_j = a_{j+1} - h_{j+1} - (a_j+h_j) > 0$ and thickness $2d_j$, where $0 < d_j < h_{j+1}$. In order for Ω to be bounded, we require that $\sum_{j=1}^\infty (2h_j + l_j) < \infty$.

Let $A_1, A_2 \in M^{2\times 2}$ with $\operatorname{rank}(A_1 - A_2) > 1$, for example $A_1 = 0, A_2 = \mathbf{1}$. Thus by Example 4 the sets $K_1 = \{A_1\}, K_2 = \{A_2\}$ are incompatible. We define $y^{(j)} : \Omega \to \mathbb{R}^2$ by

$$y^{(j)}(x) = \begin{cases} A_1 x & \text{if } x \in \Omega_j, \\ \frac{x_1 - a_{j-1} - h_{j-1}}{l_{j-1}} A_2 x + \left(1 - \frac{x_1 - a_{j-1} - h_{j-1}}{l_{j-1}}\right) A_1 x & \text{if } x \in C_{j-1}, \\ A_2 x & \text{if } x \in Q_j, \\ \frac{x_1 - a_j - h_j}{l_j} A_1 x + \left(1 - \frac{x_1 - a_j - h_j}{l_j}\right) A_2 x & \text{if } x \in C_j, \end{cases}$$

where $\Omega_j = \Omega \backslash (C_{j-1} \cup Q_j \cup C_j)$. Thus in the corridor C_{j-1}

$$|Dy^{(j)}(x)| \le c_0 + \frac{c_1}{l_{j-1}},$$

while in the corridor C_j

$$|Dy^{(j)}(x)| \le c_0 + \frac{c_1}{l_j},$$

where c_0, c_1 are constants independent of j. Thus taking $\varepsilon = 0$, we have

$$\begin{aligned} \int_{T_0(y^{(j)})} (1 + |Dy^{(j)}|^p)\, dx &= \int_{C_{j-1} \cup C_j} (1 + |Dy^{(j)}|^p)\, dx \\ &\le 2 l_{j-1} d_{j-1} \left[1 + \left(c_0 + \frac{c_1}{l_{j-1}}\right)^p\right] \\ &\quad + 2 l_j d_j \left[1 + \left(c_0 + \frac{c_1}{l_j}\right)^p\right], \end{aligned}$$

while

$$\min(\mathcal{L}^2(\Omega_{1,0}(y^{(j)})), \mathcal{L}^2(\Omega_{2,0}(y^{(j)})) = \mathcal{L}^2(Q_j) = 4h_j^2.$$

Thus, fixing the sequences h_j and l_j and letting $d_j \to 0$ sufficiently rapidly as $j \to \infty$, we violate the conclusion (4.2) of Theorem 14 for any choice of γ.

We do have a way of giving a lower bound on the constant γ in Theorems 13, 14, 16, 17. The proof of Theorem 14 gives such a lower bound (see (4.20)), but only in terms of the constant γ_0 which occurs in Lemma 15, this lemma being proved by contradiction. To estimate γ_0 would presumably require some elliptic estimates. However we can obtain upper bounds on γ by considering explicit test functions. We illustrate this in the next example for the case when $m = n$, $p = 2$, Ω is a ball and $K_1 = \{\lambda \mathbf{1}\}, K_2 = \{\mu \mathbf{1}\}$ with $\lambda \ne \mu$.

Example 12 Let $m = n > 1$, $\Omega = B(0,1)$, $A_1 = \lambda \mathbf{1}$, $A_2 = \mu \mathbf{1}$, $\lambda \ne \mu$. We consider for $k > 1$ and sufficiently small $\varepsilon > 0$ the radial mapping

$$y_\varepsilon(x) = \frac{r_\varepsilon(R)}{R} x, \tag{4.27}$$

where $R = |x|$ and

$$r_\varepsilon(R) = \begin{cases} \lambda R & \text{if } 0 \le R \le \varepsilon, \\ \mu R & \text{if } k\varepsilon \le R < 1. \end{cases} \tag{4.28}$$

For $\varepsilon < R < k\varepsilon$ we choose r_ε so that it is continuous and minimizes

$$\int_{\{\varepsilon<|x|<k\varepsilon\}} (1 + |Dy|^2)\, dx. \tag{4.29}$$

Noting that

$$|Dy|^2 = (n-1)\Big(\frac{r}{R}\Big)^2 + (r')^2, \tag{4.30}$$

the Euler-Lagrange equation for the functional

$$\int_\varepsilon^{k\varepsilon} R^{n-1} \left(1 + (n-1)\left(\frac{r}{R}\right)^2 + (r')^2\right) dR \tag{4.31}$$

has linearly independent solutions $r = R$ and $r = R^{1-n}$. Choosing constants A, B so that $r(R) = AR + BR^{1-n}$ satisfies $r(\varepsilon) = \lambda\varepsilon, r(k\varepsilon) = \mu k\varepsilon$, we find that for the optimal transition layer

$$r_\varepsilon(R) = \left(\frac{k^n\mu - \lambda}{k^n - 1}\right) R + \frac{(\lambda-\mu)(\varepsilon k)^n}{k^n - 1} R^{1-n}, \text{ if } \varepsilon < R < k\varepsilon. \tag{4.32}$$

(In fact by uniqueness of solutions to Laplace's equation this radial solution is the minimizer of (4.29) among all (not necessarily radial) maps matching the boundary conditions at $R = \varepsilon, k\varepsilon$.) Denoting by $T = \{\varepsilon < |x| < k\varepsilon\}$ the transition layer, we calculate using (4.30) that the ratio

$$\rho = \frac{\int_T (1 + |Dy|^2)\, dx}{\mathcal{L}^n(\{x : Dy(x) = \lambda x\})}$$

is given by

$$\begin{aligned} \rho &= \frac{1}{\varepsilon^{n-1}\omega_n} \int_\varepsilon^{k\varepsilon} R^{n-1} \left[1 + n\left(\frac{k^n\mu-\lambda}{k^n-1}\right)^2 + n(n-1)\left(\frac{\lambda-\mu}{k^n-1}\right)^2 \left(\frac{\varepsilon k}{R}\right)^{2n}\right] dR \\ &= \int_1^k s^{n-1} \left[1 + n\left(\frac{k^n\mu-\lambda}{k^n-1}\right)^2 + n(n-1)\left(\frac{\lambda-\mu}{k^n-1}\right)^2 \left(\frac{k}{s}\right)^{2n}\right] ds \\ &= \frac{k^n-1}{n} + \frac{(k^n\mu-\lambda)^2}{k^n-1} + (n-1)\frac{(\lambda-\mu)^2 k^n}{k^n-1}. \end{aligned}$$

Here ω_n denotes the $(n-1)$-dimensional measure of the unit sphere in $\mathbb{R}^n$. To find the optimal width of the transition layer, we minimize ρ over $k > 1$. Setting $\tau = k^n$ and minimizing over $\tau > 1$ we find that the minimum value $\rho_{\min}$ is achieved when $\tau = 1 + \frac{n}{\sqrt{1+n\mu^2}}|\lambda-\mu|$, and that

$$\rho_{\min} = (n-1)(\lambda-\mu)^2 + 2(\sqrt{1+n\mu^2} - \operatorname{sign}(\lambda-\mu))|\lambda-\mu|.$$

Interchanging λ and μ we deduce finally that the constant γ satisfies

$$\gamma \leq (n-1)(\lambda-\mu)^2 + 2h(\lambda,\mu)|\lambda-\mu|, \tag{4.33}$$

where $h(\lambda,\mu) = \min(\sqrt{1+n\mu^2} - \mathrm{sign}\,(\lambda-\mu), \sqrt{1+n\lambda^2} - \mathrm{sign}\,(\mu-\lambda))$. Of course this upper bound tends to zero as $\lambda \to \mu$. Note that the upper bound is proportional to $|\lambda-\mu|$ when both λ and μ are near one.

5 Local Minimizers and Metastability

In this section we apply the transition layer estimate to prove that certain maps or microstructures (in the *parent phase*) are local minimizers of the corresponding energy, the mechanism being that the values of the gradient that could potentially lower the energy (those of the *product phase*) are incompatible with those of the parent phase, so that the gain in energy due to the resulting transition layer is greater than the loss of energy in using the gradients of the product phase. In applications to materials undergoing solid phase transformations this provides a mechanism for *incompatibility induced hysteresis*.

The basic integral we consider is

$$I(y) = \int_\Omega W(Dy(x))\,dx, \tag{5.1}$$

where $\Omega \subset \mathbb{R}^n$ is a bounded domain that is C-connected. We assume that

(H1) $W : M^{m\times n} \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous,
(H2) There exist constants $c_0 \in \mathbb{R}, c_1 > 0, p > 1$ such that

$$W(A) \geq c_0 + c_1|A|^p \text{ for all } A \in M^{m\times n}. \tag{5.2}$$

More generally we will consider the extension (relaxation) of (5.1) to gradient Young measures

$$I(\nu) = \int_\Omega \int_{M^{m\times n}} W(A)\,d\nu_x(A)\,dx, \tag{5.3}$$

where $\nu = (\nu_x)_{x\in\Omega}$ is the Young measure corresponding to a sequence $Dy^{(j)}$ that is bounded in $L^p(\Omega; M^{m\times n})$. The functional (5.1) corresponds to the special case when $\nu_x = \delta_{Dy(x)}$ for some $y \in W^{1,p}(\Omega;\mathbb{R}^m)$.

We suppose that the parent and product phases are represented by the compact sets $K_1, K_2 \subset M^{m\times n}$ respectively, where K_1, K_2 are incompatible. Let $\varepsilon_0 = \varepsilon_0(E(C), K_1, K_2, p)$ be as in Theorem 14, and fix ε with $0 < \varepsilon < \varepsilon_0$. We assume that

(H3) $\min_{A\in N_{\varepsilon/2}(K_1)} W(A) = 0,\ W(A) \geq 0$ for all $A \in N_\varepsilon(K_1)$,
(H4) $W(A) \geq -\delta$ for all $A \in N_\varepsilon(K_2)$ and some $\delta > 0$,
(H5) $W(A) \geq \alpha$ for all $A \in [N_\varepsilon(K_1) \cup N_\varepsilon(K_2)]^c$ and some $\alpha > 0$.

Thus W has a local minimizer near the well K_1, with minimum value zero, and a possibly lower local minimizer near the well K_2. We will assume later that $\delta > 0$ is sufficiently small, while $\alpha > 0$ remains fixed. The hypotheses (H1)-(H5) are satisfied if W is a small perturbation of some W_0 which has local minimizers with equal minimum value zero at the wells K_1, K_2, as we now show.

Proposition 20 *Assume that*

(H1)′ $W_\tau : M^{m\times n} \to \mathbb{R}\cup\{+\infty\}$ *is lower semicontinuous in* $(\tau, A) \in [0,1]\times M^{m\times n}$*, with* $W_\tau(A)$ *continuous in* τ *for all* $A \in M^{m\times n}$*,*

(H2)′ $W_0(A) \geq 0$ *for all* $A \in M^{m\times n}$*, and* $W_0^{-1}(0) = K_1 \cup K_2$*,*

(H3)′ $\min_{A\in N_\varepsilon(K_1)} W_\tau(A) = 0$ *for all* $\tau \in [0,1]$*,*

(H4)′ $W_\tau(A) \geq c_0 + c_1|A|^p$ *for all* $\tau \in [0,1], A \in M^{m\times n}$*.*

Then, for sufficiently small $\tau > 0$*,* W_τ *satisfies* (H1) − (H5) *for some fixed* $\alpha > 0$ *and* $\delta = \delta(\tau)$ *satisfying*

$$\lim_{\tau\to 0+} \delta(\tau) = 0. \tag{5.4}$$

Proof Clearly W_τ satisfies (H1), (H2). To prove (H3) note that by (H3)′ there exists $A_\tau \in N_\varepsilon(K_1)$ with $W_\tau(A_\tau) = 0$. We claim that $A_\tau \in N_{\varepsilon/2}(K_1)$ for τ sufficiently small. If not, there would exist $\tau_j \to 0$ with $\operatorname{dist}(A_{\tau_j}, K_1) > \varepsilon/2$ for all j, and we can suppose that $A_{\tau_j} \to A \notin N_{\varepsilon/4}(K_1)$. But then by (H1)′

$$0 = \liminf_{j\to\infty} W_{\tau_j}(A_{\tau_j}) \geq W_0(A), \tag{5.5}$$

and so by (H2)′ $A \in K_1$, a contradiction.

To prove (H4) note that by (H1′), (H4′), W_τ attains a minimum on $N_\varepsilon(K_2)$ at some B_τ, so that $W_\tau(A) \geq -\delta(\tau)$ for $A \in N_\varepsilon(K_2)$, where

$$\delta(\tau) = \max\{-W_\tau(B_\tau), \tau\} > 0.$$

Letting $\tau \to 0+$ we have by (H1′) that $0 \leq W_0(B) \leq \liminf_{\tau\to 0+} W_\tau(B_\tau)$ for some $B \in N_\varepsilon(K_2)$ and so $\lim_{\tau\to 0+}\delta(\tau) = 0$.

To prove (H5) note that by (H1′), (H4′), W_τ attains a minimum on the closure of $[N_\varepsilon(K_1)\cup N_\varepsilon(K_2)]^c$ at some C_τ, where C_τ is bounded for sufficiently small τ. If (H5) were false then there would exist a sequence $\tau_j \to 0+$ with $W_{\tau_j}(C_{\tau_j}) \leq 1/j$ and we may assume that $C_j \to C \notin K_1 \cup K_2$. But then (H2′) and (H1′) imply that $0 < W_0(C) \leq \liminf_{j\to\infty} W_{\tau_j}(C_j) \leq 0$, a contradiction.

Theorem 21 *Let* Ω *be* C*-connected, and let* W *satisfy* (H1) − (H5) *with* δ *sufficiently small, so that* $0 < \delta < \delta_0$*, where* δ_0 *is a constant depending only on* $K_1, K_2, p, E(C), \mathcal{L}^n(C)/\mathcal{L}^n(\Omega), \varepsilon, c_0, c_1$ *and* α*. Let* $\nu^* = (\nu_x^*)_{x\in\Omega}$ *be an* L^p *gradient Young measure with* $\operatorname{supp}\nu_x^* \subset \{A \in N_\varepsilon(K_1) : W(A) = 0\}$ *and* $\bar{\nu}_x^* = Dy^*(x)$*, where* $y^* \in W^{1,p}(\Omega;\mathbb{R}^m)$*. Then there exists* $\sigma > 0$*, depending on the above quantities and* $L^n(\Omega)$*, such that*

$$I(\nu) \geq I(\nu^*) \tag{5.6}$$

for any L^p gradient Young measure $\nu = (\nu_x)_{x\in\Omega}$ with $\bar{\nu}_x = Dy(x)$ and

$$\|y - y^*\|_{L^1(\Omega;\mathbb{R}^m)} < \sigma. \tag{5.7}$$

The inequality in (5.6) *is strict unless* $\operatorname{supp}\nu_x \subset \{A \in N_\varepsilon(K_1) : W(A) = 0\}$ *for a.e. $x \in \Omega$.*

We will use the following lemmas.

Lemma 22 *Let Ω be C-connected, and let $E(C) = E$. There exist $\Delta > 0$ depending only on $K_1, K_2, p, E, \varepsilon$ and $\mathcal{L}^n(C)/\mathcal{L}^n(\Omega)$, and $\beta > 0$ depending only on E and $\mathcal{L}^n(C)/\mathcal{L}^n(\Omega)$, such that if $\nu = (\nu_x)_{x\in\Omega}$ is an L^p gradient Young measure with $\bar{\nu}_x = Dy(x)$ for $y \in W^{1,p}(\Omega;\mathbb{R}^m)$ and*

$$\int_\Omega \int_{[N_\varepsilon(K_1)\cup N_\varepsilon(K_2)]^c} (1 + |A|^p)\, d\nu_x(A)\, dx + \int_\Omega \nu_x(N_\varepsilon(K_1))\, dx < \Delta\mathcal{L}^n(\Omega) \tag{5.8}$$

then

$$\|y - z\|_{L^1(\Omega;\mathbb{R}^m)} > \beta\Delta\mathcal{L}^n(\Omega)^{\frac{n+1}{n}} \tag{5.9}$$

for all $z \in W^{1,p}(\Omega;\mathbb{R}^m)$ with $Dz(x) \in N_\varepsilon(K_1)^{\mathrm{qc}}$ a.e. in Ω.

Proof We first claim that it suffices to prove the existence of Δ in the special case when Ω is the open ball $B = B(0, r_n) = B(0, (n/\omega_n)^{\frac{1}{n}})$ for which $\mathcal{L}^n(B) = 1$, with $\beta = 1$. Indeed suppose this has been proved with corresponding $\Delta = \Delta_B$ and let Ω be C-connected with $E(C) = E$ and $\mathcal{L}^n(C) = \kappa\mathcal{L}^n(\Omega)$. Then since Ω is C-filled, Ω contains an open ball of radius $\frac{1}{2}r(C)$, and since $R(C) \geq \left(\frac{n\mathcal{L}^n(C)}{\omega_n}\right)^{\frac{1}{n}} = \left(\frac{n\kappa\mathcal{L}^n(\Omega)}{\omega_n}\right)^{\frac{1}{n}}$, Ω contains an open ball $B_\rho = a + \rho B(0,1)$ of radius

$$\rho = \frac{1}{2}\left(\frac{n\kappa\mathcal{L}^n(\Omega)}{\omega_n}\right)^{\frac{1}{n}} (1 - E^2)^{\frac{1}{2}}.$$

Therefore if (5.8) holds with Δ given by $\Delta(E,\kappa) = 2^{-n}\kappa(1 - E^2)^{\frac{n}{2}}\Delta_B$ then

$$\int_{B_\rho} \int_{[N_\varepsilon(K_1)\cup N_\varepsilon(K_2)]^c} (1 + |A|^p)\, d\nu_x(A)\, dx + \int_{B_\rho} \nu_x(N_\varepsilon(K_1))\, dx < 2^{-n}\kappa(1 - E^2)^{\frac{n}{2}}\Delta_B\mathcal{L}^n(\Omega). \tag{5.10}$$

Define $\mu = (\mu_x)_{x\in B}$ by $\mu_x = \nu_{a + \frac{\rho}{r_n}x}$ and $\tilde{y}(x) = \frac{r_n}{\rho}y(a + \frac{\rho}{r_n}x)$. Then $D\tilde{y}(x) = \bar{\mu}_x$. Hence

$$\int_B \int_{[N_\varepsilon(K_1)\cup N_\varepsilon(K_2)]^c} (1 + |A|^p)\, d\mu_x(A)\, dx + \int_B \mu_x(N_\varepsilon(K_1))\, dx < \Delta_B. \tag{5.11}$$

If $z \in W^{1,p}(\Omega;\mathbb{R}^m)$ with $Dz(x) \in N_\varepsilon(K_1)^{\mathrm{qc}}$ a.e. and $\tilde{z}(x) = \frac{r_n}{\rho}z(a + \frac{\rho}{r_n}x)$ we have that $D\tilde{z}(x) = Dz(a + \frac{\rho}{r_n}x) \in N_\varepsilon(K_1)^{\mathrm{qc}}$ a.e. $x \in B$. Since we are assuming the result holds for $\Omega = B$ and $\beta = 1$ we deduce that

$$\|\tilde{y} - \tilde{z}\|_{L^1(B;\mathbb{R}^m)} > \Delta_B,$$

which implies that

$$\|y-z\|_{L^1(\Omega;\mathbb{R}^m)} \geq \|y-z\|_{L^1(B_\rho;\mathbb{R}^m)} > \beta(\kappa,E)\Delta(\kappa,E)\mathcal{L}^n(\Omega)^{\frac{n+1}{n}},$$

where $\beta(\kappa,E)=\frac{1}{2}\kappa^{\frac{1}{n}}(1-E^2)^{\frac{1}{2}}$, proving the claim.

Suppose then that the result is false for $\Omega=B$ and $\beta=1$, so that it is false for $\Delta=\frac{1}{j}$ for every j. Then there exist a sequence of L^p gradient Young measures $\nu^{(j)}=(\nu_x^{(j)})_{x\in B}$, and mappings $y^{(j)}\in W^{1,p}(B;\mathbb{R}^m)$ with $\bar{\nu}_x^{(j)}=Dy^{(j)}(x)$, $z^{(j)}\in W^{1,p}(B;\mathbb{R}^m)$ with $Dz^{(j)}(x)\in N_\varepsilon(K_1)^{\mathrm{qc}}$ a.e. in B, such that

$$\int_B\int_{[N_\varepsilon(K_1)\cup N_\varepsilon(K_2)]^c}(1+|A|^p)d\nu_x^{(j)}(A)\,dx+\int_B \nu_x^{(j)}(N_\varepsilon(K_1))\,dx<j^{-1} \tag{5.12}$$

and

$$\|y^{(j)}-z^{(j)}\|_{L^1(B;\mathbb{R}^m)}\leq j^{-1}. \tag{5.13}$$

It follows from (5.12) and the boundedness of $N_\varepsilon(K_1), N_\varepsilon(K_2)$ that

$$\int_B\int_{M^{m\times n}}(1+|A|^p)d\nu_x^{(j)}(A)\,dx\leq M<\infty \tag{5.14}$$

for all j. We may suppose without loss of generality that $\int_B y^{(j)}(x)\,dx=0$. We use the inequality (see Morrey [54, p. 82] for similar results and proofs)

$$\int_B|u|^p\,dx\leq C\left(\int_B|Du|^p\,dx+\left|\int_B u\,dx\right|^p\right)\text{ for all }u\in W^{1,p}(B;\mathbb{R}^m), \tag{5.15}$$

where C is a constant. Applying (5.15) to $y^{(j)}$, using $\bar{\nu}_x^{(j)}=Dy^{(j)}(x)$ and Hölder's inequality, we deduce that $y^{(j)}$ is bounded in $W^{1,p}(B;\mathbb{R}^m)$. Extracting a subsequence (not relabelled) if necessary, we may assume that $\nu^{(j)}\overset{*}{\rightharpoonup}\nu$ in $L_w^\infty(B;C_0(M^{m\times n})^*)$, and hence by Sychev [68, Proposition 4.5] $\nu=(\nu_x)_{x\in B}$ is an L^p gradient Young measure. Thus $\bar{\nu}_x=Dy(x)$ a.e. for some $y\in W^{1,p}(B;\mathbb{R}^m)$ with $\int_B y\,dx=0$. We claim that $y^{(j)}\rightharpoonup y$ in $W^{1,p}(B;\mathbb{R}^m)$. To this end let $\theta_k:[0,\infty)\to[0,1]$ satisfy $\theta_k(s)=1$ for $s\in[0,k]$, $\theta_k(s)=0$ for

$s \in [k+1, \infty)$. Then if $\psi \in C_0^\infty(\Omega)$ we have that

$$\begin{aligned}
&\limsup_{j\to\infty} \left| \int_B \psi(x)(Dy^{(j)}(x) - Dy(x))\,dx \right| \\
&\qquad = \limsup_{j\to\infty} \left| \int_B \psi(x) \int_{M^{m\times n}} A\, d(\nu_x^{(j)} - \nu_x)(A)\,dx \right| \\
&\qquad \le \limsup_{j\to\infty} \left| \int_B \psi(x) \int_{M^{m\times n}} \theta_k(|A|) A\, d(\nu_x^{(j)} - \nu_x)(A)\,dx \right| \\
&\qquad\quad + \limsup_{j\to\infty} \left| \int_B \psi(x) \int_{|A|\ge k} (1-\theta_k(|A|)) A\, d(\nu_x^{(j)} - \nu_x)(A)\,dx \right| \\
&\qquad \le \limsup_{j\to\infty} \left| \int_B |\psi(x)| \left(\int_{|A|\ge k} |A|\, d(\nu_x^{(j)} + \nu_x)(A) \right) \right|, \\
&\qquad \le \frac{C}{k^{p-1}},
\end{aligned}$$

for some constant C, where we have used $\nu^{(j)} \stackrel{*}{\rightharpoonup} \nu$ in $L_w^\infty(B; C_0(M^{m\times n})^*)$, (5.14) and relation (iii) of Theorem 1. Letting $k \to \infty$ we deduce that $Dy^{(j)} \rightharpoonup Dy$ in $L^p(B; M^{m\times n})$, from which the claim follows since $\int_B y^{(j)}\,dx = \int_B y\,dx = 0$. By the compactness of the embedding we have that $y^{(j)} \to y$ strongly in $L^p(B; \mathbb{R}^m)$.

Note that by (5.12) we have that

$$\int_B (1 - \nu_x^{(j)}(N_\varepsilon(K_2))\,dx \le \frac{1}{j}. \tag{5.16}$$

Let $\varphi_k \in C_0(M^{m\times n})$, with $0 \le \varphi_k(A) \le 1$, $\varphi_{k+1}(A) \le \varphi_k(A)$ and $\lim_{k\to\infty} \varphi_k(A) = \chi_{N_\varepsilon(K_2)}(A)$ for all $A \in M^{m\times n}$, where $\chi_{N_\varepsilon(K_2)}$ is the characteristic function of $N_\varepsilon(K_2)$. Then by (5.16) we have that

$$\lim_{j\to\infty} \int_B \int_{M^{m\times n}} (1 - \varphi_k(A))\, d\nu_x^{(j)}(A)\,dx = 0,$$

and so by the weak* convergence of $\nu^{(j)}$ we deduce that

$$\int_B \int_{M^{m\times n}} (1 - \varphi_k(A)) d\nu_x(A)\,dx = 0.$$

Passing to the limit $k \to \infty$ using monotone convergence we obtain

$$\int_B [1 - \nu_x(N_\varepsilon(K_2))]\,dx = 0.$$

Thus $\operatorname{supp} \nu_x \subset N_\varepsilon(K_2)$ a.e. in Ω. In particular $Dy(x) \in N_\varepsilon(K_2)^{\mathrm{qc}}$ a.e. in B.

But from (5.13) we deduce that $z^{(j)} \to y$ in $L^1(B; \mathbb{R}^m)$. Since $Dz^{(j)} \in N_\varepsilon(K_1)^{\mathrm{qc}}$ it follows that $Dz^{(j)} \stackrel{*}{\rightharpoonup} Dy$ in $L^\infty(B; M^{m\times n})$ and thus $Dy(x) \in N_\varepsilon(K_1)^{\mathrm{qc}}$. But $N_\varepsilon(K_1)^{\mathrm{qc}}$ and $N_\varepsilon(K_2)^{\mathrm{qc}}$ are disjoint by Corollary 9, giving the desired contradiction.

Lemma 23 *Let W satisfy* (H2) *and* (H5). *Then*

$$W(A) \geq K(1+|A|^p) \text{ for all } A \in [N_\varepsilon(K_1) \cup N_\varepsilon(K_2)]^c, \tag{5.17}$$

where

$$K = \begin{cases} c_1 & \text{if } c_0 \geq c_1 \\ c_0 & \text{if } \alpha \leq c_0 < c_1 \\ \frac{\alpha c_1}{\alpha + c_1 - c_0} & \text{if } \alpha > c_0, c_1 > c_0. \end{cases}$$

Proof This is elementary.

Proof of Theorem 21 With Δ, β, K chosen as in Lemmas 22, 23 respectively, and $\gamma > 0$ the constant in the transition layer estimate (4.22), choose $\delta > 0$ with

$$\delta < \frac{K}{2} \min\left(\gamma, \Delta \min(1,\gamma)\right), \tag{5.18}$$

and let $\sigma = \beta \Delta \mathcal{L}^n(\Omega)^{\frac{n+1}{n}}$.

For ν, ν^* as in the statement of the theorem we have that

$$\begin{aligned} I(\nu) - I(\nu^*) &= I(\nu) - 0 \\ &= \int_\Omega \int_{N_\varepsilon(K_1)} W(A)\, d\nu_x(A)\, dx + \int_\Omega \int_{N_\varepsilon(K_2)} W(A)\, d\nu_x(A)\, dx \\ &\qquad + \int_\Omega \int_{[N_\varepsilon(K_1) \cup N_\varepsilon(K_2)]^c} W(A)\, d\nu_x(A)\, dx \\ &\geq 0 - \delta \int_\Omega \nu_x(N_\varepsilon(K_2))\, dx \\ &\qquad + K \int_\Omega \int_{[N_\varepsilon(K_1) \cup N_\varepsilon(K_2)]^c} (1+|A|^p)\, d\nu_x(A)\, dx \\ &\geq -\delta \int_\Omega \nu_x(N_\varepsilon(K_2))\, dx \\ &\quad + \frac{K}{2} \int_\Omega \int_{[N_\varepsilon(K_1) \cup N_\varepsilon(K_2)]^c} (1+|A|^p)\, d\nu_x(A)\, dx \\ &\quad + \frac{K}{2} \gamma \min\left(\int_\Omega \nu_x(N_\varepsilon(K_1))\, dx, \int_\Omega \nu_x(N_\varepsilon(K_2))\, dx \right). \end{aligned} \tag{5.19}$$

If $\int_\Omega \nu_x(N_\varepsilon(K_1))\, dx \leq \int_\Omega \nu_x(N_\varepsilon(K_2))\, dx$ then, since $Dy^*(x) \in N_\varepsilon(K_1)^{\mathrm{qc}}$, by Lemma 22 we have that

$$\int_\Omega \int_{[N_\varepsilon(K_1) \cup N_\varepsilon(K_2]^c} (1+|A|^p) d\nu_x(A)\, dx + \int_\Omega \nu_x(N_\varepsilon(K_1))\, dx \geq \Delta \mathcal{L}^n(\Omega),$$

and hence by (5.18), (5.19)

$$I(\nu) - I(\nu^*) \geq -\delta \int_\Omega \nu_x(N_\varepsilon(K_2))\, dx + \frac{K}{2} \min(1,\gamma) \Delta \mathcal{L}^n(\Omega) > 0. \tag{5.20}$$

On the other hand if $\int_\Omega \nu_x(N_\varepsilon(K_2))\,dx \le \int_\Omega \nu_x(N_\varepsilon(K_1))\,dx$ then

$$\begin{aligned} I(\nu) - I(\nu^*) \ge (\frac{K}{2}\gamma - \delta) \int_\Omega \nu_x(N_\varepsilon(K_2)\,dx \\ +\frac{K}{2}\int_\Omega \int_{[N_\varepsilon(K_1)\cup N_\varepsilon(K_2)]^c} (1+|A|^p)\,d\nu_x(A)\,dx \ge 0. \quad (5.21) \end{aligned}$$

From (5.20),(5.21) we see that $I(\nu) = I(\nu^*)$ if and only if $\operatorname{supp}\nu_x \subset \{A \in N_\varepsilon(K_1) : W(A) = 0\}$, completing the proof.

6 Applications

In this section we discuss the application of the results given above to materials that undergo diffusionless phase transformations involving a change of shape, usually called martensitic phase transformations.

6.1 Variant rearrangement under biaxial stress

The original motivation for this paper were experiments of Chu & James on the response of single crystal plates of martensitic material to biaxial stress. The experimental details are presented elsewhere [23,24]. In the design of these experiments attention was paid to the design of the loading device so as to correspond to the total free energy

$$E(y) = \int_\Omega \varphi(Dy(x),\theta) - T\cdot Dy(x)\,dx, \quad (6.1)$$

where $y : \Omega \to \mathbb{R}^3$, Ω is a thin rectangular plate-like domain in $\mathbb{R}^3$, $\theta > 0$ is the temperature, and $T = \sigma_1 e_1\otimes e_1 + \sigma_2 e_2 \otimes e_2,\ \sigma_1 > 0, \sigma_2 > 0$ with $e_1, e_2 \in \mathbb{R}^3$ (the orthonormal "machine basis"). In the experiments described here the temperature was held fixed at a value θ_0 below the phase transformation temperature. For this reason we henceforth drop θ from the notation. The assigned $\sigma_1 > 0, \sigma_2 > 0$ are interpreted as the tractions (per unit reference area) applied to the edges of the specimen in the directions e_1, e_2, respectively. These were varied either incrementally or continuously during the tests. The material was the alloy Cu-14wt.%Al-4.0wt.%Ni having a cubic-to-orthorhombic phase transformation, leading to six variants of martensite at the test temperature. These are modeled as energy wells of φ of the form

$$\varphi(A) \ge 0, \quad \varphi(A) = 0 \iff A \in \mathcal{M} = SO(3)U_1 \cup \cdots \cup SO(3)U_6. \quad (6.2)$$

with

$$U_1 = \begin{pmatrix} \frac{\alpha+\gamma}{2} & \frac{\alpha-\gamma}{2} & 0 \\ \frac{\alpha-\gamma}{2} & \frac{\alpha+\gamma}{2} & 0 \\ 0 & 0 & \beta \end{pmatrix}, \quad U_2 = \begin{pmatrix} \frac{\alpha+\gamma}{2} & \frac{\gamma-\alpha}{2} & 0 \\ \frac{\gamma-\alpha}{2} & \frac{\alpha+\gamma}{2} & 0 \\ 0 & 0 & \beta \end{pmatrix},$$
$$U_3 = \begin{pmatrix} \frac{\alpha+\gamma}{2} & 0 & \frac{\alpha-\gamma}{2} \\ 0 & \beta & 0 \\ \frac{\alpha-\gamma}{2} & 0 & \frac{\alpha+\gamma}{2} \end{pmatrix}, \quad U_4 = \begin{pmatrix} \frac{\alpha+\gamma}{2} & 0 & \frac{\gamma-\alpha}{2} \\ 0 & \beta & 0 \\ \frac{\gamma-\alpha}{2} & 0 & \frac{\alpha+\gamma}{2} \end{pmatrix}, \tag{6.3}$$
$$U_5 = \begin{pmatrix} \beta & 0 & 0 \\ 0 & \frac{\alpha+\gamma}{2} & \frac{\alpha-\gamma}{2} \\ 0 & \frac{\alpha-\gamma}{2} & \frac{\alpha+\gamma}{2} \end{pmatrix}, \quad U_6 = \begin{pmatrix} \beta & 0 & 0 \\ 0 & \frac{\alpha+\gamma}{2} & \frac{\gamma-\alpha}{2} \\ 0 & \frac{\gamma-\alpha}{2} & \frac{\alpha+\gamma}{2} \end{pmatrix},$$

all expressed in an orthonormal basis $\hat{e}_1, \hat{e}_2, \hat{e}_3$ (the "material basis"). The measured values of α, β, γ for this alloy are $\alpha = 1.0619$, $\beta = 0.9178$ and $\gamma = 1.0230$ (Duggin and Rachinger [31], Otsuka and Shimizu [57]). The deviation of the material basis from the machine basis measures the orientation of the specimen. Several orientations were tested.

For many purposes, including the design of the orientations of crystals used in the tests, a simpler *constrained theory* was used, valid in the regime that $|T|/\kappa$ is small[1], κ being the minimum eigenvalue of the linearized elasticity tensor, linearized about U_1. The constrained theory is based on the total free energy

$$E(\nu) = \begin{cases} -\int_\Omega \int_{\mathcal{M}} T \cdot A \, d\nu_x(A) dx, & \text{if } \operatorname{supp} \nu_x \subset \mathcal{M} \text{ a.e. } x \in \Omega \\ +\infty, & \text{otherwise,} \end{cases} \tag{6.4}$$

defined on the set of L^∞ gradient Young measures $\nu = (\nu_x)_{x\in\Omega}$. The constrained theory has been justified as a limiting theory for Young measures of low energy sequences by Forclaz [34] using Γ-convergence, but under assumptions not allowing $W(A) \to \infty$ as $\det A \to 0+$; the proof is based on replacing φ by $k\varphi$ in (6.1) and letting $k \to \infty$ (a similar procedure to letting $|T|/\kappa \to 0$ but which does not require additional smoothness assumptions on φ). A more general Γ-convergence analysis including the austenite energy well and allowing $W(A) \to \infty$ as $\det A \to 0+$ is given by [15, Proposition 1].

The design of orientations was based on the minimization of (6.4), which can be done in the following way by minimizing its integrand (see Chu [23], Chu & James [24]). The machine basis was chosen in all cases such that, for all values of $\sigma_1 > 0, \sigma_2 > 0$,

$$\min_{A\in SO(3)U_1\cup SO(3)U_2} -T\cdot A < \min_{A\in SO(3)U_3\cup\cdots\cup SO(3)U_6} -T\cdot A. \tag{6.5}$$

In fact, the minimizer is unique for all points in this open quadrant, except those on a smooth, strictly monotonically increasing curve $\mathcal{C} : \sigma_2 = f(\sigma_1)$,

[1] Using measured moduli of Yasunaga et al. [73,74] for this alloy gives $\kappa \sim 15$ GPa. A typical maximum value of $|T|$ in the tests was 15 MPa, yielding $|T|/\kappa \sim 15$ MPa/15 GPa $= 10^{-3}$.

$f \in C^\infty(0,\infty)$, which is nearly a straight line in the range of σ_1, σ_2 tested. In fact, there exist functions $R_i \in C^\infty((0,\infty)\times(0,\infty); SO(3))$, $i = 1,2$, such that $A = R_1(\sigma_1,\sigma_2)U_1$ is the unique minimizer of $-T\cdot A$, $A \in \mathcal{M}$, for $\sigma_2 < f(\sigma_1)$ and $A = R_2(\sigma_1,\sigma_2)U_2$ is its unique minimizer on $\mathcal{M}$ for $\sigma_2 > f(\sigma_1)$. The functions R_1, R_2 can and will be taken as the unique minimizers of $-T\cdot A$ on their respective energy wells $SO(3)U_1$, $SO(3)U_2$ on the full quadrant $\sigma_1 > 0, \sigma_2 > 0$. There are precisely two equi-minimizers of $-T\cdot A$, $A \in \mathcal{M}$, on $\mathcal{C}$ given by $R_1(\sigma_1, f(\sigma_1))U_1$ and $R_2(\sigma_1, f(\sigma_1))U_2$. The tests consisted of crossing the curve $\sigma_2 = f(\sigma_1)$ by various loading programs $(\sigma_1(t), \sigma_2(t)), t > 0$, and measuring the volume fractions of the subregions on the specimen where $Dy \in SO(3)U_1$ (variant 1) vs. $Dy \in SO(3)U_2$ (variant 2).

The key point for this paper is that, by direct calculation of the functions R_1, R_2,

$$\operatorname{rank}(R_2(\sigma_1, f(\sigma_1))U_2 - R_1(\sigma_1, f(\sigma_1))U_1) > 1 \tag{6.6}$$

for all $\sigma_1 > 0$ and all orientations tested. Thus, fixing $\sigma_1 = \sigma_1^\circ \in (0,\infty)$, we let $K_1 = \{R_1(\sigma_1^\circ, f(\sigma_1^\circ))U_1\}$, and $K_2 = \{R_2(\sigma_1^\circ, f(\sigma_1^\circ))U_2\}$. By Example 4, K_1 and K_2 are L^p incompatible for $p > 1$. Letting $T_\tau = \sigma_1^\circ e_1 \otimes e_1 + (c_2\tau + f(\sigma_1^\circ))e_2 \otimes e_2$ and $R_1^\tau = R_1(\sigma_1^\circ, c_2\tau + f(\sigma_1^\circ))$ for some $c_2 > 0$, a suitable function W_τ satisfying the hypotheses of Proposition 20 for $m = n = 3$ can be defined as follows:

$$W_\tau(A) = \begin{cases} -T_\tau\cdot(A - R_1^\tau U_1), & A \in \mathcal{M}, \\ \infty, & A \in \mathcal{M}^c. \end{cases} \tag{6.7}$$

W_τ clearly satisfies (H1)$'$, (H2)$'$ and (H4)$'$, while (H3)$'$ is satisfied by choosing $c_2 > 0$ sufficiently small that $R_1^\tau U_1 \in N_\varepsilon(K_1)$ for $0 \le \tau \le 1$. The region occupied by the specimen was approximately a thin rectangular plate, so we assume Ω is a rectangular solid. In particular Ω is Ω-connected. The energy density W_τ differs from that of the constrained theory by a trivial additive constant. Theorem 21 then implies that the Young measure $\nu_\tau^* = \delta_{R_1^\tau U_1}$ is metastable for sufficiently small $\tau > 0$ in the sense given there.

In this formulation we have used σ_2 as the parameter that moves the wells up and down. One could equally well use a parameterization of any other curve that crosses $\mathcal{C}$ transversally.

Experimentally, transformation occurred by a sudden avalanche of transformation from variant 1 to variant 2 or *vice-versa*. The transformation was sufficiently abrupt that a point in the σ_1, σ_2 plane could be associated with the transformation. The series of points obtained in this way from diverse monotonic loading programmes, including those for which $\sigma_1(t) = const.$, or $\sigma_2(t) = const.$, or $\sigma_1(t) + \sigma_2(t) = const.$, all starting from a point $\sigma_1(0), \sigma_2(0)$ satisfying $\sigma_2(0) \ll f(\sigma_1(0))$, at which the specimen was observed to be in variant 1, gave abrupt transformation to variant 2 at points lying very near a line $\mathcal{C}^+ : \sigma_2 = f^+(\sigma_1) > f(\sigma_1)$, $0 < a < \sigma_1 < b$. Similarly, the same kinds of loading programmes but run backwards, beginning from variant 2, led to transformation to variant 1 near a line $\mathcal{C}^- : \sigma_2 = f^-(\sigma_1) < f(\sigma_1)$, $0 < a < \sigma_1 < b$.

For all orientations tested, the three curves $\mathcal{C}, \mathcal{C}^+, \mathcal{C}^-$ were nearly parallel, but the "width of the hysteresis", $\operatorname{dist}(\mathcal{C}^+, \mathcal{C}^-)$, varied significantly with orientation.

The concept developed in this paper is consistent with the behaviour described above. We can examine this further by seeking an upper bound on the value of τ in (6.7) beyond which $\nu_\tau^* = \delta_{R_1^\tau U_1}$ ceases to be metastable in the sense of Theorem 21. As $\tau > 0$ increases, there are more and more matrices $A \in SO(3)U_2$ with a negative value of the integrand $W_\tau(A)$. Suppose a value τ^+ is reached such that for $\tau \gtrsim \tau^+$, that is $\tau \geq \tau^+$ with $\tau - \tau^+$ sufficiently small, there is a matrix $B \in SO(3)U_2$ with $\operatorname{rank}(B - R_1^\tau U_1) = 1$, such that $W_\tau(B) < W_\tau(R_1^\tau U_1)$. Then $\nu_\tau^* = \delta_{R_1^\tau U_1}$ ceases to be metastable in the sense of Theorem 21. In fact, it fails to be metastable even if L^1 in (5.7) is replaced by L^∞. In the case that $B - R_1^{\tau_1} U_1 = a \otimes n$, $\tau_1 \gtrsim \tau^+$, the counterexample is the family of competitors $\nu_x = \delta_{Dy_\xi(x)}$, $\xi > 0$, defined for $x_0 \in \Omega$ by the $W^{1,\infty}(\Omega, \mathbb{R}^3)$ mapping

$$y_\xi(x) = \begin{cases} R_1^{\tau_1} U_1(x - x_0), & \{x \in \Omega : (x - x_0) \cdot n < 0\}, \\ B(x - x_0), & \{x \in \Omega : 0 \leq (x - x_0) \cdot n \leq \xi\}, \\ R_1^{\tau_1} U_1(x - x_0) + \xi a, & \{x \in \Omega : (x - x_0) \cdot n > \xi\}. \end{cases}$$

Since $\|y_\xi - R_1^{\tau_1} U_1(x - x_0)\|_{L^1(\Omega, \mathbb{R}^3)} \leq C\xi|a|$ for a constant $C = C(\Omega)$, then ν can be made to fall into any preassigned neighbourhood of ν_τ^* in the sense of (5.7) of Theorem 21 by making ξ sufficiently small, and this competitor also works in the L^∞ case. But clearly, since $W_\tau(B) < W_\tau(R_1^\tau U_1)$ we have that $E(\nu) < E(\nu_\tau^*)$, so ν_τ^* is not metastable for $\tau \gtrsim \tau^+$.

This qualitative argument for the sequence stable-metastable-unstable as τ increases, in the sense discussed here, is complete if we can show that there exists B with the properties given above. This is true by direct calculation for all the orientations tested. This is done by first calculating explicitly $R_1^\tau U_1$, and then noticing that the wells $SO(3)U_1$ and $SO(3)U_2$ are compatible. That is, even though (6.6) holds, there are precisely two matrices $\hat{R}_a^\tau U_2, \hat{R}_b^\tau U_2 \in SO(3)U_2$ that differ from $R_1^\tau U_1$ by a matrix of rank 1 for $\tau > 0$, and there exists a smallest value $\tau^+ > 0$ such that for $\tau > \tau^+$, $W_\tau(B) < W_\tau(R_1^\tau U_1)$ where B is either $\hat{R}_a^\tau U_2$ or $\hat{R}_b^\tau U_2$.

Unless the orientation is special, the two matrices $\hat{R}_a^\tau U_2$ or $\hat{R}_b^\tau U_2$ do not give the same value of W_τ, suggesting a preference for one of them, assuming that these examples deliver the point of first loss of metastability. Let us suppose for definiteness that the preference is for $\hat{R}_a^\tau U_2$, so $\hat{R}_a^\tau U_2 - R_1^\tau U_1 = a_\tau \otimes n_\tau$ and $W_\tau(\hat{R}_a^\tau U_2) \leq W_\tau(R_1^\tau U_1)$ for $\tau \geq \tau^+$ with equality precisely at $\tau = \tau^+$. Combining these two conditions, we have

$$a_{\tau^+} \cdot T_{\tau^+} n_{\tau^+} = 0. \tag{6.8}$$

This is formally equivalent to the well-known Schmid law (with Schmid constant 0) [62]. The left hand side of (6.8) is usually interpreted as the "critical resolved stress on the twin plane", but in that case Tn is interpreted as the actual Piola-Kirchhoff traction on a pre-existing twin plane with unit normal

n and $a = (F^+ - F^-)n$, where $F^\pm$ are local limiting values of the deformation gradient. The Schmid law prescribes a critical value of $a \cdot Tn$ at which this plane begins to move. The emergence of (6.8) here has apparently nothing to do with stress in the specimen at all, which is expected to be extremely complicated once bands of the second variant appear, but rather concerns the loading device energy.

In fact, as discussed in [10] and [34], for a suitable C-connected domain Ω with corners, these simple counterexamples to metastability do not deliver the points of first loss of metastability. More complicated microstructures still in an L^∞ local neighborhood, which are not simply Dirac masses, serve as counterexamples to metastability at values of $\tau \in (\tau_1^+, \tau^+)$ for some $0 < \tau_1^+ < \tau^+$. The experimentally observed microstructure at transition (i.e., near $\mathcal{C}^+$) is still somewhat more complicated than these, and is clearly not a simple laminate. If we accept that the basis of the Schmid law is metastability as noted above, these more complicated examples call into question the validity of that law in this context and also indicate a dependence of hysteresis on the shape of the domain. The latter is also expected based on Example 11.

A detailed comparison of these upper bounds, either the one associated to τ^+ or to τ_1^+, with the experimentally measured width of the hysteresis is difficult. Experimentally, it is easiest to identify $\mathcal{C}^+$ with a possible loss of metastability, but the shoulder of the hysteresis loop is not perfectly sharp, and some bands appear before reaching $\mathcal{C}^+$, as τ is increased. Because of this ambiguity, it is unclear where one should declare that the homogeneous variant has begun to transform. However, the overall impression one gets when attempting this comparison is that the upper bounds associated to both τ^+ or to τ_1^+ underestimate the size of the hysteresis. Nevertheless there is rather good qualitative agreement, in the sense that, for two specimens of different orientation having widths of the hysteresis $\operatorname{dist}(\mathcal{C}^+, \mathcal{C}^-)$ differing by a factor of 2, the corresponding upper bounds for the two cases also differ by a factor of about 2.

6.2 Dilatational transformation strain

Martensitic transformations having a pure dilatational transformation strain are rare, but some examples are known in diffusional transformations, which involve shape change and short or long range diffusion, depending on the overall composition of the alloy. The best known example is perhaps the ordering transformation from a disordered FCC phase to an $L2_1$ phase in Ni_3Al [71], for which the ideas given above may be relevant.

As a general treatment of dilatational transformation strains, consider two compact disjoint subsets k_1, k_2 of $(0, \infty)$, and corresponding energy wells $K_1 = k_1 SO(3)$ and $K_2 = k_2 SO(3)$, where $k_i SO(3) = \{kSO(3) : k \in k_i\}$. That K_1 and K_2 are incompatible follows from [8, Theorem 4.4] and Lemma 1, and also follows from the construction below, as we will indicate.

We will construct a polyconvex function W_0 that vanishes exactly on $K_1 \cup K_2$. This construction will enable us to embed W_0 in a family $W_\tau, 0 \le \tau \le 1$, for which we will prove metastability in the sense of Theorem 21.

Following an observation of [16] (see also [4,6]), let $1 < \alpha < 3$ and let $\bar{h} : \mathbb{R} \to [0, \infty]$ be continuous with $\bar{h} = \infty$ on $(-\infty, 0]$, $\bar{h} \in C^2(0, \infty)$ and $\bar{h}^{-1}(0) = \{k^3 : k \in k_1 \cup k_2\}$. We assume that $\bar{h}$ is convex outside a compact subset $[a, b] \subset (0, \infty)$ containing $\bar{h}^{-1}(0)$, so that there exists $\gamma > 0$ such that $\bar{h}'' \ge -\gamma$ on $(0, \infty)$. Let a convex function $\tilde{h} \in C^2(\mathbb{R})$ satisfy

$$\tilde{h}(t) = \begin{cases} -3c_1 t^{\alpha/3}, & a < t < b, \\ -3c_1(b+1)^{\alpha/3}, & t > b+1. \end{cases} \tag{6.1}$$

Such a convex function exists because the tangent at $t = b$ to $-3c_1 t^{\alpha/3}$ lies below the constant function $-3c_1(b+1)^{\alpha/3}$ at $t = b+1$.

Define $h(t) = \bar{h}(t) + \tilde{h}(t)$. Since $\bar{h}'' \ge -\gamma$, $1 < \alpha < 3$, and $\bar{h}$ is convex outside $[a, b]$, there is $c_1 > 0$ such that

$$h''(t) = \bar{h}''(t) + \frac{1}{3} c_1 \alpha (3 - \alpha) t^{-2+\alpha/3} > 0 \tag{6.2}$$

on $[a, b]$ and so h is convex on $\mathbb{R}$ and bounded below by $c_0 = -3c_1(b+1)^{\alpha/3}$.

Define an energy density for an isotropic elastic material by

$$W_0(A) = c_1(\lambda_1^\alpha + \lambda_2^\alpha + \lambda_3^\alpha) + h(\lambda_1\lambda_2\lambda_3), \tag{6.3}$$

where $\lambda_1, \lambda_2, \lambda_3$ are the eigenvalues of $\sqrt{A^T A}$. Because h is convex and $1 < \alpha < 3$, W_0 is polyconvex by [5, Theorem 5.1].

Now we observe that W_0 has strict minima on $K_1 \cup K_2$. Indeed, since h is bounded below and $h(0) = \infty$, the function $\sum_i c_1 \lambda_i^\alpha + h(\lambda_1\lambda_2\lambda_3)$ attains a minimum for $\lambda_1 > 0, \lambda_2 > 0, \lambda_3 > 0$, where

$$c_1 \alpha \lambda_i^\alpha = -h'(\lambda_1\lambda_2\lambda_3)\lambda_1\lambda_2\lambda_3. \tag{6.4}$$

Hence $\lambda_1 = \lambda_2 = \lambda_3 = t^{1/3}$, where $c_1 \alpha t^{\alpha/3} = -h'(t)t$. These values of t are critical points of the function $\bar{h}(t) = 3c_1 t^{\alpha/3} + h(t) = W_0(t^{1/3} I)$, which has minimizers precisely on the set $\bar{h}^{-1}(0)$ by construction. Hence, $W_0(A)$ has minimizers precisely on $K_1 \cup K_2$, where $W_0(A) = 0$.

Since h is bounded below by c_0, the energy density W_0 satisfies the growth condition

$$W_0(A) = c_1(\lambda_1^\alpha + \lambda_2^\alpha + \lambda_3^\alpha) + h(\lambda_1\lambda_2\lambda_3) \ge c_0 + c_1|A|^\alpha, \tag{6.5}$$

so that W_0 satisfies conditions (H1) and (H2) of Section 5 for $p = \alpha$.

To show that K_1, K_2 are incompatible we can consider the special case $\alpha = 2$, when

$$W_0(A) = c_1|A|^2 + h(\det A).$$

If $\nu = (\nu_x)_{x \in \Omega}$ is an L^∞ gradient Young measure with $\operatorname{supp} \nu_x \subset K_1 \cup K_2$ a.e., we have that

$$0 = \langle \nu_x, W_0 \rangle = c_1 \langle \nu_x, |A|^2 \rangle + \langle \nu_x, h(\det A) \rangle.$$

Applying Jensen's inequality for the quasiconvex functions $|A|^2$ and $h(\det A)$, we have that

$$\langle \nu_x, |A|^2 \rangle \geq |\bar{\nu}_x|^2, \ \ \langle \nu_x, h(\det A) \rangle \geq h(\det \bar{\nu}_x).$$

But $c_1|\bar{\nu}_x|^2 + h(\det \bar{\nu}_x) = W_0(\bar{\nu}_x) \geq 0$. Hence $\langle \nu_x, |A|^2 \rangle = |\bar{\nu}_x|^2$, so that $\langle \nu_x, |A - \bar{\nu}_x|^2 \rangle = 0$ and hence $\nu_x = \delta_{Dy(x)}$ with $\bar{\nu}_x = Dy(x)$. But $Dy(x) \in K_1 \cup K_2$ a.e., so that y is a $W^{1,\infty}$ conformal mapping in 3 dimensions. By classic results of Reshetnyak [60] all such mappings are smooth and therefore Dy cannot be supported nontrivially on disjoint closed sets. Thus, K_1, K_2 are incompatible.

The energy density W_0 can easily be extended to a family W_τ satisfying the hypotheses (H1)$'$-(H4)$'$ of Proposition 20. Let ε_0, γ be as in the transition layer estimate (Theorem 17) and let $0 < \varepsilon < \varepsilon_0$ be fixed. Since $N_\varepsilon(K_1)$ and $N_\varepsilon(K_2)$ are disjoint, we can let

$$W_\tau(A) = W_0(A) - \tau H(\det A), \tag{6.6}$$

where $H : \mathbb{R} \to [0,1]$ is a smooth function satisfying

$$H(t) = \begin{cases} 1, \text{ if } t \in \{k^3 : k \in k_2\} := N_2 \\ 0, \text{ if } \operatorname{dist}(t, N_2) > \rho(\varepsilon), \end{cases}$$

where $\rho(\varepsilon) > 0$ is sufficiently small. Clearly, W_τ satisfies the hypotheses (H1)$'$-(H4)$'$ with $p = \alpha$. Therefore, any L^p gradient Young measure $\nu^* = (\nu^*_x)_{x \in \Omega}$ satisfying $\operatorname{supp} \nu^*_x \subset \{A \in N_\varepsilon(K_1) : W(A) = 0\}$ is metastable in the sense of Theorem 21 for sufficiently small $\tau > 0$, even though $W_\tau(A) = 0$ for $A \in K_1$ and $W_\tau(A) = -\tau$ for $A \in K_2$. In [16, Theorem 3.5] it is shown that such a result for free-energy functions of the form (6.3) is not valid if the second energy well is arbitrarily deep.

Depending on the structure of K_1 the form of these metastable Young measures is strongly restricted by Reshetnyak's theorem, but $\nu^*_x = \delta_{Dy^*(x)}$, where y^* is a conformal mapping, is a possibility.

Although it is interesting that pure dilatational phase transformations can be described by polyconvex free-energy functions, the functions W_τ also serve as lower bounds for free-energy functions for which metastability in the sense of Theorem 21 also holds. For example, by multiplying through the metastability estimate by a sufficiently small positive constant, W_τ can be a lower bound for a variety of non-polyconvex energy densities, with various choices of positive-definite linear elastic moduli. Of course, this modification also decreases γ, including the largest value of γ for which there is an $\varepsilon_0 > 0$ satisfying the metastability theorem. In this sense, softening a material, but keeping the wells the same, lowers the barrier for metastabilty.

6.3 Terephthalic acid

Terephthalic acid [27,3] is an interesting example in this context, since, among all reversible structural transformations, it has an exceptionally large transformation strain. It is the largest strain in a nominally reversible transformation in terms of dist (K_1, K_2) of which we are aware in a material that has no rank-one connections between K_1 and K_2, that is, no solutions $A, B \in M^{3\times3}$ of $\mathrm{rank}(B - A) = 1$, $A \in K_1, B \in K_2$. The clearly visible large change-of-shape shown by Davey et al. [27] is remarkable.

Terephthalic acid undergoes the transformation from Form I to Form II between 80°C and 100°C [27]. The transformation is reversible upon cooling to 30°C, at least for a subset of crystallites; the application of a slight stress aids the reverse transformation. The crystal structure and lattice parameter measurements of the I-II transformation have been determined by Bailey et al. [3]. Knowledge of these two structures and lattice parameters does not imply a unique transformation stretch matrix due to the existence of infinitely many linear transformations that take a lattice to itself. The transformation stretch matrix

$$U = \begin{pmatrix} 0.970 & 0.038 & -0.121 \\ 0.038 & 0.835 & -0.017 \\ -0.121 & -0.017 & 1.298 \end{pmatrix}, \tag{6.1}$$

is the one delivered by an algorithm [21] designed to give the smallest distortion measured by an appropriate norm. The associated lattice correspondence of the two phases (i.e., which vector is transformed to which vector) agrees with descriptions of the transformation [3] and, semi-quantitatively, with photographs of crystals of the two phases [27]. The eigenvalues of U are 1.339, 0.939, 0.825. Nominally, there are two wells $K_1 = SO(3), K_2 = SO(3)U$. In fact twinning is observed in the Form I, but this appears to be growth twinning [27], and not produced during transformation. (Both phases are triclinic, so there is no lowering of symmetry during transformation.) Since the middle eigenvalue of U is not 1, there are no rank-one connections between K_1 and K_2 [13].

The best sufficient conditions known that two wells K_1 and K_2 of this form are incompatible are due to Dolzmann, Kirchheim, Müller & Šverák [30]. Condition (ii) of their Theorem 1.2 is satisfied by U. Therefore, K_1 and K_2 are incompatible, and our metastability theorem applies to this case.

7 Perspective on metastability and hysteresis

In recent years different but related concepts of metastability have appeared in the literature [38,25,78,44,28,20,45,79] motivated by some experimental results on a dramatic lattice parameter dependence of the sizes of hysteresis loops. These observations call for new mathematical concepts of metastability whose form is not at all clear.

Typical martensitic materials have energy wells of the form $K_1 = SO(3)$ and $K_2 = SO(3)U_1 \cup \cdots \cup SO(3)U_n$, with $n \geq 1$, and positive-definite, symmetric matrices $U_1, \ldots, U_n \in M^{3\times3}$ satisfying $\{U_1, \ldots, U_n\} = \{QU_1Q^T : Q \in G\}$, where G is a finite group of orthogonal matrices (cf., (6.3)). Modulo the comments in Section 6.3 on the difficulties of determining the transformation stretch matrix, U_1 for a particular material can be inferred from X-ray measurements. All first order martensitic phase transformations have some amount of thermal hysteresis, which refers to the fact that the transformation path on cooling differs from that on heating. A measurement of the fraction of the sample that has transformed vs. temperature during a heating/cooling cycle gives a loop, called the hysteresis loop, whose width is a typical measure of the hysteresis. While indicative of dissipation, the hysteresis loop does not collapse to zero as the loop is traversed more and more slowly, and so is apparently not due to thermally activated processes, or dissipative mechanisms like viscosity or viscoelasticity.

The matrix U_1 can be changed by changing the composition of the material. Suppose the ordered eigenvalues of U_1 are $\lambda_1 \leq \lambda_2 \leq \lambda_3$. The main experimental observation underlying the analysis of hysteresis in the papers listed above is that, if a family of alloys is prepared having a sequence of values of λ_2 approaching 1, the hysteresis gets dramatically small. Experimental graphs [25] of hysteresis vs. λ_2 show an apparent cusp-like singularity at $\lambda_2 = 1$, i.e., an extreme sensitivity of the size of the hysteresis to $|\lambda_2 - 1|$. Very careful changes of composition in increments of 1/4 % lead to alloys with exceptionally low hysteresis of $2-3°$C in a variety of systems [20, 75]. Since $\lambda_2 = 1$ is a necessary and sufficient condition that there is a rank-one connection between K_1 and K_2, these results indicate that the removal of stressed transition layers by strengthening conditions of compatibility is relevant to hysteresis.

A strict application of the ideas in this paper does not explain this behaviour. That is because, in all of these cases that have been studied experimentally, K_1 and K_2 are compatible even in the starting alloys for which λ_2 is relatively far from 1. In fact, all of these cases support solutions of the crystallographic theory of martensite [72, 13], implying that there exist $A, B \in K_2$ and $C \in K_1$, such that $\operatorname{rank}(B - A) = 1$ and $\operatorname{rank}(\lambda B + (1 - \lambda)A - C) = 1$ for some $0 < \lambda < 1$. This series of rank one connections implies the existence of a Young measure $(\nu_x)_{x \in \Omega}$ supported nontrivially on K_1, K_2, consisting of a laminate of two martensite variants $\ldots A/B/A/B \ldots$ meeting the austenite C phase across a vanishingly small planar transition layer. In fact, the laminated martensite can be confined between two such parallel planes which can be arbitrarily close together (see [11] for details). This family of test measures then provide a counterexample to the metastability of say $\nu^* = \delta_C$ in the sense of Theorem 21, even if L^1 in (5.7) is replaced by L^∞.

A special family of test functions y_ε of the type just described - a laminate $\ldots A/B/A/B \ldots$ confined between parallel planes at the distance ε and interpolated with C in a layer near these planes – can be constructed explicitly. Its energy can then be calculated by using a bulk energy of the type studied in this paper with a suitable elastic energy density W_τ, together with a interfacial energy per unit area (taken as constant) on the A/B boundaries. In this case $-\tau$ is interpreted as the temperature and $\tau = 0$ is the transformation temperature. This has been done in [78] and improved by Zwicknagl [79]. A graph of total energy vs. ε gives a barrier whose height is very sensitive to $|\lambda_2 - 1|$, and decreases with decreasing temperature $-\tau$. If a critical value $\varepsilon = \varepsilon_{\rm crit}$ is introduced (modelling a pre-existing martensite nucleus of this type), and the temperature $\theta_c = -\tau$ is calculated at which $\varepsilon = \varepsilon_{\rm crit}$, then the resulting graph of $0 - \theta_c$ vs. λ_2, all else fixed, has a singularity at $\lambda_2 = 1$ and a shape similar to the experimental graph of hysteresis vs. λ_2.

A related idea for a geometrically linear theory of the cubic-to-tetragonal transformation and a sharp interface model of interfacial energy is presented by Knüpfer, Kohn & Otto [45] (see also [44]). They show that the minimal bulk + interfacial energy of an inclusion of martensite of volume V scales as the maximum of $V^{2/3}, V^{9/11}$. Minimal assumptions are made on the shape of the inclusion. If a bulk term is added to this energy of the form $-c\tau V$, $c > 0$, modelling a lowering of the martensite wells as the temperature $-\tau$ is decreased below transformation temperature, then their result gives an energy barrier of the type described above. They note that it would be interesting to do a similar analysis of an austenite inclusion in martensite, and they conjecture a higher energy barrier for the reverse transformation. This is open, as is a similar analysis for the cubic-to-orthorhombic case, where it would be interesting to investigate the dependence of the predicted barrier on λ_2.

Recently, even stronger conditions of compatibility called the *cofactor conditions* [38, 20] have been closely satisfied in the ZnCuAu system by compositional changes, leading to the alloy $Zn_{45}Au_{30}Cu_{25}$. The cofactor conditions imply not only $\lambda_2 = 1$ but also a variety of other microstructures with zero elastic energy. The alloy $Zn_{45}Au_{30}Cu_{25}$ has a transformation strain $|U - I|$ comparable to that of the alloys tuned to satisfy only $\lambda_2 = 1$, but shows still smaller hysteresis than the lowest achieved by the $\lambda_2 = 1$ alloys, and also exceptional reversibility [63]. This example may indicate that metastability in phase transformations is not only sensitive to the wells being gradient compatible, but also to the presence of a variety of different functions whose gradients are nontrivially supported on K_1, K_2. Another possibly relevant hypothesis is that metastability is influenced by a possible sudden increase of the size of the quasiconvex hull of the energy wells when the cofactor conditions are satisfied.

An apparently obvious reconciliation of these concepts is to retain the idea of metastability, quantified by local minimization, but to include a contribution for interfacial energy. Accepted models of this type fall into two classes: sharp interface models and gradient models. However, when combined with accepted notions of local minimization, neither of these models give the behaviour described above. Before commenting on these two cases, we first note that concepts of linearized stability are not relevant: most measured values of linearized elastic moduli do soften as temperature is lowered to the phase transformation temperature, but the limiting value of the minimum eigenvalue of the elasticity tensor is clearly positive at transition in most cases, and this is the rule for strongly first order phase transformations.

A typical sharp interface model assigns an energy per unit area to the jump set of Dy. A comparative discussion of the energy minimisation problem for several versions of these models is discussed in [17]. Consider the simple but relevant case of deciding whether a linear deformation $y^*(x) = Ax, x \in \Omega$, is metastable in some sense, where $A \in K_1$, $W_\tau(A) = 0$ and $W_\tau(K_2) = -\tau$, with K_1 and K_2 independent of τ. Suppose we have favoured the low hysteresis situation by tuning the material as described above so that there exists $B \in K_2$ such that $B - A = a \otimes n$. Putting aside linearized stability, relevant concepts of local minimizer have the property that competitors can have gradients on or near K_2, at least on sufficiently small sets. Trivially, if the underlying function space allows us to smooth jumps of Dy, then a mollified version of the continuous function given for $x_0 \in \Omega$ by

$$y_\varepsilon(x) = \begin{cases} B(x - x_0), & \text{if } 0 < (x - x_0) \cdot n < \varepsilon, \\ A(x - x_0), & \text{otherwise}, \end{cases} \tag{7.1}$$

defeats metastability in L^∞ as soon as $\tau > 0$, predicting zero hysteresis. Thus, of course, we have to prevent smoothing. This is easily done by forcing a jump, by restricting the domain of W_τ to, say, $N_\varepsilon(K_1) \cup N_\varepsilon(K_2)$ with ε sufficiently small. However, in that case, the prototypical test function (7.1) for ε sufficiently small has positive energy regardless how big is the value of τ. Thus, apparently for any of the accepted notions of local minimizer, infinite hysteresis is predicted.

This dominance of interfacial energy at small scales, which overstabilizes linear deformations, also occurs when gradient models of interfacial energy are combined with the bulk energies studied here, as shown in [11]. Consider a frame-indifferent energy density $W_\tau \in C^2(M_+^{3\times 3})$, continuous in τ and satisfying $W_\tau(A) \to \infty$ as $\det A \to 0$, and having positive-definite linearized elasticity tensor at I. Suppose $W_\tau(K_1) = 0$ and $W_\tau(K_2) = -\tau$, for disjoint sets $K_1 = SO(3)$ and $K_2 = SO(3)U_1 \cup \cdots \cup SO(3)U_n$, and assume a total energy of the form

$$I(y) = \int_\Omega W_\tau(Dy) + \alpha |D^2 y|^2 \, dx \tag{7.2}$$

with $\alpha > 0$. In [11] it is shown that $y^*(x) = Rx + c, R \in SO(3), c \in \mathbb{R}^3$ is a local minimizer of I in L^1 *for every* $\tau > 0$. Again, infinite hysteresis is predicted. Note that there may or may not be rank-one connections between K_1 and K_2. It is probable that the the model introduced in [17], that includes contributions from both sharp and diffuse interfacial energies, also leads to a metastability result similar to that in [11], though this has not been checked.

This inevitability of either zero hysteresis or infinite hysteresis, or, in the case of linearized stability, predicted hysteresis that is too large, is avoided in models with interfacial energy if, instead of using the standard approach to local minimization, one uses a fixed neighbourhood of the proposed metastable deformation y^*, e.g., $\| y - y^* \|_{L^1} \leq \varepsilon_{\text{crit}}$. This is similar in spirit to the introduction of the critical nucleus size above (also called $\varepsilon_{\text{crit}}$). While this ultimately requires the formulation of an additional theory to predict $\varepsilon_{\text{crit}}$, it would nevertheless be interesting to know whether this approach is consistent with the observed lattice parameter dependence of hysteresis, as mentioned above.

Exotic models of interfacial energy that decrease the interfacial energy contribution when two interfaces get close together could also restore finite hysteresis. These are not widely accepted.

A better accepted idea, that is related to the introduction of the fixed neighbourhood using $\varepsilon_{\rm crit}$, is that, above transformation temperature, there are a variety of small nuclei of martensite, stabilized by defects, waiting to grow, and there are similar islands of austenite below transformation temperature. While this consistent with the (usually mild) dependence of hysteresis on preliminary processing, it is puzzling how this could yield hysteresis that is observed to be quite reproducible from alloy to alloy, given similar processing. However, such thinking is based on the idea of a single "most dangerous" nucleus determining transformation. If, on the other hand, macroscopic transformation arises from a collective interaction among many defects, so that something like the law of large numbers is applicable, then one can imagine a reproducible size of the hysteresis. This kind of collective nucleation around defects, modelled by a position dependent dissipation rate, can be seen in the recent numerical simulations of DeSimone & Kružík [29].

Once metastability is lost, complex dissipative dynamic processes take place, involving interface motion, microstructural evolution, and creation and annihilation of microstructure. There is currently insufficient information to formulate such dynamic laws, and the mathematical theory in general of the dynamics of microstructure is primitive. There are a number of known possible approaches, including constitutive modelling, the sharp interface kinetics of Abeyaratne & Knowles [1] and the method of quasistatic evolution of Mielke & Theil [53]. All of these are reasonable based on general principles, but the latter seems to be the only one at present that can deal with sufficient complexity of microstructure to begin to contemplate faithful dynamic predictions [29]. It is not yet known if these would be consistent with the sensitivity to conditions of compatibility mentioned above.

The surprising influence of conditions like $\lambda_2 = 1$ suggest that simple kinematic approaches are valuable. Their simplicity lies in the observation that the conditions for loss of metastabilty seem to be much simpler than the description of the dynamic process that takes place once metastability is lost. From the perspective of this paper, and the apparent success of the cofactor conditions, it would be interesting to have methods of quantifying the possibility of having many functions whose gradients are supported nontrivially on K_1 and K_2, especially those having finite area of the jump set of the gradient. A step in this direction is taken in recent work of Rüland [61].

Acknowledgements

The research of JMB was supported by by the EC (TMR contract FMRX - CT EU98-0229 and ERBSCI**CT000670), by EPSRC (GRlJ03466, the Science and Innovation award to the Oxford Centre for Nonlinear PDE EP/E035027/1, and EP/J014494/1), the European Research Council under the European

Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no 291053 and by a Royal Society Wolfson Research Merit Award. The research of RDJ was supported by NSF-PIRE (OISE-0967140) and the MURI project Managing the Mosaic of Microstructure (FA9550-12-1-0458, administered by AFOSR). We are especially grateful to Jan Kristensen and Vladimir Šverák for suggestions and advice, and to David Kinderlehrer, Bob Kohn, Pablo Pedregal and Gregory Seregin for their interest and helpful comments.

References

1. R. Abeyaratne and J. K. Knowles. *Evolution of Phase Transitions: A Continuum Theory.* Cambridge University Press, 2011.
2. R. Aumann and S. Hart. Bi-convexity and bi-martingales. *Israel J. Math.*, 54:159–180, 1986.
3. M. Bailey and C. J. Brown. The crystal structure of terephthalic acid. *Acta Crystallographica*, 22:387–391, 1967.
4. J. M. Ball. Constitutive inequalities and existence theorems in nonlinear elastostatics. In R. J. Knops, editor, *Nonlinear Analysis and Mechanics, Heriot-Watt Symposium, Vol. 1.* Pitman, 1977.
5. J. M. Ball. Convexity conditions and existence theorems in nonlinear elasticity. *Arch. Rat. Mech. Anal.*, 63:337–403, 1977.
6. J. M. Ball. Discontinuous equilibrium solutions and cavitation in nonlinear elasticity. *Phil. Trans. Royal Soc. London A*, 306:557–611, 1982.
7. J. M. Ball. A version of the fundamental theorem for Young measures. In M. Rascle, D. Serre, and M. Slemrod, editors, *Proceedings of conference on 'Partial differential equations and continuum models of phase transitions'*, pages 3–16. Springer Lecture Notes in Physics. No. 359, 1989.
8. J. M. Ball. Sets of gradients with no rank-one connections. *J. Math. Pures et Appliquèes*, 69:241–259, 1990.
9. J. M. Ball. Some open problems in elasticity. In *Geometry, Mechanics, and Dynamics*, pages 3–59. Springer, New York, 2002.
10. J. M. Ball, C. Chu, and R. D. James. Hysteresis during stress-induced variant rearrangement. *J. de Physique IV*, C8:245–251, 1995.
11. J. M. Ball and E. C. M. Crooks. Local minimizers and planar interfaces in a phase-transition model with interfacial energy. *Calc. Var. Partial Differential Equations*, 40(3-4):501–538, 2011.
12. J. M. Ball and R. D. James. Varying volume fractions of gradient Young measures. In preparation.
13. J. M. Ball and R. D. James. Fine phase mixtures as minimizers of energy. *Arch. Rat. Mech. Anal.*, 100:13–52, 1987.
14. J. M. Ball and R. D. James. Local minimizers and phase transformations. *Zeitschrift für Angewandte Mathematik und Mechanik*, 76 (Suppl. 2):389–392, 1996.
15. J. M. Ball and K. Koumatos. Quasiconvexity at the boundary and the nucleation of austenite. In preparation.
16. J. M. Ball and J. E. Marsden. Quasiconvexity at the boundary, positivity of the second variation, and elastic stability. *Arch. Rat. Mech. Anal.*, 86:251–277, 1984.
17. J. M. Ball and C. Mora-Corral. A variational model allowing both smooth and sharp phase boundaries in solids. *Communications on Pure and Applied Analysis*, 8:55–81, 2009.
18. J. M. Ball and F. Murat. $W^{1,p}$-quasiconvexity and variational problems for multiple integrals. *J. Functional Analysis*, 58:225–253, 1984.
19. K. Bhattacharya, N. B. Firoozye, R. D. James, and R. V. Kohn. Restrictions on microstructure. *Proc. Royal Soc. Edinburgh*, 124A:843–878, 1994.
20. X. Chen, Y. Song, V. Dabade, and R. D. James. Study of the cofactor conditions: Conditions of supercompatibility between phases. *Journal of the Mechanics and Physics of Solids*, 61:2566 – 2587, 2013.

21. X. Chen, Y. Song, and R. D. James. An algorithm for determining the transformation stretch tensor for structural phase transformations in complex lattices. preprint, 2014.
22. M. Chlebík and B. Kirchheim. Rigidity for the four gradient problem. *J. Reine Angew. Math.*, 551:1–9, 2002.
23. C. Chu. *Hysteresis and microstructure: a study of biaxial loading on compound twins of copper-aluminium-nickel single crystals*. PhD thesis, Department of Aerospace Engineering and Mechanics, University of Minnesota, 1993.
24. C. Chu and R. D. James. Biaxial loading experiments on Cu-Al-Ni single crystals. In *Experiments in Smart Materials and Structures*, pages 61–69. ASME, 1993. AMD-Vol. 181.
25. J. Cui, Y. S. Chu, O. Famodu, Y. Furuya, J. Hattrick-Simpers, R. D. James, A. Ludwig, S. Thienhaus, M. Wuttig, Z. Zhang, and I. Takeuchi. Combinatorial search of thermoelastic shape memory alloys with extremely small hysteresis width. *Nature Materials*, pages 286–290, 2006.
26. B. Dacorogna. *Direct methods in the calculus of variations*, volume 78 of *Applied Mathematical Sciences*. Springer, New York, second edition, 2008.
27. R. J. Davey, S. J. Maginn, S. J. Andrews, A. M. Buckley, D. Cottler, P. Dempsey, J. E. Rout, D. R. Stanley, and A. Taylor. Stabilization of a metastable phase by twinning. *Nature*, 366:248–250, 1993.
28. R. Delville, S. Kasinathan, Z. Zhang, V. Humbeeck, R. D. James, and D. Schryvers. A transmission electron microscopy study of phase compatibility in low hysteresis shape memory alloys. *Philosophical Magazine*, pages 177–195, 2010.
29. A. DeSimone and M. Kružík. Domain patterns and hysteresis in phase-transforming solids: analysis and numerical simulations of a sharp interface dissipative model via phase-field approximation. *Networks and Heterogeneous Media*, 8:481–489, 2013.
30. G. Dolzmann, B. Kirchheim, S. Müller, and V. Šverák. The two-well problem in three dimensions. *Calc. Var.*, 10:21–40, 2000.
31. M. J. Duggin and W. A. Rachinger. The nature of the martensitic transformation in a copper-nickel-aluminum alloy. *Acta metall.*, 12:529–535, 1964.
32. D. Faraco and L. Székelyhidi. Tartar's conjecture and localization of the quasiconvex hull in $\mathbb{R}^{2\times 2}$. *Acta Math.*, 200(2):279–305, 2008.
33. N. Firoozye. *Optimal Translations and Relaxations of some Multiwell Energies*. PhD thesis, Courant Institute, New York University, 1990.
34. A. Forclaz. Local minimizers and the Schmid law in corner-shaped domains. *Arch. Rational Mech. Anal.*, 211:555–591, 2014.
35. L. E. Fraenkel. On regularity of the boundary in the theory of Sobolev spaces. *Proc. London Math. Soc. (3)*, 39(3):385–427, 1979.
36. Y. Grabovsky and T. Mengesha. Direct approach to the problem of strong local minima in calculus of variations. *Calculus of Variations and PDE*, 29:59–83, 2007. Erratum: *ibid.* 32:407-408, 2008.
37. S. Heinz. On the structure of the quasiconvex hull in planar elasticity. *Calc. Var.*, 50:481–489, 2014.
38. R. D. James and Z. Zhang. A way to search for multiferroic materials with unlikely combinations of physical properties. In A. Planes, L. Manõsa, and A. Saxena, editors, *Magnetism and Structure in Functional Materials, Springer Series in Materials Science*, volume 9, pages 159–175. Springer-Verlag, Berlin, 2005.
39. D. Kinderlehrer and P. Pedregal. Characterizations of Young measures generated by gradients. *Arch. Rational Mech. Anal.*, 115:329–365, 1991.
40. D. Kinderlehrer and P. Pedregal. Gradient Young measures generated by sequences in Sobolev spaces. *J. Geom. Anal.*, 4:59–90, 1994.
41. B. Kirchheim. Deformations with finitely many gradients and stability of quasiconvex hulls. *C. R. Acad. Sci. Paris Sér. I Math.*, 332:289–294, 2001.
42. B. Kirchheim. *Rigidity and geometry of microstructures*. Habilitation, University of Leipzig, 2003.
43. B. Kirchheim and L. Székelyhidi, Jr. On the gradient set of Lipschitz maps. *J. reine angew. Math.*, 625:215–229, 2008.
44. H. Knüpfer and R. V. Kohn. Minimal energy for elastic inclusions. *Proc. Royal Soc. Lond. Ser. A: Math. Phys. Eng. Sci.*, 2127:695–717, 2011.

45. H. Knüpfer, R. V. Kohn, and F. Otto. Nucleation barriers for the cubic to tetragonal phase transformation. *Communications on Pure and Applied Mathematics*, 66:867–904, 2013.
46. R. V. Kohn, V. Lods, and A. Haraux. Some results about two incompatible elastic strains. Unpublished mansuscript, 2000.
47. R. V. Kohn and P. Sternberg. Local minimizers and singular perturbations. *Proc. Royal Soc. Edinburgh*, 111A:69–84, 1989.
48. J. Kristensen. *Lower Semicontinuity of Variational Integrals.* PhD thesis, Technical University of Lyngby, 1994.
49. K. Kuratowski and C. Ryll-Nardzewski. A general theorem on selectors. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.*, 13:397–403, 1965.
50. J. P. Matos. Young measures and the absence of fine microstructure in a class of phase transitions. *European J. Applied Maths*, 3:31–54, 1992.
51. V. Maz'ya. *Sobolev spaces with applications to elliptic partial differential equations*, volume 342 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, augmented edition, 2011.
52. E. J. McShane and T. A. Botts. *Real Analysis.* van Nostrand, Princeton, N.J., 1959. reprinted Dover, 2005.
53. A. Mielke and F. Theil. On rate-independent hysteresis models. *Nonlinear Differential Equations and Applications*, 11:151–189, 2004.
54. C. B. Morrey. *Multiple Integrals in the Calculus of Variations.* Springer, 1966.
55. S. Müller. A sharp version of Zhang's theorem on truncating sequences of gradients. *Trans. Amer. Math. Soc.*, 351(11):4585–4597, 1999.
56. S. Müller. Variational methods for microstructure and phase transitions. In *Calculus of variations and geometric evolution problems*, volume 1713 of *Lecture Notes in Math.*, pages 85–210. Springer, Berlin, 1999.
57. K. Otsuka and K. Shimizu. Morphology and crystallography of thermoelastic Cu-Al-Ni martensite analyzed by the phenomenological theory. *Trans. Japan Inst. Metals*, 15:103–108, 1974.
58. K. R. Parthasarathy. *Probability measures on metric spaces.* Probability and Mathematical Statistics, No. 3. Academic Press, Inc., New York-London, 1967.
59. P. Pedregal. Jensen's inequality in the calculus of variations. *Differential Integral Equations*, 7:57–72, 1994.
60. Y. G. Reshetnyak. Liouville's theorem on conformal mappings under minimal regularity assumptions. *Siberian Math. J.*, 8:631–653, 1967.
61. A. Rüland. The cubic-to-orthorhombic phase transition - rigidity and non-rigidity properties in the linear theory of elasticity. *Arch. Rational Mech. Anal.* submitted.
62. E. Schmid and W. Boas. *Plasticity of Crystals (translation of the 1935 text in German).* F. A. Hughes, London, 1950.
63. Y. Song, X. Chen, V. Dabade, T. W. Shield, and R. D. James. Enhanced reversibility and unusual microstructure of a phase-transforming material. *Nature*, 502:85–88, 2013.
64. E. M. Stein. *Singular Integrals and Differentiability Properties of Functions.* Princeton University Press, Princeton, New Jersey, 1970.
65. G. Strang. The width of a chair. *Amer. Math. Monthly*, 89:529–534, 1982.
66. V. Šverák. On regularity for the Monge-Ampère equation without convexity assumptions. Heriot-Watt University preprint, 1991.
67. V. Šverák. New examples of quasiconvex functions. *Arch. Rational Mech. Anal.*, 119:293–300, 1992.
68. M. A. Sychev. A new approach to Young measure theory, relaxation and convergence in energy. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 16(6):773–812, 1999.
69. L. Tartar. Some remarks on separately convex functions. In *Proceedings of conference on Microstructures and phase transitions, IMA, Minneapolis, 1990*, 1993.
70. D. H. Wagner. Survey of measurable selection theorems. *SIAM J. Control Optimization*, 15(5):859–903, 1977.
71. J. C. Wang, M. Osawa, T. Yokokawa, H. Harada, and M. Enomoto. Modeling the microstructural evolution of ni-base superalloys by phase field method combined with {CALPHAD} and {CVM}. *Computational Materials Science*, 39:871 – 879, 2007.
72. M. S. Wechsler, D. S. Lieberman, and T. A. Read. On the theory of the formation of martensite. *Trans. AIME J. Metals*, 197:1503–1515, 1953.

73. M. Yasunaga, Y. Funatsu, S. Kojima, K. Otsuka, and T. Suzuki. Ultrasonic velocity near the martensitic transformation temperature. *J. de Physique*, C4:603–608, 1982.
74. M. Yasunaga, Y. Funatsu, S. Kojima, K. Otsuka, and T. Suzuki. Measurement of elastic constants. *Scripta Met.*, 17:1091–1094, 1983.
75. R. Zarnetta, R. Takahashi, M. L. Young, A. Savan, Y. Furuya, S. Thienhaus, B. Maaß, M. Rahim, J. Frenzel, H. Brunken, Y. S. Chu, V. Srivastava, R. D. James, I. Takeuchi, G. Eggeler, and A. Ludwig. Identification of quaternary shape memory alloys with near zero thermal hysteresis and unprecedented functional stability. *Advanced Functional Materials*, pages 1917–1923, 2010.
76. K. Zhang. Rank 1 connections and the three "well" problem. Unpublished manuscript, 1991.
77. K. Zhang. A construction of quasiconvex functions with linear growth at infinity. *Ann. Scuola. Norm. Sup. Pisa*, 19:313–326, 1992.
78. Z. Zhang, R. D. James, and S. Müller. Energy barriers and hysteresis in martensitic phase transformations. *Acta Materialia (Invited Overview)*, 57:2332–4352, 2009.
79. B. Zwicknagl. Microstructures in low-hysteresis shape memory alloys: scaling regimes and optimal needle shapes. *Arch. Rational Mech. Anal.*, 213:355–421, 2014.