CONTEMPORARY MATHEMATICS

238

Nonlinear Partial Differential Equations

International Conference on
Nonlinear Partial Differential Equations
and Applications
March 21–24, 1998
Northwestern University

Gui-Qiang Chen
Emmanuele DiBenedetto
Editors

American Mathematical Society
Providence, Rhode Island

Contents

Preface

This volume is a collection of refereed original research papers and expository articles and stems from the scientific program of the 1997-98 Nonlinear PDE Emphasis Year at Northwestern University, which was jointly sponsored by Northwestern University and the National Science Foundation. Most of the papers presented are from the distinguished mathematicians who spoke at the International Conference on Nonlinear Partial Differential Equations, March 21-24, 1998, Evanston, IL.

The book is a cross-section of the most significant recent advances and current trends and directions in nonlinear partial differential equations and related topics. Contributions range from modern approaches to the classical theory in elliptic and parabolic equations to nonlinear hyperbolic systems of conservation laws and their numerical treatment.

The general guiding idea in editing this volume has been twofold. On one hand, we have solicited the papers that contribute in a substantial way to the general analytical treatment of the theory of nonlinear partial differential equations. On the other hand, we have attempted to collect the contributions to computational methods and applications, originating from significant realistic mathematical models of natural phenomena, to seek synergistic links between theory and modeling and computation and to underscore current research trends in partial differential equations. The borderline between these two aspects of mathematical research is rather fuzzy. We have also selected a set of papers that would bridge them.

Examples of the first kind of contributions include new insights into the role of the Harnack inequality in the theory of fully nonlinear elliptic equations, new results on the local behavior of degenerate parabolic equations, a treatment of the complex eikonal equations, and the solvability of implicit degenerate elliptic systems and motion by curvature.

Included in this broad category also are the papers establishing the regularity, large-time behavior, and L^1 stability of entropy solutions, the analysis of nonclassical shocks, and the convergence of the vanishing viscosity method for initial-boundary value problems for nonlinear hyperbolic systems of conservation laws, as well as the asymptotic stability of diffusion waves for the multidimensional nonlinear wave equations and the structural stability of steady-state solutions of the Navier-Stokes equations.

Contributions of the second kind range from numerical methods for predicting the large-scale dynamics and multidimensional simple front tracking algorithms to mathematical aspects of turbulent convection, geometry of crystal shapes, singularities in relativity and plasma dynamics, and high field kinetic semiconductor models.

This volume would not have been possible without the help and support of a number of people and institutions. First, we would like to thank the American Mathematical Society, especially, Edward G. Dunne (Editor of Book Program), Christine M. Thivierge (Acquisitions Assistant), Deborah Smith (Production Editor), and the technical support group for their prompt and professional assistance and their patience with our slow pace.

Karen Townsend deserves our special thanks for her assistance in the review process.

We are also grateful to the referees for their constructive criticisms and suggestions, and to Konstantina Trivisa and Mikhail Feldman for their invaluable help with the editorial work.

Finally, we wish to acknowledge the financial support of the National Science Foundation through grant DMS-9708261 and Northwestern University, more specially, the Department of Mathematics and the Office of the Vice President for Research of Northwestern University.

The purchaser of this volume is entitled to the online version of this book by the AMS. To gain access, follow the instructions given on the form found in the back of this volume.

Gui-Qiang Chen and Emmanuele DiBenedetto
Evanston, Illinois

Contemporary Mathematics
Volume **238**, 1999
B 0-8218-1196-7-03536-2

Nonclassical Shocks and the Cauchy Problem: General Conservation Laws

Paolo Baiti, Philippe G. LeFloch and Benedetto Piccoli

ABSTRACT. In this paper we establish the existence of *nonclassical* entropy solutions for the Cauchy problem associated with a conservation law having a *nonconvex* flux-function. Instead of the classical Oleinik entropy criterion, we use a single *entropy inequality* supplemented with a *kinetic relation*. We prove that these two conditions characterize a unique *nonclassical Riemann solver*. Then we apply the wave-front tracking method to the Cauchy problem. By introducing a new total variation functional, we can prove that the corresponding approximate solutions converge strongly to a nonclassical entropy solution.

1. Introduction

In this paper we establish a new existence theorem for weak solutions of the Cauchy problem associated with a nonlinear hyperbolic conservation law,

$$\partial_t u + \partial_x f(u) = 0, \qquad u(x,t) \in \boldsymbol{R} \quad x \in \boldsymbol{R},\ t > 0, \tag{1.1}$$

$$u(x,0) = u_0(x), \qquad x \in \boldsymbol{R}. \tag{1.2}$$

The flux-function $f : \boldsymbol{R} \to \boldsymbol{R}$ is *nonconvex* and the initial data $u_0 : \boldsymbol{R} \to \boldsymbol{R}$ is a function with bounded total variation. We are interested in weak solutions that are of bounded total variation and additionally satisfy the fundamental *entropy inequality*

$$\partial_t U(u) + \partial_x F(u) \leq 0 \tag{1.3}$$

for a (fixed) strictly convex entropy $U : \boldsymbol{R} \to \boldsymbol{R}$. As usual, the entropy-flux is defined by $F'(u) = U'(u)f'(u)$. We refer to Lax [**21, 22**] for these fundamental notions.

1991 *Mathematics Subject Classification.* Primary 35L65; Secondary 76L05.

Key words and phrases. conservation law, hyperbolic, entropy solution, nonclassical shock, kinetic relation, wave-front tracking.

Completed in September 1998.

The authors were supported in part by the European Training and Mobility Research project HCL # ERBFMRXCT960033. The second author was also supported by the Centre National de la Recherche Scientifique and a Faculty Early Career Development Award (CAREER) from the National Science Foundation under grant DMS 95-02766.

This self-contained paper is part of a series [**3, 5, 6**] devoted to proving the existence of nonclassical solutions for the *Cauchy problem* (1.1)–(1.2) supplemented with a single entropy inequality, (1.3), and a "kinetic relation" (see below). The paper [**3**] treated the case of a cubic flux $f(u) = u^3$ and placed a rather strong assumption on the kinetic function. Our purpose here is to provide an existence result for a *large class* of fluxes and kinetic relations covering all the examples arising in the applications. We will also provide examples where the total variation blows up when our assumptions are violated.

It is well-known since the works of Kružkov [**20**] and Volpert [**33**] that the problem (1.1)-(1.2) admits a unique (classical) entropy solution satisfying *all* of the entropy inequalities (1.3). In the present work we are interested in weak solutions constrained by a *single* entropy inequality. This question is motivated by zero diffusion-dispersion limits like

$$\partial_t u + \partial_x f(u) = \epsilon\, u_{xx} + \gamma \varepsilon^2\, u_{xxx}, \qquad \varepsilon \to 0 \ \text{ with } \ \gamma \ \text{ fixed.} \tag{1.4}$$

Hayes and LeFloch [**13, 14, 15**] observed that limiting solutions given by (1.4) and many similar *continuous* or *discrete* models satisfy the single entropy inequality (1.3) for a *particular choice* of entropy U, induced by the regularization terms. As is well-known, when the flux is convex the entropy inequality (1.3) singles out a *unique* weak solution of (1.1)-(1.2). However when the flux lacks convexity, this is no longer true and there is room for an additional selection criterion. It appears that weak solutions of the Cauchy problem (1.1)–(1.3) may exhibit *undercompressive, nonclassical shocks* which are the source of *non-uniqueness.* In [**13, 14**] it was proposed to further constrain the *entropy dissipation* of a nonclassical shock in order to uniquely determine its propagation speed. The corresponding relation is called a *kinetic relation.*

Jacobs, McKinney and Shearer [**17**] and then Hayes and LeFloch [**13**] (also [**16**]) observed that limits of diffusive-dispersive regularizations like (1.4) depend on the parameter γ and may fail to coincide with the classical entropy solutions of Kružkov-Volpert's theory. The sign of the parameter γ turns out to be critical. The corresponding kinetic function has been determined for several examples analytically and numerically.

The concept of a kinetic relation was introduced earlier in the material science literature, in the context of propagating phase transitions in solids undergoing phase transformations. James [**18**] recognized that weak solutions satisfying the standard entropy inequality were not unique. Abeyaratne and Knowles [**1, 2**] and Truskinovsky [**31, 32**] were pioneers in studying the Riemann problem and the properties of shock waves in phase dynamics. The kinetic relation was placed in a mathematical perspective by LeFloch in [**23**]. Earlier works on the Riemann problem with phase transitions include the papers by Slemrod [**30**] (where a model like (1.4) was introduced) and Shearer [**29**] (where the Riemann problem was solved using Lax entropy inequalities).

The papers [**13, 16, 17**] are concerned with the existence and properties of the traveling wave solutions associated with nonclassical shocks. The implications of a single entropy inequality for nonconvex equations and for non-genuinely nonlinear systems were discovered in [**13, 14**]. The numerical computation of nonclassical shocks via finite difference schemes was tackled in [**15, 25**]. Finally, for a review of these recent results we refer the reader to [**24**].

In [**3**], where the cubic case $f(u) = u^3$ is considered, it is proved that starting from a nonclassical Riemann solver, a front-tracking algorithm (Dafermos [**8**], DiPerna [**9**], Bressan [**7**], Risebro [**28**], Baiti and Jenssen [**4**]) applied to the Cauchy problem (1.1)-(1.2) converges to a weak solution satisfying the entropy condition (1.3), provided the initial data have bounded total variation.

The main difficulty in [**3**] was to derive a *uniform bound* on the total variation of the approximate solutions since nonclassical solutions do not satisfy the standard Total Variation Diminishing (TVD) property. Due to the presence of nonclassical shocks one was forced to introduce a *new functional*, equivalent to the total variation, which was decreasing in time for approximate solutions. This was achieved by estimating the strengths of waves across each type of interaction.

In the present paper we generalize [**3**] in two different directions: on one hand we consider general fluxes having one inflection point. The study of this case is required before tackling the harder case of systems [**5,6**]. On the other hand we relax the hypotheses imposed in [**3**] on the kinetic function, especially the somehow restrictive assumption that shocks with small strength were always classical.

As already pointed out, the difficult part in the convergence proof is finding a *modified measure of total variation.* In the cubic case [**3**] elementary properties of the (cubic) flux were used, in particular its symmetry with respect to 0. In the case of nonsymmetric fluxes it happens that an explicit form of the modified total variation can not be easily derived. To accomplish the same purpose here, we use a fixed-point argument on a suitable function space (see Sections 4 and 5). This approach should also clarify the choices made in [**3**] (see Section 6).

The paper is organized as follows. In Section 2 we start by listing our hypotheses and in Section 3 investigate how to solve the Riemann problem in the class of nonclassical solutions. In particular we prove that, under mild assumptions, every Riemann solver generating an $\mathbf{L}^1$-continuous semigroup of entropy solutions must be of the form considered here. Sections 4 to 6 are devoted to the definition and construction of the modified total variation. Finally, in Section 7 we present examples of blow-up of the total variation in cases when our hypotheses fail.

We also mention two companion papers which treat the uniqueness of nonclassical solutions [**5**] and the existence of nonclassical solutions for systems [**6**], respectively.

2. Assumptions

This section displays the assumptions required on the flux-function f and on the kinetic function φ. We assume that f is a smooth function of the variable u and admits a *single non-degenerate inflection point.* In other words, with obvious normalization, we make the following two assumptions:

(A1) $f(0) = 0, \;\; f'(u) > 0, \;\; u\, f''(u) > 0$ for all $u \neq 0$.

(A2) For some $p \geq 1$, f has the following Taylor expansion at $u = 0$

$$f(u) = H\, u^{2p+1} + o(u^{2p+1}) \quad \text{for some } H \neq 0.$$

The results of this paper extends to the case where $u\, f''(u) < 0$ holds. Note that (A1) implies

$$\lim_{u \to \pm\infty} f(u) = \pm\infty.$$

Consider the graph of the function f in the (u, f)-plane. For any $u \neq 0$ there exists a unique line that passes through the point with coordinates $(u, f(u))$ and is tangent to the graph at a point $\big(\tau(u), f(\tau(u))\big)$ with $\tau(u) \neq u$. In other words

$$f'\big(\tau(u)\big) = \frac{f(u) - f\big(\tau(u)\big)}{u - \tau(u)}. \tag{2.1}$$

Note that $u\,\tau(u) < 0$ and set also $\tau(0) = 0$. Thanks to the assumption (A1) on f, the map $\tau : \boldsymbol{R} \to \boldsymbol{R}$ is monotone decreasing and onto, and so is invertible. The inverse function satisfies

$$f'(u) = \frac{f(u) - f\big(\tau^{-1}(u)\big)}{u - \tau^{-1}(u)} \qquad \text{for all} \quad u \neq 0. \tag{2.2}$$

For any $u \neq 0$, define the point $\varphi^*(u) \neq u$ by the relation

$$\frac{f(u)}{u} = \frac{f\big(\varphi^*(u)\big)}{\varphi^*(u)}, \tag{2.3}$$

so that the points with coordinates

$$\big(\varphi^*(u), f(\varphi^*(u))\big), \qquad (0,0), \qquad \big(u, f(u)\big)$$

are aligned. Again from the assumptions (A1) above, it follows that $\varphi^* : \boldsymbol{R} \to \boldsymbol{R}$ is monotone decreasing and onto. Finally observe that

$$u\,\tau^{-1}(u) \leq u\,\varphi^*(u) \leq u\,\tau(u) \qquad \text{for all } u. \tag{2.4}$$

In Section 3 we shall prove that, in order to have uniqueness for the Riemann problem, for every left state u one has to single out a unique right state $\varphi(u)$ that can be connected to u with a nonclassical shock. The function $\varphi : \boldsymbol{R} \mapsto \boldsymbol{R}$ is called a *kinetic function* and depends on the regularization adopted for (1.1).
Given φ, we define the function $\alpha : \boldsymbol{R} \mapsto \boldsymbol{R}$ by the relation

$$\frac{f(u) - f\big(\alpha(u)\big)}{u - \alpha(u)} = \frac{f(u) - f\big(\varphi(u)\big)}{u - \varphi(u)}, \tag{2.5}$$

so that the points with coordinates

$$\big(\varphi(u), f(\varphi(u))\big), \qquad \big(\alpha(u), f(\alpha(u))\big), \qquad \big(u, f(u)\big)$$

are aligned.

In the whole of this paper a strictly convex entropy-entropy flux pair (U, F) is fixed to serve in the entropy inequality (1.3). In Proposition 3.1 we shall prove that for any $u_l \neq 0$ there exists a point $\varphi^\sharp(u_l)$ (depending on u_l and on the choice of (U, F)) such that the discontinuity (u_l, u_r) is admissible with respect to (1.3) iff $u_l\,\varphi^\sharp(u_l) \leq u_r u_l \leq u_l^2$. Finally, we shall denote by $g^{[k]}$ the k-th iterate of a map g.

Now select a kinetic function $\varphi : \boldsymbol{R} \mapsto \boldsymbol{R}$ satisfying the following set of properties:

- [H1] $u\,\varphi^\sharp(u) \leq u\,\varphi(u) \leq u\,\tau(u)$ for all u;
- [H2] φ is monotone decreasing;
- [H3] φ is Lipschitz continuous;
- [H4] $u\,\alpha(u) \leq 0$ for all u;

[H5] there exists $\varepsilon_0 > 0$ such that the Lipschitz constant η of the function $\varphi^{[2]}$ on the interval $I_0 := [-\varepsilon_0, \varepsilon_0]$ is less than 1. Moreover

$$\sup_{u \neq 0} \frac{\varphi^{[2]}(u)}{u} < 1. \tag{2.6}$$

The kinetic function describes the set of all *admissible nonclassical shock waves* to be used shortly in Section 3. In the rest of the present section we discuss each of the above assumptions and demonstrate that they are *"almost optimal."*

The condition [H1] means that the jump connecting u to $\varphi(u)$ is a nonclassical shock satisfying the entropy inequality (1.3) (cfr. Proposition 3.1 in Section 3). See Figure 2.1. The regularity properties [H2]-[H3] are basic, having here in mind the examples arising in the applications [**13, 17**].

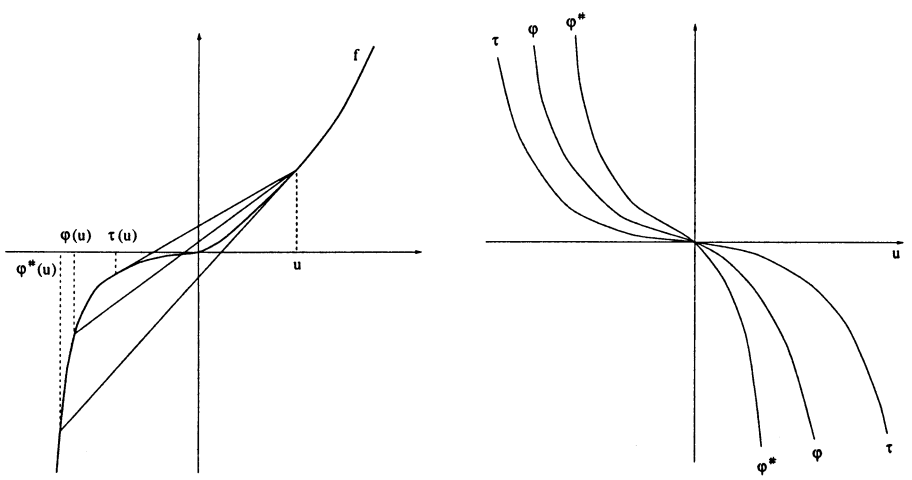

FIGURE 2.1

Interestingly the entropy inequality [H1] *implies* that

$$\alpha\big(\varphi(u)\big) \geq \alpha(u) \tag{2.7}$$

or equivalently

$$0 < \operatorname{sgn}(u)\, \varphi\big(\varphi(u)\big) \leq |u| \quad \text{for all } u \neq 0. \tag{2.8}$$

See (3.12)-(3.15). This will guarantee the solvability of every Riemann problem using *at most two waves.*

Our requirement [H4] is somewhat stronger than (2.7) and will ensure that the solution of the Riemann problem is *classical* as long as the left and the right state have the *same sign,* that is, lie in the same region of convexity. Note that the condition [H4] also forces φ to take its values on a smaller interval:

$$u\,\varphi^*(u) \leq u\,\varphi(u) \leq u\,\tau(u) \qquad \text{for all } u \neq 0. \tag{2.9}$$

Using (2.9) for u and also for $\varphi(u)$ evidently implies that φ satisfies (2.8).

Finally [H5] restricts the behavior of $\varphi^{[2]}$ (hence of φ) close to 0. It is worth pointing out that (2.6) is simply a strengthened version of (2.8) in which we are

just excluding the case of equality. Moreover [H5] excludes only the case of equality in d) of Lemma 2.1 below.

For concreteness, in the case where φ is smooth, then [H5] is equivalent to saying $\varphi'(0) > -1$ and $|\varphi^{[2]}(u)| < |u|$ for all $u \neq 0$. If φ is only Lipschitz continuous [H5] is indeed more general than these two conditions.

Let us derive some properties for the above functions near the inflection point.

LEMMA 2.1. *Under the assumptions* (A1)-(A2) *made on on the flux f, the functions τ and φ^* satisfy*

a) $\tau'(0) \in (-1, 0)$.

b) $|\tau(u)| < |u|$ *for small* u.

c) $\big(\varphi^*\big)'(0) = -1$.

d) *If* (2.8) *holds and φ is differentiable at $u = 0$ then* $\varphi'(0) \in [-1, 1]$.

PROOF. By hypothesis we have $f(u) = Hu^{2p+1} + o(u^{2p+1})$. By the definition (2.1), $\tau = \tau(u)$ satisfies

$$\big(H(2p+1)\tau^{2p} + o(\tau^{2p})\big)(u - \tau) = Hu^{2p+1} + o(u^{2p+1}) - H\tau^{2p+1} - o(\tau^{2p+1}).$$

By a bifurcation analysis it follows that τ is differentiable at $u = 0$. So, if we expand $\tau(u) = Cu + o(u)$, then it follows

$$Hu^{2p+1}\Big(2pC^{2p+1} - (2p+1)C^{2p} + 1\Big) + o(u^{2p+1}) = 0,$$

hence

$$h(C) := 2pC^{2p+1} - (2p+1)C^{2p} + 1 = 0.$$

By studying the zeroes of the function h, it follows that $\tau'(0) = C \in (-1, 0)$. (To illustrate this, note that for $f(u) = u^3$ we have $\tau(u) = -u/2$ and $\tau'(0) = -1/2$.) Hence a) holds as well as b).

By our hypotheses on the flux and the definition (2.3) of φ^* it follows that

$$Hu^{2p} = H\big(\varphi^*(u)\big)^{2p} + o\left(u^{2p}\right) + o\left(\big(\varphi^*(u)\big)^{2p}\right). \tag{2.10}$$

Writing $\varphi^*(u) = C'u + o(u)$, (2.10) yields

$$Hu^{2p} = H(C'u)^{2p} + o(u^{2p}),$$

hence $(C')^{2p} = 1$ which, together with $u\varphi^*(u) < 0$, implies $C' = -1$ and c) is proven.

Finally, assume that φ is differentiable so $\varphi(u) = C''\,u + o(u)$. In view of (2.8)

$$\operatorname{sgn}(u)\,\varphi\big(\varphi(u)\big) = (C'')^2\,|u| + o(u) \leq |u|,$$

thus $C'' \in [-1, 1]$. Hence d) follows. □

3. General Nonclassical Riemann Solver

A nonclassical Riemann solver is now defined from the kinetic function φ given in Section 2. The classical entropy solutions (Oleinik [**27**], Liu [**26**]) are recovered with the trivial choice $\varphi = \tau$. We also prove that our construction is essentially the unique possible one as long as the fundamental entropy inequality (1.3) is enforced (Assumption [H1]).

It is well-known that the Oleinik entropy criterion [**27**] states that a shock connecting u_- to u_+ is (Oleinik)-admissible iff

$$\frac{f(w)-f(u_-)}{w-u_-} \geq \frac{f(u_+)-f(u_-)}{u_+-u_-}, \tag{3.1}$$

for all w between u_- and u_+. An easy consequence of (3.1) is that the chord connecting the points $\big(u_-, f(u_-)\big)$ and $\big(u_+, f(u_+)\big)$ does not cross the graph of the flux f.

PROPOSITION 3.1. *Consider the conservation law* (1.1) *in the class of weak solutions satisfying the entropy inequality* (1.3) *for some strictly convex entropy* U.

Then for every u *there exists a point* $\varphi^\sharp(u)$ *such that a shock wave connecting a left state* u_- *to a right state* u_+ *satisfies the entropy inequality iff*

$$u_-\,\varphi^\sharp(u_-) \leq u_-\,u_+ \leq u_-^2. \tag{3.2}$$

Moreover we have

$$u_-\,\tau^{-1}(u_-) < u_-\,\varphi^\sharp(u_-). \tag{3.3}$$

PROOF. Let $\lambda = \lambda(u_-, u_+)$ be the shock speed and consider the entropy dissipation

$$D(u_-, u_+) := -\lambda\big(U(u_+)-U(u_-)\big) + F(u_+) - F(u_-).$$

We easily calculate that

$$\begin{aligned} D(u_-, u_+) &= \int_{u_-}^{u_+} \big(f'(m)-\lambda\big)\,U'(m)\,dm \\ &= -\int_{u_-}^{u_+} \big(f(m)-f(u_-)-\lambda\,(m-u_-)\big)\,U''(m)\,dm. \end{aligned} \tag{3.4}$$

The Rankine-Hugoniot relation for (1.1) yields λ:

$$\lambda = \frac{f(u_+)-f(u_-)}{u_+-u_-}.$$

Suppose for definiteness that $u_- > 0$. When $u_+ > u_-$, since f is convex in the region $m \in (u_-, u_+)$ we have

$$f(m) - f(u_-) - \lambda\,(m-u_-) < 0 \tag{3.5}$$

and therefore $D(u_-, u_+) > 0$. Moreover, it follows from (3.4) and the concavity/convexity properties of f, that the entropy dissipation $u \mapsto D(u_-, u)$ achieves a minimum negative value at $u = \tau(u_-)$ and vanishes at exactly two points (see an argument in [**14**]):

$$\begin{aligned} &D(u_-, \cdot) \text{ is monotone decreasing for } u < \tau(u_-), \\ &D(u_-, \cdot) \text{ is monotone increasing for } u > \tau(u_-), \\ &D\big(u_-, \tau(u_-)\big) < 0, \\ &D(u_-, u_-) = 0, \\ &D\big(u_-, \varphi^\sharp(u_-)\big) = 0. \end{aligned} \tag{3.6}$$

Hence (3.2) follows. On the other hand when $u_+ \leq \tau^{-1}(u_-)$ it is geometrically clear that the part of the graph of f corresponding to $m \in (u_+, u_-)$ lies *above* the chord

connecting the points $(u_-, f(u_-))$ and $(u_+, f(u_+))$. This means that the opposite sign holds now in (3.5). But since $u_+ < u_-$ we again obtain that $D(u_-, u_+) > 0$. This implies that $u_- \, \tau^{-1}(u_-) < u_- \, \varphi^\sharp(u_-)$. □

The shocks satisfying

$$u_- \, \tau(u_-) \leq u_- \, u_+ \leq u_-^2 \tag{3.7}$$

are Oleinik-admissible and will be referred to as *classical shocks*. On the other hand for entropy admissible *nonclassical shocks*, (3.1) is violated, i.e.,

$$u_- \, \varphi^\sharp(u_-) \leq u_- \, u_+ \leq u_- \, \tau(u_-). \tag{3.8}$$

This establishes that the condition [H1] in Section 2 is in fact a *consequence* of the entropy inequality (1.3).

From now on we rely on the kinetic function φ selected in Section 2 and we solve the Riemann problem (1.1),

$$u(x,0) = u_0(x) = \begin{cases} u_l, & \text{for } x < 0, \\ u_r, & \text{for } x > 0, \end{cases} \tag{3.9}$$

where u_l and u_r are constants. We restrict attention to the case $u_l > 0$, the other case being completely similar. To define the *nonclassical Riemann solver* we distinguish between four cases:

(i) If $u_r \geq u_l$, the solution u is a (Lipschitz continuous) rarefaction wave connecting monotonically u_l to u_r.

(ii) If $u_r \in \big[\alpha(u_l), u_l\big)$, the solution is a classical shock wave connecting u_l to u_r.

(iii) If $u_r \in \big(\varphi(u_l), \alpha(u_l)\big)$, the solution contains a (slower) nonclassical shock connecting u_l to $\varphi(u_l)$ followed by a (faster) classical shock connecting to u_r.

(iv) If $u_r \leq \varphi(u_l)$, the solution contains a nonclassical shock connecting u_l to $\varphi(u_l)$ followed by a rarefaction connecting to u_r.

For $u_l = 0$, the Riemann problem is a single rarefaction wave, connecting monotonically u_l to u_r. The function u will be called the *φ-admissible nonclassical solution* of the Riemann problem. Clearly different choices for φ yield different weak solutions u. This is natural as we already pointed out that limits given by (1.4) and similar models do depend on the parameter γ.

The above construction is essentially unique, as we show with the following two theorems.

THEOREM 3.2. *Consider the Riemann problem* (1.1)-(3.9) *in the class of piecewise smooth solutions satisfying the entropy inequality* (1.3) *for some strictly convex entropy U.*

Then either the Riemann problem admits a unique solution or else there exists a one-parameter family of solutions containing at most two (shock or rarefaction) waves.

Next for any nonclassical shock connecting some states u_- and u_+ with the speed λ, we impose the kinetic relation

$$D(u_-,u_+)=\begin{cases}\Phi^-(\lambda) & \text{if } u_+<u_-,\\ \Phi^+(\lambda) & \text{if } u_+>u_-,\end{cases} \tag{3.10}$$

where the kinetic functions are Lipschitz continuous and satisfy

$$\begin{aligned}&\Phi^\pm(0)=0,\\ &\Phi^\pm \text{ is monotone decreasing,}\\ &\Phi^\pm(\lambda)\geq D^\pm(\lambda).\end{aligned} \tag{3.11}$$

In the latter condition the lower bound $D^\pm$ is the maximum negative value of the entropy dissipation

$$D^\pm(\lambda):=D\big(\tau^{-1}(u),u\big),\quad \lambda=f'(u)\qquad \text{for } \pm u\geq 0.$$

Then (3.10) *selects a unique nonclassical solution in the one-parameter family of solutions.*

Observe that given $\lambda>0$ there are exactly one positive value and one negative value u such that $\lambda=f'(u)$. This property led us to define kinetic functions $\Phi^\pm$ for nonclassical shocks corresponding to decreasing and to increasing jumps.

Proof. The inequalities in Proposition 3.1 restrict the range of values taken by nonclassical shocks. First of all we show here that *at most two waves* can be combined together.

We now claim that

$$\varphi^\sharp\big(\varphi^\sharp(u_-)\big)=u_- \quad \text{for all } u_-. \tag{3.12}$$

Indeed we have by definition

$$D\big(u_-,\varphi^\sharp(u_-)\big)=0,\quad u_-\neq\varphi^\sharp(u_-)$$

and

$$D\left(\varphi^\sharp(u_-),\varphi^\sharp\big(\varphi^\sharp(u_-)\big)\right)=0,\quad \varphi^\sharp(u_-)\neq\varphi^\sharp\big(\varphi^\sharp(u_-)\big).$$

The conclusion follows immediately from the fact that the entropy dissipation has a single "nontrivial" zero; see (3.6).

We want to prove that the function $u\mapsto\varphi^\sharp(u)$ is decreasing. Again, by a bifurcation argument it follows that $\varphi^\sharp$ is differentiable. Now notice that $D(u,v)=-D(v,u)$ hence

$$\partial_u D\big(u_-,\varphi^\sharp(u_-)\big)=-\partial_v D\big(\varphi^\sharp(u_-),u_-\big). \tag{3.13}$$

From (3.6) we have that $\operatorname{sgn}\big(\partial_v D\big(\varphi^\sharp(u_-),u_-\big)\big)=-\operatorname{sgn}\big(\partial_v D\big(u_-,\varphi^\sharp(u_-)\big)\big)$, hence it follows that

$$\operatorname{sgn}\left(\partial_u D\big(u_-,\varphi^\sharp(u_-)\big)\right)=\operatorname{sgn}\left(\partial_v D\big(u_-,\varphi^\sharp(u_-)\big)\right). \tag{3.14}$$

Taking the total differential of the identity $D\big(u_-,\varphi^\sharp(u_-)\big)=0$ with respect to u_- and using (3.14) gives $d\varphi^\sharp/du_-<0$ for all u_-.

Consider a nonclassical shock connecting u_- to $\varphi(u_-)$. By hypothesis $u_-\varphi^\sharp(u_-)\leq u_-\varphi(u_-)$ hence by the monotonicity of $\varphi^\sharp$ and (3.12) it follows that $u_-\varphi^\sharp\big(\varphi(u_-)\big)\leq u_-\varphi^\sharp\big(\varphi^\sharp(u_-)\big)=(u_-)^2$. This prevents us to combine together more than two

waves. Indeed since the speeds of the (rarefaction or shock) must be ordered (increasing) along a combination of waves, it is easily checked geometrically that the only possible wave patterns are:

1. a rarefaction wave,
2. a classical shock wave,
3. a nonclassical shock followed by a classical shock,
4. or else a nonclassical shock followed by a rarefaction.

Finally we discuss the selection of nonclassical shocks. It is enough to prove that for each fixed u_- there is a unique nonclassical connection to a state u_+ satisfying both the jump relation and the kinetic relation.

Suppose $u_- > 0$ is *fixed* and regard the entropy dissipation as a function of the speed λ:

$$\Psi(\lambda) = D\big(u_-, u_+(\lambda)\big), \quad \lambda = \frac{f\big(u_+(\lambda)\big) - f(u_-)}{u_+(\lambda) - u_-}.$$

It is not hard to see that

$$\begin{aligned}
&\Psi \text{ is increasing for } \lambda \in \big[f'(\tau(u_-)), f'(u_-)\big],\\
&\Psi\big(f'(\tau(u_-))\big) = D^+\big(f'(\tau(u_-))\big) \leq \Phi\big(f'(\tau(u_-))\big),\\
&\Psi\big(f'(u_-)\big) = 0 \geq \Phi\big(f'(u_-)\big).
\end{aligned}$$

In view of the assumptions made on Ψ it is clear that the equation

$$\Psi(\lambda) = \Phi(\lambda)$$

admits exactly one solution. This completes the proof that the nonclassical wave is unique. □

The property (3.12) implies that

$$0 < \operatorname{sgn}(u)\, \varphi\big(\varphi(u)\big) \leq |u| \quad \text{for all } u \neq 0, \tag{3.15}$$

which is (2.8).

We have already seen that a kinetic relation is sufficient to select a unique way of solving the Riemann Problem and the solution was described earlier. Now we want to prove that this is essentially the *unique expression* a Riemann Solver can have.

More precisely, assume the following are given:

- a set $\mathcal{A}$ of admissible waves satisfying the entropy inequality (1.3) for a fixed, strictly convex pair (U, F);
- for every pair of states (u_l, u_r), a way of solving the associated Riemann problem, using only admissible waves in $\mathcal{A}$. Denote by $\mathcal{R}(u_l, u_r)$ the Riemann solution;
- an $\mathbf{L}^1$-continuous semigroup of solution for (1.1)-(1.2), compatible with the above Riemann solutions. (Note that in [**5**] it is proven that, if such a semigroup exists, then there is a unique way of solving the Riemann problem associated with any pair of states u_l, u_r.)

Any collection of $\{\mathcal{R}(u_l, u_r); u_l, u_r \in \boldsymbol{R}\}$ satisfying the above assumptions will be called here a *basic $\mathcal{A}$-admissible Riemann Solver*. We are going to prove that

$\mathcal{R}(u_l, u_r)$ coincides with (i)-(iv) for some choice of the function φ. This completely justifies our study of the Nonclassical Riemann Solver made in the present paper.

The admissibility criterion imposed by $\mathcal{A}$ could be recovered by the analysis of the limits of some regularizations of (1.1) like (1.4), or by a kinetic relation as in this paper (see also [**13,14**]). But it could also be given *a priori* by some physical or mathematical argument.

THEOREM 3.3. *Every basic $\mathcal{A}$-admissible Riemann Solver coincides with a Nonclassical Riemann Solver for a suitable choice of the function φ.*

PROOF. In the previous discussion it was observed that there are only four possible wave patterns, namely a single shock, a single rarefaction wave or else a nonclassical shock followed by either a shock or a rarefaction. Without loss of generality, assume $u_l > 0$. Any state $u_r > u_l$ can be connected to the right of u_l only by a rarefaction wave, hence $\mathcal{R}(u_l, u_r)$ must coincide with this rarefaction.

In the following we shall consider all the shocks connecting u_l to $\tau(u_l)$ to be nonclassical. Since u_l can be connected by a single classical wave only to points $u_r > \tau(u_l)$, then u_l must be connected by a nonclassical shock to at least one right state $u_r \leq \tau(u_l)$.

Let us see that this right point is unique. By contradiction, assume there exist points $\tilde{u} < \bar{u} < 0$ such that u_l can be connected to both of them by a nonclassical shock. By hypothesis $\bar{u}$ and $\tilde{u}$ are connected by an (admissible) rarefaction. Hence the Riemann problem $(u_l, \tilde{u})$ can be solved either by a single nonclassical shock or by a nonclassical shock to $\bar{u}$ followed by a rarefaction to $\tilde{u}$. This contradicts the uniqueness of the Riemann solver $\mathcal{R}(u_l, u_r)$. It follows that u_l can be connected with a nonclassical shock to exactly one right state, call it $\varphi(u_l)$.

By uniqueness, this implies immediately that all the states $u_r < \varphi(u_l)$ are connected to the right of u_l by the nonclassical shock to $\varphi(u_l)$ followed by a rarefaction to u_r.

Introduce now the point $\alpha(u_l)$ as in (2.5). The points in the interval $[\alpha(u_l), u_l)$ can not be reached neither by a rarefaction, nor by a wave pattern containing a (single) nonclassical shock. Hence they must be reached by a classical shock. Now, if $\varphi(u_l) = \alpha(u_l) = \tau(u_l)$ then we are done and the Riemann solution $\mathcal{R}(u_l, u_r)$ coincides with the Liu solution. Otherwise $\varphi(u_l) < \tau(u_l) < \alpha(u_l)$ and the points u_r in the interval $\big[\varphi(u_l), \tau(u_l)\big)$ are reached by the nonclassical shock followed by a classical shock, since this is the only way to connect u_l and u_r. It remains to cover $\big[\tau(u_l), \alpha(u_l)\big)$. The points in this interval can be reached either by a single classical shock or by the nonclassical shock followed by a classical one. So, let $\bar{u}_l := \sup\{u_r \geq \varphi(u_l)$ that are connected to the left of u_l by the nonclassical shock followed by a classical one $\}$. Then $\bar{u}_l \leq \alpha(u_l)$ and every $u > \bar{u}_l$ is connected to left of u_l by a single classical shock. By the $\mathbf{L}^1$-continuity property and an analysis of the wave-speeds it follows that the solution of the Riemann problem $(u_l, \bar{u}_l)$ with a nonclassical shock followed by a classical shock and the one with a single classical shock must coincide, hence $\bar{u}_l = \alpha(u_l)$. It follows that $\mathcal{R}(u_l, u_r)$ coincides with the nonclassical Riemann solver for this choice of φ. □

4. New Total Variation Functional

A classical way to prove convergence of approximate schemes for conservation laws is to give uniform bounds on the $\mathbf{L}^\infty$ and $\mathbf{BV}$ norms of the approximate

solutions and then pass to the limit by using Helly's compactness theorem. Unfortunately, in contrast to the classical case, the total variation of the approximate solutions can increase across interactions due to the creation or interaction of nonclassical shocks. Hence a careful analysis is needed, of how the strengths of waves change across interactions. In the classical case of systems [**12**] the so-called interaction potential Q is used to compensate a (possible) increase of the total variation. In our case, however, it appears that if two fronts of strength σ and σ' interact at time t (here strength means the size of the jump in the discontinuity) then there are cases in which the variation of the total variation is linear in the strength of the incoming waves, i.e. $\Delta\mathbf{TV}(t) \sim C\big(|\sigma| + |\sigma'|\big)$. This implies that we cannot use the potential Q to control the increase in the total variation since Q is a quadratic functional (see [**12**]).

Our approach is to construct a modified total variation functional which decreases in time along suitable wave-front tracking approximations of (1.1)-(1.2), and which is equivalent to the usual total variation, i.e. we are looking for a functional $\mathbf{V}$ such that for every piecewise constant approximate solution $v(t,x)$ constructed by front-tracking we have $\Delta\mathbf{V}\big(v(t,\cdot)\big) \leq 0$ for every $t > 0$ and there exist positive constants C_1, C_2, depending only on the $\mathbf{L}^\infty$ and $\mathbf{BV}$ norms of the initial data u_0, such that $C_1\,\mathbf{V}(v) \leq \mathbf{TV}(v) \leq C_2\,\mathbf{V}(v)$ (see [**3**]). The definition of $\mathbf{V}$ can be regarded as a generalization of the standard distance $|u_r - u_l|$.

Now, let $u : \boldsymbol{R} \mapsto \boldsymbol{R}$ be a piecewise constant function and let x_α, $\alpha = 1, \ldots, N$, be the points of discontinuity of u. Define

$$\mathbf{V}(u) := \sum_{\alpha=1}^{N} \sigma\big(u(x_\alpha-), u(x_\alpha+)\big), \tag{4.1}$$

where $\sigma(u_l, u_r)$ measures the strength of the wave connecting the left state u_l to the right state u_r. Notice that if $\sigma(u_l, u_r) = |u_r - u_l|$, then $\mathbf{V}(u) = \mathbf{TV}(u)$. So, a new definition of the strength $\sigma(u_l, u_r)$ is necessary. More precisely, we set

$$\sigma(u_l, u_r) := \begin{cases} \big(\psi(u_r) - \psi(u_l)\big)\,\mathrm{sgn}\,(u_r - u_l)\,\mathrm{sgn}\,(u_l) & \text{if } \big(u_r - \varphi(u_l)\big)\,\mathrm{sgn}\,(u_l) \geq 0, \\ \psi(u_r) + \psi(u_l) - 2\,\psi\big(\varphi(u_l)\big) & \text{if } \big(u_r - \varphi(u_l)\big)\,\mathrm{sgn}\,(u_l) \leq 0. \end{cases} \tag{4.2}$$

where $\psi : \boldsymbol{R} \mapsto \boldsymbol{R}$ is a continuous function that is increasing (resp. decreasing) for u positive (resp. negative). It is also assumed that $\psi(0) = 0$.

The wave strength σ depends on the kinetic function φ as well as on the function ψ to be determined in Section 5. Observe that the function $u_r \mapsto \sigma(u_l, u_r)$ is a piecewise linear function in term of $\psi(u_r)$ resembling the letter W. It achieves a *local minimum value* at $u_r = u_l$ and at $u_r = \varphi(u_l)$, the latter corresponding of course to the nonclassical shock. Therefore the strength of the nonclassical shock is counted *less* than what it would be with the standard total variation. This choice is made to *compensate for the increase* of the standard total variation that arises in certain wave interactions involving nonclassical shocks.

Let u_ν be the sequence of piecewise constant solutions of (1.1)-(1.2) constructed via wave-front tracking from an approximation of the initial data u_0, following [**3**]. We replace the data u_0 with a piecewise constant approximation $u_\nu(0)$ such that

$$u_\nu(0) \to u_0 \quad \text{in the } \mathbf{L}^1 \text{ norm}, \quad \mathbf{TV}(u_\nu(0)) \to \mathbf{TV}(u_0). \tag{4.3}$$

Based on the nonclassical Riemann solver of Section 3, we approximately solve the corresponding Cauchy problem for small time. Let δ_ν be a sequence of positive numbers converging to zero. For each ν, the approximate solution u_ν is constructed as follows. Solve approximatively the Riemann problem at each discontinuity point of u_ν. This is obtained by approximating the solution given by the nonclassical Riemann solver: every shock or nonclassical shock travels with the correct shock speed, while the rarefaction fans are approximated by rarefaction fronts. More precisely, every rarefaction wave connecting the states u_l and u_r, say, with $\sigma(u_l, u_r) > \delta_\nu$ is approximated by a finite number of small jumps traveling with speed equal to the right characteristic speed and with strength less than or equal to δ_ν.

When two wave-front meet, we again use the nonclassical approximate Riemann solver and continue inductively in time. The main aim is to estimate the total variation, that is to prove that there exists a positive constant C such that

$$\mathbf{TV}\big(u_\nu(t)\big) \le C, \quad t \ge 0 \tag{4.4}$$

uniformly in ν.

From now on we assume that a kinetic function satisfying [H1]-[H5] is fixed. First of all notice that under these hypotheses the interaction patterns for all couples of waves are analogous to those considered and listed in Section 2 of [**3**]. We shall rely on this classification in the rest of the present section. To prove that u_ν is well-defined, it is sufficient to show that the above construction can be carried on for all positive times.

PROPOSITION 4.1. *Assume that the function*

$$u \mapsto \operatorname{sgn}(u)\big(\psi(u) - \psi\big(\varphi(u)\big)\big)$$

is monotone increasing. Then the approximate solutions $u_\nu(t)$ are well-defined for all times $t \ge 0$ and satisfy

$$\|u_\nu(t)\|_{\mathbf{L}^\infty(\boldsymbol{R})} \le \max\big\{c, |\varphi(c)|\big\}, \qquad c := \|u_\nu(0)\|_{\mathbf{L}^\infty(\boldsymbol{R})}. \tag{4.5}$$

PROOF. As in [**3**] it is sufficient to prove that the total number of waves does not increase in time, so it can be bounded uniformly in t (for fixed ν). Since only two waves may leave after the interaction of two waves, it is sufficient to prove that the rarefactions do not increase their strength across interaction. Denote by σ the strength of rarefactions and $\Delta\sigma$ the change across the interaction. Referring to the cases of wave interactions listed in [**3**], we have (recalling that we assume $u_l > 0$):

Case 1. Trivial case: $\Delta\sigma < 0$.

Case 4. The variation of the strength across the interaction is computed by

$$\begin{aligned}\Delta\sigma &= \big(\psi(u_r) - \psi\big(\varphi(u_l)\big)\big) - \big(\psi(u_m) - \psi(u_l)\big)\\ &\le \psi\big(\varphi(u_m)\big) - \psi(u_m) - \big(\psi\big(\varphi(u_l)\big) - \psi(u_l)\big) \le 0.\end{aligned}$$

Case 6. This is a limiting case of Case 4.

$$\Delta\sigma = \psi\big(\varphi(u_m)\big) - \psi(u_m) - \psi\big(\varphi(u_l)\big) + \psi(u_l) \le 0.$$

Case 17. Now the variation is given by

$$\Delta\sigma = \big(\psi\big(\varphi(u_l)\big) - \psi(u_r)\big) - \big(\psi(u_m) - \psi(u_r)\big) < 0.$$

So the approximate solutions are well-defined for all positive times.

We now prove (4.5). It is obvious that the only interactions that can increase the $\mathbf{L}^\infty$-norm are those in which a nonclassical shock is involved. Let $\mathcal{R}(u)$ be the range of a piecewise constant function u. For every approximate solution u_ν, across an interaction at time t we have

$$\mathcal{R}\big(u_\nu(t+,\cdot)\big) \subseteq \mathcal{R}\big(u_\nu(t-,\cdot)\big) \cup \mathcal{R}\big(\varphi\big(u_\nu(t-,\cdot)\big)\big), \tag{4.6}$$

as follows from the definition of the Riemann solver in Section 3. It is clear that (4.5) holds for $t = 0+$. Now fix ν and assume that for a positive time t we have

$$M(t) := \|u_\nu(t,\cdot)\|_{\mathbf{L}^\infty} > \|u_\nu(0,\cdot)\|_{\mathbf{L}^\infty}.$$

Then by (4.6), there exists $\tilde u \in \mathcal{R}\big(u_\nu(0,\cdot)\big)$ and a positive integer n such that

$$M(t) = \big|\varphi^{[n]}(\tilde u)\big|.$$

Recall that $\big|\varphi^{[2]}(u)\big| \le |u|$. Hence n must be odd, otherwise by induction

$$M(t) = \big|\varphi^{[n]}(\tilde u)\big| \le |\tilde u| \le \|u_\nu(0,\cdot)\|_{\mathbf{L}^\infty},$$

which is a contradiction. So $n = 2q+1$ and again by induction it follows that

$$M(t) = \big|\varphi^{[2q]}\big(\varphi(\tilde u)\big)\big| \le \big|\varphi(\tilde u)\big| \le \big|\varphi\big(\|u_\nu(0,\cdot)\|_{\mathbf{L}^\infty}\big)\big|.$$

Hence (4.5) follows. This completes the proof of Proposition 4.1. □

Assuming now that the approximate initial data satisfy

$$\|u_\nu(0)\|_{\mathbf{L}^\infty(\boldsymbol{R})} \le C\ \|u_0\|_{\mathbf{L}^\infty(\boldsymbol{R})}, \tag{4.7}$$

we conclude from (4.5) that

$$\|u_\nu(t)\|_{\mathbf{L}^\infty(\boldsymbol{R})} \le C' \quad \text{for all } t \ge 0, \tag{4.8}$$

uniformly in ν.

We next derive a uniform **BV** bound or, more precisely, we prove that **V** decreases along approximate solutions.

PROPOSITION 4.2. *Assume that the function*

$$u \mapsto \operatorname{sgn}(u)\big(\psi(u) - \psi\big(\varphi(u)\big)\big) \quad \textit{is monotone increasing.} \tag{4.9}$$

Then for the approximate solutions,

$$t \mapsto \mathbf{V}\big(u_\nu(t)\big) \quad \textit{is monotone decreasing.} \tag{4.10}$$

PROOF. The function $t \mapsto \mathbf{V}\big(u_\nu(t)\big)$ is piecewise constant with discontinuities located only at interaction times. Hence it suffices to show that $\mathbf{V}$ decreases across every collision. Assume that the three states u_l, u_m and u_r, are separated by two interacting wave fronts of strength σ_1^- and σ_2^-, each being generated by a (nonclassical) Riemann solution. Each front can be either a classical, nonclassical shock or a rarefaction wave.

The complete list of interaction patterns can be found in Section 2 of [**3**], but Cases 5 and 12 therein never occur because of our assumption [H2]. In [**3**] a case by case analysis was developed. Here thanks to the general definition (4.1)-(4.2) we have some simplifications.

Any outgoing pattern is made of at most two waves, with strength σ_1^+ and (possibly) σ_2^+. Hence the variation of $\mathbf{V}$ across the interaction is given by $\Delta\mathbf{V} = (\sigma_1^+ + \sigma_2^+) - (\sigma_1^- + \sigma_2^-) := \Sigma^+ - \Sigma^-$.

The function σ introduced in (4.2) satisfies the following key properties:

(i) σ is *additive on ordered waves*, in the sense that if u_1, u_2, u_3 are three states such that $u_1 < u_2 < u_3$ and such that $\operatorname{sgn}(u_i)\big(u_j - \varphi(u_i)\big) \geq 0$ for all $i > j$, then
$$\sigma(u_1, u_3) = \sigma(u_1, u_2) + \sigma(u_2, u_3),$$
$$\sigma(u_3, u_1) = \sigma(u_3, u_2) + \sigma(u_2, u_1);$$

(ii) if $\operatorname{sgn}(u_1) = \operatorname{sgn}(u_2)$ then $\sigma(u_1, u_2) = \sigma(u_2, u_1)$;

(iii) for every outgoing pattern we have $\Sigma^+ = \sigma(u_l, u_r)$.

These properties can be checked from the definition of σ. In particular (iii) implies that $\Delta\mathbf{V} \leq 0$ iff $\Sigma^- \geq \sigma(u_l, u_r)$.

The interaction cases can be split in four families.

CASES 8, 9, 10, 11, 15, 16, 17. The states before the interaction are ordered. Hence by (i) it follows that $\Sigma^- = \sigma(u_l, u_r)$ and so $\Delta\mathbf{V} = 0$.

CASES 1, 2, 3, 4, 7. The states before the interaction are not ordered. Then a cancellation takes place, but the canceled wave lives on one region of convexity. Using (i)-(ii) it follows that $\Sigma^- > \sigma(u_l, u_r)$, hence $\Delta\mathbf{V} < 0$.

CASES 13, 14, 18, 19. As in the previous case but now the canceled wave cross the state 0. Indeed all of them are special cases of 14. For the latter, it easy to see that $\Sigma^+ = \psi(u_l) - \psi(u_r)$ and $\Sigma^- = \big(\psi(u_l) - \psi(u_m)\big) - \big(\psi(u_m) - \psi(u_r)\big)$, hence $\Delta\mathbf{V} = 0$.

It remains to check only Case 6.

CASE 6. This is the only case which requires condition (4.9). Indeed, $u_l(u_l - u_m) > 0$ and it follows that

$$\begin{aligned}\Delta\mathbf{V} &= \psi(u_l) - \psi\big(\varphi(u_l)\big) + \psi\big(\varphi(u_m)\big) - \psi\big(\varphi(u_l)\big) + \\ &\quad - \big(\psi(u_m) - \psi(u_l)\big) - \big(\psi(u_m) - \psi\big(\varphi(u_m)\big)\big) \\ &= 2\Big[\big(\psi(u_l) - \psi\big(\varphi(u_l)\big)\big) - \big(\psi(u_m) - \psi\big(\varphi(u_m)\big)\big)\Big] \leq 0.\end{aligned}$$

This completes the proof. □

The existence of a function ψ satisfying the condition (4.9) will be established in Section 5. The equivalence between $\mathbf{TV}$ and $\mathbf{V}$ will be proved there, too. Now we are ready to conclude with the main result of the present paper.

THEOREM 4.3. *Consider the conservation law* (1.1) *together with the nonclassical Riemann solver characterized by the function* $\varphi : \boldsymbol{R} \to \boldsymbol{R}$. *Suppose that* φ *satisfies the assumptions* [H1]–[H5] *listed in Section* 2.

Given an initial data u_0 *with bounded total variation, there exists a positive constant* $\tilde{C}$ *depending only on* φ *and the* $\mathbf{L}^\infty$*-norm of* u_0 *such that the approximate solutions* $u_\nu(t)$ *(constructed by wave-front tracking) satisfy*

$$\mathbf{TV}\big(u_\nu(t)\big) \leq \tilde{C}\, \mathbf{TV}(u_0) \tag{4.11}$$

for all times $t \geq 0$.

A subsequence of u_ν *converges in the* $\mathbf{L}^1$ *norm toward a weak solution of the conservation law* (1.1)-(1.2) *which satisfies the entropy inequality* (1.3).

PROOF. By (4.8)-(4.10), the approximate solutions constructed above have uniformly bounded $\mathbf{L}^\infty$-norm and total variation. We can apply Helly's theorem to find a (sub)sequence which converges in $\mathbf{L}^1{}_{loc}$ to a function u. Since the modified and the usual strengths of waves are equivalent (see (5.10)) u is a nonclassical weak solution of (1.1)-(1.2) satisfying also the entropy inequality (1.3). □

5. Construction of the Function ψ

In this section we prove the existence of a function ψ satisfying (4.9) needed in Propositions 4.1 and 4.2. This will be accomplished by a fixed-point argument in a suitable function space X defined below.

Denote by $\mathrm{Lip}_I(\psi)$ the Lipschitz constant of a function ψ defined on some interval I. Let $M > 0$ be a constant greater than the $\mathbf{L}^\infty$-norm of u_0 and define $J_M := [-M, M] \cup \big[\varphi(M), \varphi(-M)\big]$. Finally, let L_M be the Lipschitz constant of φ in the set J_M. Introduce the space

$$X := \Big\{\psi \in C\big(J_M; \boldsymbol{R}\big) \,:\, \psi(0) = 0,\ \|\psi\|_X < \infty\Big\},$$

endowed with the norm

$$\|\psi\|_X := \sup_{u \neq 0} \left| \frac{\psi(u)}{u - \varphi(u)} \right|.$$

Then define the subset $Y \subset X$ by

$$Y := \left\{ \psi \in X :\ \ \psi \text{ is } \begin{matrix}\text{increasing} \\ \text{decreasing}\end{matrix} \text{ for } u \gtrless 0;\ \ \mathrm{Lip}_{I_0}(\psi) \leq K \right\},$$

where $K \geq (1 + L_M)/(1 - \eta)$ is a fixed constant and η is the Lipschitz constant introduced in [H5].

The reason why we consider J_M instead of $[-M, M]$ is that we need a φ-invariant set and actually $\varphi(J_M) \subseteq J_M$ while in general this is not true for $[-M, M]$.

LEMMA 5.1. $\big(X, \|\cdot\|_X\big)$ *is a Banach space and* Y *is a closed subset.*

PROOF. It is clear that X is a normed space. Let us see that it is complete. Let $\psi_n \in X$, $n = 1, 2, \ldots$ be a Cauchy sequence in the norm $\|\cdot\|_X$. By definition, for every $\varepsilon > 0$, there exists $\bar{n}$ such that for all $m, n \geq \bar{n}$ we have

$$\big|\psi_n(u) - \psi_m(u)\big| \leq |u - \varphi(u)|\, \varepsilon \tag{5.1}$$

for all $u \neq 0$, but also for $u = 0$. Hence the sequence ψ_n is also Cauchy in the space $C\big(J_M; \boldsymbol{R}\big)$ with the sup-norm, and so it converges to a continuous function ψ. Moreover, by passing pointwise to the limit, we see that $\psi(0) = 0$. Finally by letting $m \to \infty$ in (5.1) we see that the convergence holds actually in the space X.

Finally let us see that Y is closed. Take $\psi_n \to \psi$ in X with $\psi_n \in Y$ for all n. First of all, by passing pointwise to the limit, it follows that ψ satisfies the monotonicity properties. By hypothesis we have

$$\left| \frac{\psi_n(u) - \psi_n(v)}{u - v} \right| \leq K \tag{5.2}$$

for all n and all $u, v \in I_0$ with $u \neq v$. Since ψ_n converges to ψ pointwise, by passing to the limit in (5.2) we get $\mathrm{Lip}_{I_0}(\psi) \leq K$, hence $\psi \in Y$ and Y is closed. □

Now define the map $T : X \mapsto X$ by the relation

$$\big(T\psi\big)(u) := \psi\big(\varphi(u)\big) + |u|, \qquad u \in \boldsymbol{R}. \tag{5.3}$$

THEOREM 5.2. *T maps X into X and is a contraction.*

PROOF. Let $\psi \in X$ be fixed. It is clear that $\big(T\psi\big)(0) = 0$. Let us see that T is a contraction. For all $\psi, \bar{\psi} \in X$ and $u \neq 0$ we have

$$\left| \frac{T\psi(u) - T\bar{\psi}(u)}{u - \varphi(u)} \right| = \left| \frac{\varphi(u) - \varphi\big(\varphi(u)\big)}{u - \varphi(u)} \right| \left| \frac{\psi\big(\varphi(u)\big) - \bar{\psi}\big(\varphi(u)\big)}{\varphi(u) - \varphi\big(\varphi(u)\big)} \right|,$$

hence

$$\big\|T\psi - T\bar{\psi}\big\|_X \leq \sup_{u \neq 0} \left| \frac{\varphi(u) - \varphi\big(\varphi(u)\big)}{u - \varphi(u)} \right| \cdot \big\|\psi - \bar{\psi}\big\|_X,$$

and by taking $\bar{\psi} \equiv 0$ in this last inequality, it follows that $\|T\psi\|_X < \infty$ and T maps X into itself. Now, it is easy to see (i.e. geometrically) that

$$\left| \frac{\varphi(u) - \varphi\big(\varphi(u)\big)}{u - \varphi(u)} \right| < 1, \quad u \neq 0, \tag{5.4}$$

or even more

$$\left| \frac{\varphi(u) - \varphi\big(\varphi(u)\big)}{u - \varphi(u)} \right| = \frac{\varphi\big(\varphi(u)\big) - \varphi(u)}{u - \varphi(u)} = 1 - \frac{1 - \varphi\big(\varphi(u)\big)/u}{1 - \varphi(u)/u}, \tag{5.5}$$

for all $u \neq 0$. By (2.6) and (5.5) it follows that

$$\sup_{u \neq 0} \left| \frac{\varphi(u) - \varphi\big(\varphi(u)\big)}{u - \varphi(u)} \right| < 1.$$

Hence T is a contraction. □

By the contraction principle the map T has a unique fixed point in X. Denote it by $\psi : J_M \to \boldsymbol{R}$. By construction the function $\psi(u) - \psi\big(\varphi(u)\big)$ is monotone increasing (resp. monotone decreasing) for u positive (resp. negative). More precisely in view of (5.3) and $T(\psi) = \psi$, we have

$$\begin{aligned} &\psi(u) - \psi\big(\varphi(u)\big) = u, && \text{for } u > 0, \\ &\psi(u) - \psi\big(\varphi(u)\big) = -u, && \text{for } u < 0. \end{aligned} \tag{5.6}$$

Therefore the assumption of Propositions 4.1 and 4.2 holds together with the uniform $\mathbf{L}^\infty$ bound (4.8) and the bound for the new functional $\mathbf{V}\big(u_\nu(t)\big)$, i.e. (4.10).

At this point it seemed we could not say anything about the regularity of ψ close to 0. And we will need ψ to be Lipschitz continuous on I_0 to prove equivalence between $\mathbf{TV}$ and $\mathbf{V}$.

Let us consider the second iterate of $T : X \mapsto X$.

LEMMA 5.3. *$T^{[2]} : X \mapsto X$ and is a contraction. Moreover $T^{[2]}$ maps Y into itself.*

PROOF. The first assertion is trivial. Take $\psi_0 \in Y$. By our definition and [H2] it follows that $T^{[2]}\psi$ is increasing (resp. decreasing) for u positive (resp. negative). Iterating (5.3), we get that $T^{[2]}$ is defined by

$$T^{[2]}\psi(u) = \psi\big(\varphi^{[2]}(u)\big) + \operatorname{sgn}(u)\big(u - \varphi(u)\big), \quad u \in \boldsymbol{R}. \tag{5.7}$$

The relation (5.7) together with $\varphi^{[2]}\big(I_0\big) \subset I_0$, imply

$$\operatorname{Lip}_{I_0}\big(T^{[2]}\psi_0\big) \leq 1 + \operatorname{Lip}_{I_0}(\varphi) + \operatorname{Lip}_{\varphi^{[2]}(I_0)}(\psi_0)\cdot\operatorname{Lip}_{I_0}\big(\varphi^{[2]}\big) \leq 1 + L_M + K\eta \leq K, \tag{5.8}$$

by the choice of K. Hence $T^{[2]}\psi_0 \in Y$. □

Now, $T^{[2]}$ is a contraction on X, hence it admits a unique fixed point. Since $T^{[2]}$ maps Y into Y and Y is closed, it follows that this fixed point belongs to Y. Every fixed point of T is also a fixed point of $T^{[2]}$, hence $T^{[2]}$ and T have the *same* fixed point. Thus the fixed point of T belongs to Y and so it is Lipschitz continuous on a neighborhood of 0 and satisfies the monotonicity properties.

REMARK 5.4. The operator T does not map Y into Y. Nevertheless, since $T^{[2]}$ maps Y into Y and φ is Lipschitz, it follows that, for every $\psi \in X$, also the Lipschitz constant of $T^{[2n+1]}\psi$ cannot grow too much as $n \to \infty$.

We point out that if ψ_0 were a fixed point of $T^{[2]}$ only, then we could not recover the relations (5.6). So we need ψ to be a fixed point of *both* T and $T^{[2]}$.

Finally we prove that the functional $\mathbf{V}$ is equivalent to the usual total variation.

LEMMA 5.5. *Given $M > 0$, there exist positive constants C_1, C_2 such that*

$$C_1\,\mathbf{V}(u) \leq \mathbf{TV}(u) \leq C_2\,\mathbf{V}(u) \tag{5.9}$$

for any piecewise constant function u with $\|u\|_{\mathbf{L}^\infty} \leq M$.

PROOF. It is sufficient to prove that

$$C_1\,\sigma(u_l, u_r) \leq |u_r - u_l| \leq C_2\,\sigma(u_l, u_r), \tag{5.10}$$

for all u_l, u_r with $|u_l|, |u_r| \leq M$. Without loss of generality we can assume $u_l > 0$. For all $u_r > 0$, by the monotonicity of ψ and φ we have

$$\big|\psi(u_r) - \psi(u_l)\big| = \big|\psi\big(\varphi(u_r)\big) - \psi\big(\varphi(u_l)\big)\big| + |u_r - u_l| \geq |u_r - u_l|,$$

hence

$$\left|\frac{u_r - u_l}{\sigma(u_l, u_r)}\right| = \left|\frac{u_r - u_l}{\psi(u_r) - \psi(u_l)}\right| \leq 1. \tag{5.11}$$

If, instead, $u_r < 0$ we have

$$\left|\frac{u_r - u_l}{\sigma(u_l, u_r)}\right| \leq \left|\frac{u_l - \varphi(u_l)}{\psi(u_l) - \psi\big(\varphi(u_l)\big)}\right| = \left|\frac{u_l - \varphi(u_l)}{u_l}\right| \leq (1 + L_M) =: C_2, \tag{5.12}$$

since $|\varphi(u)| \leq L_M|u|$ for all $|u| \leq M$.

Next we prove that ψ is Lipschitz continuous on $I := [-M, M]$ (hence also on J_M). First of all, we can assume $M > \varepsilon_0$. Since ψ is a fixed point of $T^{[2]}$ it follows that

$$\psi(u) = \psi\big(\varphi^{[2]}(u)\big) + \operatorname{sgn}(u)\big(u - \varphi(u)\big),$$

which implies

$$\operatorname{Lip}_I(\psi) \leq \operatorname{Lip}_I\big(\varphi^{[2]}\big) \cdot \operatorname{Lip}_{\varphi^{[2]}(I)}(\psi) + \operatorname{Lip}_I(\varphi) + 1. \tag{5.13}$$

Note that by (5.13) the Lipschitz constant of ψ on the interval $[-M, M]$ can be controlled by that on the (strictly) smaller interval $\big[\varphi^{[2]}(-M), \varphi^{[2]}(M)\big]$.

More precisely, even though $\operatorname{Lip}_I(\varphi^{[2]})$ may be greater than 1, it happens that the function $\varphi^{[2]}$ has only one fixed point on $(-\infty, +\infty)$, namely $u = 0$. Hence, having fixed $M > \varepsilon_0$, there exists an integer p such that the iterates $\varphi^{[2p]}(u) \in [-\varepsilon_0, \varepsilon_0]$ for all $|u| \in [\varepsilon_0, M]$, where p depends only on ε_0 and M. By iterating (5.13), this implies that

$$\operatorname{Lip}_I(\psi) \leq K_1 \cdot \operatorname{Lip}_{I_0}(\psi) + K_2, \tag{5.14}$$

where K_1, K_2 are constants depending only on M, ε_0 and the Lipschitz constant of φ. Since $\operatorname{Lip}_{I_0}(\psi) \leq K$, (5.14) says that ψ is Lipschitzian.

Then the conclusion holds with

$$C_1 := \Big(K_1 \cdot \operatorname{Lip}_{I_0}(\psi) + K_2\Big)^{-1}.$$

□

6. Remarks on the Construction

The present result is stronger than the one presented in [**3**]. On one hand we consider a more general flux-function; moreover we drop both the assumption that the solution should coincide with the classical one in a small neighborhood of 0 (see (H2) in [**3**]), and the assumption that α should be decreasing. Concerning this last hypothesis, notice that in the cubic-flux case with the choice $\psi(u) = |u|$ (as we considered in [**3**]) we have

$$\operatorname{sgn}(u)\big(\psi(u) - \psi\big(\varphi(u)\big)\big) = -\alpha(u).$$

So, α is decreasing iff (4.9) holds. This means that the monotonicity request on α comes out by the *particular* choice $\psi(u) = |u|$. The assumption can be drop just by carefully choosing the function ψ.

The choice (4.2) appears to be a sort of nonlinear generalization of the definition of $\sigma(u_l, u_r)$ given in [**3**], the latter corresponding to the case $\psi(u) = |u|$. Unfortunately this last choice does not work in the general case mainly because the flux-function f is not symmetric.

The case $\varphi \equiv \tau$ corresponds to the classical case in which the Oleinik-Liu solutions [**26,27**] are selected. Notice that in view of Lemma 2.1 hypotheses [H1]-[H5] are automatically satisfied. So, we expects that a sufficient condition for the nonclassical solution to be in **BV** is that φ and τ have the same behavior near $u = 0$, roughly speaking $\varphi'(0) = \tau'(0)$. In fact, we could prove existence in **BV** under the weaker hypothesis [H5].

The function $|u|$ on the right-hand side of (5.3) can be replaced by a more general Lipschitz continuous function $G(u)$, i.e. we can look for a function ψ satisfying the equation

$$\psi(u) = \psi\big(\varphi(u)\big) + G(u),$$

with G increasing (resp. decreasing) for u positive (resp. negative), and behaving like $|u|$ for u close to 0. The corresponding function ψ obtained by a fixed-point argument similar to the one presented in the previous section, depends on G and, in general, is *nonlinear*. Indeed, if one tries to use a *piecewise linear* ψ of the form

$$\psi(u) := \begin{cases} \lambda^+ u, & \text{for } u > 0, \\ \lambda^- u, & \text{for } u < 0, \end{cases} \tag{6.1}$$

for some positive λ^+ and negative λ^-, then the condition $\sigma \geq 0$ (more precisely $\operatorname{sgn}(u)\big(\psi(u) - \psi\big(\varphi(u)\big)\big) \geq 0$) implies that $m := \lambda^-/\lambda^+$ must satisfy

$$\sup_{v<0} \left| \frac{\varphi(v)}{v} \right| =: A^- \leq |m| \leq A^+ := \inf_{u>0} \left| \frac{u}{\varphi(u)} \right|.$$

So a necessary condition is $A^- \leq A^+$. If $\varphi(u) = -\alpha u + o(u)$, then the previous condition is violated as long as there exists a state w such that $|\varphi(w)| > \alpha^{-1}|w|$, and this could be the case when the flux is not symmetric. Nevertheless the choice (6.1) works for (1.1) with a symmetric flux function, and in this case one can take $\lambda^+ = -\lambda^- = 1$, provided that $\varphi'(u) > -1$ for all u.

If we are interested only in *small* data it is possible to choose $\psi(u) = |u|$ even for general fluxes and regular φ. Indeed, if $\varphi \in C^1$ and $\varphi'(0) > -1$, then (5.6) reduces to

$$\big[\operatorname{sgn}(u)(\psi - \psi(\varphi))\big]'(u) = \big(1 + \varphi'(u)\big),$$

which is positive for u close to zero.

Finally, our hypothesis [H5] seems to be unavoidable, as there are counterexamples (see Section 7) in which $\varphi'(0) = -1$ and the total variation of the solution blows up in finite time.

7. Examples of Blow-Up of the Total Variation

In this section we present two examples in which hypothesis [H5] does not hold and the total variation of the *exact* nonclassical solution blows up in finite time. For a recent important result about blow-up for systems of conservation laws, see Jenssen [**19**].

Example 7.1. Consider the equation (1.1) with the following flux-function

$$f(u) := \begin{cases} u^h, & u \geq 0, \\ u^k, & u < 0, \end{cases}$$

with $k > h$ odd and greater than 1. It should be stressed that this function does not satisfy our regularity conditions since it is only Lipschitz continuous at the

origin. Nevertheless, the example presented now is of interest since it shows new features not encountered in the classical case. We recall that, when the classical Oleinik entropy condition is enforced, the solution of the Cauchy problem (1.1)-(1.2) with Lipschitz continuous flux has bounded variation and in fact is total variation diminishing.

In the context of nonclassical solutions, we will produce an example where the initial data is in **BV** but the total variation of the solution blows up *instantaneously* at $t = 0$. This actually happens for a particular choice of the kinetic function φ for which $\varphi'(0+) = -\infty$. It should be noticed that also $\tau'(0+) = -\infty$, nevertheless the classical solution exists and is in **BV**. This means that in the case of a Lipschitz continuous flux-function, whether the total variation of the solution blows up or not, is not determined by the value of $\varphi'(0)$ but, as we shall see, can be related to the behavior of the function $\alpha - \varphi$ near $u = 0$.

It is not difficult to see that for u positive $\varphi^*(u) = -u^\gamma$, where $\gamma = \frac{h-1}{k-1} < 1$, and that $\tau(u) < \tau_k(u)$ where $\tau_k(u)$ satisfies (2.1) with $f(u) = u^k$ for all u. Hence $\tau(u) < -C_k u$ for a positive constant C_k depending only on k, and so $\tau(u) < -2u^2$ if for example $0 < 2u < C_k$. Choose now $\alpha(u) = -u^2 > \tau(u)$ for all $u \geq 0$. It follows that $\varphi(u) = -u^\gamma\big(1 + O(u)\big)$, hence

$$\alpha(u) - \varphi(u) \geq -u^2 + \frac{1}{2}u^\gamma,$$

if u is sufficiently small.

Choose now an integer n_0 such that $1/n_0^\beta < C_k$ where $\beta = 1/\gamma > 1$. Take the initial data of the form

$$u_0(x) := \begin{cases} 1/n_0^\beta, & \text{if } x \in (-\infty, 2n_0], \\ 1/n^\beta, & \text{if } x \in (2n, 2n+1),\ n \geq n_0, \\ -2/n^{2\beta}, & \text{if } x \in \big(2n+1, 2(n+1)\big),\ n \geq n_0. \end{cases}$$

An easy estimate implies that

$$\mathbf{TV}(u_0) \leq 4 \sum_{n=n_0}^{\infty} \frac{1}{n^\beta} < \infty.$$

For small positive t, the solution is obtained just by solving the Riemann problems at each discontinuity point in u_0. Notice that $-2/n^{2\beta} > \tau(1/n^\beta)$, hence the Riemann problem with data $(1/n^\beta, -2/n^{2\beta})$ is solved by a nonclassical shock from $1/n^\beta$ to $\varphi\big(1/n^\beta\big)$ followed by a classical shock from $\varphi\big(1/n^\beta\big)$ to $-2/n^{2\beta}$. In particular it follows that

$$\begin{aligned} \Delta\mathbf{TV}(0) &\geq \sum_{n=n_0}^{\infty} \big(\alpha(1/n^\beta) - \varphi(1/n^\beta)\big) \\ &\geq \frac{1}{2} \sum_{n=n_0}^{\infty} \frac{1}{n^{\gamma\beta}} - \sum_{n=n_0}^{\infty} \frac{1}{n^{2\beta}} \geq \frac{1}{2} \sum_{n=n_0}^{\infty} \frac{1}{n}. \end{aligned}$$

This implies that $\Delta\mathbf{TV}(0) = +\infty$ hence $\mathbf{TV}(0+) = +\infty$.

Finally, notice that $u(t, \cdot) \notin \mathbf{BV}$ but $u(t, \cdot) \in \mathbf{BV}_{loc}$.

EXAMPLE 7.2. Now we will take $f(u) = u^3$, so our hypotheses on the flux-function are satisfied. Since the total variation of the solution of the Riemann problem (u_l, u_r) depends in a Lipschitz continuous way on $|u_l - u_r|$, it appears that in this case the total variation can not blow up instantaneously. In fact, we shall prove that for suitable initial data u_0 and choice of the kinetic function φ, there exists a time $\bar{t}$ such that

$$\mathbf{TV}\big(u(t,\cdot)\big) = +\infty,$$

for all $t \geq \bar{t}$, where u is the solution of (1.1)-(1.2).

We shall consider the case $\varphi(u) = \varphi^*(u) = -u$ for all u, hence $\varphi(u)$ does not satisfy [H5]. In this situation *every* Riemann problem with $u_l u_r < 0$ generates a nonclassical shock; more precisely the solution is given by a nonclassical shock connecting u_l to $-u_l$ followed by a classical shock connecting $-u_l$ to u_r, no matter how small u_r is. This means that arbitrarily small oscillations near 0 can produce nonclassical shocks of arbitrarily large strength. For related results connected with the study of radially symmetric systems, see the works of Freistühler, for instance in [**10, 11**].

Now let us construct initial data for which the total variation of the solution blows up. We define $u_0(x)$ to be equal to 1 for $x < 0$ and equal to 0 for $0 < x < x_0 := 1$. In x_0 a rarefaction will originate. First of all, the Riemann problem in $x = 0$ is solved by a single classical shock traveling with speed $\lambda_0 := \lambda(1, 0) = 1$. We want to define inductively points x_n, y_n and states u_n such that $x_{n-1} < y_n < x_n$ for all n and u_0 is given by

$$u_0(x) := \begin{cases} u_n, & \text{if } x_{n-1} < x < y_n, \\ 0, & \text{if } y_n < x < x_n, \\ 0, & \text{if } x > \sup_n x_n. \end{cases} \tag{7.1}$$

The idea is the following: start at x_0 and take u_1 small and negative to be defined later. The Riemann problem at x_0 is solved by a rarefaction wave which will interact with the original shock outgoing from the origin, at the point $P_0 := (1, 1)$ in the (x, t)-plane. This interaction will produce a slower nonclassical shock connecting 1 to -1 and a faster classical shock which will interact with the rarefaction until the point P_1 (see Figure 7.1). Let x_1 be the x-coordinate of point P_1. Now, draw back the line with slope $\lambda(u_1, 0)$ passing through P_1. Let y_1 be the x-coordinate of the intersection point between this line and the x-axis. Notice that $0 < \lambda(u_1, 0) < \lambda(u_1)$ hence we have $x_0 < y_1 < x_1$. Moreover the Riemann problem at y_1 is actually solved by the shock traveling with speed $\lambda(u_1, 0)$. Since P_1 depends only on the speed at the right of the rarefaction (that is $\lambda(u_1) = 3u_1^2$), then it is clear that once u_1 is known, so x_1, y_1 are.

Let us now proceed inductively: assume points x_n, y_n and value u_{n+1} are given and assume that the rarefaction originating at x_n interacts with the shock originating at y_n at the point P_n, producing a nonclassical shock connecting $(-1)^n$ to $(-1)^{n+1}$ traveling with speed 1 and a classical shock interacting with the previous rarefaction until point P_{n+1}. As before let x_{n+1}, y_{n+1} be the x-coordinates of the point P_{n+1} and the intersection-point between the x-axis and the line with slope $\lambda(u_{n+1})$ passing through P_{n+1}, respectively. Again $x_n < y_{n+1} < x_{n+1}$.

We notice that each interaction at P_n generates a nonclassical shock between the states 1 and -1, traveling with speed 1. Hence these fronts will never interact

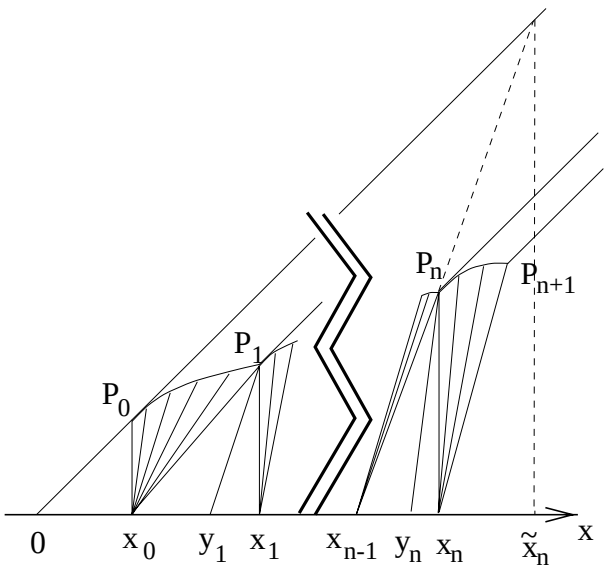

FIGURE 7.1

in the future. If we can generate infinitely many nonclassical shocks in finite time, then the total variation of the solution will blow up.

This is achieved by suitably choosing the states u_n in such a way that the sequence x_n converges to some finite $\bar{x}$. We define u_n inductively by letting

$$u_{n+1} := (-1)^{n+1} \cdot \left| \lambda^{-1} \left(\frac{1}{1 + 2^n x_n} \right) \right|.$$

Indeed, let $\tilde{x}_n$ be the x-coordinate of the intersection point between the nonclassical shock originating at P_0 and the line with slope $\lambda(u_n)$ and originating at $(x_{n-1}, 0)$ (see Figure 7.1). Then an easy computation gives

$$x_{n+1} - x_n \leq \tilde{x}_{n+1} - x_n = \frac{\lambda(u_{n+1})}{1 - \lambda(u_{n+1})} x_n = \frac{1}{2^n}.$$

This implies that $x_n \to \bar{x} \leq \sum_{n=0}^{\infty} 1/2^n = 2$. It is easy to see that the points P_n will converge to a point $\bar{P} = (\bar{t}, \bar{x})$ with $\bar{t} \leq \bar{x}$. By construction, at time $\bar{t}$ the solution will have infinitely many nonclassical shocks connecting the states 1 and -1, hence $\mathbf{TV}\big(u(\bar{t}, \cdot)\big) = \infty$, and since they will never interact in the future, this is true even for all $t > \bar{t}$. On the other hand we have

$$\mathbf{TV}\big(u(0, \cdot)\big) = 1 + 2 \sum_{n=1}^{\infty} \left| \lambda^{-1} \left(\frac{1}{1 + 2^n x_n} \right) \right| \leq 1 + \frac{2}{\sqrt{3}} \sum_{n=1}^{\infty} \left(\frac{1}{\sqrt{2}} \right)^n < \infty.$$

REMARK 7.3. It is possible to construct an example similar to Example 7.2, when we request only the existence of a single point $\bar{u} > 0$ such that $\varphi^{[2]}(\bar{u}) = \bar{u}$, and even if [H4] does not hold.

References

[1] R. Abeyaratne and J. Knowles, *Kinetic relations and the propagation of phase boundaries in solids*, Arch. Rational Mech. Anal. **114** (1991), 119–154.

[2] R. Abeyaratne and J. Knowles, *Implications of viscosity and strain gradient effects for the kinetics of propagating phase boundaries in solids*, SIAM J. Appl. Math. **51** (1991), 1205–1221.

[3] D. Amadori, P. Baiti, P.G. LeFloch and B. Piccoli, *Nonclassical shocks and the Cauchy problem for nonconvex conservation laws*, J. Differential Equa. **151** (1999), 345–372.

[4] P. Baiti and H.K. Jenssen, *On the front-tracking algorithm*, J. Math. Anal. Appl. **217** (1998), 395–404.

[5] P. Baiti, P.G. LeFloch and B. Piccoli, *Uniqueness of classical and nonclassical solutions for nonlinear hyperbolic systems*, preprint.

[6] P. Baiti, P.G. LeFloch and B. Piccoli, *Nonclassical shocks and the Cauchy problem: Strictly hyperbolic systems*, in preparation.

[7] A. Bressan, *Global solutions to systems of conservation laws by wave-front tracking*, J. Math. Anal. Appl. **170** (1992), 414–432.

[8] C. Dafermos, *Polygonal approximations of solutions of the initial value problem for a conservation law*, J. Math. Anal. Appl. **38** (1972), 33–41.

[9] R.J. DiPerna, *Global existence of solutions to nonlinear hyperbolic systems of conservation laws*, J. Differential Equa. **20** (1976), 187–212.

[10] H. Freistühler, *Non-uniformity of vanishing viscosity approximation*, Appl. Math. Lett. **6** (1993), 35–41.

[11] H. Freistühler, *On the Cauchy problem for a class of hyperbolic systems of conservation laws*, J. Differential Equa. **112** (1994), 170–178.

[12] J. Glimm, *Solutions in the large for nonlinear hyperbolic systems of equations*, Comm. Pure Appl. Math. **18** (1965), 697–715.

[13] B.T. Hayes and P.G. LeFloch, *Nonclassical shocks and kinetic relations: Scalar conservation laws*, Arch. Rational Mech. Anal. **139** (1997), 1–56.

[14] B.T. Hayes and P.G. LeFloch, *Nonclassical shocks and kinetic relations: Strictly hyperbolic systems*, SIAM J. Math. Anal., to appear.

[15] B.T. Hayes and P.G. LeFloch, *Nonclassical shocks and kinetic relations: Finite difference schemes*, SIAM J. Numer. Anal. **35** (1998), 2169–2194.

[16] B.T. Hayes and M. Shearer, *Undercompressive shocks for scalar conservation laws with nonconvex fluxes*, Preprint, North Carolina State Univ., 1997.

[17] D. Jacobs, W.R. McKinney and M. Shearer, *Traveling wave solutions of the modified Korteweg-deVries Burgers equation*, J. Differential Equa. **116** (1995), 448–467.

[18] R.D. James, *The propagation of phase boundaries in elastic bars*, Arch. Rational Mech. Anal. **73** (1980), 125–158.

[19] H.K. Jenssen, *Blow-up for systems of conservation laws*, Preprint, 1998.

[20] S.N. Kružkov, *First order quasilinear equations in several independent variables*, Math. USSR Sbornik **10 (2)** (1970), 217–243.

[21] P.D. Lax, *Hyperbolic systems of conservation laws II*, Comm. Pure Appl. Math. **10** (1957), 537–566.

[22] P.D. Lax, *Shock waves and entropy*, Contributions to Nonlinear Functional Analysis, ed. E.A. Zarantonello, Academic Press, New York, 1971, pp. 603–634.

[23] P.G. LeFloch, *Propagating phase boundaries: formulation of the problem and existence via the Glimm scheme*, Arch. Rational Mech. Anal. **123** (1993), 153–197.

[24] P.G. LeFloch, *An introduction to nonclassical shocks of systems of conservation laws*, Proceedings of the "International School on Theory and Numerics for Conservation Laws", Freiburg/Littenweiler (Germany), 20–24 October 1997, ed. D. Kröner, M. Ohlberger and C. Rohde, Lecture Notes in Computational Science and Engineering (1998), pp. 28–72.

[25] P.G. LeFloch and C. Rohde, *High-order schemes, entropy inequalities and nonclassical shocks*, in preparation.

[26] T.P. Liu, *Admissible solutions of hyperbolic conservation laws*, Memoir Amer. Math. Soc. 240, 1981.

[27] O. Oleinik, *Discontinuous solutions of nonlinear differential equations*, Usp. Mat. Nauk. 12 (1957), 3–73 (in Russian); English transl. in Amer. Math. Soc. Transl. Ser. 2 (26), 95–172.

[28] N.H. Risebro, *A front-tracking alternative to the random choice method*, Proc. Amer. Math. Soc. **117** (1993), 1125–1139.
[29] M. Shearer, *The Riemann problem for a class of conservation laws of mixed type*, J. Differential Equa. **46** (1982), 426–443.
[30] M. Slemrod, *Admissibility criteria for propagating phase boundaries in a van der Waals fluid*, Arch. Rational Mech. Anal. **81** (1983), 301–315.
[31] L. Truskinovsky, *Dynamics of non-equilibrium phase boundaries in a heat conducting nonlinear elastic medium*, J. Appl. Math. Mech. (PMM) **51** (1987), 777–784.
[32] L. Truskinovsky, *Kinks versus shocks*, Shock induced transitions and phase structures in general media, R. Fosdick, E. Dunn, and M. Slemrod ed., IMA Vol. Math. Appl. 52, Springer-Verlag, 1993.
[33] A.I. Volpert, *The space BV and quasilinear equations*, Math. USSR Sb. **2** (1967), 225–267.

Università di Padova, Via G.B. Belzoni 7, 35131 Padova, Italy.
E-mail address: baiti@math.unipd.it

Centre de Mathématiques Appliquées & Centre National de la Recherche Scientifique, U.M.R. 7641 Ecole Polytechnique, 91128 Palaiseau Cedex, France.
E-mail address: lefloch@cmapx.polytechnique.fr

S.I.S.S.A, Via Beirut 4, 34014 Trieste, Italy.
E-mail address: piccoli@sissa.it

Contemporary Mathematics
Volume **238**, 1999
B 0-8218-1196-7-03537-4

The Harnack Inequality and Non-Divergence Equations

Luis A. Caffarelli

The Harnack inequality is one of the central properties of solutions of linear (or appropriate nonlinear) second order elliptic equations in divergence or nondivergence form.

It is of particular interest when the coefficients are merely measurable, i.e., a solution u must satisfy

$$(*) \qquad D_i a_{ij}(x) D_j u = 0,$$

or

$$(**) \qquad a_{ij}(x) D_{ij} u = 0 \ ,$$

and

$$A(x) = a_{ij}(x)$$

is assumed only to be a measurable strictly elliptic matrix, i.e.,

$$\lambda \|y\|^2 \leq y^T A y \leq \Lambda \|y\|^2$$

for all values of x.

The importance of $A(x)$ being only measurable, is that constitutes in itself a class invariant under dilation. No matter how we try to "blow up" our solution, by dilations

$$u(x) = \mu u(\lambda x)$$

we remain in the same class, always far, from say, constant coefficients.

That is why such equations (and the Harnack inequality) played such an important rôle in the theory of non-linear (far from linear) equations, where there is no hope of "freezing the coefficients".

The Harnack inequality then states

HARNACK INEQUALITY. Let u be an appropriate weak solution of $(*)$, resp. $(**)$.

If $u \geq 0$ in $B_r(x_0)$, then

$$\sup_{B_{r/2}(x_0)} u \leq C \inf_{B_{r/2}(x_0)} u \ ,$$

1991 *Mathematics Subject Classification.* Primary 35B45, 35J15, 35J60.
Partially supported by an NSF grant.

where $B_r(x)$ is the ball of radius r and center x.

In case $(*)$, if u is a solution, $u-b$ is also a solution for any constant b.

In case $(**)$, if u is a solution $u-\ell$ is also a solution for any linear function ℓ.

In [**C**] (see also [**CC**]), following the Crandall-Lions theory of viscosity solutions to non divergence equations, we defined the class $\mathcal{S}=\mathcal{S}_\Lambda$ of "all viscosity solutions to some elliptic non divergence equations of ellipticity Λ".

More precisely,

DEFINITION 1. The continuous function $u\in \underline{\mathcal{S}}_\Lambda$ (i.e., is a "subsolution") iff whenever a quadratic polynomial P touches u by above at some point x_0,

$$\|[D^2P]^-(x_0)\|\le \Lambda\|[D^2P]^+(x_0)\|$$

and u belongs to $\mathcal{S}$, if both $u,-u$ belong to $\underline{\mathcal{S}}$.

Here touches by above, means $P\ge u$ in a neighborhood of x_0, and $P(x_0)=u(x_0)$. Also $(D^2P)^\pm$ denotes the positive (negative) part of the symmetric matrix D^2P. Using the Krylov-Safanov technique ([**K-S**]) we proved

THEOREM 1. *If $u\in\mathcal{S}$, it satisfies the Harnack inequality with a constant $C=C(\Lambda)$.*

Since, if $u\in\mathcal{S}_\Lambda$, $u-\ell\in\mathcal{S}_\Lambda$ for any linear function ℓ, actually, $u-\ell$ satisfies the Harnack inequality. The converse is elementary.

THEOREM 2. *Assume that u is continuous and for any linear function ℓ, $u-\ell$ satisfies the Harnack inequality, with constant C. Then*

$$u\in\mathcal{S}_\Lambda$$

with $\Lambda=\Lambda(C)$.

PROOF. Let P be a quadratic polynomial, tangent by below to u at the origin. By subtracting the linear part, ℓ, we may assume that

$$P=\sum_{\alpha_i>0}\alpha_i x_i^2+\sum_{\beta_i<0}\beta_i x_i^2$$

and $P\le u$ in $B_h(0)$.

In particular

$$v=u-\min\beta_i h^2\ge 0\ \text{ in }\ B_h(0)\ .$$

Thus, by Harnack inequality,

$$\sup_{B_{h/2}} v\le Cv(0)=C[-\min(\beta_i)h^2]\ .$$

But, always in B_h,

$$v=u-\min\beta_i h^2\ge P-\min\beta_i h^2\ .$$

In particular

$$\alpha_{\max}\left(\frac{h}{2}\right)^2\le(1+C)[-\min\beta_i]h^2 \qquad \square$$

A consequence of the argument above is the following remark:

COROLLARY 1. *Assume $u-\ell$ satisfies the Harnack inequality for every ℓ. If u has in $B_h(X_0)$ a tangent polynomial by below, i.e.,*

a) $u - P \geq 0$ *on* $B_h(X_0)$

b) $u - P = 0$ *at* X_0.

Then u has a tangent polynomial by above in $B_{h/2}(X_0)$,

$$Q = \ell + A|X - X_0|^2$$

with $A \leq C\|D^2P\|$, and ℓ the linear part of P at X_0.

PROOF. In the ball of radius ρ $\rho < h$), $v = u-\ell(X)+\|D^2P\|\rho^2$, is a nonnegative function that satisfies the Harnack inequality. Since $v(0) = \|D^2P\|\rho^2$, $\sup_{B_{\rho/2}} v \leq C\|D^2P\|\rho^2$.

We would like to discuss now a more interesting case, that arises for instance in homogenization (see [**C1**]).

In that case, one has a function u_0, that, being a uniform limit of solutions u_ε to highly oscillatory equations, has inherited the following two properties

a) $(u_0 - \ell)$ satisfies the Harnack inequality for every ℓ,

b) $u_0(X) - u_0(X - X_0)$ satisfies the Harnack inequality for every X_0.

We want to show that in this case, u_0 satisfies a fully non linear uniformly elliptic equation

$$F(D^2u, \nabla u) = 0 \ .$$

Although not necessary for this discussion, we start by pointing out the following Theorem, due to Cabre and Caffarelli.

THEOREM 3. (see [**C-C**], Lemma 5.6) *Assume that, for any translation X_0,*

$$v(X) = u(X) - u(X - X_0)$$

satisfies the Harnack inequality. Then u is locally $C^{1,2}$.

We would like now to prove

THEOREM 4. *Let u be a continuous function in the unit ball B_1, such that*

a) $|u| \leq 1$

b) $u \in \mathcal{S}_\Lambda$, *and for any X_0, for any constant C,*

$$v_{X_0} = u(X) - u(X - X_0) + C$$

satisfies the Harnack inequality.

Then, there exists a second order non linear operator $F(D^2\omega, D\omega)$, with $F(0, P) \equiv 0$, uniformly elliptic such that $u|_{B_1}$ is a viscosity solution of

$$F(D^2u, Du) = 0.$$

PROOF. The proof consists in building such an operator. Of course F is not unique: a linear function, ℓ, satisfies every possible non linear operator with $F(0, p) = 0$.

We start by recalling a basic Pucci extremal operator: For θ large, to be fixed, depending on Λ we define

$$\mathcal{P}(D^2u) = \sum_{\lambda_j>0} \lambda_j + \theta \sum_{\lambda_j<0} \lambda_j,$$

where λ_j denote the eigenvalues of D^2u.

For any value of the vector $\vec{p}$ in R^n consider now the set of quadratic polynomials P, tangent by above to u at some point X_0, with $\nabla P(X_0) = \vec{p}$. That is:

$$T^+(\vec{p}) = \begin{cases} \text{i) } P \text{ quadratic,} \\ P/ \quad \text{ii) } P \text{ tangent by above to } u \text{ at a point } X_0 \in B_1, \\ \text{iii) } \nabla P(X_0) = p \ . \end{cases}$$

Similar definition for $T^-(\vec{p})$.

Define

$$\widetilde{F}(D^2u, Du) = \sup_{P \in T^+(\nabla u)} \mathcal{P}(D^2u - D^2p).$$

We first note that $\widetilde{F}$ is finite.

Indeed $\mathcal{P}$ is a Lipschitz function on D^2u, with $\|\mathcal{P}\|_{\text{Lip}} = \theta$ so we just need to pin down $\widetilde{F}$ at zero. But we recall, from the proof of Theorem 2, that P being tangent by above to u, implies

$$|\lambda_{\min}(P)| \leq C|\lambda_{\max}(P)| \ .$$

Therefore

$$\mathcal{P}(-D^2P) \leq 0$$

if $\theta \geq nC$.

This makes, for every q,

$$\widetilde{F}(0, \vec{q}) \leq 0 \ .$$

To make $\widetilde{F}(0, \vec{q}) = 0$, we modify it to

$$F(D^2u, Du) = \max(\widetilde{F}(D^2u, Du), \mathcal{P}(D^2u)) \ .$$

This makes of F a uniformly elliptic a function, Lipschitz on D^2u, with

a) $\|F\|_{\text{Lip}} \leq \theta$

b) $F(0), \ ) = 0$.

Let us check that u is a viscosity solution.

If P is tangent by above to u at x_0,

$$\mathcal{P}(D^2P - D^2P) = 0 \ .$$

Thus $F \geq 0$.

If Q is tangent to u be below at X_1 we make two observations

a) Always from Theorem 2,

$$|D^2P)^+| \leq C|D^2P^-| \ .$$

So, since we chose $\theta > nC$,

$$\mathcal{P}(D^2Q) \leq 0 \ .$$

b) If $\nabla Q(X_1) = \vec{q}$ and $P \in T^+(q)$, i.e., is tangent to u, by above, at some point X_0 and $\nabla P(X_0)$ is also $\vec{q}$, then

$$\widetilde{P} = P(X - X_0) - Q(X - X_1)$$

is tangent to

$$u(X - X_0) - u(X - X_1)$$

by above at the origin, and $\nabla \widetilde{P}(0) = 0$.

Again, from the proof of Theorem 2

$$|D^2(P-Q)^-| < C|D^2(P-Q)^+|.$$

Therefore

$$\mathcal{P}(Q-P) \leq 0$$

for any Q in $T^-(\vec{q})$ and any P in $T^+(\vec{q})$.

Therefore

$$F(D^2Q, DQ) \leq 0$$

for any Q in T.

This completes the proof that u is a viscosity solution of

$$F(D^2w, Dw) = 0 \ .$$

Finally, we add the possibility of subtracting linear functions from $u_0(X) - u_0(X - X_0)$.

THEOREM 5. *Assume that*

$$u_0(X) - u_0(X - X_0) - \ell(X)$$

satisfies the Harnack inequality for any linear function ℓ. Then $u_0(X)$ is a viscosity solution of a fully nonlinear, uniformly elliptic equation

$$F(D^2u) = 0.$$

REMARK. The necessary condition in Theorem 5 is also sufficient (see [**C-C**] Theorem 5.3).

Further $u_0 \in C^{1,\alpha}$ (see Corollary 5.7 of [**C-C**]).

PROOF OF THEOREM 5. The proof is the same as that of Theorem 4, but we now construct

$$\widetilde{F}(D^2u) = \sup_{P \in T^+} \mathcal{P}(D^2u - D^2P)$$

where the sup is now taken over all possible P in T^+, disregarding the value of ∇P at the contact point.

The last characterization we would like to discuss concerns solution to equations

$$F(D^2u) = 0$$

with F uniformly elliptic and concave.

It is shown in [**C-C**], that viscosity solutions, u of such equations have the property that any convex combination of its translations ($\lambda_i \geq 0$, $\sum \lambda_i = 1$)

$$v = \sum \lambda_i u(x - x_i)$$

is a viscosity subsolution of the same equation

$$F(D^2v) \geq 0$$

It is also shown that the difference of a sub and supersolution, in particular, $v - u$, belongs to the class $\underline{\mathcal{S}}$ of subsolutions to some elliptic operator. (Heuristically, pure second derivatives of u belong to $\underline{\mathcal{S}}$.)

We prove the inverse characterization.

THEOREM 6. *Let u be a continuous function in B_1 such that*

a) $u_0 \in \mathcal{S}$.

b) $u_0(x+x_0) - u_0(x) \in \mathcal{S}$, *for any* x_0.

c) *For any convex combination*

$$v(x) = \Big[\sum \lambda_i u(x - x_i) \Big] - u(x)$$

belongs to $\underline{\mathcal{S}}$.

Then u is a $(C^{2,\alpha})$ solution of an equation

$$F(D^2 u) = 0$$

with F concave.

PROOF. We will first construct the convex, uniformly elliptic level surface

$$F(M) = 0.$$

The full F can then be constructed by, for instance, linear extension in the direction of the identity.

More precisely, for every P in T^+ consider the cone

$$\Gamma_P = \{M : \mathcal{P}(M - D^2 P) \geq 0\}$$

with $\mathcal{P}$, as before defined by

$$\mathcal{P}(M) = \sum_{\lambda_i \geq 0} \lambda_i + \theta \sum_{\lambda_i \leq 0} \lambda_i$$

for some large $\theta(n)$ to be chosen later.

Next, let

$$\widetilde{\Gamma} = \Big(\bigcup_{P \in T^+} \Gamma_P \Big).$$

And finally, D, the convex envelope

$$D = \Big\{ M : M = \sum_1^{n \times n} p_i Q_i \text{ with } Q_i \in \widetilde{\Gamma} \text{ and } p_i \geq 0 \ , \ \sum p_i = 1 \Big\} \ .$$

Since all of the $\partial \Gamma_P$ define uniformly elliptic operators, so does $\partial \widetilde{\Gamma}$ and ∂D.

Let us show that

a) If $P \in T^+$, $D^2 P \in D$,

b) If $P \in T^-$, $D^2 P \notin D$.

That $D^2 P$, for any $P \in T^+$, belong to D happens by definition, since they belong to $\widetilde{\Gamma}$.

Assume that $Q \in T^-$ and $D^2 Q \in D$. Let us get a contradiction. Indeed, there are $Q_i \in \Gamma_{P_i}$, such that

$$Q = \sum_1^{n \times n} p_i Q_i$$

Let x_i be the points where P_i is tangent by above to u, and consider the function

$$v = \sum p_i u(x - x_i) - u(x - x_Q)$$

(with x_Q the point where Q is tangent to u). Then v has

$$\sum_{P_i} p_i P_i - Q$$

as a polynomial tangent by above at zero, that is

$$\sum_{1}^{n\times n} p_i(P_i - Q_i)$$

must satisfy, v belonging to the class $\bar{S}$, that the positive part of $D^2(\sum_1^{n\times n} p_i(P_i - Q_i)$ controls the negative part, i.e.

$$\|[D^2(\sum p_i(P_i - Q_i)]^-\| \leq C\|[D^2(\sum p_i(P_i - Q_i)]^+\|.$$

But according to the definition of $\mathcal{P}$,

$$\|[D^2(P_i - Q_i)]^-\| \geq \theta\|D^2(P_i - Q_i)\|^+$$

with θ large to be chosen. Since the number of matrices involved in the sum is a fixed one, $(n \times n)$, we have that

$$\|[\sum D^2 p_i(P_i - Q_i)]^-\| \geq \|[D^2(p_1(P_1 - Q_1))]^-\| - \sum \|D^2(p_i(P_i - Q_i)^+\|$$

If we choose $p_1(P_1 - Q_1)$ the one for which

$$\|[D^2(p_i(P_i - Q_i)]^-\|$$

is maximum, we get from the bounds above, that

$$\begin{aligned}\|[D^2(\sum p_i(P_i - Q_i))]^-\| &\geq \Big(1 - \frac{n\times n}{\theta}\Big)\|[D^2(p_1(P_1 - Q_1))]^-\| \\ &\geq \frac{\theta}{n\times n}\Big(1 - \frac{n\times n}{\theta}\Big)\|[D^2(\sum p_i(P_i - Q_i))]^+\|\end{aligned}$$

or

$$\|[D^2 \sum p_i(P_i - Q_i)]^-\| \geq \Big(\frac{\theta}{n\times n} - 1\Big)\|[D^2 \sum p_i(P_i - Q_i)]^+\|,$$

a contradiction to the fact that $v \in \underline{\mathcal{S}}$ if we choose θ large.

Finally, let us point out that if we choose P^+,

$$P^{\pm} = \pm\|u\|_{L^\infty(B_1)}|x|^2$$

an appropriate vertical translation of $P^{\pm}$ is tangent to u by above (resp. below) at some point.

Thus the Lipschitz graph ∂D is controlled by above and below at the origin. This completes the proof.

References

[C] Caffarelli, L.A., *Interior a priori estimates for solutions of fully nonlinear equations*, Annals of Math. **130** (1989), 189–213.

[C1] Caffarelli, L.A., *A note on nonlinear homogenization*, CPAM (to appear).

[C-C] Caffarelli, L.A. and Cabre, X., *Fully nonlinear equations*, Colloquium Publications, Vol.43, AMS, 1995.

[KS] Krylov, N.V. and Safonov, M.V., *An estimate of the probability that a diffusion process hits a set of positive measure*, Dokl. Akad. Nauk. SSSR **245** (1979).

Department of Mathematics, The University of Texas at Austin, Austin, TX 78712-1082

Contemporary Mathematics
Volume **238**, 1999
B 0-8218-1196-7-03538-6

Vanishing Viscosity Limit for Initial-Boundary Value Problems for Conservation Laws

Gui-Qiang Chen and Hermano Frid

Abstract. The convergence of the vanishing viscosity method for initial-boundary value problems is analyzed for nonlinear hyperbolic conservation laws through several representative systems. Some techniques are developed to construct the global viscous solutions and establish the H^{-1} compactness of entropy dissipation measures for the convergence of the viscous solutions with general initial-boundary conditions. The representative examples considered include the systems of isentropic gas dynamics, nonlinear elasticity, and chromatography.

1. Introduction

We are concerned with the convergence of the vanishing viscosity method for initial-boundary value problems for nonlinear hyperbolic systems of conservation laws. Physical motivation is the vanishing viscosity limit from viscous compressible fluids to the inviscid ones with initial-boundary conditions, which is a natural way to determine entropy solutions for the inviscid equations in the interior of fluid domains under consideration. Our analysis focuses on several representative examples of nonlinear systems including the models for isentropic gas dynamics, nonlinear elasticity, and chromatography. The main objective in addressing the particular systems is to expose a general procedure of analysis for such a problem with general boundary conditions, especially *nonhomogeneous* ones. In order to make the main steps clearer and to avoid superfluous technicalities, we restrict our analysis here to the standard domains of form $Q_T = (0,T)\times(0,1)$ or $Q = (0,\infty)\times(0,1)$. Many other examples may be treated by following the same procedure.

We start with a general system of conservation laws in one space variable:

$$\partial_t u + \partial_x f(u) = 0, \qquad (t,x)\in Q, \tag{1.1}$$

1991 *Mathematics Subject Classification.* Primary: 35L65, 35L50; Secondary: 35B25, 35L67.
Key words and phrases. Initial-boundary value problems, vanishing viscosity limit, conservation laws, convergence, compactness, entropy solutions, estimates.

with $u \in U \subset \mathbf{R}^m$, $f \in C^1(U;\mathbf{R}^m)$, for some domain $U \subset \mathbf{R}^m$. We consider the following initial-boundary conditions for (1.1):

$$u|_{t=0} = u_0(x), \tag{1.2}$$

$$u|_{x=0} = a_0(t), \qquad u|_{x=1} = a_1(t). \tag{1.3}$$

We assume that the initial-boundary data satisfy

$$u_0 \in L^\infty((0,1);\mathbf{R}^m), \qquad a_0, a_1 \in L^\infty((0,\infty);\mathbf{R}^m). \tag{1.4}$$

Since the boundary layers generally exist for arbitrarily given $a_i, i = 1, 2$, our focus here is to expose a procedure to construct weak entropy solutions in the interior of Q such that the solutions obtained are natural for the case that there is no boundary layer (see [**5**]). With this in mind, a definition of entropy solutions for the initial-boundary value problem was given in [**5**] for general multidimensional systems of conservation laws in more general (not necessarily cylindrical) domains, motivated from [**1**] for the scalar case. Some discussions about the initial-boundary problem in different contexts, related to [**1**], have been made for hyperbolic systems of conservation laws (*e.g.* [**2, 13, 19, 20, 30, 34, 36**]).

We say that $\eta \in C^1(\mathbf{R}^m)$ is an *entropy* for (1.1), with associated *entropy flux* $q \in C^1(\mathbf{R}^m)$, if

$$\nabla q(u) = \nabla \eta(u) \nabla f(u). \tag{1.5}$$

We call $F(u) = (\eta(u), q(u))$ an *entropy pair*. If $\eta(u)$ is convex, we say $F(u)$ is a convex entropy pair. An entropy pair $\mathcal{F}(u,v) = (\alpha(u,v), \beta(u,v))$ is called a *boundary entropy pair* if, for each fixed $v \in \mathbf{R}^m$, $\alpha(u,v)$ is convex with respect to u, and

$$\alpha(v,v) = \beta(v,v) = \partial_u \alpha(v,v) = 0. \tag{1.6}$$

We say that $\mathcal{F}(u,v) = (\alpha(u,v), \beta(u,v))$ is a *generalized boundary entropy pair* if it is the uniform limit of a sequence of boundary entropy pairs over compact sets.

Definition 1.1. We say that $u \in L^\infty(Q_T;\mathbf{R}^m)$ is a *weak entropy solution* of (1.1)-(1.3) in Q_T if it satisfies

- Conservation Laws (1.1): For all $\phi \in C_0^\infty(Q_T)$, $\phi \geq 0$, and any convex entropy pair (η, q),

$$\iint_{Q_T} (\eta(u)\partial_t \phi + q(u)\partial_x \phi)\, dx\, dt \geq 0; \tag{1.7}$$

- Initial Condition (1.2):

$$\operatorname*{ess\,lim}_{t\to 0+} \int_0^1 |u(t,x) - u_0(x)|\, dx = 0; \tag{1.8}$$

- Boundary Condition (1.3): For any $\gamma(t) \in L^1(0,T)$, $\gamma(t) \geq 0$, a.e., and any boundary entropy pair $\mathcal{F} = (\alpha, \beta)$,

$$\operatorname*{ess\,lim}_{x\to 0+} \int_0^T \beta(u(t,x), a_0(t))\gamma(t)\, dt \leq 0, \quad \operatorname*{ess\,lim}_{x\to 1-} \int_0^T \beta(u(t,x), a_1(t))\gamma(t)\, dt \geq 0. \tag{1.9}$$

To illustrate some features of the above definition, we consider a simple example. Let (1.1) be strictly hyperbolic, which means that the Jacobian ∇f is diagonalizable and all eigenvalues are real and distinct. In (1.2)-(1.3), let a_0, u_0, and a_1 be constant states u_l, u_m, and u_r, respectively, such that u_l and u_m are connected by a shock with a negative speed, while u_m and u_r are connected by a shock with a positive speed. To simplify, we assume that both shocks belong to genuinely nonlinear families, that is, the corresponding eigenvalues of ∇f are monotone along the integral curves of the associated right-eigenvectors (*cf.* [**21**]). Then it is easy to prove that $u(t,x) \equiv u_m$, $(t,x) \in Q$, is an entropy solution for the corresponding initial-boundary value problem, according to the above definition, provided that u_l, u_m, and u_r are sufficiently close to each other.

Indeed, (1.7) and (1.8) are trivially satisfied. For (1.9), it suffices to check the inequality for the left boundary, $x = 0$, since the other is similar. Then this reduces to showing that $\beta(u_m, u_l) \le 0$. Now, from the Lax shock condition, it follows that $\beta(u_m, u_l) \le s\alpha(u_m, u_l)$, for $|u_m - u_l|$ sufficiently small (see [**22**]). Now, from the properties of boundary entropies, one has $\alpha(u,v) \ge 0$, and hence the desired inequality follows since $s < 0$.

This solution is actually consistent with the natural physical solution in the interior of Q, with boundary layers on the boundaries, via the characteristic analysis. It would be interesting to analyze systematically the uniqueness of entropy solutions in the sense of (1.7)-(1.9) for such problems.

We next recall an important fact about the solutions of (1.1)-(1.3) according to (1.7)-(1.9), established in [**5**], which holds even for the general multidimensional case in general (not necessarily cylindrical) domains.

THEOREM 1.1. *Assume that* (1.1) *is endowed with a strictly convex entropy. A function* $u(t,x) \in L^\infty(Q_T; \mathbf{R}^m)$ *satisfies* (1.7)-(1.9) *if and only if the following conditions hold:*

(a). $u(t,x)$ *satisfies* (1.1) *in the sense of distributions in* Q_T*;*

(b). *Given any boundary entropy pair* $(\alpha(u,v), \beta(u,v))$*, there exists a constant* $M > 0$ *such that, for any nonnegative* $\phi(t,x) \in C_0^\infty\big((-\infty, T) \times \mathbf{R}\big)$ *and any* $v \in \mathbf{R}^m$*,*

$$\begin{aligned} &\int_0^T \int_0^1 (\alpha(u(t,x),v)\partial_t\phi + \beta(u(t,x),v)\partial_x\phi)\,dx\,dt \\ &\qquad + \int_0^1 \alpha(u_0(x),v)\phi(0,x)\,dx + M\int_\Gamma \alpha(u^b,v)\phi\,dt \ge 0, \end{aligned} \tag{1.10}$$

where $\Gamma = \cup_{j=0}^1 \{x = j, t > 0\}$*, and* $u^b(t) = a_i(t), i = 0, 1$.

In the subsequent sections, we solve problem (1.1)–(1.3), in the sense of Definition 1.1, for the representative systems of nonlinear elasticity, chromatography, and isentropic gas dynamics, according to the following scheme. In Section 2, we establish some general results for the parabolic systems obtained from (1.1) with an additional viscosity term in its right-hand side. In particular, we obtain a useful uniform estimate (2.22) for the derivative of the viscous solutions, which is essential in order to establish the H^{-1}-compactness of entropy dissipation measures for *nonhomogeneous* boundary conditions addressed here (see §3.1). In Section 3, we apply the results of Sections 1-2 and the compensated compactness methods to obtain the existence of entropy solutions when (1.1) is either the 2×2 system of

nonlinear elasticity, or the $m \times m$ system of chromatography in Langmuir coordinates, or some other systems mentioned therein. Finally, in Section 4, we analyze the convergence of the viscous approximate solutions of the initial-boundary value problem for the system of isentropic Euler equations for gas dynamics with the aid of the results in Sections 1-2, especially estimate (2.22).

2. Parabolic Systems

In this section we consider the initial-boundary value problem for the parabolic system obtained from (1.1) with an additional viscosity term. Namely, we are concerned with the following initial-boundary value problem:

$$\partial_t u + \partial_x f(u) = \varepsilon \partial_{xx} u, \qquad (t,x) \in Q, \tag{2.1}$$

$$u|_{x=0} = a_{0,\varepsilon}(t), \qquad u|_{x=1} = a_{1,\varepsilon}(t), \quad t > 0, \tag{2.2}$$

$$u|_{t=0} = u_{0,\varepsilon}(x), \qquad x \in (0,1). \tag{2.3}$$

To simplify the statements of the results, we assume that $a_{0,\varepsilon}, a_{1,\varepsilon}$ are smooth and

$$\sup_{\varepsilon>0} \|(a_{0,\varepsilon}, \varepsilon a'_{0,\varepsilon}, a_{1,\varepsilon}, \varepsilon a'_{1,\varepsilon})\|_{L^\infty(0,\infty)} < \infty, \tag{2.4}$$

and $u_{0,\varepsilon} \in C_0^\infty(0,1)$ with

$$\sup_{\epsilon>0} \|u_{0,\varepsilon}\|_{L^\infty(0,1)} < \infty, \tag{2.5}$$

and the compatibility conditions:

$$a_{0,\varepsilon}^{(k)}(0) = a_{1,\varepsilon}^{(k)}(0) = 0, \qquad \text{for all } k \in \mathbf{N}. \tag{2.6}$$

For the purposes of the applications given in this paper, it suffices to consider the situation in which f satisfies:

$$f \in C^3(\mathbf{R}^m; \mathbf{R}^m) \text{ is globally Lipschitz, and } f(-u) = f(u). \tag{2.7}$$

Denote by $K^\varepsilon(t,x)$ the fundamental solution of the heat equation $\partial_t v = \varepsilon \partial_{xx} v$, that is,

$$K^\varepsilon(t,x) = \frac{1}{\sqrt{4\pi\varepsilon t}} e^{-\frac{x^2}{4\varepsilon t}}.$$

We will make use of the fact that

$$\|K^\varepsilon(t)\|_1 = 1, \qquad \|K^\varepsilon_x(t)\|_1 = \frac{1}{\sqrt{\varepsilon \pi t}}, \tag{2.8}$$

where $\| \ \|_1$ denotes the norm of $L^1(\mathbf{R})$.

For a function $\zeta(x)$ defined in $(0,1)$, denote by $\tilde{\zeta}$ the function defined in $\mathbf{R}$ such that

$$\begin{cases} \tilde{\zeta}(x) = \zeta(x), & 0 < x < 1, \\ \tilde{\zeta}(-x) = -\tilde{\zeta}(x), & x \in \mathbf{R}, \\ \tilde{\zeta}(x+2n) = \tilde{\zeta}(x), & x \in \mathbf{R},\ n \in \mathbf{Z}. \end{cases} \tag{2.9}$$

Set $h_\varepsilon(t,x) = (1-x)a_{0,\varepsilon}(t) + x a_{1,\varepsilon}(t)$, $x \in (0,1)$, $t \geq 0$. If $u(t,x)$ is a smooth solution of (2.1)–(2.3) in $[0,T] \times (0,1)$, then $w(t,x) = u(t,x) - h_\varepsilon(t,x)$ is a smooth

solution of the initial-boundary value problem:

$$\partial_t w - \varepsilon \partial_{xx} w = -\partial_x f(u) - \partial_t h_\varepsilon, \qquad (t,x) \in Q, \tag{2.10}$$

$$w|_{x=0} = 0, \qquad w|_{x=1} = 0, \qquad t > 0, \tag{2.11}$$

$$w|_{t=0} = u_{0,\varepsilon}(x), \qquad x \in (0,1). \tag{2.12}$$

Hence, we expect that $u(t,x)$ satisfies

$$u(t) = K^\varepsilon(t) * \tilde{u}_{0,\varepsilon} - \int_0^t K^\varepsilon(t-s) * (\widetilde{\partial_x f(u)}(s) + \widetilde{\partial_t h_\varepsilon}(s))\, ds + h_\varepsilon(t), \tag{2.13}$$

for $(t,x) \in Q$.

To see this fact, we denote the right-hand side of (2.10) by $r(t,x)$ and approximate it in $L^1_{loc}(Q)$ by a sequence $r^n(t,x)$ in $C_0^\infty(Q)$. If $w^n(t,x)$ is the corresponding solution for the modified (2.10) with conditions (2.11)-(2.12), then we clearly have

$$w^n(t) = K^\varepsilon(t) * \tilde{u}_{0,\varepsilon} - \int_0^t K^\varepsilon(t-s) * \widetilde{r^n}(s)\, ds.$$

Now, letting $n \to \infty$, we see that w^n converges in $L^1_{loc}(Q)$ to a certain w satisfying

$$w(t) = K^\varepsilon(t) * \tilde{u}_{0,\varepsilon} - \int_0^t K^\varepsilon(t-s) * \tilde{r}(s)\, ds.$$

From the properties of the heat kernel and the function $\tilde{r}(t,x)$, we deduce that w satisfies (2.10) in the sense of distributions in Q. Then $w - (u - h_\varepsilon)$ satisfies the homogeneous heat equation. By the standard regularity theory, w is then smooth. Since (2.11) and (2.12) are easily verified from (2.13), the uniqueness of the solution of (2.10)–(2.12) implies that $w = u - h_\varepsilon$. On the other hand, if u satisfies (2.13) and has continuous derivatives of first order in t and up to second order in x, throughout Q, then applying the heat operator $\partial_t - \varepsilon \partial_{xx}^2$ to both sides of (2.13), for $(t,x) \in Q$, yields that u satisfies (2.1) in Q in the sense of distributions and, hence, in the classical sense. Conditions (2.2)-(2.3) are also immediately deduced from (2.13).

Then, for smooth solutions in $[0,T] \times (0,1)$, (2.13) is equivalent to

$$\begin{aligned} u(t) = K^\varepsilon(t) * \tilde{u}_{0,\varepsilon} &- \int_0^t \partial_x K^\varepsilon(t-s) * f(\tilde{u})(s)\, ds \\ &- \int_0^t K^\varepsilon(t-s) * \widetilde{\partial_t h_\varepsilon}(s)\, ds + h_\varepsilon(t), \end{aligned} \tag{2.14}$$

where we have used the definition of the extension $\widetilde{\partial_x f(u)}(s)$ of $\partial_x f(u)(s)$, from $(0,1)$ to $\mathbf{R}$ according to (2.9). Indeed, to pass the derivative from $\widetilde{\partial_x f(u)}(s)$ to the heat kernel to obtain (2.14) from (2.13), we write the convolution as a sum of integrals in $(0,1)$, using (2.9), then apply integration by parts, and observe that the sums like

$$\sum_{n \in \mathbf{Z}} \int_0^t \{K^\varepsilon(t-s, x-2n-1) - K^\varepsilon(t-s, x-2n+1)\} f(a_{1,\varepsilon}(s))\, ds,$$

resulting also from this process, vanish identically. Hence a possible strategy for solving (2.1)–(2.3) is to obtain first a solution of (2.14), and then to prove that it possesses the required regularity.

Let $\mathcal{G}_T = L^\infty((0,T);L^\infty(0,1))$ and define the operator $\mathcal{L}:\mathcal{G}_T\to\mathcal{G}_T$ by

$$\mathcal{L}(v)(t) = K^\varepsilon(t)*\tilde{u}_{0,\varepsilon} - \int_0^t \partial_x K^\varepsilon(t-s)*f(\tilde{v})(s)\,ds - \int_0^t K^\varepsilon(t-s)*\widetilde{\partial_t h_\varepsilon}(s)\,ds + h_\varepsilon(t). \tag{2.15}$$

Let $\|\nabla f(u)\|_{L^\infty} = C_0$. We have

$$\|\mathcal{L}(v_1)(t) - \mathcal{L}(v_2)(t)\|_{L^\infty} \le C_0\sqrt{\frac{T}{\varepsilon\pi}}\|v_1 - v_2\|_{L^\infty},$$

and so $\mathcal{L}$ is a contraction in $\mathcal{G}_T$ as long as

$$C_0\sqrt{\frac{T}{\varepsilon\pi}} < 1. \tag{2.16}$$

By the Banach Fixed Point Theorem, there exists a unique $u\in\mathcal{G}_T$, which satisfies $\mathcal{L}(u) = u$ when $0\le t\le T$.

LEMMA 2.1. *Let u be the unique fixed point of $\mathcal{L}$ in $\mathcal{G}_T$ with $T = \alpha_0\varepsilon$ for $\alpha_0 << 1$ independent of ε, such that* (2.16) *holds. Then there exists $C_1 > 0$ such that,*

$$\|\partial_x u(t)\|_{L^\infty} \le \frac{C_1}{\sqrt{\varepsilon t}}, \qquad 0 < t \le T, \tag{2.17}$$

with C_1 independent of ε. Furthermore, there exists a constant C_2, depending on ε, such that

$$\|\partial_x u(t)\|_{L^\infty} \le C_2, \qquad 0 < t \le T. \tag{2.18}$$

PROOF. The proof of (2.17) reduces to proving the following assertion: there exists a constant $C_1 > 0$ such that, if $v\in\mathcal{G}_T\cap C((0,T]\times(0,1))$, $v|_{x=0} = a_{0,\varepsilon}$, $v|_{x=1} = a_{1,\varepsilon}$, and

$$\|\partial_x v(t)\|_{L^\infty} \le \frac{C_1}{\sqrt{\varepsilon t}}, \tag{2.19}$$

then $\mathcal{L}(v)$ also satisfies these properties.

Indeed, since u is the unique fixed point of $\mathcal{L}$ in $\mathcal{G}_T$, we have $u = \lim_{k\to\infty} u^k$ in $\mathcal{G}_T$, where $u^1 = h$, $u^{k+1} = \mathcal{L}(u^k)$. Hence, from the assertion, we have that (2.19) is satisfied for $v = u^k$, $u^k|_{x=0} = a_{0,\varepsilon}(t)$, and $u^k|_{x=1} = a_{1,\varepsilon}(t)$, for all $k\in\mathbf{N}$. Therefore, the standard arguments yield that u must also satisfy these properties.

We now pass to the proof of the assertion. From the hypothesis on v, one has

$$\partial_x\mathcal{L}(v)(t) = \partial_x K^\varepsilon(t)*\tilde{u}_{0,\varepsilon} - \int_0^t \partial_x K^\varepsilon(t-s)*\widetilde{\partial_x f(v)}(s)\,ds - \int_0^t \partial_x K^\varepsilon(t-s)*\partial_t\widetilde{h_\varepsilon}(s)\,ds + \partial_x h_\varepsilon(t),$$

and so

$$\begin{aligned}\|\partial_x\mathcal{L}(v)(t)\|_{L^\infty} \leq & \frac{\|\tilde{u}_{0,\varepsilon}\|_{L^\infty}}{\sqrt{\pi\varepsilon t}} + \frac{CC_1}{\varepsilon} + 2\sqrt{\frac{T}{\pi\varepsilon}}\|\partial_t h_\varepsilon\|_{L^\infty} + \|\partial_x h_\varepsilon\|_{L^\infty} \\ \leq & \frac{1}{\sqrt{\pi\varepsilon t}}\big(\|u_{0,\varepsilon}\|_{L^\infty} + CC_1\sqrt{\frac{\pi T}{\varepsilon}} + 2T\|\partial_t h_\varepsilon\|_{L^\infty} + \sqrt{\pi\varepsilon T}\|\partial_x h_\varepsilon\|_{L^\infty}\big) \\ \leq & \frac{C_1}{\sqrt{\varepsilon t}},\end{aligned}$$

provided that

$$C_1 \geq \frac{\|u_{0,\varepsilon}\|_{L^\infty} + 2T\|\partial_t h_\varepsilon\|_{L^\infty} + \sqrt{\varepsilon\pi T}\|\partial_x h_\varepsilon\|_{L^\infty}}{\sqrt{\pi}(1 - C\sqrt{\frac{T}{\varepsilon}})}, \tag{2.20}$$

where $C > 0$ depends only on C_0, independent of ε and v. Since $T = \alpha_0\varepsilon$ for $\alpha_0 << 1$ independent of ε, the fact that C_1 can be taken independent of ε is clearly seen from (2.20), because of (2.4) and (2.5).

The second part of the statement follows similarly. We only need to prove that there exists a constant C_2 such that, if $\|\partial_x v\|_{L^\infty} \leq C_2$, $v|_{x=0} = a_{0,\varepsilon}$ and $v|_{x=1} = a_{1,\varepsilon}$, then $\mathcal{L}(v)$ has also these properties. To this end, we observe that we may write

$$\begin{aligned}\partial_x\mathcal{L}(v)(t) = K^\varepsilon(t) * \widetilde{\partial_x u_{0,\varepsilon}} - \int_0^t \partial_x K^\varepsilon(t-s) * \widetilde{\partial_x f(v)}(s)\,ds \\ - \int_0^t \partial_x K^\varepsilon(t-s) * \partial_t\widetilde{h_\varepsilon}(s)\,ds + \partial_x h_\varepsilon(t).\end{aligned}$$

Hence

$$\|\partial_x\mathcal{L}(v)\|_{L^\infty} \leq \|\widetilde{\partial_x u_{0,\varepsilon}}\|_{L^\infty} + CC_2\sqrt{\frac{T}{\varepsilon}} + 2\sqrt{\frac{T}{\pi\varepsilon}}\|\partial_t h_\varepsilon\|_{L^\infty} + \|\partial_x h_\varepsilon\|_{L^\infty} \leq C_2, \tag{2.21}$$

provided that

$$C_2 \geq \frac{\|\partial_x u_{0,\varepsilon}\|_{L^\infty} + 2\sqrt{T/(\pi\varepsilon)}\|\partial_t h_\varepsilon\|_{L^\infty} + \|\partial_x h_\varepsilon\|_\infty}{1 - C\sqrt{T/\varepsilon}},$$

where C depends only on C_0. This concludes the proof. □

Let u be the unique fixed point of $\mathcal{L}$ in $\mathcal{G}_T$. By Lemma 2.1, $\partial_x u$ is bounded in $[0,T]\times(0,1)$. Clearly, u is a weak solution of (2.1)–(2.3) in the sense that u belongs to the space

$$W(T) = \{v \in L^2([0,T]; W^{1,2}(0,1)) \mid \partial_t v \in L^2([0,T]; W^{-1,2}(0,1))\},$$

satisfying

$$\langle\partial_t u(t), \phi\rangle + \varepsilon\int_0^1 \partial_x u\partial_x\phi\,dx = \int_0^1 f(u)(t)\partial_x\phi\,dx,$$

for almost all $t \in [0,T]$ and all $\phi \in W_0^{1,2}(\Omega)$, $u(0) = u_{0,\varepsilon}$, and $u(t) - h_\varepsilon(t) \in W_0^{1,2}(0,1)$, for almost all $t \in [0,T]$. The fact that $\partial_t u \in L^2([0,T]; W^{-1,2}(0,1))$ follows the observation that u is the limit in L^∞ of a sequence u^n satisfying

$$\partial_t u^{n+1} - \varepsilon\partial_{xx}u^{n+1} = -\partial_x f(u^n),$$

with $\|\partial_x u^n\|_{L^\infty}$ uniformly bounded in n. Therefore, $\partial_t u^n(t,x)$ is uniformly bounded in $L^2([0,T];W^{-1,2}(0,1))$ and must converge weakly to $\partial_t u(t,x)$.

Applying the regularity theory for parabolic equations (see [**17, 23, 28**]), one deduces that u is a smooth solution of (2.1)–(2.3) in $[0,T]\times(0,1)$. It is easy to see that the hypothesis that f is globally Lipschitz allows one to repeat the above procedures step by step in time, indefinitely, to obtain a global smooth solution to (2.1)–(2.3) satisfying (2.17)-(2.18).

If system (2.1) is endowed with a bounded invariant region (see [**8**]), then the global smooth solution u is uniformly bounded. We now prove that this allows us to obtain a useful estimate for $\partial_x u$.

THEOREM 2.1. *Let u be the unique smooth solution of* (2.1)-(2.3). *Assume that u is uniformly bounded in $[0,\infty)\times(0,1)$, independently of ε, and that* (2.4)-(2.6) *hold. Then, for all $\delta>0$ sufficiently small, there exists a positive constant M, independent of ε, such that*

$$\begin{cases} \|\varepsilon\partial_x u(t)\|_{L^\infty}\le M, & \text{for } t>\varepsilon\delta,\\ \|\varepsilon\partial_x u(t)\|_{L^\infty}\le M\sqrt{\frac{\varepsilon}{t}}, & \text{for } 0<t\le\varepsilon\delta. \end{cases} \tag{2.22}$$

The same result holds for the smooth solution of the Cauchy problem.

PROOF. For any $t_0>0$, consider the operator $\mathcal{L}$ in $\mathcal{G}_T=L^\infty([t_0,t_0+T];L^\infty(\Omega))$, given by

$$\begin{aligned} \mathcal{L}(v)(t)=&K^\varepsilon(t-t_0)*\tilde u(t_0)-\int_{t_0}^t \partial_x K^\varepsilon(t-s)*f(\tilde v)(s)\,ds\\ &-\int_{t_0}^t K^\varepsilon(t-s)*\widetilde{\partial_t h_\varepsilon}(s)\,ds+h_\varepsilon(t). \end{aligned} \tag{2.23}$$

Exactly as above, we easily see that $\mathcal{L}$ is a contraction mapping in $\mathcal{G}_T$ if (2.16) is satisfied. Also, identically as in the proof of Lemma 2.1, we prove the assertion that, for C_1 satisfying (2.20) with $\|u_{0,\varepsilon}\|_{L^\infty}$ replaced by $\|u(t_0)\|_{L^\infty}$, if $v\in\mathcal{G}_T\cap C((t_0,t_0+T]\times\Omega)$, $v|_{x=0}=a_{0,\varepsilon}$, $v|_{x=1}=a_{1,\varepsilon}$, and

$$\|\partial_x v(t)\|_{L^\infty}\le\frac{C_1}{\sqrt{\varepsilon(t-t_0)}}, \tag{2.24}$$

then $\mathcal{L}(v)$ also satisfies these properties. Thus, using the fact that u is the unique fixed point of $\mathcal{L}$, we deduce from the standard arguments that u must satisfy (2.24). Take $T=2\varepsilon\delta$, for any $\delta>0$ such that (2.16) holds. Then, for any $t>\varepsilon\delta$, we take some $t_0=t-T/2$ in (2.24) to obtain

$$\|\varepsilon\partial_x u(t)\|_{L^\infty}\le\frac{C_1}{\sqrt{\delta}}. \tag{2.25}$$

On the other hand, for $0<t\le\varepsilon\delta$, we have that (2.17) holds. From Lemma 2.1, C_1 can be taken independent of ε. Therefore, taking $M=C_1/\sqrt{\delta}$, we conclude the proof of (2.22).

The proof of estimate (2.22) for the smooth solution of the Cauchy problem is completely similar. □

In order to prove (1.10) for hyperbolic systems, we need to get a corresponding inequality for the associated parabolic systems. To this end, we will make use of a

construction as in [**28, 30**]. For $\delta > 0$ sufficiently small, define

$$d(x) = \begin{cases} x, & 0 < x < \delta, \\ \delta, & \delta < x < 1-\delta, \\ 1-x, & 1-\delta < x < 1, \end{cases}$$

and, for some $M > 0$, set

$$\xi_\varepsilon(x) \equiv 1 - e^{-\frac{M}{\varepsilon} d(x)}.$$

For any $\varphi \in C_0(\mathbf{R})$, $\varphi \geq 0$, the function $\xi_\varepsilon(x)$ satisfies

$$M \int_0^1 |\xi_\varepsilon'(x)| \varphi(x)\, dx \leq \varepsilon \int_0^1 \xi_\varepsilon'(x)\varphi'(x)\, dx + M(\varphi(0) + \varphi(1)). \tag{2.26}$$

Indeed,

$$\begin{aligned} \int_0^1 \xi_\varepsilon'(x)\varphi'(x)\, dx &= \int_{\{0<d<\delta\}} \xi_\varepsilon'(x)\varphi'(x)\, dx \\ &= -\int_{\{0<d<\delta\}} \xi_\varepsilon''(x)\varphi(x)\, dx - \frac{M}{\varepsilon}(\varphi(1) + \varphi(0)) \\ &\quad + \frac{M}{\varepsilon} e^{-\frac{M\delta}{\varepsilon}} (\varphi(1-\delta) + \varphi(\delta)) \\ &\geq \frac{M^2}{\varepsilon^2} \int_{\{0<d<\delta\}} e^{-\frac{M d(x)}{\varepsilon}} \varphi(x)\, dx - \frac{M}{\varepsilon}(\varphi(1) + \varphi(0)) \\ &= \frac{M}{\varepsilon} \int_O^1 |\xi_\varepsilon'(x)| \varphi(x)\, dx - \frac{M}{\varepsilon}(\varphi(1) + \varphi(0)), \end{aligned}$$

which immediately give (2.26).

THEOREM 2.2. *Let u be the smooth solution of* (2.1)–(2.3)*, and let $(\alpha(u,v), \beta(u,v))$ be a boundary entropy pair for* (1.1)*. Then there exists a constant $M > 0$ such that, for all $\phi \in C_0^\infty((-\infty, T) \times \mathbf{R})$, $\phi \geq 0$, and $v \in \mathbf{R}^m$,*

$$\begin{aligned} &-\textstyle\int_0^T \int_0^1 \{\alpha(u,v)\partial_t\phi + \beta(u,v)\partial_x\phi + \varepsilon\alpha(u,v)\partial_{xx}\phi\}\, \xi_\varepsilon\, dx\, dt \\ &\leq \textstyle\int_0^1 \alpha(u_{0,\varepsilon}, v)\phi(x,0)\xi_\varepsilon\, dx + M \int_\Gamma \alpha(u_\varepsilon^b, v)\phi\, dt \\ &\quad + 2\varepsilon \textstyle\int_0^T \int_0^1 \alpha(u,v)\xi_\varepsilon'\partial_x\phi\, dx\, dt, \end{aligned} \tag{2.27}$$

where $\Gamma = \cup_{j=0}^1 \{x = j, t > 0\}$, and $u_\varepsilon^b = a_{i,\varepsilon}, i = 0, 1$.

PROOF. Denote $\eta(u) = \alpha(u,v)$, $q(u) = \beta(u,v)$. By the convexity of α with respect to u and (1.6), we easily see that there must exist a constant $M > 0$, independent of v, such that $|q(u)| \leq M\eta(u)$. Now, multiplying (2.1) by $\nabla\eta(u)$, one obtains

$$\partial_t \eta(u) + \partial_x q(u) \leq \varepsilon \partial_{xx}\eta(u). \tag{2.28}$$

Then, multiplying (2.28) by $\xi_\varepsilon\phi$, integrating in $Q_T = (0,T)\times(0,1)$, and using integration by parts, we obtain

$$\begin{aligned} & -\int_0^T\int_0^1 \{\eta(u)\partial_t\phi + q(u)\partial_x\phi + \varepsilon\eta(u)\partial_{xx}\phi\}\,\xi_\varepsilon\,dx\,dt \\ \leq\ & \int_0^1 \eta(u_{0,\varepsilon})\phi(x,0)\xi_\varepsilon\,dx + M\int_0^T\int_0^1 \eta(u)\phi|\xi_\varepsilon'|\,dx\,dt \\ & -\varepsilon\int_0^T\int_0^1 \partial_x(\eta(u)\phi)\xi_\varepsilon'\,dx\,dt + 2\varepsilon\int_0^T\int_0^1 \eta(u)\xi_\varepsilon'\partial_x\phi\,dx\,dt, \end{aligned}$$

where we have used $|q(u)|\leq M\eta(u)$. Applying (2.26) with $\eta(u)\phi$ replacing φ in the inequality displayed above, we then obtain (2.27). □

3. Nonlinear Elasticity, Chromatography, and Other Systems

In this section we apply the results in Sections 1-2 to solving the initial-boundary value problem for two specific systems: the one arising in one-dimensional nonlinear elasticity and the other appearing in chromatography with Langmuir coordinates. We also discuss other applications which follow in a similar fashion.

3.1. Nonlinear Elasticity. Consider the one-dimensional nonlinear elasticity system:

$$\begin{cases} \partial_t u_1 - \partial_x\sigma(u_2) = 0, \\ \partial_t u_2 - \partial_x u_1 = 0, \end{cases} \tag{3.1}$$

where σ is a smooth function satisfying $\sigma'(\tau)>0$, and $\tau\sigma''(\tau)>0$ if $\tau\neq 0$. Then, in this case, $f(u) = (-\sigma(u_2), -u_1)^\top$. System (3.1) is endowed with the following strictly convex entropy:

$$\eta_*(u) = u_1^2 + \int_0^{u_2}\sigma(\tau)\,d\tau,$$

with entropy-flux:

$$q_*(u) = u_1\sigma(u_2).$$

Given a convex entropy $\eta(u)$, a boundary entropy pair $(\alpha(u,v),\beta(u,v))$ can be defined by taking the quadratic part of η and its associated flux. That is,

$$\begin{aligned} \alpha(u,v) &= \eta(u) - \eta(v) - \nabla\eta(v)(u-v), \\ \beta(u,v) &= q(u) - q(v) - \nabla\eta(v)(f(u)-f(v)). \end{aligned}$$

Also, system (3.1) is endowed with a pair of independent Riemann invariants (*i.e.* the functions whose gradient are left-eigenvectors of ∇f) given by

$$w_1 = u_1 + \int_0^{u_2}\sqrt{\sigma'(\tau)}\,d\tau, \qquad w_2 = u_1 - \int_0^{u_2}\sqrt{\sigma'(\tau)}\,d\tau.$$

The regions given by

$$R = \{u\in\mathbf{R}^2\,|\,|w_1|<M,\ |w_2|<M\},$$

for any $M>0$, are invariant under the flow of the parabolic system (2.1) corresponding to (3.1) (cf. [**8, 11, 14**]). Given uniformly bounded initial-boundary data, we take a region R like the above with M large enough so that the initial-boundary data assume values in R. In order to have the flux function f of (3.1) satisfying condition (2.7), we first change from the coordinates u to

$$\bar u = u - u_*, \tag{3.2}$$

where $u_* = (0, -A), A > 0$, is any point of the axis $u_1 = 0$ which does not belong to R. Then we replace f by

$$\bar{f}(\bar{u}) = \varphi(\bar{u}) f(\bar{u} + u_*), \text{ if } \bar{u} > 0; \qquad \bar{f}(-\bar{u}) = \bar{f}(\bar{u}), \tag{3.3}$$

with $\varphi \in C_0^\infty(\mathbf{R}^2)$ and $\varphi(\bar{u}) = 1$, if $\bar{u} \in R - u_*$, and such that $(0, -A)$ does not belong to the support of φ. The function $\bar{f}$ satisfies (2.7) and coincides with f in the invariant region R. Now, by the invariant region arguments (cf. [**8**]), any smooth solution of system (2.1) associated with (3.1), with $\bar{f}$ replacing f, takes its values in R, as long as the initial-boundary data take values in R. Hence, replacing f by $\bar{f}$ has no real effect, and the solution of the modified system is also a solution of the original one.

For given initial-boundary data (1.2)-(1.4), we can find smooth approximate functions $a_{0,\varepsilon}$, $a_{1,\varepsilon}$, and $u_{0,\varepsilon}$, which converge to a_0, a_1, and u_0, respectively, in $L^1_{loc}(0,\infty)$ and $L^1(0,1)$ and satisfy (2.4)-(2.6), using the standard techniques of cutoff and mollification. We now consider the compactness of the smooth solution sequence u^ε of the viscous systems (2.1) corresponding to (3.1). First, this sequence is uniformly bounded in $L^\infty(Q; \mathbf{R}^2)$, because all of the functions u^ε assume values in R, which is a bounded region of $\mathbf{R}^2$. That is,

$$\|u^\varepsilon\|_{L^\infty(Q)} \le B_1, \tag{3.4}$$

for some $B_1 > 0$ independent of ε. To apply DiPerna's compactness result in [**11**], it suffices to verify the following:

$$\partial_t \eta(u^\varepsilon) + \partial_x q(u^\varepsilon) \quad \text{lies in a compact subset of } H^{-1}_{loc}(Q), \tag{3.5}$$

for any smooth entropy-entropy flux pair (η, q).

With the aid of our estimate (2.22), property (3.5) can be seen as follows. We first multiply (2.1) by $\nabla\eta(u)$ to obtain

$$\partial_t \eta(u^\varepsilon) + \partial_x q(u^\varepsilon) = \varepsilon \partial_{xx} \eta(u^\varepsilon) - \varepsilon (\partial_x u^\varepsilon)^\top \nabla^2 \eta(u^\varepsilon) \partial_x u^\varepsilon. \tag{3.6}$$

If η is strictly convex (*e.g.* $\eta = \eta_*$ given above), integrating (3.6) in Q_T with any $T > 0$, we have

$$\begin{aligned} c_0 \iint_{Q_T} \varepsilon (\partial_x u^\varepsilon)^2 dx dt \le & \varepsilon \int_0^T (\eta'(u^\varepsilon) \partial_x u^\varepsilon)|_0^1 dt - \int_0^1 \eta(u^\varepsilon)|_0^T dx - \int_0^T q(u^\varepsilon)|_0^1 dt \\ \le & A_1 \int_0^\varepsilon \sqrt{\frac{\varepsilon}{t}} dt + A_2 \int_\varepsilon^T M dt + A_3 \le B_2, \end{aligned}$$

using estimate (2.22), where $A_i, i = 1, 2, 3$, and B_2 are independent of ε. Thus, we have

$$\sqrt{\varepsilon} \|\partial_x u^\varepsilon\|_{L^2(Q_T)} \le B_3, \tag{3.7}$$

for some constant $B_3 > 0$, depending on T, but independent of ε.

Now, from (3.7), we obtain as usual that, for any smooth entropy η,

$$\varepsilon (\partial_x u^\varepsilon)^\top \nabla^2 \eta(u^\varepsilon) \partial_x u^\varepsilon$$

is uniformly bounded in $\mathcal{M}(Q_T)$, the space of signed Radon measures in Q_T. Therefore, by Sobolev's embeddings, it is compact in $W^{-1,p}(Q_T)$, for $1 < p < 2$. Also, from (3.7), we obtain that, for any smooth entropy $\eta(u)$,

$$\varepsilon \partial_{xx} \eta(u^\varepsilon)$$

is compact in $W^{-1,2}(Q_T)$ (in fact it converges to 0). Thus the right-hand side of (3.6) is compact in $W^{-1,p}(Q_T)$, using again Sobolev's embeddings. Now, because of (3.4), the left-hand side of (3.6) is uniformly bounded in $W^{-1,\infty}(Q_T)$. Then, an interpolation argument gives (3.5) (see [**10, 29**]).

Once we have proved (3.5), we can use DiPerna's compactness result in [**11**] to conclude the compactness of the sequence u^ε in $L^1_{loc}(Q)$. Let u be the limit of a subsequence u^{ε_k} in $L^1_{loc}(Q)$ with $\varepsilon_k \to 0$ as $k \to \infty$. Hence, from (2.27) in Theorem 2.2, we obtain (1.10), using the fact that ξ_ε and $\varepsilon\xi'_\varepsilon$ are uniformly bounded and converge pointwise to 1 and 0, respectively.

Thus, Theorem 1.1 can be applied to conclude that u is an entropy solution of the initial-boundary value problem for (3.1).

THEOREM 3.1. *Let a_0, a_1, and u_0 satisfy* (1.4). *Then there exists a global entropy solution of the initial-boundary problem* (3.1) *and* (1.2)-(1.3) *in the sense of* (1.7)-(1.9).

3.2. Chromatography: The $m \times m$ chromatography system for Langmuir isotherms (cf. [**31**]) is given by

$$\partial_t u_i + \partial_x \left(\frac{k_i u_i}{1 + \sum_{j=1}^m u_j} \right) = 0, \quad x \in \mathbf{R}, \quad t \geq 0, \quad 1 \leq i \leq m, \tag{3.8}$$

where $0 < k_1 < k_2 < \cdots < k_m$ are given numbers. It is well known (cf. [**18**]) that (3.8) is endowed with m linearly independent Riemann invariants $w_1, \ldots, w_m$, which have the property that the level surfaces $w_i = const.$ are affine hyperplanes in $\mathbf{R}^m$ (also see Temple [**38**]). For these systems, using the maximum principle (see [**33**]), it is easy to show that the regions

$$R = \{u \in \mathbf{R}^m \mid |w_i(u) - \bar{w}_i| \leq M_i, \ i = 1, \cdots, m\}$$

are invariant under the flow of the associated parabolic system (2.1), where $\bar{w} = (\bar{w}_1, \cdots, \bar{w}_n)$ is a constant state in $\mathbf{R}^m$ and $M_i > 0$ are arbitrary constants, as long as they are contained in the domain $\{u \in \mathbf{R}^m \mid u_i \geq 0, i = 1, \cdots, m\}$. Then, the same procedures as the one for the system of nonlinear elasticity can yield the existence of entropy solutions of the initial-boundary value problem for (3.8), where we apply the compactness theorem of James-Peng-Perthame [**18**], with the aid of Theorem 2.1 (i.e. (2.22)).

THEOREM 3.2. *Let a_0, a_1, and u_0 satisfy* (1.4). *Then there exists a global entropy solution of the initial-boundary problem* (3.8) *and* (1.2)-(1.3) *in the sense of* (1.7)-(1.9).

3.3. Other Systems: The same techniques can be used to prove the corresponding results for other systems such as the quadratic systems with umbilic degeneracy studied in [**6**], the class of conjugate type systems considered in [**15**], and the systems addressed in [**32**]. All of these systems have bounded invariant regions over which the flux functions are smooth, say, at least C^3 in the interior of the invariant regions and C^2 up to the boundaries.

4. System of Isentropic Euler Equations

The system of isentropic Euler equations reads

$$\begin{cases} \partial_t \rho + \partial_x m = 0, \\ \partial_t m + \partial_x \big(\frac{m^2}{\rho} + p(\rho) \big) = 0, \end{cases} \tag{4.1}$$

where ρ represents the density, m is the momentum, and $p(\rho)$ is the pressure. The behavior of the pressure function $p(\rho)$ depends on the fluids under consideration. We assume at the onset that $p(\rho)$ satisfies

$$p'(\rho) > 0 \text{ (hyperbolicity)}, \quad \rho p''(\rho) + 2p'(\rho) > 0 \text{ (genuine nonlinearity)}, \tag{4.2}$$

away from the vacuum $\rho = 0$ and, when $\rho \to 0+$,

$$p(\rho) \approx \kappa \rho^\gamma (1 + P(\rho)), \qquad |P^{(n)}(\rho)| \le C \rho^{1-n}, \ 0 \le n \le 4, \tag{4.3}$$

for some $\gamma > 1$. This means that, when $\rho \to 0$, the pressure law $p(\rho)$ has the same principal singularity as the γ-law, but allows additional singularities in the derivatives.

System (4.1) is endowed with a pair of independent Riemann invariants given by

$$w = \frac{m}{\rho} + \int_0^\rho \frac{1}{s} \sqrt{p'(s)}\, ds, \quad z = \frac{m}{\rho} - \int_0^\rho \frac{1}{s} \sqrt{p'(s)}\, ds. \tag{4.4}$$

Given positive constants M_i, $i = 1, 2$, consider the region R of the plane ρ–m given by $-z \le M_1$, $w \le M_2$, that is,

$$R = \big\{ (\rho, m) \,|\, -M_1 \rho + \rho \int_0^\rho \frac{1}{s} \sqrt{p'(s)}\, ds \le m \le M_2 \rho - \rho \int_0^\rho \frac{1}{s} \sqrt{p'(s)}\, ds \big\}.$$

Then the region is invariant under smooth flows of the parabolic system (2.1) associated with (4.1) (cf. [**8**]), provided that we can show

$$\rho^\varepsilon(t, x) \ge \delta^\varepsilon(t), \quad 0 < t < \infty, \tag{4.5}$$

where $\delta^\varepsilon(t) > 0$ depends on ε and t. Thus, we assume that the initial-boundary data (2.2)-(2.3), for the viscous systems (2.1) associated with (4.1), take values in R, for large $M_i, i = 1, 2$.

We notice that, in the region R, the flux function of (4.1) is only Lipschitz continuous because of the singularity in $\rho = 0$. In order to have the flux function $f(\rho, m) = (m, m^2/\rho + p(\rho))^\top$ partially satisfying (2.7), we artificially extend it to the half-plane $\{\rho < 0\}$ as an even function. The resultant function is smooth only away from the vacuum line $\{\rho = 0\}$. Hence local (in time) smooth solutions of the problem (2.1)–(2.3), corresponding to (4.1), can be extended only while they stay in the region $\rho > 0$. This is the main difference between the analyses for system (4.1) and for the systems in Section 3.

For given initial-boundary data a_0, a_1, and u_0 satisfying (1.4) and

$$\begin{cases} \rho_0(x) \ge 0, \quad |m_0(x)| \le C_0 \rho_0(x), \quad C_0 > 0, \\ \rho(t, i) \ge 0, \quad |m(t, i)| \le C_0 \rho(t, i), \quad i = 0, 1, \end{cases} \tag{4.6}$$

there exists sequences $a_{0,\varepsilon}$, $a_{1,\varepsilon}$, and $u_{0,\varepsilon}$ that converge to a_0, a_1, and u_0, respectively, in $L^1_{\rm loc}(0, \infty)$ and $L^1(0, 1)$ and that satisfy (2.4)-(2.6), and

$$\rho_{0,\varepsilon}(x), \rho^\varepsilon(t, 0), \rho^\varepsilon(t, 1) \ge \alpha^\varepsilon > 0, \tag{4.7}$$

where $\alpha^\varepsilon \to 0$ as $\varepsilon \to 0$. Under (4.7), there exists a unique global smooth solution of the Cauchy problem for the viscous system satisfying (4.5) (see [**12, 4**]). Actually, the key point to construct such a solution is to show that any local smooth solution, assuming values in $\{\rho > 0\}$, defined up to a certain time $T > 0$ must satisfy,

$$\rho^\varepsilon(t,x) \geq \delta^\varepsilon(T) > 0, \quad \text{for } 0 \leq t < T, \tag{4.8}$$

for all $x \in (0,1)$ and some $\delta^\varepsilon(T) > 0$ depending on both ε and T. The proof of (4.8) in [**12, 4**] can be easily adapted for the initial-boundary value problem with the help of an obvious version of Theorem 2.1 for local smooth solutions. Nevertheless, we will give an alternate proof here for (4.8) for our initial-boundary problem.

Consider the equation

$$\partial_t \rho + \partial_x(\rho u) = \varepsilon \partial_{xx} \rho. \tag{4.9}$$

Multiplying (4.9) by $v'(\rho)$ with $v(\rho) = 1/\rho$, we obtain

$$\partial_t v - \varepsilon \partial_{xx} v = \partial_x(uv) + v''(\rho u \partial_x \rho - \varepsilon(\partial_x \rho)^2) \leq \partial_x(uv) + \frac{2vu^2}{2\varepsilon}. \tag{4.10}$$

Consider the equation

$$\partial_t g - \varepsilon \partial_{xx} g = \partial_x(ug) + \frac{2gu^2}{\varepsilon}, \quad x \in (0,1), \tag{4.11}$$

together with the conditions:

$$g|_{x=0} = v^\varepsilon(t,0), \quad g|_{x=1} = v^\varepsilon(t,1), \tag{4.12}$$

$$g|_{t=0} = v^\varepsilon(0,x), \quad x \in (0,1). \tag{4.13}$$

If g is a smooth solution of (4.11)–(4.13) defined for $0 \leq t < T$, the maximum principle, applied to the difference $v - g$, gives $v(t,x) \leq g(t,x)$, for $(t,x) \in Q_T$. Thus, to get (4.8), all we have to do is to prove that

$$g(t,x) \leq N^\varepsilon(T), \quad \text{for } (t,x) \in Q_T, \tag{4.14}$$

for some positive number $N^\varepsilon(T)$, depending on both ε and T.

Now take any $t_0 \in (0,T)$ and consider the operator $\mathcal{L}$ in $\mathcal{G}_\tau = L^\infty((t_0, t_0+\tau) \times (0,1))$, with $t_0 + \tau \leq T$, given by

$$\begin{aligned} \mathcal{L}(h) =& K^\varepsilon(t-t_0) * \tilde{g}(t_0) + \int_{t_0}^t K^\varepsilon(t-s) * \big(2\frac{\widetilde{hu^2}(s)}{\varepsilon} - \partial_t \tilde{\zeta}(s)\big)\, ds \\ &- \int_{t_0}^t \partial_x K^\varepsilon(t-s) * (\tilde{u}\tilde{h})\, ds + \zeta(t), \end{aligned} \tag{4.15}$$

where $\zeta(t,x) = (1-x)g(t,1) + xg(t,0)$, for $0 < x < 1$, $t > 0$, and the $\tilde{}$ has the same meaning as in Section 2. This operator is a contraction mapping in $\mathcal{G}_\tau$ if

$$2\max\{2M^2, MC_0, 1\}\sqrt{\frac{\tau}{\varepsilon}} < 1, \tag{4.16}$$

as one can easily verify, where M is a constant such that $|u| \leq M$ in R. In this case g is its unique fixed point. Now take $t_0 = T - \tau_0$ with

$$\tau_0 = \varepsilon/[8(\max\{2M^2, MC_0, 1\})^2]$$

and $\tau = \tau_0$. Define

$$N(t_0) = \sup_{0 \leq t \leq t_0} \|g(s)\|_{L^\infty(0,1)} + \|\zeta\|_{L^\infty(Q)}.$$

We now show that there exists a constant $N(T) > N(t_0)$ such that, if $h \in \mathcal{G}_\tau$ satisfies

$$\|h(t)\|_{L^\infty} \le N(T), \qquad 0 < t < T, \tag{4.17}$$

then $\mathcal{L}(h)$ also satisfies this inequality. We can seen this as follows.

$$\begin{aligned}\|\mathcal{L}(h)(t)\|_{L^\infty} \le & N(t_0) + 2M^2\frac{\tau_0}{\varepsilon}N(T) + \tau_0\|\partial_t\zeta\|_{L^\infty(Q)} + MN(T)\sqrt{\frac{\tau_0}{\varepsilon}} \\ \le & N(t_0) + \tau_0\|\partial_t\zeta\|_{L^\infty(Q)} + N(T)\bar{M}\sqrt{\frac{\tau_0}{\varepsilon}},\end{aligned}$$

where $\bar{M} = 2\max\{2M^2, MC_0, 1\}$. Therefore, one deduces that the assertion is true, provided

$$N(T) \ge \frac{N(t_0) + \tau_0\|\partial_t\zeta\|_{L^\infty(Q)}}{1 - \bar{M}\sqrt{\frac{\tau_0}{\varepsilon}}}. \tag{4.18}$$

Since $\mathcal{L}$ is a contraction mapping in $\mathcal{G}_\tau$, bound (4.17) must also hold for g, which then proves (4.8).

Once we have proven (4.8), we can easily show the existence of a unique global smooth solution of problem (2.1)–(2.3), corresponding to (4.1). The remaining of the proof of the existence of a solution to problem (1.1)–(1.3) follows the same procedure as the one for the systems in Section 3. In the polytropic case $p = \kappa\rho^\gamma$, after proving (3.5) with the help of Theorem 2.1, we may use the results in [**12**] ($\gamma = 1 + 2/(2k+1)$, $k > 1$), [**3**] ($1 < \gamma \le 5/3$), [**24**] ($\gamma \ge 3$), and [**25**] ($5/3 < \gamma < 3$) for the reduction of the Young measures to Dirac measures. The same can be done for more general pressure law $p(\rho)$ satisfying (4.2)-(4.3), by using the reduction procedure in the recent paper [**7**].

THEOREM 4.1. *Let a_0, a_1, and u_0 satisfy* (1.4) *and* (4.6). *Then there exists a global entropy solution of the initial-boundary problem* (4.1)-(4.3) *and* (1.2)-(1.3) *in the sense of* (1.7)-(1.9) *and*

$$\rho(t,x) \ge 0, \quad |m(t,x)| \le C\rho(t,x), \qquad \text{for some } C > 0 \text{ independent of } t.$$

Acknowledgments

Gui-Qiang Chen's research was supported in part by the National Science Foundation grants DMS-9623203 and DMS-9708261, and by an Alfred P. Sloan Foundation Fellowship. Hermano Frid's research was supported in part by CNPq-Brazil, proc. 352871/96-2.

References

[1] Bardos, C., Le Roux, A. Y., and Nedelec, J. C., *First order quasilinear equations with boundary conditions*, Comm. Partial Diff. Eqs. **4** (1979), 1017–1034.

[2] Benabdallah, A. and Serre, D., *Problèmes aux limites pour les systèmes hyperboliques non-linéaires de équations à une dimension d'espace*, C.R. Acad. Sc. Paris, Série I, **305** (1987), 677–680.

[3] Chen, G.-Q., *Convergence of the Lax-Friedrichs scheme for isentropic gas dynamics (III)*, Acta Mathematica Scientia **8** (1988), 243–276 (in Chinese); **6** (1986), 75–120.

[4] Chen, G.-Q., *Remarks on the paper "Convergence of the viscosity method for isentropic gas dynamics"* , Proc. Amer. Math. Soc. **125** (1997), 2981–2986.

[5] Chen, G.-Q. and Frid, H., *Divergence-measure fields and hyperbolic conservation laws*, Arch. Rat. Mech. Anal. (1999) (to appear).
[6] Chen, G.-Q. and Kan, P. T., *Hyperbolic conservation laws with umbilic degeneracy I*, Arch. Rational Mech. Anal. **130** (1995), 231–276.
[7] Chen, G.-Q., and LeFloch, Ph., *Compressible Euler equations with general pressure law*, Preprint, April 1998 (submitted).
[8] Chueh, K. N., Conley, C. C., and Smoller, J. A., *Positively invariant regions for systems of nonlinear diffusion equations*, Ind. Univ. Math. J. **26** (1977), 372–411.
[9] Dafermos, C. M., *Hyperbolic systems of conservation laws*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994), 1096–1107, Birkhäuser, Basel, 1995.
[10] Ding, X., Chen, G.-Q., and Luo, P., *Convergence of the Lax-Friedrichs scheme for isentropic gas dynamics (I),(II)*, Acta Mathematica Scientia, **5** (1985), 415–432, 433-472 (in English); **7** (1987), 467–480, **8** (1988), 61–94 (in Chinese).
[11] DiPerna, R., *Convergence of approximate solutions to conservation laws*, Arch. Rational Mech. Anal. **82** (1983), 27–70.
[12] DiPerna, R., *Convergence of the viscosity method for isentropic gas dynamics*, Commun. Math. Phys. **91** (1983), 1–30.
[13] Dubois, F. and LeFloch, Ph. G., *Boundary conditions for nonlinear hyperbolic systems of conservation laws*, J. Diff. Eqs. **71** (1988), 93–122.
[14] Frid, H., Compacidade Compensada e Aplicações às Leis de Conservação, Lecture Notes for the 19^{th} Brazilian Colloquium of Mathematics (in Portuguese), IMPA (1993).
[15] Frid, H. and Santos, M. M., *Nonstrictly hyperbolic systems of conservation laws of the conjugate type*, Commun. Partial Diff. Eqs. **19** (1994), 27–59.
[16] Glimm, J., *Solutions in the large for nonlinear hyperbolic systems of equations*, Comm. Pure Appl. Math. **18** (1965), 95–105.
[17] Evans, L. C., Partial Differential Equations, Amer. Math. Soc.: Providence, RI, 1998.
[18] James, F., Peng, Y.-J., and Perthame, B., *Kinetic formulation for chromatography and some other hyperbolic systems*, J. Math. Pures Appl. **74** (1995), 367-385.
[19] Joseph, K. T. and LeFloch, P., *Boundary layers in weak solutions to hyperbolic conservation laws*, preprint CMAP, # 341, Ecole Polytechnique, France (1998).
[20] Kan, P.-T., Santos, M., and Xin, Z., *Initial-boundary value problem for conservation laws*, Commun. Math. Phys. **186** (1997), 701-730.
[21] Lax, P. D., *Hyperbolic systems of conservation laws II*, Comm. Pure Appl. Math. **10** (1957), 537–566.
[22] Lax, P. D., *Shock waves and entropy*, In: Contributions to Functional Analysis, ed. E. A. Zarantonello, Academic Press, New York, 1971, pp. 603–634.
[23] Lions, J. L. and Magenes, E., Non-Homogeneous Boundary Value Problems and Applications, 2 Vols., Springer-Verlag (1972).
[24] Lions, P. L., Perthame, B., and Tadmor, E., *Kinetic formulation of the isentropic gas dynamics and p-system*, Commun. Math. Phys. **163** (1994), 415–431.
[25] Lions, P. L., Perthame, B., and Souganidis, P. E., *Existence of entropy solutions for the hyperbolic systems of isentropic gas dynamics in Eulerian and Lagrangian coordinates*, Comm. Pure Appl. Math. **49** (1995), 599–638.
[26] Liu, T.-P., *Initial-boundary value problems for gas dynamics*, Arch. Rational Mech. Anal. **64** (1977), 137–168.
[27] Majda, A., Compressible Fluid Flow and Systems of Conservation Laws in Several Space Variables, Applied Mathematical Sciences, 53. Springer-Verlag: New York-Berlin, 1984.
[28] Málec, J., Nečas, J., Rokyta, M., and Ružička, M., Weak and Measure-valued Solutions to Evolutionary PDEs, Chapman & Hall, London, 1996.
[29] Murat, F., *L'injection du cone positif de H^{-1} dans $W^{-1,q}$ est compacte pour tout $q < 2$*, J. Math. Pures Appl. **60** (1981), 309–322.
[30] Otto, F., First order equations with boundary conditions. Preprint no. 234, SFB 256, Univ. Bonn. 1992.
[31] Rhee, H.-K, Aris, R., and Amundson, N. R., *On the theory of multicomponent chromatography*, Philos. Trans. Roy. Soc. London, **A267** (1970), 419–455.
[32] Rubino, B., *On the vanishing viscosity approximation to the Cauchy problem for a 2×2 system of conservation laws*, Anal. Non Linéaire **10** (1993), 627–656.

[33] Serre, D., *Richness and the classification of quasilinear hyperbolic systems*, In: Multidimensional Hyperbolic Problems and Computations, ed. J. Glimm and A. J. Majda, IMA Vol. 29, Springer-Verlag, New York, 1991, 315–333.
[34] Serre, D., Systems of Conservation Laws, Fondations, Diderot Editeur, Paris, 1996.
[35] Smoller, J., Shock Waves and Reaction-Diffusion Equations, Springer-Verlag, New York, 1983.
[36] Szepessy, A., *Measure-valued solutions to scalar conservation laws with boundary conditions*, Arch. Rat. Mech. Anal. (1989), 181–193.
[37] Tartar, L., *Compensated compactness and applications to partial differential equations*, In: Research Notes in Mathematics, Nonlinear Analysis and Mechanics ed. R. J. Knops, **4**(1979), Pitman Press, New York, 136–211.
[38] Temple, B., *Systems of conservation laws with invariant submanifolds*, Trans. Amer. Math. Soc. **280** (1983), 781–795.

(Gui-Qiang Chen) Department of Mathematics, Northwestern University, 2033 Sheridan Road, Evanston, IL 60208-2730, USA
E-mail address: gqchen@math.nwu.edu

(Hermano Frid) Instituto de Matemática, Universidade Federal do Rio de Janeiro, C. Postal 68530, Rio de Janeiro, RJ 21945-970, Brazil
E-mail address: hermano@lpim.ufrj.br

Contemporary Mathematics
Volume **238**, 1999
B 0-8218-1196-7-03539-8

On the Prediction of Large-Scale Dynamics Using Unresolved Computations

Alexandre J. Chorin, Anton P. Kast, and Raz Kupferman

Abstract. We present a theoretical framework and numerical methods for predicting the large-scale properties of solutions of partial differential equations that are too complex to be properly resolved. We assume that prior statistical information about the distribution of the solutions is available, as is often the case in practice. The quantities we can compute condition the prior information and allow us to calculate mean properties of solutions in the future. We derive approximate ways for computing the evolution of the probabilities conditioned by what we can compute, and obtain ordinary differential equations for the expected values of a set of large-scale variables. Our methods are demonstrated on two simple but instructive examples, where the prior information consists of invariant canonical distributions

1. Introduction

There are many problems in science that can be modeled by a set of differential equations, but where the solution of these equations is so complicated that it cannot be found in practice, either analytically or numerically. For a numerical computation to be accurate the problem must be well resolved, i.e, enough variables (or "degrees of freedom") must be represented in the calculation to capture all the relevant features of the solution; insufficient resolution yields sometimes disastrous results. A well-known example in which good resolution cannot be achieved is turbulent flow, where one has to resolve all scales ranging from the size of the system down to the dissipation scale—a prohibitively expensive requirement. One is then compelled to consider the question of how to predict complex behavior when the number of variables that can be used in the computation is significantly less than needed for full resolution. This is the question considered in the present paper; part of the theoretical framework and methods have already been briefly discussed in [**CKK98**].

Studies on underresolved problems exist in a wide range of different contexts, along with a large amount of literature that describes problem-specific methods. In turbulence, for example, there are various modeling methods for large eddy simulations. In all cases one needs to make additional assumptions about the relation

1991 *Mathematics Subject Classification.* Primary 65M99.

This work was supported in part by the US Department of Energy under contract DE-AC03-76-SF00098, and in part by the National Science Foundation under grant DMS94-14631.

between those degrees of freedom that are represented in the computation and the "hidden", or "invisible" degrees of freedom that are discarded from the computation. A number of interesting attempts have been made over the years to fill in data from coarse grids in difficult computations so as to enhance accuracy without refining the grid (see e.g. [**SM97, MW90**]). Indeed, nothing can be done without some information regarding the unresolved degrees of freedom. Such additional assumptions are usually motivated by intuitive reasoning and their validity is usually assessed by comparing the resulting predictions to experimental measurements.

In many problems the lack of resolution is due primarily to the insufficiency and sometimes also the inaccuracy of the measurements that provide initial conditions for the system of equations. This is the case for example in weather forecasting, where the initial information consists of local weather measurements collected at a relatively small number of meteorological stations. The problem of insufficient and sometimes noisy data is not considered in the present paper. We focus here on the case where underresolution is imposed by computational limitations. Initial data will be assumed to be available at will, and this assumption will be fully exploited by allowing us to select the set of degrees of freedom that are represented in the computation at our convenience. Another issue that often arises in the modeling of complex systems is uncertainty regarding the equations themselves. This important question is also beyond the scope of this paper; the adequacy of the system of equations to be solved is taken for granted.

We now define the problem and introduce some of the nomenclature: We consider a system described by a differential equation of the form

$$u_t = F(u), \tag{1.1}$$

where t is time, subscripts denote differentiation, $u(x,t)$ is the dependent variable, and $F(u) = F(u, u_x, u_{xx}, \dots)$ is a (generally nonlinear) function of its arguments; the spatial coordinate x and the dependent variable u can be of arbitrary dimensionality.

To solve an equation of the form (1.1) on a computer one ordinarily discretizes the dependent variable $u(x,t)$ both in space and time and replaces the differential equation by an appropriate relation between the discrete variables. As described, the solution to the discrete system may approximate the solution of the differential equation well only if the discretization is sufficiently refined. It is our basic assumption that we cannot afford such a refined discretization, and must therefore be content with a much smaller number of variables. One still has the liberty to choose the degrees of freedom that are retained in the computation; those will be chosen, for convenience, to be linear functionals of the dependent variable $u(x,t)$:

$$U_\alpha[u(\cdot,t)] \equiv (g_\alpha(\cdot), u(\cdot,t)) \equiv \int g_\alpha(x) u(x,t)\, dx, \tag{1.2}$$

where α is an index that enumerates the selected degrees of freedom. Variables of the form (1.2) will be referred to as *collective variables*; every collective variable U_α is defined by a *kernel* $g_\alpha(x)$. Point values of $u(x)$ at a set of points x_α, and spectral components of $u(x)$ for a set of modes k_α are two special cases of collective variables; in the first case the corresponding kernels are delta functions, $g_\alpha(x) = \delta(x - x_\alpha)$, whereas in the second case the kernels are spectral basis functions, $\exp(ik_\alpha \cdot x)$. We assume that our computational budget allows us to operate on a set of at most N collective variables, so that $\alpha = 1, \dots, N$. The question is, what can be predicted

about the state of the system at a future time t given the values of the collective variables U_α at an initial time $t = 0$?

Suppose that we know at time $t = 0$ that the collective variables U_α assume a set of values V_α. (We will denote by $U = (U_1, \ldots, U_N)^T$ and $V = (V_1, \ldots, V_N)^T$ the vectors whose entries are the collective variables and their initial values, respectively.) Our postulate that the number of collective variables N does not suffice to resolve the state of the system implies that the initial data, V, do not determine sharply enough the initial condition, $u(x, 0)$. A priori, every function $u(x, 0)$ that is compatible with the given values of the collective variables, that is, belongs to the set

$$\mathcal{M}(V) = \{v(x) : U_\alpha[v(\cdot)] = V_\alpha, \quad \alpha = 1, \ldots, N\}. \tag{1.3}$$

is a plausible initial condition. One could define underresolution in terms of the set of functions (1.3); the problem is underresolved if this set is non-trivial. Clearly, the state of the system at future times depends on the particular initial condition; in many cases it is even very sensitive to small variations in the initial condition. One wonders then in what sense the future can be predicted when the initial condition is not known with certainty.

The essence of our approach is the recognition that underresolution necessarily forces one to consider the evolution of a set, or ensemble, of solutions, rather than a single initial value problem. This requires the replacement of equation (1.1) by a corresponding equation for a probability measure defined on the space of the solutions of (1.1). The prediction of the future state of the system can then be reinterpreted as the prediction of most likely, or mean, properties of the system. Loosely stated, in cases where sufficient resolution cannot be achieved the original task of solving an initial value problem has to be replaced by a more modest one—the determination of "what is most likely to happen given what is initially known."

At first, there seems to be no practical progress in the above restatement of the problem. First, the statistical problem also requires initial conditions; a measure defined on the space of initial conditions $u(x, 0)$ must be provided for the statistical problem to be well-defined. Second, the high-dimensional Liouville equation that describes the flow induced by (1.1) is not easier to solve than the original initial value problem. It turns out that in many problems of interest there exists a natural measure μ that characterizes the statistical properties of the system; what is meant by "natural" has to be clarified; an important class of such measures are invariant ones. We are going to use this information to partially cure the two aforementioned difficulties: First, this measure will define the initial statistical state of the system by being interpreted as a "prior" measure—a quantification of our beliefs regarding the state of the system prior to the specification of any initial condition. The initial values of the collective variables are constraints on the set of initial states and induce on μ a conditional measure that constitutes an initial condition for the Liouville equation. Second, the existence of a distinguished statistical measure suggests a way to generate a hierarchy of approximations to the Liouville equation, examples of which will be described in the following sections.

The rest of this paper is organized as follows: In Section 2 we present our theory, and provide a recipe (2.11) for approximating the mean evolution of a set of collective variables. In Section 3 we derive formulas for the calculation of conditional expectations in the case of Gaussian prior measures; these are necessary for the evaluation of the right-hand side of equation (2.11). In Sections 4 and 5

we demonstrate the power of our theory by considering two examples: a linear Schrödinger equation and a nonlinear Hamiltonian system. Conclusions are presented in Section 6.

2. Presentation of the Theory

Our starting point is a general equation of motion of the form (1.1), and a set of collective variable U_α defined by (1.2) for a set of kernels $g_\alpha(x)$; the question of what constitutes a good choice of kernels will be discussed below.

In many problems of interest there exists a measure on the space of solutions of (1.1) that is invariant under the flow induced by (1.1); a measure that has this property is referred to as an *invariant measure*. Invariant measures are known to play a central role in many problems; macroscopic systems (that is, systems that have a very large number of degrees of freedom) whose macroscopic properties do no change in time, often exhibit an invariant statistical state. By that we mean the following: when the large scale observable properties of the system remain constant in time, the likelihood of the microscopic degrees of freedom to be in any particular state is distributed according to a measure that is invariant in time. We will assume that such an invariant measure μ_0 exists and that we know what it is. The measure μ_0 will then be postulated to be the *prior measure*, i.e, it describes the probability distribution of initial conditions before any measurement has been performed. We will denote averages with respect to the invariant measure μ_0 by angle brackets $\langle\cdot\rangle$; let $O[u(\cdot)]$ be a general functional of u, then

$$\langle O\rangle = \int O[u(\cdot)]\, d\mu_0, \tag{2.1}$$

where the integration is over an appropriate function space. We shall write formally,

$$d\mu_0 = f_0[u(\cdot)]\,[du], \tag{2.2}$$

as if the measure μ_0 were absolutely continuous with respect to a Lebesgue measure, where $f_0[u]$ is the invariant probability density, and $[du]$ is a formal product of differentials.

We next assume that a set of measurements has been carried out and has revealed the values V_α of the collective variables U_α at time $t = 0$. This information can be viewed as a set of constraints on the set of initial conditions, which is now given by (1.3). Constraints on the set of functions $u(x)$ automatically induce on μ_0 a *conditional measure*, which we denote by μ_V. In a physicist's notation,

$$d\mu_V = f_V[u(\cdot)]\,[du] = c f_0[u(\cdot)]\,[du] \times \prod_{\alpha=1}^{N} \delta\left(U_\alpha[u(\cdot)] - V_\alpha\right), \tag{2.3}$$

where $f_V[u(\cdot)]$ is the conditional probability density, and c is an appropriate normalization factor. The conditional probability density is equal, up to a normalization, to the prior probability density projected on the space of functions $\mathcal{M}(V)$ that are compatible with the initial data. Note that the conditional measure μ_V is, in general, not invariant. Averages with respect to the conditional measure will be denoted by angle brackets with a subscript that symbolizes the constraints imposed on the set of functions,

$$\langle O\rangle_V \equiv \int O[u(\cdot)] f_V[u(\cdot)]\,[du]. \tag{2.4}$$

The dynamics have not been taken into consideration so far, except for the fact that the measure μ_0 was postulated to be invariant. Let $f[u(\cdot),t]$ be the probability density of the solutions of (1.1) at time t, that is, the probability density that evolves from the initial probability density $f_V[u(\cdot)]$ under the flow induced by (1.1); it satisfies the Liouville equation [**Ris84**]

$$f_t + \left(\frac{\delta f}{\delta u}(\cdot), F(u(\cdot))\right) = 0, \tag{2.5}$$

where $\frac{\delta f}{\delta u}$ denotes a functional derivative. An equivalent statement is that if S_t denotes the time evolution operator induced by (1.1), i.e., $S_t : u(x,0) \to u(x,t)$, then

$$f[u(\cdot),t] = f[S_t^{-1}u(\cdot),0] = f_V[S_t^{-1}u(\cdot)], \tag{2.6}$$

where S_t^{-1} is the operator inverse to S_t, which we assume to exist.

The objective that has been defined in the introductory section is to calculate the expectation value of observables $O[u(\cdot)]$ at time t, given the initial data V. In terms of the notations introduced above this is given by

$$\langle O[u(\cdot),t]\rangle_V = \langle O[S_t u(\cdot)]\rangle_V \tag{2.7}$$

(operators are generally treated as function of the dependent variable and time, $O[u(\cdot),t]$; when no reference to time is being made the expression refers to the initial time).

We next make the following observations: (i) The initial probability measure (2.3) is completely determined by the N numbers V_α. (ii) By the invariance of $f_0[u]$ and by equation (2.6), the probability density at later time t can still be represented as the invariant density projected on a set of N conditions; specifically,

$$f[u(\cdot),t] = c\, f_0[u(\cdot)] \prod_{\alpha=1}^{N} \delta\left[\left(g_\alpha(\cdot), S_t^{-1}u(\cdot)\right) - V_\alpha\right]. \tag{2.8}$$

Note however that the set of functions that support this measure at time t is generally not of the form (1.3), that is, the observable $\left(g_\alpha(\cdot), S_t^{-1}u(\cdot)\right)$ is not a linear functional of u.

These observations suggest an approximate procedure for solving the Liouville equation (2.5). We propose an ansatz in which the N conditions that are imposed on μ_0 remain for all times conditions on the values of the collective variables U; namely, the probability density is specified by a time-dependent vector of N numbers $V_\alpha(t)$, such that

$$f[u(\cdot),t] \approx c\ f_0[u(\cdot)] \prod_{\alpha=1}^{N} \delta\left[U_\alpha[u(\cdot)] - V_\alpha(t)\right]. \tag{2.9}$$

One has still to specify the time evolution of the vector $V(t)$. Suppose that the distribution of solutions is indeed given by (2.9) at time t, and consider a later time $t+\Delta t$. The value of the observable $U_\alpha[u(\cdot)]$ at the later time will, in general, not be uniform throughout the ensemble of solutions. The ansatz (2.9) projects the distribution back onto a set of solutions $\mathcal{M}(V(t+\Delta t))$. A natural choice for $V_\alpha(t+\Delta t)$ is the expectation value of the collective variable $U_\alpha[u(\cdot)]$ given that the

distribution at time t was (2.9):

$$V_\alpha(t+\Delta t) \approx \langle U_\alpha[u(\cdot), t+\Delta t]\rangle_{V(t)} = \\ = \langle U_\alpha[u(\cdot)]\rangle_{V(t)} + \Delta t\ \langle (g_\alpha(\cdot), F(u(\cdot)))\rangle_{V(t)} + O(\Delta t^2). \tag{2.10}$$

Taking the limit $\Delta t \to 0$ we finally obtain,

$$\frac{dV_\alpha}{dt} = \langle (g_\alpha(\cdot), F(u(\cdot)))\rangle_{V(t)}\,. \tag{2.11}$$

Equation (2.11) is our main tool in the present paper and we will next discuss its implications:

- Equation (2.11) constitutes a closed set of N ordinary differential equations, which by our postulate is within the acceptable computational budget.
- The central hypothesis in the course of the derivation was that the distribution of solutions can be approximated by (2.9). This approximation assumes that for all times t the collective variable U_α has a uniform value V_α *for all the trajectories in the ensemble of solutions*. This assertion is initially correct (by construction) at time $t = 0$, but will generally not remain true for later times. The approximation is likely to be a good one as long as the above assertion is approximately true, that is, as long as the distribution of values assumed by the collective variables remains sufficiently narrow. In many cases it is possible to guarantee a small variance by a clever selection of collective variables (i.e., of kernels). Note furthermore that the smallness of the variance can be verified self-consistently from the knowledge of the probability density (2.9).
- Equation (2.11) still poses the technical problem of computing its right-hand side. This issue is the subject of the next section.
- The case where the equations of motion (1.1) are linear, i.e,

$$u_t = Lu, \tag{2.12}$$

with L being a linear operator, can be worked out in detail. Using the fact that $S_t = \exp(Lt)$, the solution to the Liouville equation (2.8) can be rearranged as

$$f[u(\cdot), t] = c f_0[u(\cdot)] \prod_{\alpha=1}^{N} \delta\left[\left(e^{-L^\dagger t} g_\alpha(\cdot), u(\cdot)\right) - V_\alpha\right], \tag{2.13}$$

where $L^\dagger$ is the linear operator adjoint to L. Thus, the probability density for all times is f_0 projected on the set of functions for which a set of N linear functionals of u have the values V; note that V here is not time dependent, but is the vector of initial values of the collective variables U. The kernels that define these functionals are time dependent, and evolve according to the dual equation

$$\frac{dg_\alpha}{dt} = -L^\dagger g_\alpha. \tag{2.14}$$

If the kernels g_α are furthermore eigenfunctions of the dual operator $L^\dagger$ with eigenvalues λ_α, the ansatz (2.9) is exact, with $V_\alpha(t) = V_\alpha(0)\, e^{\lambda_\alpha t}$. Hald [**Hal**] shows that by selecting kernels that are *approximate* eigenfunctions of $L^\dagger$, one can bound the error introduced by the ansatz (2.9), while retaining the simplicity of the procedure.

- The two alternatives of evolving either the values V_α or the kernels $g_\alpha(x)$ are analogous to Eulerian versus Lagrangian approaches in fluid mechanics, or Schrödinger versus Heisenberg approaches in quantum mechanics. For nonlinear equations one has a whole range of intermediate possibilities; for example one may split the operator F in equation (1.1) as $F = L+Q$, where L is linear. The kernels can be evolved according to the linear operator, while the values of the collective variables can be updated by the remaining nonlinear operator. The art is to find partitions $F = L + Q$ that minimize the variance of the distribution of values assumed by the collective variables.
- Equation (2.11) should be viewed as a first approximation to the solution of the Liouville equation, where the only information that is updated in time is the mean value of a fixed set of collective variables. In principle, one could also update higher moments of those variables, and use this additional information to construct a better approximation. For example, equipped with the knowledge of means and covariances one could find new kernels and new values for the corresponding collective variables, such that the distribution obtained by conditioning the invariant distribution with those new constraints is compatible with the calculated means and covariances. Thus, one could imagine an entire hierarchy of schemes that take into account an increasing number of moments of the resolved variables.

3. Conditional Expectation with Gaussian Prior

Equation (2.11) is a closed set of equations for the vector $V(t)$, which requires the computation of a conditional average on its right-hand side. To have a fully constructive procedure, we need to evaluate conditional averages $\langle O[u(\cdot)]\rangle_V$, where O is an arbitrary observable, and V denotes as before the vector of values of a set of collective variables U of the form (1.2). In this section we present three lemmas that solve this problem for the case where the prior measure μ_0 is Gaussian. In the two examples below, the prior measure is either Gaussian or can be viewed as a perturbation of a Gaussian measure.

The random function $u(x)$ has a Gaussian distribution if its probability density is of the form

$$f_0[u(\cdot)] = Z^{-1} \exp\left(-\frac{1}{2}\iint u(x)a(x,y)u(y)\,dx\,dy + \int b(x)u(x)\,dx\right), \tag{3.1}$$

where $a(x,y)$ and $b(x)$ are (generalized) functions, and Z is a normalizing constant. The functions $a(x,y)$ and $b(x)$ are related to the mean and the covariance of $u(x)$ by

$$\langle u(x)\rangle = \left(a^{-1}(x,\cdot), b(\cdot)\right), \tag{3.2}$$

and

$$\mathrm{Cov}\,[u(x),u(y)] \equiv \langle u(x)u(y)\rangle - \langle u(x)\rangle\,\langle u(y)\rangle = a^{-1}(x,y), \tag{3.3}$$

where the generalized function $a^{-1}(x,y)$ is defined by the integral relation

$$\left(a(x,\cdot), a^{-1}(\cdot,y)\right) = \left(a^{-1}(x,\cdot), a(\cdot,y)\right) = \delta(x-y). \tag{3.4}$$

To compute the expectation value of higher moments of u one can use Wick's theorem [**Kle89**]:

$$\langle (u_{i_1} - \langle u_{i_1} \rangle) \cdots (u_{i_l} - \langle u_{i_l} \rangle) \rangle = \begin{cases} 0, & l \text{ odd} \\ \sum \mathrm{Cov}\left[u_{i_{p_1}}, u_{i_{p_2}}\right] \cdots \mathrm{Cov}\left[u_{i_{p_{l-1}}}, u_{i_{p_l}}\right], & l \text{ even} \end{cases}, \tag{3.5}$$

with summation over all possible pairings of $\{i_1, \ldots, i_l\}$.

Next, suppose that the random function $u(x)$ is drawn from a Gaussian distribution, and a set of measurements reveal the vector of values V for a set of collective variables U of the form (1.2). This information changes the probability measure μ_0 into a conditional measure μ_V with density f_V given by (2.3). Conditional averages of operators $O[u(\cdot)]$ can be calculated by using the following three lemmas:

LEMMA 3.1. *The conditional expectation of the function $u(x)$ is a linear form in the conditioning data V:*

$$\langle u(x) \rangle_V = \langle u(x) \rangle + \sum_{\alpha=1}^{N} c_\alpha(x) \left\{ V_\alpha - \langle U_\alpha[u(\cdot)] \rangle \right\}, \tag{3.6}$$

where the vector of functions $c_\alpha(x)$ is given by

$$c_\alpha(x) = \sum_{\beta=1}^{N} \left(a^{-1}(x, \cdot), g_\beta(\cdot) \right) m^{-1}_{\beta\alpha}, \tag{3.7}$$

and where the $m^{-1}_{\beta\alpha}$ are the entries of an $N \times N$ matrix M^{-1} whose inverse M has entries

$$m_{\beta\alpha} = Cov\left[U_\beta[u(\cdot)], U_\alpha[u(\cdot)]\right] = \iint g_\beta(x) a^{-1}(x, y) g_\alpha(y)\, dx\, dy. \tag{3.8}$$

PROOF. Given the prior measure μ_0 and the values V of the collective variables U, we define a *regression function* (an approximant to $u(x)$) of the form

$$R(x) = \sum_{\alpha=1}^{N} r_\alpha(x) V_\alpha + s(x), \tag{3.9}$$

where the functions $r_\alpha(x)$ and $s(x)$ are chosen such to minimize the mean square error,

$$E(x) = \left\langle e^2(x) \right\rangle \equiv \left\langle \left[u(x) - \sum_{\alpha=1}^{N} r_\alpha(x) U_\alpha[u(\cdot)] - s(x) \right]^2 \right\rangle \tag{3.10}$$

for all x. Note that this is an *unconditional* average with respect to μ_0.

Minimization with respect to $s(x)$ implies that

$$\frac{\partial E(x)}{\partial s(x)} = \langle e(x) \rangle = \left\langle u(x) - \sum_{\alpha=1}^{N} r_\alpha(x) U_\alpha[u(\cdot)] - s(x) \right\rangle = 0, \tag{3.11}$$

which, combined with (3.9), yields

$$R(x) = \langle u(x) \rangle + \sum_{\alpha=1}^{N} r_\alpha(x) \left\{ \langle U_\alpha[u(\cdot)] \rangle - V_\alpha \right\}. \tag{3.12}$$

Minimization with respect to $r_\alpha(x)$ implies:

(3.13)
$$\frac{\partial E(x)}{\partial r_\alpha(x)} = \langle e(x)\, U_\alpha[u(\cdot)]\rangle = \left\langle \left[u(x) - \sum_{\beta=1}^{N} r_\beta(x) U_\beta[u(\cdot)] - s(x) \right] U_\alpha[u(\cdot)] \right\rangle = 0.$$

Equation (3.13) can be rearranged by substituting equations (3.3) and (3.11) into it, and using the fact that $U_\alpha[u(\cdot)] = (g_\alpha(\cdot), u(\cdot))$:

$$\sum_{\beta=1}^{N} \mathrm{Cov}\left[U_\alpha[u(\cdot)], U_\beta[u(\cdot)]\right] r_\beta(x) = \left(g_\alpha(\cdot), a^{-1}(x,.)\right). \tag{3.14}$$

One readily identifies the functions $r_\alpha(x)$ as satisfying the definition (3.7) of the functions $c_\alpha(x)$. Comparing (3.12) with (3.6), the regression function is nothing but the right-hand side of equation (3.6).

It remains to show that the regression curve equals also the left-hand side of (3.6). Consider equation (3.13): it asserts that the random variable $e(x)$ is statistically orthogonal to the random variables $U_\alpha[u(\cdot)]$. Note that both $e(x)$ and the collective variables U_α are linear functionals of the Gaussian function $u(x)$, and are therefore jointly Gaussian. Jointly Gaussian variables that are statistically orthogonal are independent, hence, the knowledge of the value assumed by the variables $U_\alpha[u(\cdot)]$ does not affect the expectation value of $e(x)$,

(3.15)
$$\left\langle u(x) - \sum_{\alpha=1}^{N} r_\alpha(x) U_\alpha[u(\cdot)] - s(x) \right\rangle_V = \left\langle u(x) - \sum_{\alpha=1}^{N} r_\alpha(x) U_\alpha[u(\cdot)] - s(x) \right\rangle.$$

The function $s(x)$ is not random and $\langle U_\alpha[u(\cdot)]\rangle_V = V_\alpha$, from which immediately follows that

$$\langle u(x)\rangle_V = \langle u(x)\rangle + \sum_{\alpha=1}^{N} r_\alpha(x) \left\{ V_\alpha - \langle U_\alpha[u(\cdot)]\rangle \right\}. \tag{3.16}$$

This completes the proof. □

Lemma 3.2. *The conditional covariance of the function $u(x)$ differs from the unconditional covariance by a function that depends on the kernels $g_\alpha(x)$, without reference to the conditioning data V:*

$$Cov\left[u(x), u(y)\right]_V = Cov\left[u(x), u(y)\right] - \sum_{\alpha=1}^{N} c_\alpha(x) \left(g_\alpha(\cdot), a^{-1}(\cdot, y)\right). \tag{3.17}$$

Proof. The proof follows the same line as the second part of the proof of Lemma 3.1. Consider the following expression:

(3.18)
$$e(x)e(y) = \left[u(x) - \sum_{\alpha=1}^{N} r_\alpha(x) U_\alpha[u(\cdot)] - s(x) \right] \left[u(y) - \sum_{\beta=1}^{N} r_\beta(y) U_\beta[u(\cdot)] - s(y) \right].$$

Both $e(x)$ and $e(y)$ are independent of the collective variables U. It is always true that if A_1, A_2 and A_3 are random variables with A_3 being independent of A_1 and

A_2, then $\langle A_1 A_2 \rangle_{A_3} = \langle A_1 A_2 \rangle$. Hence,

$$\langle e(x)e(y) \rangle_V = \langle e(x)e(y) \rangle \,, \tag{3.19}$$

from which (3.17) follows after straightforward algebra. □

LEMMA 3.3. *Wick's theorem extends to conditional expectations:*

$$\left\langle (u_{i_1} - \langle u_{i_1} \rangle_V) \cdots (u_{i_l} - \langle u_{i_l} \rangle_V) \right\rangle_V = \tag{3.20}$$
$$\begin{cases} 0, & l \text{ odd} \\ \sum Cov\left[u_{i_{p_1}}, u_{i_{p_2}}\right]_V \cdots Cov\left[u_{i_{p_{l-1}}}, u_{i_{p_l}}\right]_V, & l \text{ even} \end{cases},$$

where again the summation is over all possible pairings of $\{i_1, \ldots, i_l\}$.

PROOF. Using the fact that a delta function can be represented as the limit of a narrow Gaussian function, the conditional expectation of any list of observables, $O_1[u(\cdot)], \ldots, O_p[u(\cdot)]$, can be expressed as

$$\langle O_1[u(\cdot)] \cdots O_p[u(\cdot)] \rangle_V = \lim_{\Delta \to 0} \int O_1[u(\cdot)] \cdots O_p[u(\cdot)] \, f_V^\Delta[u(\cdot)][du], \tag{3.21}$$

where

$$f_V^\Delta[u(\cdot)] = c_\Delta \, f_0[u(\cdot)] \prod_{\alpha=1}^{N} \frac{1}{\sqrt{\pi}\Delta} \exp\left[-\frac{(U_\alpha[u(\cdot)] - V_\alpha)^2}{\Delta^2} \right], \tag{3.22}$$

the coefficient c_Δ is a normalization, and the order of the limit $\Delta \to 0$ and the functional integration has been interchanged. Note that the exponential in (3.22) is quadratic in $u(x)$, hence the finite-Δ probability density $f_V^\Delta[u(\cdot)]$ is Gaussian, Wick's theorem applies, and the limit $\Delta \to 0$ can finally be taken. □

The conditional expectation of any observable $O[u(\cdot)]$ can be deduced, in principle, from a combination of Lemmas 3.1-3.3.

In the examples considered below, the dependent variable $u(x,t)$ is a vector; let $u^i(x,t)$ denote the i'th component of the d-dimensional vector $u(x,t)$. All the above relations are easily generalized to the vector case. To keep notations as clear as possible, we denote indices associated with the collective variables by Greek subscripts, and indices associated with the components of u by Roman superscripts. The probability density $f_0[u(\cdot)]$ is Gaussian if it is of the following form,

$$f_0[u(\cdot)] = \frac{1}{Z} \exp\left(-\frac{1}{2} \sum_{i,j=1}^{d} \iint u^i(x) a^{ij}(x,y) u^j(y) \, dx \, dy + \sum_{i=1}^{d} \int b^i(x) u^i(x) \, dx \right), \tag{3.23}$$

where $a^{ij}(x,y)$ are now the entries of a $d \times d$ matrix of functions, and $b^i(x)$ are the entries of a vector of functions. These functions are related to the mean and the covariance of the vector $u(x)$ by

$$\left\langle u^i(x) \right\rangle = \sum_{j=1}^{d} \left([a^{-1}(x,\cdot)]^{ij}, b^j(\cdot) \right), \tag{3.24}$$

and

$$\mathrm{Cov}\left[u^i(x), u^j(y)\right] = [a^{-1}(x,y)]^{ij}, \tag{3.25}$$

where $[a^{-1}(x,y)]^{ij}$ is defined by

$$\sum_{j=1}^{d} \left([a^{-1}(x,\cdot)]^{ij}, a^{jk}(\cdot,y) \right) = \delta(x-y)\,\delta_{ik}. \tag{3.26}$$

Suppose now that a set of measurements reveals the values V_α^i of a matrix of collective variables of the form,

$$U_\alpha^i[u(\cdot)] = \left(g_\alpha(\cdot), u^i(\cdot) \right), \tag{3.27}$$

where $\alpha = 1, \ldots, N$ and $i = 1, \ldots, d$. The conditional expectation and covariance of $u^i(x)$ are given by straightforward generalizations of Lemmas 3.1 and 3.2:

$$\left\langle u^i(x) \right\rangle_V = \left\langle u^i(x) \right\rangle + \sum_{\alpha=1}^{N} \sum_{j=1}^{d} c_\alpha^{ij}(x) \left\{ V_\alpha^j - \left\langle U_\alpha^j[u(\cdot)] \right\rangle \right\}, \tag{3.28}$$

and

$$\mathrm{Cov}\left[u^i(x), u^j(y)\right]_V = \mathrm{Cov}\left[u^i(x), u^j(y)\right] - \sum_{\alpha=1}^{N} \sum_{k=1}^{d} c_\alpha^{ik}(x) \left(g_\alpha(\cdot), [a^{-1}(\cdot,y)]^{kj} \right), \tag{3.29}$$

where

$$c_\alpha^{ij}(x) = \sum_{\beta=1}^{N} \sum_{k=1}^{d} \left([a^{-1}(x,\cdot)]^{ik}, g_\beta(\cdot) \right) [m^{-1}]_{\beta\alpha}^{kj}, \tag{3.30}$$

and where the $[m^{-1}]_{\beta\alpha}^{ij}$ are the entries of an $N \times N \times d \times d$ tensor M^{-1} whose inverse M has entries

$$m_{\beta\alpha}^{ij} = \iint g_\beta(x) [a^{-1}(x,y)]^{ij} g_\alpha(y)\, dx\, dy. \tag{3.31}$$

4. A Linear Schrödinger Equation

The equations of motion. The first example is a linear Schrödinger equation that we write as a pair of real equations:

$$\begin{aligned} p_t &= -q_{xx} + m_0^2 q \\ q_t &= +p_{xx} - m_0^2 p \end{aligned}, \tag{4.1}$$

where $p(x,t)$ and $q(x,t)$ are defined on the domain $(0, 2\pi]$, m_0 is a constant, and periodic boundary conditions are assumed. Equations (4.1) are the Hamilton equations of motion for the Hamiltonian [**FH65**],

$$H[p(\cdot), q(\cdot)] = \frac{1}{2} \int_0^{2\pi} \left[(p_x)^2 + (q_x)^2 + m_0^2 (p^2 + q^2) \right] dx, \tag{4.2}$$

with $p(x)$ and $q(x)$ being the canonically conjugate variables.

The prior measure. Equation (4.1) preserves any density that is a function of the Hamiltonian. We will assume that the prior measure is given by the canonical density,

$$f_0[p(\cdot), q(\cdot)] = \exp\left\{ -H[p(\cdot), q(\cdot)] \right\}, \tag{4.3}$$

where the temperature has been chosen equal to one.

The measure defined by equation (4.3) is absolutely continuous with respect to a Wiener measure [**McK95**], and its samples are, with probability one, almost

nowhere differentiable. The corresponding solutions of the equations of motion are weak and hard to approximate numerically.

The Hamiltonian (4.2) is quadratic in p and q, hence the probability density (4.3) is Gaussian. By symmetry we see that the unconstrained means $\langle p(x)\rangle$ and $\langle q(x)\rangle$ are zero. To extract the matrix of covariance functions A^{-1}, we write the Hamiltonian (4.2) as a double integral:

$$H[p(\cdot),q(\cdot)] = \iint \Big[p_x(x)\delta(x-y)p_x(y) + q_x(x)\delta(x-y)q_x(y) + \\ + m_0^2\, p(x)\delta(x-y)p(y) + m_0^2\, q(x)\delta(x-y)q(y)\Big]\, dx\, dy. \tag{4.4}$$

Integration by parts shows that the entries of the matrix of functions A are

$$a^{ij}(x,y) = \left[-\delta''(x-y) + m_0^2\,\delta(x-y)\right]\,\delta_{ij}, \tag{4.5}$$

where the indices i and j represent either p or q, and $\delta''(\cdot)$ is a second derivative of a delta function. The integral equation for the inverse operator A^{-1} can be solved by Fourier series. The result is a translation-invariant diagonal matrix

$$[a^{-1}(x,y)]^{ij} = \frac{1}{2\pi}\delta_{ij}\sum_{k=-\infty}^{\infty}\frac{e^{ik(x-y)}}{k^2+m_0^2}. \tag{4.6}$$

The collective variables. We assume that the initial data for equations (4.1) are drawn from the distribution (4.3), and that $2N$ measurements have revealed the values of the $2N$ collective variables,

$$\begin{aligned} U_\alpha^p[p(\cdot),q(\cdot)] &\equiv (g_\alpha(\cdot),p(\cdot)) = V_\alpha^p \\ U_\alpha^q[p(\cdot),q(\cdot)] &\equiv (g_\alpha(\cdot),q(\cdot)) = V_\alpha^q, \end{aligned} \tag{4.7}$$

for $\alpha = 1,\dots,N$. The kernels $g_\alpha(x)$ are translates of each other, $g_\alpha(x) = g(x-x_\alpha)$, and the points $x_\alpha = 2\pi\alpha/N$ form a regular mesh on the interval $(0,2\pi]$. We choose

$$g(x) = \frac{1}{\sqrt{\pi}\sigma}\sum_{\tau=-\infty}^{\infty}\exp\left[-\frac{(x-2\pi\tau)^2}{\sigma^2}\right], \tag{4.8}$$

i.e., the kernel is a normalized Gaussian whose width is σ, with suitable images to enforce periodicity. The Fourier representation of $g(x)$ is

$$g(x) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{ikx}e^{-\frac{1}{4}k^2\sigma^2}. \tag{4.9}$$

We could have trivialized this example by choosing as kernels a set of trigonometric functions, which are eigenfunctions of the evolution operator. The goal here is to demonstrate what one could do when an exact representation of the eigenfunctions is not known.

Conditional expectation. We now demonstrate the application of the Lemmas derived in the previous section. Given the initial data, V^p and V^q, we may calculate the expectation of the functions $p(x)$ and $q(x)$; these conditional averages are given by equation (3.28). Because the unconditional averages of $p(x)$, $q(x)$, U_α^p and U_α^q all vanish, and the unconditional covariance $[a^{-1}(x,y)]^{ij}$ is diagonal with

respect to i and j (p and q are independent), equation (3.28) reduces to a simpler expression; the conditional average of $p(x)$, for example, is

$$\langle p(x)\rangle_V = \sum_{\alpha=1}^{N} c_\alpha^{pp}(x) V_\alpha^p, \tag{4.10}$$

where

$$c_\alpha^{pp}(x) = \sum_{\beta=1}^{N} \left([a^{-1}(x,\cdot)]^{pp}, g_\beta(\cdot)\right) [m^{-1}]_{\beta\alpha}^{pp} = c_\alpha^{qq}(x), \tag{4.11}$$

and $[m^{-1}]_{\beta\alpha}^{pp}$ are the entries of an $N \times N$ matrix M^{-1} (the upper indices p are considered as fixed) whose inverse M has entries

$$m_{\beta\alpha}^{pp} = \iint g_\beta(x)[a^{-1}(x,y)]^{pp} g_\alpha(y)\, dx\, dy = m_{\beta\alpha}^{qq}. \tag{4.12}$$

Substituting the Fourier representations of A^{-1} (4.6) and g (4.9), we obtain

$$c_\alpha^{pp}(x) = \frac{1}{2\pi} \sum_{\alpha=1}^{N} \sum_{k=-\infty}^{\infty} \frac{e^{-\frac{1}{4}k^2\sigma^2}}{k^2+m_0^2} \exp\left[ik(x-x_\beta)\right] [m^{-1}]_{\beta\alpha}^{pp}, \tag{4.13}$$

and

$$m_{\beta\alpha}^{pp} = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \frac{e^{-\frac{1}{2}k^2\sigma^2}}{k^2+m_0^2} \exp\left[ik(x_\alpha - x_\beta)\right]. \tag{4.14}$$

The regression function (4.10) can be viewed as an "optimal interpolant"; it is the expectation value of the function $p(x)$ given what is known. Examples of regression functions are plotted in Figure 1 for a mesh of $N = 5$ points. The open circles represent the values of the five collective variables V_α^p; the abscissa is the location of the point x_α around which the average is computed, and the ordinate is the value of the corresponding collective variable. The three curves represent the interpolating function (4.10) for three different values of the kernel width: $\sigma = \Delta x = 2\pi/N$ (solid line), $\sigma = 0.5\,\Delta x$ (dashed line), and $\sigma = 0.1\,\Delta x$ (dash-dot line). The parameter m_0 was taken to be one.

Time evolution. We next consider the time evolution of the mean value of the collective variables U^p and U^q, first based on the approximating scheme (2.11). The equation for V_α^p, for example, is

$$\begin{aligned} \frac{dV_\alpha^p}{dt} &= \left\langle \left(g_\alpha(\cdot), -q_{xx}(\cdot) + m_0^2 q(\cdot)\right)\right\rangle_V \\ &= -\left(g_\alpha(\cdot), \frac{\partial^2}{\partial x^2}\langle q(\cdot)\rangle_V\right) + m_0^2 \left(g_\alpha(\cdot), \langle q(\cdot)\rangle_V\right). \end{aligned} \tag{4.15}$$

Substituting the regression function (4.10) we find:

$$\frac{dV_\alpha^p}{dt} = \sum_{\gamma=1}^{N} \left\{ \sum_{\beta=1}^{N} \left(g_\alpha(\cdot), g_\beta(\cdot)\right) [m^{-1}]_{\beta\gamma}^{qq} \right\} V_\gamma^q. \tag{4.16}$$

A similar equation is obtained for V_α^q by the symmetry transformation $V_\alpha^p \to V_\alpha^q$ and $V_\alpha^q \to -V_\alpha^p$. Equation (4.16) represents a set of $2N$ ordinary differential equations that approximate the mean evolution of the collective variables. These equations are easy to solve with standard ODE solvers. Note that the matrix elements in braces need to be computed only once to define the scheme.

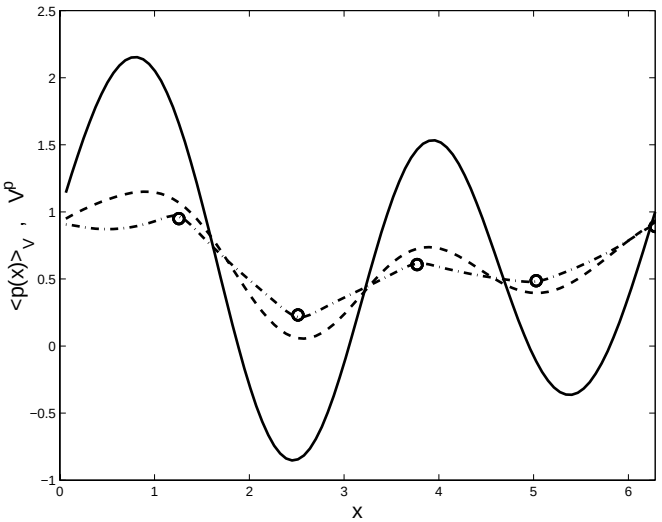

FIGURE 1. Example of regression functions for the linear Schrödinger equation. Values for five collective variables were chosen, representing local averages of $p(x)$ on a uniformly spaced grid. The kernels are translates of each other and have Gaussian profiles of width σ centered at the grid points. The lines represent the regression function, or optimal interpolant $\langle p(x)\rangle_V$ given by equation (4.10) for $\sigma = \Delta x$ (solid), $\sigma = 0.5\,\Delta x$ (dashed), and $\sigma = 0.1\,\Delta x$ (dash-dot).

We next calculate the *exact* mean value of the collective variables, U^p and U^q, at time t, conditioned by the initial data, V^p and V^q, at time $t = 0$, so that they can be compared with the result $V(t)$ of the scheme we just presented. We are able to do so in the present case because the equations are linear, and a simple representation of the evolution operator can be found.

The solution to the initial value problem (4.1) can be represented by Fourier series,

$$
\begin{aligned}
p(x,t) &= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\int e^{ik(x-y)}\left[p(y)\,\cos\omega t + q(y)\,\sin\omega t\right]\,dy \\
q(x,t) &= \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\int e^{ik(x-y)}\left[q(y)\,\cos\omega t - p(y)\,\sin\omega t\right]\,dy
\end{aligned}
\tag{4.17}
$$

where $p(y)$ and $q(y)$ are the (random) initial conditions, and $\omega = k^2 + m_0^2$.

The expectation values of the collective variables U^p_α and U^q_α are obtained by averaging the scalar products $(p(\cdot,t), g_\alpha(\cdot))$ and $(q(\cdot,t), g_\alpha(\cdot))$ with respect to the initial distribution. Because equations (4.17) are linear in the random variables

$p(y)$ and $q(y)$ this gives

(4.18)

$$\langle U_\alpha^p[p(\cdot),q(\cdot),t]\rangle_V = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\int e^{ik(x_\alpha-y)-\frac{1}{4}k^2\sigma^2}\left[\langle p(y)\rangle_V\cos\omega t+\langle q(y)\rangle_V\sin\omega t\right]dy$$

$$\langle U_\alpha^q[p(\cdot),q(\cdot),t]\rangle_V = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\int e^{ik(x_\alpha-y)-\frac{1}{4}k^2\sigma^2}\left[\langle q(y)\rangle_V\cos\omega t-\langle p(y)\rangle_V\sin\omega t\right]dy.$$

Note that in the linear case averaging and time evolution commute; equation (4.18) would have also been obtained if we first computed the mean initial state, $\langle p(y)\rangle_V$ and $\langle q(y)\rangle_V$, evolved it in time according to (4.17), and finally computed the collective variables by taking the appropriate scalar products.

To complete the calculation, we substitute the linear regression formula (4.10) for $\langle p(y)\rangle_V$ and $\langle q(y)\rangle_V$ and obtain:

(4.19)
$$\begin{aligned}\langle U_\alpha^p[p(\cdot),q(\cdot),t]\rangle_V &= \sum_{\beta,\gamma=1}^{N}\left\{c^C_{\alpha\beta}(t)[m^{-1}]^{pp}_{\beta\gamma}V^p_\gamma + c^S_{\alpha\beta}(t)[m^{-1}]^{qq}_{\beta\gamma}V^q_\gamma\right\}\\ \langle U_\alpha^q[p(\cdot),q(\cdot),t]\rangle_V &= \sum_{\beta,\gamma=1}^{N}\left\{c^C_{\alpha\beta}(t)[m^{-1}]^{pp}_{\beta\gamma}V^q_\gamma - c^S_{\alpha\beta}(t)[m^{-1}]^{qq}_{\beta\gamma}V^p_\gamma\right\}\end{aligned},$$

where

$$c^C_{\alpha\beta}(t) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\frac{\cos\omega t}{\omega}e^{ik(x_\alpha-x_\beta)}e^{-\frac{1}{2}k^2\sigma^2}, \tag{4.20}$$

and

$$c^S_{\alpha\beta}(t) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\frac{\sin\omega t}{\omega}e^{ik(x_\alpha-x_\beta)}e^{-\frac{1}{2}k^2\sigma^2}. \tag{4.21}$$

Results. We now compare the exact formula (4.19) for the future expectation value of the collective variables to the approximation (4.16). Figures 2*a*–2*c* compare between the two evolutions for $N = 5$ and randomly selected initial data, V^p_α and V^q_α. The graphs show the mean time evolution of the collective variable $U^p_1[p(\cdot),q(\cdot)]$. The same set of initial values was used in the three plots; the difference is in the width σ of the kernels $g_\alpha(x)$: $\sigma = \Delta x$ (Figure 2*a*), $\sigma = 0.5\,\Delta x$ (Figure 2*b*), and $\sigma = 0.1\,\Delta x$ (Figure 2*c*). In the first case, in which the kernel width equals the grid spacing, the approximation is not distinguishable from the exact solution on the scale of the plot for the duration of the calculation. The two other cases show that the narrower the kernel is, the sooner the curve deviates from the exact solution.

5. A Nonlinear Hamiltonian System

The equations of motion. The method demonstrated in the preceding section can be generalized to a nonlinear Schrödinger equation. However, we want to exhibit the power of our method by comparing the solutions that it yields to exact solutions; in the nonlinear case, exact solutions of problems with random initial conditions are hard to find, so we resort to a stratagem. Even though our method applies to nonlinear partial differential equations, we study instead a finite

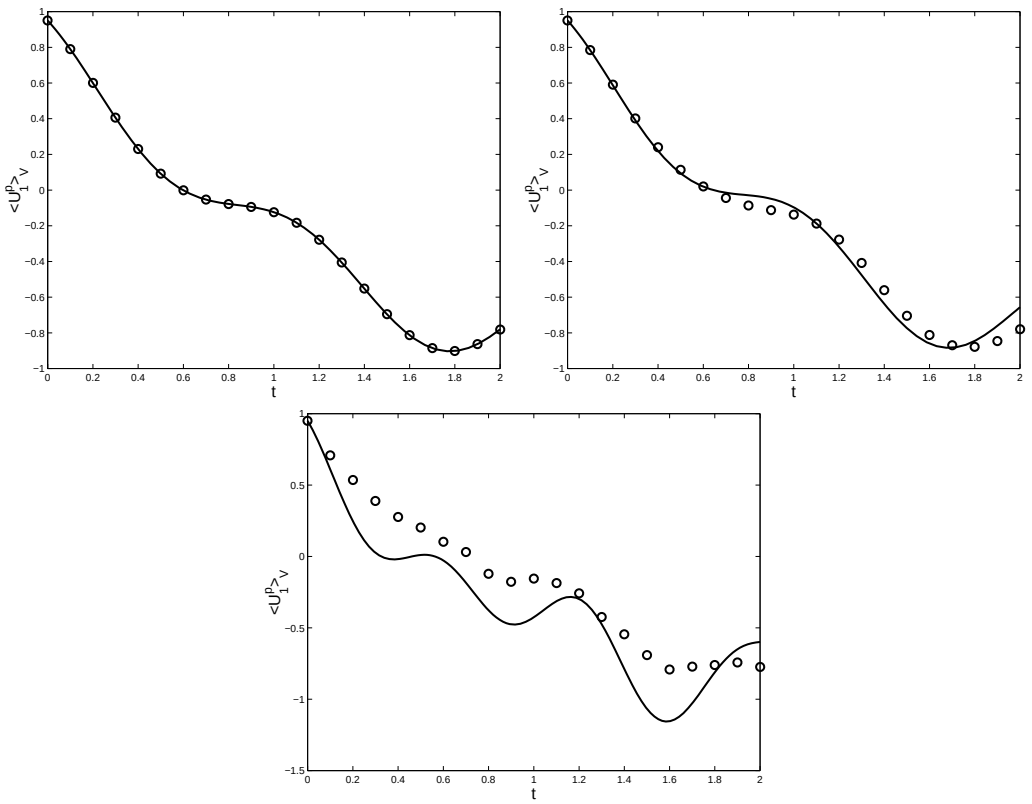

FIGURE 2. Mean evolution of the collective variable $U_1^p[p(\cdot), q(\cdot)]$ for $N = 5$, and a random choice of the initial data V^p and V^q. The open dots represent the exact solution (4.19), whereas the lines represent the approximate solution obtained by an integration of the set of 10 ordinary differential equations (4.16). The three graphs are for different values of the kernel width σ: (a) $\sigma = \Delta x$, (b) $\sigma = 0.5\,\Delta x$, and (c) $\sigma = 0.1\,\Delta x$.

dimensional system of $2n$ ordinary differential equations that is formally a finite difference approximation of a nonlinear Schrödinger equation:

$$\begin{aligned} \frac{dp(j)}{dt} &= -\frac{q(j-1) - 2q(j) + q(j+1)}{\Delta x^2} + q^3(j) \\ \frac{dq(j)}{dt} &= +\frac{p(j-1) - 2p(j) + p(j+1)}{\Delta x^2} - p^3(j) \end{aligned} \qquad j = 1, \ldots, n, \tag{5.1}$$

where $\Delta x = 1/n$ is the mesh spacing, and periodicity is enforced with $p(0) \equiv p(n)$, $p(n+1) \equiv p(1)$, etc; this system is non-integrable for $n > 1$. The approximation is only formal because we shall be considering non-smooth data which give rise to weak solutions that cannot be readily computed by difference methods.

We shall pretend that n is so large that the system (5.1) cannot be solved on a computer, and shall therefore seek an approximation that requires a computation with fewer variables. In practice we shall pick an n small enough so that the results of the approximate procedure can be compared to an *ensemble* of exact solutions.

The prior measure. The system of equations (5.1) is the Hamilton equations of motion for the Hamiltonian

(5.2)
$$H[p,q] = \frac{1}{2}\sum_{j=1}^{n}\left\{\left[\frac{p(j+1)-p(j)}{\Delta x}\right]^2 + \left[\frac{q(j+1)-q(j)}{\Delta x}\right]^2 + \frac{1}{2}\left[p^4(j)+q^4(j)\right]\right\},$$

where $p \equiv (p(1),\ldots,p(n))$ and $q \equiv (q(1),\ldots,q(n))$. The differential equations (5.1) preserve the canonical density

$$f_0[p,q] = \exp\left\{-H[p,q]\right\}, \tag{5.3}$$

which we postulate, as before, to be the prior probability density.

The prior density (5.3) is not Gaussian, which raises a technical difficulty in computing expectation values. We adopt here an approximate procedure where the density (5.3) is approximated by a Gaussian density that yields the same first and second moments (means and covariances) of the vectors p and q. The means are zero by symmetry:

$$\langle p(j)\rangle = \langle q(j)\rangle = 0 \tag{5.4}$$

(positive and negative values of these have equal weight). Also all p's and q's are uncorrelated:

$$\langle p(j_1)q(j_2)\rangle = 0, \tag{5.5}$$

since the density factors into a product of a density for the p's and a density for the q's. Thus $\langle p(j_1)p(j_2)\rangle = \langle q(j_1)q(j_2)\rangle$ are the only non-trivial covariances. Finally, since the Hamiltonian is translation invariant, these covariances depend only on the separation between the indices j_1 and j_2, and are symmetric in $j_1 - j_2$.

To relate the present discrete problem to the continuous formalism used in the preceding section we write in analogy to (4.6)

$$\begin{aligned}\mathrm{Cov}\,[p(j_1),p(j_2)] &= [a^{-1}(j_1,j_2)]^{pp} = c(|j_1-j_2|)\\ \mathrm{Cov}\,[p(j_1),q(j_2)] &= [a^{-1}(j_1,j_2)]^{pq} = 0,\end{aligned} \tag{5.6}$$

with $j_1, j_2 = 1,\ldots,n$. We computed the numbers, $c(|j_1 - j_2|)$, for $n = 16$ and $j_1 - j_2 = 0,\ldots,15$ by a Metropolis Monte-Carlo algorithm [**BH92**]; the covariances obtained this way are shown in Figure 3. Along with the zero means, the numbers represented in Figure 3 completely determine the *approximate* prior distribution.

The collective variables. We next define a set of $2N$ collective variables ($N < n$), whose values we assume to be given at the initial time. The class of collective variables that is the discrete analog of (4.7) is of the form

$$\begin{aligned}U^p_\alpha[p,q] &= (g_\alpha(\cdot),p(\cdot)) \equiv \sum_{j=1}^{n} g_\alpha(j)p(j)\\ U^q_\alpha[p,q] &= (g_\alpha(\cdot),q(\cdot)) \equiv \sum_{j=1}^{n} g_\alpha(j)q(j)\end{aligned} \qquad \alpha = 1,\ldots,N, \tag{5.7}$$

where the g's are discrete kernels. In the calculations we exhibit we chose $n = 16$ and $N = 2$ so that we aim to reduce the number of degrees of freedom by a factor

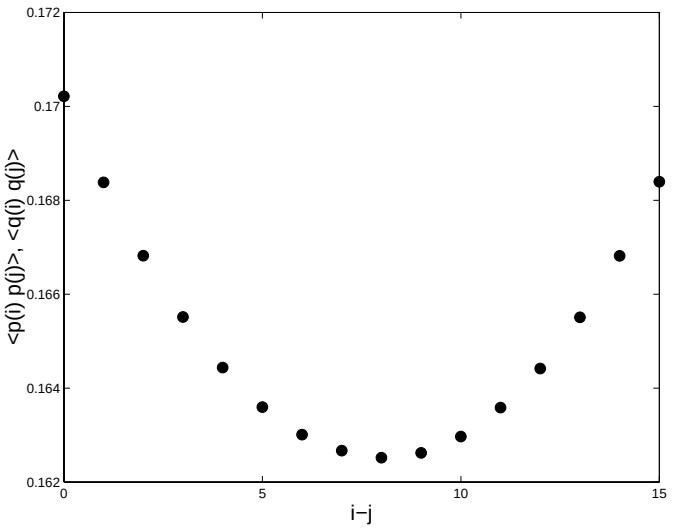

FIGURE 3. The covariance $\langle p(i)\, p(j)\rangle = \langle q(i)\, q(j)\rangle$ as function of the grid separation $i - j$ for the non-Gaussian probability distribution (5.3) with $n = 16$. These values were computed by a Metropolis Monte-Carlo simulation.

of 8. We pick as kernels discretized Gaussian functions centered at the grid points $j = 1$ and $j = 9$:

$$g_1(j) = \frac{1}{Z} \exp\left\{-\frac{d^2(1,j)}{n^2\sigma^2}\right\}$$
$$g_2(j) = \frac{1}{Z} \exp\left\{-\frac{d^2(9,j)}{n^2\sigma^2}\right\} \tag{5.8}$$

where Z is a normalizing constant, $\sigma = 0.25$, and $d(j_1, j_2)$ is a distance function over the periodic index axis, i.e., it is the minimum of $|j_1 - j_2|$, $|j_1 - j_2 - n|$, and $|j_1 - j_2 + n|$.

Conditional expectation. With the approximate measure defined by the covariances (5.6), and the collective variables (5.7), whose measured values are again denoted by V_α^p and V_α^q, we can approximate the conditional expectation of various observables $O[p, q]$. We shall need specifically the conditional expectation values of $p(j)$ and $p^3(j)$.

The approximate conditional expectation value of $p(j)$ is given by the discrete analog of equation (4.10), namely,

$$\langle p(j)\rangle_V = \sum_{\alpha=1}^{N} c_\alpha^{pp}(j) V_\alpha^p, \tag{5.9}$$

where

$$c_\alpha^{pp}(j) = \sum_{\beta=1}^{N} \left([a^{-1}(j,\cdot)]^{pp}, g_\beta(\cdot)\right) [m^{-1}]_{\beta\alpha}^{pp}, \tag{5.10}$$

and

$$m^{pp}_{\beta\alpha} = \sum_{j_1,j_2=1}^{n} g_\beta(j_1)[a^{-1}(j_1,j_2)]^{pp} g_\alpha(j_2). \tag{5.11}$$

(Again, the matrix inversion is only with respect to the lower indices α and β.)

To calculate the approximate conditional expectation value of $p^3(j)$ we first use Wick's theorem (Lemma 3.3):

$$\left\langle p^3(j)\right\rangle_V = 3\left\langle p^2(j)\right\rangle_V \langle p(j)\rangle_V - 2\langle p(j)\rangle_V^3 , \tag{5.12}$$

and then calculate the conditional second moment by using the discrete analog of equation (3.17):

$$\left\langle p^2(j)\right\rangle_V = \langle p(j)\rangle_V^2 + [a^{-1}(j,j)]^{pp} - \sum_{\alpha=1}^{N} c^{pp}_\alpha(j)\left(g_\alpha(\cdot), [a^{-1}(\cdot,j)]^{pp}\right). \tag{5.13}$$

Time evolution. The approximating scheme for calculating the mean evolution of the $2N$ collective variables U^p and U^q is derived by substituting the kernels (5.8) and the equations of motion (5.1) in the approximation formula (2.11). The equation for V^p_α, for example, is

$$\begin{aligned}\frac{dV^p_\alpha}{dt} = &-\frac{1}{\Delta x^2}\sum_{j=1}^{n} g_\alpha(j)\left[\langle q(j-1)\rangle_V - 2\langle q(j)\rangle_V + \langle q(j+1)\rangle_V\right] \\ &+ \sum_{j=1}^{n} g_\alpha(j)\left\langle q^3(j)\right\rangle_V .\end{aligned} \tag{5.14}$$

Substituting the expressions for the conditional expectations (5.9) and (5.12), and performing the summation, using the values of the covariances plotted in Figure 3, we explicitly obtain a closed set of 4 ordinary differential equations. The equation for V^p_1 is:

$$\begin{aligned}\frac{dV^p_1}{dt} = &-19.5\,(V^q_2 - V^q_1) \\ &+ \left[1.50\,(V^q_1)^3 - 0.88\,(V^q_1)^2 V^q_2 + 0.27\,V^q_1 (V^q_2)^2 + 0.11\,(V^q_2)^3\right].\end{aligned} \tag{5.15}$$

The equation for V^p_2 is obtained by substituting $1 \leftrightarrow 2$; the equations for V^q_1 and V^q_2 are obtained by the transformation $p \to q$ and $q \to -p$.

Unlike in the linear case, we cannot calculate analytically the mean evolution of the collective variables. To assess the accuracy of the approximate equation (5.15) we must compare the solution it yields with an average over an ensemble of solutions of the "fine scale" problem (5.1). To this end, we generated a large number of initial conditions that are consistent with the given values, V^p and V^q, of the collective variables. The construction of this ensemble was done by a Metropolis Monte Carlo algorithm, where new states are generated randomly by incremental changes, and accepted or rejected with a probability that ensures that for large enough samples the distribution converges to the conditioned canonical distribution. We generated an ensemble of 10^4 initial conditions; each initial state was then evolved in time using a fourth-order Runge-Kutta method. Finally, for each time level we computed the distribution of collective variables, U^p and U^q; the average of this distribution should be compared with the prediction of equations (5.15).

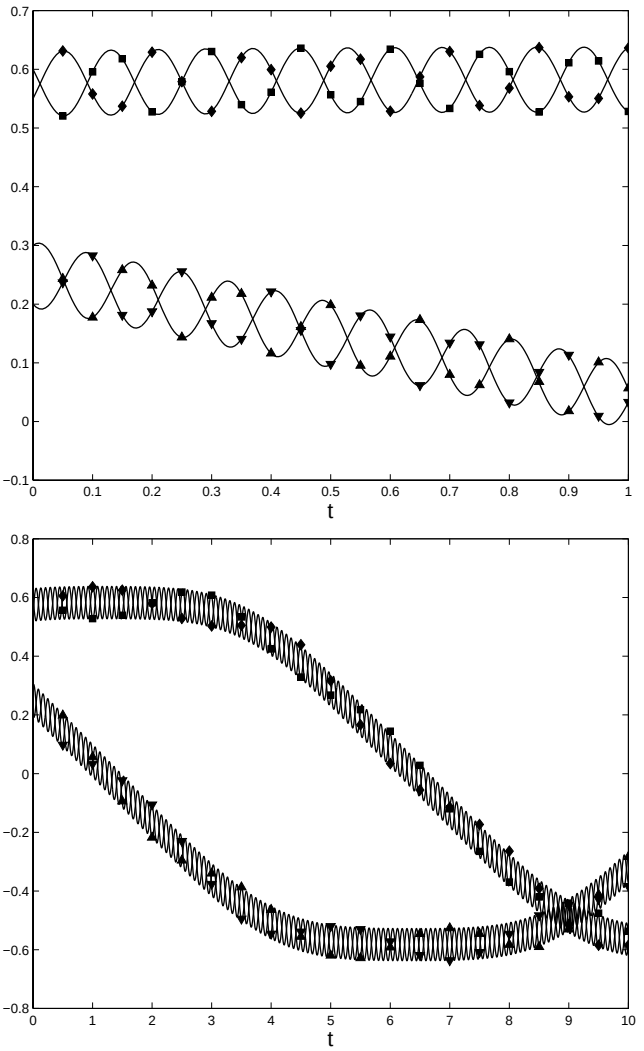

FIGURE 4. Evolution in time of the mean value of the four collective variable: V_1^p (▼), V_2^p (▲), V_1^q (■), and V_2^q (♦). The symbols represent the values of these quantities obtained by solving the 32 equations (5.1) for 10^4 initial conditions compatible with the initial data, and averaging. The solid lines are the values of the four corresponding functions obtained by integrating equation (5.15). Figures (a) and (b) are for the time intervals $[0, 1]$ and $[0, 10]$ respectively.

The comparison between the true and the approximate evolution is shown in Figure 4. Once again the reduced system of equations reproduces the average behavior of the collective variables with excellent accuracy, but at a very much smaller computational cost. Indeed, we compare one solution of 4 equations to 10^4 solutions of 32 equations.

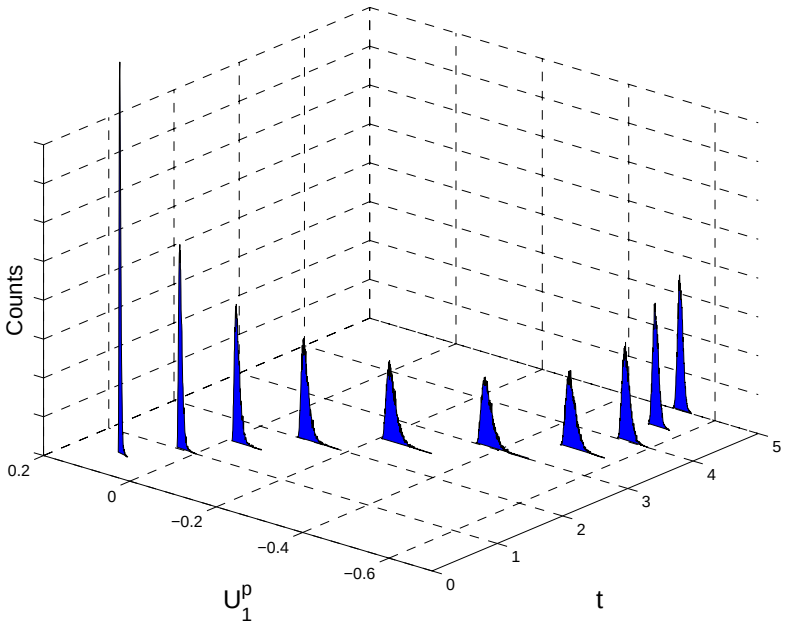

FIGURE 5. Evolution of the distribution of the collective variable U_1^p. The x-axis represents time, the y-axis represents the value of U_1^p, and the z-axis is proportional to the density of states that correspond to the same value of U_1^p at the given time.

In Figure 5 we show the evolution of the *distribution* of values assumed by the collective variable U_1^p; the data was extracted from the evolution of the ensemble. The distribution is initially sharply peaked, and spreads out as time evolves; yet, it remains sufficiently narrow throughout this computation, so that the approximation that projects that distribution back onto a sharp one is reasonable. This indicates that the choice of collective variables, or kernels, was appropriate. The use of narrow kernels, or even point values, would have yielded a distribution of value that spreads out almost instantaneously.

6. Conclusions

We have shown how to calculate efficiently, for a class of problems, the average behavior of an ensemble of solutions the individual members of which are very difficult to evaluate. The approach is reminiscent of statistical mechanics, where it is often easier to predict the evolution of a mole of particles than to predict the evolution of, say, a hundred particles, if one is content with the average behavior of a set of coarse variables (collective variables). The key step is the identification of a correspondence between underresolution and statistics; underresolved data define, together with prior statistical information, an ensemble of initial conditions, and the most one can aim for is to predict the expectation with respect to this ensemble of certain observables at future times. Our approach applies in those cases where prior statistical information is available, and is consistent with the differential equations; for example, it may consist of a measure invariant under the flow defined by the

differential equations. Fortunately, there are important classes of problems where we can find such information.

We proposed a scheme (2.11) that advances in time a set of variables that approximate the expectation values of a set of collective variables. As we explained, this scheme has to be viewed as a first approximation; more sophisticated schemes may be designed by allowing the kernels to vary in time and/or by keeping track of higher moments of the collective variables. Such refinements are the subject of ongoing research [**CKKT**].

One limitation of our present scheme can be perceived by considering the long time behavior of the nonlinear Hamiltonian system presented in Section 5. The flow induced by equations (5.1) is likely to be ergodic, hence the probability density function will approach, as $t \to \infty$, the invariant distribution. Indeed, the initial data have a decreasing influence on the statistics of the solutions as time progresses. This implies that the expectation values of the observables U^p and U^q will tend to their unconditional means, i.e., will decay to zero. On the other hand, no such decay occurs if one integrates the effective equations (5.15) for very long times. One must conclude that the present model is accurate for time intervals that are not longer than the time during which the initial data influence the outcome of the calculation.

The above discussion raises a number of questions interesting on their own: What is the range of influence, or the predictive power, of a given set of data? How much information is contained in partial data? These questions need to be formulated in a more quantitative way; they are intimately related to the question of how to choose appropriate collective variables, and their scope is beyond any particular method of solution.

Finally, a full knowledge of the prior measure is a luxury one cannot always expect. One needs to consider problems where the statistical information is only partial; for example, a number of moments may be known from asymptotics and scaling analyses (e.g., in turbulence theory [**Bar96, BC97, BC98**]). One can readily see from the nonlinear example that one can make do with the knowledge of means, covariances, and perhaps some higher-order moments. In addition, this knowledge is needed only on scales comparable with the widths of the kernels.

References

[Bar96] G.I. Barenblatt, *Scaling, self-similarity and intermediate asymptotics*, Cambridge University Press, Cambridge, 1996.

[BC97] G.I. Barenblatt and A.J. Chorin, *Scaling laws for fully developed turbulent flow in pipes*, Appl. Mech. Rev. **50** (1997), 413–429.

[BC98] G.I. Barenblatt and A.J. Chorin, *Scaling laws and vanishing viscosity limits in turbulence theory*, Proc. Symposia Appl. Math. AMS **54** (1998), 1–25.

[BH92] K. Binder and D. Heerman, *Monte-carlo simulation in statistical physics*, Springer, Berlin, 1992.

[CKK98] A.J. Chorin, A. Kast, and R. Kupferman, *Optimal prediction of underresolved dynamics*, Proc. Nat. Acad. Sci. USA **95** (1998), 4094–4098.

[CKKT] A.J. Chorin, A. Kast, R. Kupferman, and B. Turkington, *Optimal prediction of two-dimensional euler flow*, in preparation.

[FH65] R. Feynman and A. Hibbs, *Quantum mechanics and path integrals*, McGraw-Hill, New York, 1965.

[Hal] O.H. Hald, *Optimal prediction of linear systems*, in preparation.

[Kle89] H. Kleinert, *Gauge fields in condensed matter*, World Scientific, Singapore, 1989.

[McK95] H.P. McKean, *Statistical mechanics of nonlinear wave equations IV: Cubic Schrödinger equations*, Comm. Math. Phys. **168** (1995), 479–491.

[MW90] W.D. McComb and A.G. Watt, *Conditional averaging procedure for the elimination of the small-scale modes from incompressible fluid turbulence at high reynolds numbers*, Phys. Rev. Lett. **65** (1990), 3281–3284.

[Ris84] H. Risken, *The fokker-planck equation*, Springer, New York, 1984.

[SM97] A. Scotti and C. Meneveau, *Fractal model for coarse-grained partial differential equations*, Phys. Rev. Lett. **78** (1997), 867–870.

Department of Mathematics, Lawrence Berkeley National Laboratory, Mail Stop 50A-2152, 1 Cyclotron Road, Berkeley, CA 94720

E-mail address: chorin@math.berkeley.edu

URL: http://math.berkeley.edu/∼chorin

Department of Mathematics, Lawrence Berkeley National Laboratory, Mail Stop 50A-2152, 1 Cyclotron Road, Berkeley, CA 94720

E-mail address: anton@math.lbl.gov

URL: http://www.lbl.gov/∼anton

Department of Mathematics, Lawrence Berkeley National Laboratory, Mail Stop 50A-2152, 1 Cyclotron Road, Berkeley, CA 94720

Current address: Institute of Mathematics, The Hebrew University, Jerusalem, 91904 Israel

E-mail address: raz@math.lbl.gov

URL: http://www.lbl.gov/∼raz

Contemporary Mathematics
Volume **238**, 1999
B 0-8218-1196-7-03540-4

Variational Bounds in Turbulent Convection

Peter Constantin

1. Introduction

When sufficient energy is steadily supplied to a fluid, the ensuing dynamical behavior involves many spatial and temporal scales and energy is dissipated efficiently. For instance when sufficiently strong heat is supplied against the pull of gravity to a fluid, the heat flux due to fluid flow convection exceeds the heat flux due to molecular diffusion. The average of heat flux is quantified in the Nusselt number N. Numerous experiments and numerical simulations ([**1**]) under a variety of conditions report power-law behavior

$$N \sim R^q$$

where the Rayleigh number R is proportional to the amount of heat supplied externally. The exponent q is very robust and most experiments give $q = \frac{2}{7}$, while some situations produce $q = \frac{1}{3}$ for large R.

Mathematically, the description is based on the three dimensional Boussinesq equations for Rayleigh-Bénard convection ([**2**]), a system of equations coupling the three dimensional Navier-Stokes equations to a heat advection-diffusion equation. The only known rigorous upper bound for N ([**3**]) at large R is of the order $R^{\frac{1}{2}}$; the bound is valid for all weak solutions (the global existence of smooth solutions is not known). This bound is not only a mathematical uper bound: there are physical reasons why $q = \frac{1}{2}$ might be the true asymptotic value at exceedingly high R (in conditions perhaps difficult to achieve in the laboratory; nevertheless, a few recent experimental results hint also at $q = \frac{1}{2}$).

Although one can describe conditions that imply $N \sim R^q$ with $q = \frac{1}{3}$ and even $q = \frac{2}{7}$ for a range of R in the full Boussinesq system ([**4**]), there is yet no rigorous derivation of the exponent $2/7$ as the unconditional limit for large R in a different non-trivial model of convection. The scaling exponents have been discussed by several authors using physical reasoning and dimensional analysis ([**5**]) and in particular the exponent $2/7$ has been derived in several physical fashions involving somewhat different predictions.

1991 *Mathematics Subject Classification.* Primary: 76D05, 76F10.
Partially supported by NSF/DMS-9501062.

Variational methods for bounding bulk dissipation in turbulence are a classical subject. Ideas of Malkus from the fifties were followed by Howard's flux maximization results [**6**] and subsequently were developed by Busse [**7**] and many others. This classical approach starts from the Reynolds equations and assumes certain statistical symmetries.

In recent years another general variational method has been developed and applied to estimate bulk dissipation quantities in systems in which energy is supplied by boundary conditions ([**8**] - [**15**]). A connection between the classical method of Howard and Busse and a version of the background field method has been established in ([**16**]). The method starts by translating the equation in function space by a background - a time independent function that obeys the driving boundary conditions. A quadratic form is associated naturally to each background, and the method consists in selecting those backgrounds for which this quadratic form is positive semi-definite and then minimizing a certain integral of the background. The set of selected backgrounds is convex. The method has certain advantages over the classical approach – in particular, there is no need for statistical assumptions. The method is flexible enough to accommodate more partial differential information. The partial differential equation confers special properties to the functions that represent long-lived solutions. These functions belong to a large but finite dimensional set, the attractor associated to the PDE at the given values of the parameters. If one can find certain quantitative features of functions belonging to this attractor one can incorporate them in a judicious variational problem. This is how a rigorous upper bound of the form

$$N \leq 1 + C_1 R^{\frac{1}{3}} (1 + \log_+(R))^{\frac{2}{3}}$$

for arbitrary R was derived recently [**17**] for the three dimensional equations for Rayleigh-Bénard convection obtained in the limit of infinite Prandtl number. The Prandtl number is the ratio of the fluid's viscosity to the fluid's heat conduction coefficient. These equations are an example of active scalars ([**18**]); they are easier to analyze and simulate numerically than the full Boussinesq system. In the infinite Prandtl number example one can obtain more information about the long time behavior of solutions than in the finite Prandtl number equations. The additional information concerns higher derivatives. In order to exploit this additional information and deduce a better upper bound one needs to modify substantially the background field method: the quadratic form is no longer required to be semi-definite. Instead, the additional information coming from the evolution equation is incorporated in the constraints of a mini-max procedure.

There are several other examples of active scalars for which one can obtain interesting rigorous bounds for the bulk dissipation. For instance, recent results ([**19**]) on convection in a porous layer employ an improvement of the background field method ([**20**]) and agree remarkably well with the experimental data.

In this paper we will confine ourselves to the effects of rotation on heat transfer in the infinite Prandtl number cases. Not only are these systems more amenable to analysis but also the variety of physical phenomena poses a challenge to the background flow method as originally formulated. Indeed, in its original formulation the method seems insensitive to linear low order anti-symmetric perturbations such as rotation. The physical effect of very rapid rotation is to stratify the flow and to totally suppress convective heat transport. This effect has been proved recently at large but finite rotation rates in infinite Prandtl number convection in ([**21**]) using

the background field methodology. The limit of slow rotation is not singular. At fixed rotation one can recover the large R rigorous logarithmic $1/3$ upper bound ([**22**]). The situation is complicated though: moderate rotation rates may effectively increase the heat transfer. This experimental fact ([**23**]) is consistent with the fact that the logarithmic $1/3$ upper bound diverges at very high rotation rates; the best known rigorous uniform upper bound valid for all rotation rates has a higher exponent $(2/5)$ than the bound found in the absence of rotation. The uniform bound

$$N \leq\sim R^{\frac{2}{5}}$$

will be derived in this work. We start with the non-rotating case.

2. Infinite Prandtl Number Equations

The infinite Prandtl number equations for Rayleigh-Bénard convection in the Boussinesq approximation are a system of five equations for velocities (u, v, w), pressure p and temperature T in three spatial dimensions. The temperature is advected and diffuses according to the active scalar equation

$$(\partial_t + \mathbf{u} \cdot \nabla) T = \Delta T \tag{1}$$

where $\mathbf{u} = (u, v, w)$. The velocity and pressure are determined from the temperature by solving time independent non-local equations of state:

$$-\Delta u + p_x = 0, \tag{2}$$

together with

$$-\Delta v + p_y = 0 \tag{3}$$

and

$$-\Delta w + p_z = RT. \tag{4}$$

R represents the Rayleigh number. The velocity is divergence-free

$$u_x + v_y + w_z = 0. \tag{5}$$

The horizontal independent variables (x, y) belong to a basic square $Q \subset \mathbf{R}^2$ of side L. Sometimes we will drop the distinction between x and y and denote both horizontal variables x. The vertical variable z belongs to the interval $[0, 1]$. The non-negative variable t represents time. The boundary conditions are as follows: all functions $((u, v, w),\ p,\ T)$ are periodic in x and y with period L; u, v, and w vanish for $z = 0, 1$, and the temperature obeys $T = 0$ at $z = 1$, $T = 1$ at $z = 0$.

We will write

$$\|f\|^2 = \frac{1}{L^2} \int_0^1 \int_Q |f(x, y, z)|^2 dz dx dy$$

for the (normalized) L^2 norm on the whole domain. We denote by Δ_D the Laplacian with periodic-Dirichlet boundary conditions. We will denote by Δ_h the Laplacian in the horizontal directions x and y. We will use $< \cdots >$ for long time average:

$$\langle f \rangle = \limsup_{t \to \infty} \frac{1}{t} \int_0^t f(s) ds.$$

We will denote horizontal averages by an overbar:

$$\overline{f(\cdot, z)} = \frac{1}{L^2} \int_Q f(x, y, z) dx dy.$$

We will also use the notation for scalar product

$$(f,g) = \frac{1}{L^2}\int_0^1 \int_Q (fg)(x,y,z)dxdydz.$$

The Nusselt number is

$$N = 1 + \langle (w,T) \rangle . \tag{6}$$

One can prove using the equation (1) and the boundary conditions that

$$N = \left\langle \|\nabla T\|^2 \right\rangle \tag{7}$$

and using the equations of state (2 - 4) that

$$\left\langle \|\nabla u\|^2 \right\rangle = R(N-1). \tag{8}$$

This defines a Nusselt number that depends on the choice of initial data; we take the supremum of all these numbers. The system has global smooth solutions for arbitrary smooth initial data. The solutions exist for all time and approach a finite dimensional set of functions. If we think in terms of this dynamical system picture then the Nusselt number represents the maximal long time average distance from the origin on trajectories. Because all invariant measures can be computed using trajectories the Nusselt number is also the maximal expected dissipation, when one maximizes among all invariant measures.

3. Bounding the Heat Flux

We take a function $\tau(z)$ that satisfies $\tau(0) = 1$, $\tau(1) = 0$, and write $T = \tau + \theta(x,y,z,t)$. The role of τ is that of a convenient background; there is no implied smallness of θ, but of course θ obeys the same homogeneous boundary conditions as the velocity. The equation obeyed by θ is

$$\left(\partial_t + u\cdot\nabla - \Delta\right)\theta = -\tau'' - w\tau' \tag{9}$$

where we used $\tau' = \frac{d\tau}{dz}$. We are interested in the function $b(z,t)$ defined by

$$b(z,t) = \frac{1}{L^2}\int_Q w(\cdot,z)T(\cdot,z)dx.$$

Its average is related to the Nusselt number:

$$N - 1 = \left\langle \int_0^1 b(z)dz \right\rangle .$$

Note that

$$T - \overline{T} = \theta - \overline{\theta}$$

Also note that from the boundary conditions and incompressibility

$$\overline{w}(z,t) = 0$$

and therefore

$$b(z,t) = \frac{1}{L^2}\int_Q w(\cdot,z)\theta(\cdot,z)dx.$$

From the equation (9) it follows that

$$N = \left\langle -2\int_0^1 \tau'(z)b(z)dz - \|\nabla\theta\|^2 \right\rangle + \int_0^1 \left(\tau'(z)\right)^2 dz. \tag{10}$$

Now we are in a position to explain the variational method and some previous results. Consider a choice of the background τ that is "admissible" in the sense that

$$\left\langle -2\int_0^1 \tau'(z)b(z)dz - \|\nabla\theta\|^2 \right\rangle \leq 0$$

holds *for all* functions θ. Then of course

$$N \leq \int_0^1 \left(\tau'(z)\right)^2 dz.$$

The set of admissible backgrounds is not empty, convex and closed in the H^1 topology. The background method, as originally applied, is then to seek the admissible background that achieves the minimum $\int_0^1 \left(\tau'(z)\right)^2 dz$. Such an approach would predict $N \leq cR^{\frac{1}{2}}$ for this active scalar, just as in the case of the full Boussinesq system. One can do better. Let us write

$$b(z,t) = \frac{1}{L^2}\int_Q \int_0^z \int_0^{z_1} w_{zz}(x,z_2,t)\theta(x,z)dxdzdz_2dz_1. \tag{11}$$

It follows that

$$|b(z,t)| \leq z^2\left(1+\|\tau\|_{L^\infty}\right)\|w_{zz}\|_{L^\infty(dz;L^1(dx))}. \tag{12}$$

Now we will use two a priori bounds. First, one can prove using (9) and (8) that there exists a positive constant C_Δ such that

$$\left\langle \|\Delta\theta\|^2 \right\rangle \leq C_\Delta\left\{RN + \int_0^1 \left[\left(\tau''(z)\right)^2 + Rz(\tau'(z))^2\right]dz\right\} \tag{13}$$

holds. Secondly, one has the basic logarithmic bound ([**17**])

$$\|w_{zz}\|_{L^\infty} \leq CR(1+\|\tau\|_{L^\infty})[1+\log_+(R\|\Delta\theta\|)]^2. \tag{14}$$

We will describe briefly how to obtain (13) and (14) in the next section. Using (14) together with (13) in (12) one deduces from (10)

$$N \leq \int_0^1 (\tau'(z))^2 dz + CR(1+\|\tau\|_{L^\infty})^2\left[\int_0^1 z^2|\tau'|dz\right]$$

$$\left[1+\log_+\left\{RN + \int_0^1 \left[\left(\tau''(z)\right)^2 + Rz(\tau'(z))^2\right]dz\right\}\right] \tag{15}$$

Choosing τ to be a smooth approximation of $\tau(z) = \frac{1-z}{\delta}$ for $0 \leq z \leq \delta$ and $\tau = 0$ for $z \geq \delta$ and optimizing in δ one obtains

THEOREM 1. *There exists a constant C_0 such that the Nusselt number for the infinite Prandtl number equation is bounded by*

$$N \leq N_0(R)$$

where

$$N_0(R) = 1 + C_0 R^{1/3}(1+\log_+ R)^{\frac{2}{3}}$$

The associated optimization procedure consists in the mini-max suggested by (10) for functions θ that obey the constraint (13).

THEOREM 2. *The Nusselt number for the infinite Prandtl number equation can be bounded by the constrained mini-max procedure*

$$N \leq \inf_{\tau} \sup_{\theta \in C_\tau} \left\{ \left\langle -\|\nabla\theta\|^2 + 2\int_0^1 -\tau'(z)b(z)dz \right\rangle + \int_0^1 (\tau'(z))^2 \right\}$$

where C_τ is the set of smooth, time dependent functions θ that obey periodic-homogeneous Dirichlet boundary conditions and the inequality

$$\left\langle \|\Delta\theta\|^2 \right\rangle \leq C_\Delta \left\{ RN_0(R) + \int_0^1 \left[(\tau''(z))^2 + Rz(\tau'(z))^2 \right] dz \right\}.$$

The functions $b(z,t)$ are computed via

$$b(z,t) = \frac{1}{L^2} \int \int_Q w(x,y,z,t)\theta(x,y,z,t)dxdy$$

and the functions $w(x,y,z,t)$ are computed by solving

$$\Delta^2 w = -R\Delta_h \theta$$

with periodic-homogeneous Dirichlet and Neumann boundary conditions.

4. Two Inequalities

The inequalities (13) and (14) played an important role. We present here the ingredients needed to prove them because they are of more general use.

In order to prove (13) using only the bound (8) on the velocity we use the interpolation inequality

$$\|\nabla\theta\|^2_{L^4(dx)} \leq 3\|\theta\|_{L^\infty}\|\Delta\theta\|_{L^2(dx)}$$

that is valid in all dimensions (and can be proved directly by integration by parts). Multiplying (9) by $-\Delta\theta$, integrating by parts in the convective term and using the divergence-free condition one obtains after long time average the bound (13).

In order to obtain (14) we write first the equation obeyed by the pressure in view of (5):

$$\Delta p = RT_z.$$

Differentiating and substituting, the equation (4) becomes

$$\Delta^2 w = -R\Delta_h T. \tag{16}$$

In view of the incompressibility condition, the boundary conditions are

$$w(x,y,0) = w'(x,y,0) = w(x,y,1) = w'(x,y,1) = 0. \tag{17}$$

Denote by $(\Delta^2_{DN})^{-1}f$ the solution $w = (\Delta^2_{DN})^{-1}f$ of

$$\Delta^2 w = f$$

with horizontally periodic and vertically Dirichlet and Neumann boundary conditions $w = w' = 0$. Thus, in the infinite Prandtl number system

$$w_{zz} = -RB\theta$$

where

$$B = \frac{\partial^2}{\partial z^2}(\Delta^2_{DN})^{-1}\Delta_h.$$

The inequality (14) was proved as a consequence of the logarithmic L^∞ estimate for the operator B ([**17**]) given below.

THEOREM 3. *For any $\alpha \in (0,1)$ there exists a positive constant C_α such that every Hölder continuous function θ that is horizontally periodic and vanishes at the vertical boundaries satisfies*

$$\|B\theta\|_{L^\infty} \leq C_\alpha \|\theta\|_{L^\infty} \left(1 + \log_+ \|\theta\|_{C^{0,\alpha}}\right)^2 . \tag{18}$$

The spatial $C^{0,\alpha}$ norm is defined as

$$\|\theta\|_{C^{0,\alpha}} = \sup_{X=(x,y,z)\in Q\times[0,1]} |\theta(X,t)| + \sup_{X\neq Y} \frac{|\theta(X,t)-\theta(Y,t)|}{|X-Y|^\alpha}$$

The proof ([**17**]) is based on a decomposition

$$B\theta = (I - B_1 + B_2 + B_3)B_1\theta$$

where

$$B_1(\theta) = \Delta_h \left(\Delta_D\right)^{-1} \theta$$

and B_2 and B_3 are certain singular integral operators. One proves for B_j, $j = 1,2,3$ the estimates

$$\|B_j\theta\|_{L^\infty} \leq C_\alpha \|\theta\|_{L^\infty} \left(1 + \log_+ \|\theta\|_{C^{0,\alpha}}\right). \tag{19}$$

These estimates are well-known for singular integral operators of the classical Calderon-Zygmund type. The operators B_j are not translationally invariant. They have kernels K_j,

$$B_1(\theta)(x,z) = L^{-2} \int_Q \int_0^1 K_1(x-y,z,\zeta) \left(\theta(y,\zeta) - \theta(x,z)\right) dyd\zeta$$

and

$$B_2(\theta)(x,z) = L^{-2} \int_Q \int_0^1 K_2(x-y,z,\zeta) \left(\theta(y,\zeta) - \theta(y,1)\right) dyd\zeta$$

and

$$B_3(\theta)(x,z) = L^{-2} \int_Q \int_0^1 K_3(x-y,z,\zeta) \left(\theta(y,\zeta) - \theta(y,0)\right) dyd\zeta.$$

The kernels K_j can be written as oscillatory sums of exponentials. The Poisson summation formula and Poisson kernel are used to derive inequalities of the type

$$|K_1(x-y,z,\zeta)| \leq C \left(|x-y|^2 + |z-\zeta|^2\right)^{-\frac{3}{2}} \tag{20}$$

and

$$|K_2(x-y,z,\zeta)| \leq C \left(|x-y|^2 + |1-\zeta|^2\right)^{-\frac{3}{2}} \tag{21}$$

and similarly

$$|K_3(x-y,z,\zeta)| \leq C \left(|x-y|^2 + |\zeta|^2\right)^{-\frac{3}{2}} . \tag{22}$$

The inequalities (20, 21, 22) are the heart of the matter; once they are proved, the estimates (19) follow in a straightforward manner.

5. Rotation

We assume that the domain D rotates at a uniform angular rate around the z axis, and we place ourselves in a frame rotating with the domain. We will still consider the infinite Prandtl number case. The boundary conditions and the equation (1) for the temperature are the same as in the non-rotating case. In the presence of rotation the velocity is determined by the temperature through the Poincaré-Stokes equation of state:

$$\begin{array}{c} -\Delta u - E^{-1}v + p_x = 0 \\ -\Delta v + E^{-1}u + p_y = 0 \\ -\Delta w + p_z = RT. \end{array} \tag{23}$$

Here E is the Ekman number. The non-rotating case corresponds formally to $E = \infty$. The incompressibility condition (5) is maintained. We denote by ζ the vertical component of vorticity

$$\zeta = v_x - u_y. \tag{24}$$

Taking the divergence of (23) to obtain the equation for the pressure:

$$\Delta p - E^{-1}\zeta = RT_z. \tag{25}$$

Eliminating the pressure we obtain the analogue of (16)

$$\Delta^2 w - E^{-1}\zeta_z = -R\Delta_h T \tag{26}$$

together with

$$-\Delta\zeta - E^{-1}w_z = 0. \tag{27}$$

Incompressibility is used to deduce the boundary conditions

$$\begin{array}{c} w(x,y,0,t) = w(x,y,1,t) = 0 \\ w_z(x,y,0,t) = w_z(x,y,1,t) = 0 \\ \zeta(x,y,0,t) = \zeta(x,y,1,t) = 0. \end{array} \tag{28}$$

From (26) and (27) it is easy to obtain ([**21**]) bounds for the velocity and pressure that are uniform for all rotation rates E^{-1}:

$$\|\Delta w\|^2 + 2\|\nabla\zeta\|^2 \leq R^2, \tag{29}$$

$$\|p_z\|^2 \leq 4R^2 \tag{30}$$

$$\|\nabla u\|^2 + \|\nabla v\|^2 + \|\nabla w\|^2 \leq R^2. \tag{31}$$

These inequalities hold pointwise in time and are valid in the non-rotating case as well. Notice that the uniform bound (29) has a very important consequence for strongly rotating (small Ekman number) systems: the vertical acceleration w_z is suppressed. Indeed from (27) it follows that

$$(\Delta_D)^{-1} w_z = -E\zeta$$

and thus w_z tends to zero in H^{-1} as $E \to 0$ at fixed R. In order to take advantage of this observation we need to control the growth of the full gradients of the horizontal components of velocity at the boundaries. This is achieved ([**21**]) in the following manner. First we differentiate the equation for u in (23) with respect to z

$$-\Delta u_z = E^{-1}v_z - p_{zx}, \tag{32}$$

we multiply by u and integrate horizontally:

$$-\overline{u\Delta u_z} = E^{-1}\overline{v_z u} + \overline{p_z u_x}. \tag{33}$$

Secondly, we observe that

$$-\overline{u\Delta u_z} = \frac{d}{dz}\left(\frac{1}{2}\overline{|\nabla u|_2^2}\right) - \frac{d}{dz}\overline{uu_{zz}}. \tag{34}$$

Integrating (33, 34) vertically on $[0, z]$ using the Dirichlet boundary condition on u we obtain

$$\frac{1}{2}\overline{|\nabla u(\cdot,0)|_2^2} = \frac{1}{2}\overline{|\nabla u(\cdot,z)|_2^2} - \overline{uu_{zz}} - E^{-1}\int_0^z \overline{v_z u} - \int_0^z \overline{p_z u_x},$$

and integrating again with respect to z from 0 to 1 we deduce

$$\frac{1}{2}\overline{|\nabla u(\cdot,0)|^2} \le \frac{1}{2}\|\nabla u\|^2 + \|u_z\|^2 + E^{-1}\|v_z\|\|u\| + \|p_z\|\|u_x\|. \tag{35}$$

Now from (35) using the bounds (30), (31) and the Poincare inequality we obtain

$$\overline{|\nabla u(\cdot,0)|^2} \le C(1 + E^{-1})R^2. \tag{36}$$

Similar inequalities hold for v and the other boundary $z = 1$.

6. Heat Flux in a Rotating System

We recall (10)

$$N = \left\langle -2\int_0^1 \tau'(z)b(z)dz - \|\nabla\theta\|^2 \right\rangle + \int_0^1 (\tau'(z))^2\, dz \tag{37}$$

and write

$$\int_0^1 \tau'(z)b(z)dz = -(w_z, \Theta)$$

where Θ is

$$\Theta(x,y,z,t) = \int_0^z \tau'(s)\theta(x,y,s,t)ds. \tag{38}$$

Now we replace w_z using (27) in order to exhibit the small parameter E

$$\int_0^1 \tau'(z)b(z)dz = E(\Delta\zeta.\Theta) \tag{39}$$

We need to integrate by parts once and consider a boundary term:

$$\int_0^1 \tau'(z)b(z)dz = I + II \tag{40}$$

where

$$I = -E(\nabla\zeta, \nabla\Theta) \tag{41}$$

and

$$II = E\overline{\zeta_z(\cdot,1,t)\Theta(\cdot,1,t)}. \tag{42}$$

It is easy to show that

$$\|\nabla\Theta\| \le g\|\nabla\theta\| \tag{43}$$

where

$$g = \left[\int_0^1 (1-z) \left(\tau'(z)\right)^2 dz \right]^{\frac{1}{2}}. \tag{44}$$

The first term in (40) is bounded in view of (29)

$$|I| = E|(\nabla\zeta, \nabla\Theta)| \leq \frac{Eg}{\sqrt{2}} R \|\nabla\theta\|. \tag{45}$$

The second term can be written after one horizontal integration by parts as

$$II = E\overline{(u_z(\cdot,1,t)\Theta_y(\cdot,1,t) - v_z(\cdot,1,t)\Theta_x(\cdot,1,t))}. \tag{46}$$

Because Θ is an integral of θ it is easy to see that

$$\|\nabla_h \Theta(\cdot,1,t)\|_h \leq G\|\nabla\theta\|$$

where

$$G = \sup_z |\tau'(z)| \tag{47}$$

and

$$\|\nabla_h \Theta(\cdot,1,t)\|_h^2 = \overline{|\nabla_h \Theta(\cdot,1,t)|^2}$$

is the normalized horizontal L^2 norm. Using the boundary bound (36) on u_z and v_z we deduce that the contribution of the second term is estimated

$$|II| \leq CG\sqrt{E^2+E} R \|\nabla\theta\|. \tag{48}$$

Gathering (45) and (48) we obtain

$$\left| \int_0^1 \tau'(z) b(z) dz \right| \leq C \left\{ Eg + G\sqrt{E^2+E} \right\} R \|\nabla\theta\|. \tag{49}$$

We deduce

$$\left| \int_0^1 \tau'(z) b(z) dz \right| \leq C_1 \left\{ g^2 E^2 + G^2 (E^2+E) \right\} R^2 + \frac{1}{2} \|\nabla\theta\|^2 \tag{50}$$

On the other hand, it is not difficult to see using (29) and $0 \leq T \leq 1$ (maximum principle) in (11) that

$$\left| \int_0^1 \tau'(z) b(z) dz \right| \leq C_2 R \int_0^1 z^{\frac{3}{2}} |\tau'(z)| dz \tag{51}$$

This observation allows us to improve the results of ([**21**]). For any τ we may choose to apply either the bound (50) or (51) in the Nusselt number calculation (37). Let us set

$$\Gamma_\tau(E,R) = \min \left\{ 2C_1 \left[g^2 E^2 + G^2 \left(E^2 + E \right) \right] R^2; 2C_2 M R \right\} \tag{52}$$

where

$$M = \int_0^1 z^{\frac{3}{2}} |\tau'(z)| dz. \tag{53}$$

Consequently we obtain

$$N \leq \int_0^1 \left(\tau'(z)\right)^2 dz + \Gamma_\tau(E,R). \tag{54}$$

If one chooses τ to be a smooth approximation of $\tau = (1-z)\delta^{-1}$ for $0 \leq z \leq \delta$ and $\tau = 0$ for $\delta \leq z \leq 1$ then $g = O(\delta^{-\frac{1}{2}})$, $G = O(\delta^{-1})$ and $M = O(\delta^{\frac{3}{2}})$.

Optimizing in τ ([**21**], [**22**]) one obtains

THEOREM 4. *The Nusselt number for rotating infinite Prandtl-number convection is bounded by*

$$N - 1 \leq \min\left\{c_1 R^{\frac{2}{5}};\ (c_2 E^2 + c_3 E)R^2\right\}.$$

7. Discussion

Rotation has a non-trivial effect on heat transfer in the infinite Prandtl number convection. The equation determining the vertical velocity from the temperature is

$$\left(\Delta^2 + E^{-2}\partial_z \Delta_D^{-1}\partial_z\right) w = -R\Delta_h T$$

The operator $\partial_z \Delta_D^{-1}\partial_z$ is a low order perturbation of Δ^2 and both operators are non-negative in L^2. In the absence of rotation ($E = \infty$) one has a rigorous upper bound of the type $N \leq\sim R^{\frac{1}{3}}(\log R)^{\frac{2}{3}}$. However, the presently known rotation independent uniform bound has a higher exponent, $N \leq\sim R^{\frac{2}{5}}$. If rotation is increased sufficiently ($ER^{\frac{8}{5}} << 1$) for fixed R, then its effect is to dramatically laminarize the flow and the the heat transfer is due then exclusively to molecular diffusion: $N \to 1$. On the other hand, for fixed E one can recover the logarithmic $1/3$ bound for large R ([**22**]), but the bound diverges for $E \to 0$; the envelope is finite nevertheless because of the uniform $2/5$ bound. These rigorous results capture some of the complexity of the phenomena.

References

[1] Transitions to turbulence in helium gas, F. Heslot, B. Castaing, and A. Libchaber, Phys. Rev. A 36, 5870-5873 (1987), Observation of the Ultimate Regime in Rayleigh-Benard Convection, X. Chavanne, F. Chilla, B. Castaing, B. Hebral, B. Chabaud and J. Chaussy, Phys. Rev. Lett. 79, (1997), Strongly turbulent Rayleigh-Benard convection in mercury: comparison with results at moderate Prandtl number, S. Cioni, S. Ciliberto, and J. Sommeria, J. Fluid Mech. 335, 111-140 (1997), Structure of hard-turbulent convection in two dimensions: Numerical evidence, J. Werne, Phys. Rev. E 48, 1020-1035 (1993), Hard turbulence in rotating Rayleigh-Benard convection, K. Julinen, S. Legg, J. McWilliams, and J. Werne, Phys. Rev. E 53, 5557-5560 (1996).

[2] S. Chandrasekhar, Hydrodynamic and hydromagnetic stability, Oxford University Press, Oxford, 1961.

[3] L.N. Howard, Heat transport in turbulent convection, J. Fluid Mechanics **17** (1964), 405-432; C. R. Doering, P. Constantin, Variational bounds on energy dissipation in incompressible flows III. Convection, Phys. Rev E, **53** (1996), 5957-5981.

[4] P. Constantin, C. R. Doering, Heat transfer in convective turbulence, Nonlinearity **9** (1996), 1049-1060.

[5] Convection at high Rayleigh number, L.N. Howard, Applied Mechanics, Proc. 11th Cong. Applied Mech. (Ed. H. Gertler), pp. 1109-1115 (1964), Infinite Prandtl Number Turbulent Convection, S.-K. Chan, Stud. Appl. Math 50, 13-49 (1971), Scaling of hard thermal turbulence in Rayleigh-Benard convection, B, Castaing, G. Gunaratne, F. Heslot, L. Kadanoff, A. Libchaber, S. Thomae, X.-Z. Wu, S. Zaleski and G. Zanetti, JFM 204, 1-30 (1989), Heat transport in high-Rayleigh-number convection, B.I. Shraiman and E.D. Siggia, Phys. Rev. A 42, 3650-3653 (1990), Z.-S. She, Phys. Fluids A 1, 911 (1989), V. Yakhot, Phys. Rev. Lett. 69, 769 (1992), Turbulent thermal convection at arbitrary Prandtl number, R.H. Kraichnan, Phys. Fluids 5, 1374-1389 (1962), High Rayleigh number convection, E.D. Siggia, Ann. Rev. Fluid Mech. 26, 137-168 (1994).

[6] Howard, L. *J. Fluid Mech*, **17**, 405 (1963).

[7] Busse, F. *J. Fluid Mech*, **37**, 457 (1969).

[8] P. Constantin, C. R. Doering, Variational bounds in dissipative systems, Physica D **82**, (1995), 221-228.

[9] Doering, C., Constantin, P. *Phys. Rev. Lett.* **69** 1648-1651, (1992); *Phys. Rev. E* **49** 4087-4099, (1994);
[10] Constantin, P., Doering, C. *Phys. Rev. E* **51**, 3192-31-98, (1995).
[11] Doering, C., Constantin, P., *Phys. Rev E* **53**, 5957-5981, (1996).
[12] Constantin, P., Doering, C., *Nonlinearity* **9**, 1049-1060, (1996).
[13] C. Marchioro, Physica D **74** (1994), (1994), 395.
[14] R. R. Kerswell, J. Fluid Mech **321** (1996), 335.
[15] Nicodemus, R., Grossman, S., Holthaus, M. *Phys. Rev. Lett.*, **79**, 21, R4710 (1997).
[16] R. R. Kerswell, Physica D **100** (1997), 355.
[17] P. Constantin, C. Doering, Infinite Prandtl Number Convection, J. Stat. Phys, to appear.
[18] P. Constantin, Geometric Statistics in Turbulence, Siam Review **36**, 1994.
[19] C. Doering, P. Constantin, Bounds for heat transport in a porous layer, J. Fluid Mech, to appear.
[20] Nicodemus, R., Grossman, S., Holthaus, Physica D **101** (1997), 178.
[21] P. Constantin, C. Hallstrom, V. Putkaradze, Heat transport in rotating convection, Phys. D, to appear.
[22] C. Hallstrom, Ph.D. Thesis, University of Chicago, in preparation.
[23] Liu, Y., Ecke, R. *Phys. Rev. Lett*, **79**, 12, R2257 (1997).

Department of Mathematics, University of Chicago, 5734 S. University Avenue, Chicago, IL 60637

Contemporary Mathematics
Volume **238**, 1999
B 0-8218-1196-7-03541-6

On the Solvability of Implicit Nonlinear Systems in the Vectorial Case

Bernard Dacorogna and Paolo Marcellini

Abstract. We continue the study of a functional analytic method based on Baire category theorem for handling existence of almost everywhere solutions of a large class of partial differential equations and systems, which are nonlinear in the highest derivatives. We consider in this paper the solvability in the vectorial case of some implicit nonlinear systems of arbitrary order. The results have applications to the calculus of variations, nonlinear elasticity, problems of phase transitions or optimal design.

1. Introduction

The aim of this paper is to prove some existence results for systems of the form

$$\text{(1.1)} \qquad \begin{cases} u \in W^{N,\infty}(\Omega;\mathbb{R}^m) \\ F_i\left(x, D^{[N-1]}u(x), D^N u(x)\right) = 0, \quad \text{a.e. } x \in \Omega, \ \ i = 1,\dots,I \\ D^\alpha u(x) = D^\alpha \varphi(x), \quad x \in \partial\Omega, \ \ \alpha = 0,\dots,N-1, \end{cases}$$

where Ω is an open set of $\mathbb{R}^n$ $(n \geq 1)$, F_i for $i = 1,\dots,I$, $(I \geq 1)$ are given real functions and $u : \mathbb{R}^n \to \mathbb{R}^m$ $(m \geq 1)$ is the unknown in the Sobolev class $W^{N,\infty}(\Omega;\mathbb{R}^m)$. For the partial derivatives of u up to the order N $(\alpha = 1, 2, \dots, N \geq 1)$ we use the symbols (see the next section for some details more)

$$D^\alpha u = \left(\frac{\partial^\alpha u^i}{\partial x_{j_1} \dots \partial x_{j_N}}\right)^{1\leq i\leq m}_{1\leq j_1,\dots,j_N\leq n}, \quad \alpha = 1, 2, \dots, N,$$

$$D^{[N-1]}u = \left(u, \dots, D^\alpha u, \dots, D^{N-1}u\right).$$

Finally $\varphi \in W^{N,\infty}(\Omega;\mathbb{R}^m)$ is a fixed boundary datum.

Recently in a series of papers the authors exploited a method to obtain existence of solutions to differential problems of the type (1.1) in the vector valued case $m \geq 1$. In particular in [**18**], [**19**], [**20**], [**21**], the authors studied the first order case $(N = 1)$, while in [**22**] they considered the second order one $(N = 2)$. The method, based on *Baire category theorem*, was originated by A. Cellina [**11**] to

1991 *Mathematics Subject Classification.* Primary 35G30, 35R70; Secondary 34A60, 49J45.
We thank Gui-Qiang Chen and Emmanuele DiBenedetto for their hospitality at Northwestern University during the International Conference on Nonlinear Partial Differential Equations and Applications.

study existence of some ordinary differential inclusion, i.e. $n = 1$. Other researches for the scalar case, following the work of A. Cellina, were those of F.S. De Blasi and G. Pianigiani [**25**], [**26**], and A. Bressan and F. Flores [**7**]. Again in the scalar case we mention some recent result by M.A. Sychev [**49**] and S. Zagatti [**50**]. Coming back to the vectorial setting $N \geq 1$, there is an explicit construction of a solution for a particular system (1.1) by A. Cellina and S. Perrotta [**12**]. Moreover a generalization of the second order case $N = 2$, with similar assumptions than in the author's paper [**22**], have been proposed by L. Poggiolini [**46**] for the general vector-valued case with higher derivatives, i.e. $m \geq 1$ and $N \geq 1$.

In this paper we consider a general framework, with general assumption, which well fit into applications to the calculus of variations, nonlinear elasticity, problems of phase transitions or optimal design. Some model results proved in this paper are stated in Section 3; some other model results for singular values are stated below in this introduction. We have in progress a book [**23**] on this subject that will explain in details these applications, as well as some other related existence results for general systems of the type (1.1).

We propose below two theorems for *singular values*, consequence of the general results proved in this paper, respectively for first and for second order problems. For a given matrix $\xi \in \mathbb{R}^{n\times n}$ we denote by $0 \leq \lambda_1(\xi) \leq ... \leq \lambda_n(\xi)$ its singular values, i. e. the eigenvalues of the symmetric matrix $\left(\xi^t \xi\right)^{1/2}$.

THEOREM 1.1. *Let $\Omega \subset \mathbb{R}^n$ be an open set, $a_i : \overline{\Omega} \times \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, n$, be continuous functions satisfying $c \leq a_1(x,s) \leq \dots \leq a_n(x,s)$ for a positive constant c and for every $(x,s) \in \overline{\Omega} \times \mathbb{R}^n$. Let $\varphi \in C^1\left(\overline{\Omega}; \mathbb{R}^n\right)$ (or piecewise C^1) satisfy*

$$\prod_{i=\nu}^{n} \lambda_i(D\varphi(x)) < \prod_{i=\nu}^{n} a_i(x, \varphi(x)), \quad x \in \Omega, \quad \nu = 1, \dots, n.$$

Then there exists a function $u \in W^{1,\infty}(\Omega; \mathbb{R}^n)$ such that

$$\begin{cases} \lambda_i(Du(x)) = a_i(x, u(x)), & \text{a.e. } x \in \Omega, \quad i = 1, \dots, n \\ u(x) = \varphi(x), & x \in \partial\Omega. \end{cases} \tag{1.2}$$

It is interesting to see an implication of Theorem 1.1 when $n = 2$; in this case we have

$$|\xi|^2 = \sum_{i,j=1}^{2} \xi_{ij}^2 = (\lambda_1(\xi))^2 + (\lambda_2(\xi))^2, \quad |\det \xi| = \lambda_1(\xi)\lambda_2(\xi).$$

The differential problem (1.2) can be equivalently formulated in the form

$$\begin{cases} |Du(x)|^2 = a_1^2 + a_2^2, & \text{a.e. } x \in \Omega \\ |\det Du(x)| = a_1 a_2, & \text{a.e. } x \in \Omega \\ u(x) = \varphi(x), & x \in \partial\Omega. \end{cases} \tag{1.3}$$

Therefore system (1.3) can be seen as a combination of the *vectorial eikonal equation* and of the *equation of prescribed absolute value of the Jacobian determinant.*

We consider below the singular values of a *symmetric* matrix $\xi \in \mathbb{R}_s^{n\times n}$, $0 \leq \lambda_1(\xi) \leq \dots \leq \lambda_n(\xi)$, which are now, because of the symmetry of the matrix, the absolute value of the eigenvalues.

THEOREM 1.2. *Let $\Omega \subset \mathbb{R}^n$ be an open set, $a_i : \overline{\Omega} \times \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$, $i = 1, ..., n$ be continuous functions satisfying*

$$0 < c \leq a_1(x, s, p) \leq \ldots \leq a_n(x, s, p)$$

for a constant c and for every $(x, s, p) \in \overline{\Omega} \times \mathbb{R} \times \mathbb{R}^n$. Let $\varphi \in C^2_{piec}\left(\overline{\Omega}\right)$ be such that

$$\lambda_i\left(D^2\varphi(x)\right) < a_i(x, \varphi(x), D\varphi(x)), \quad a.e.\ x \in \Omega, \quad i = 1, \ldots, n \tag{1.4}$$

(in particular $\varphi \equiv 0$). Then there exists a function $u \in W^{2,\infty}(\Omega)$ such that

$$\begin{cases} \lambda_i\left(D^2u(x)\right) = a_i(x, u(x), Du(x)), & a.e.\ x \in \Omega, \quad i = 1, \ldots, n \\ u(x) = \varphi(x),\ Du(x) = D\varphi(x), & x \in \partial\Omega. \end{cases}$$

As a consequence we find that the following Dirichlet-Neumann problem (1.5) admits a solution.

COROLLARY 1.3. *Let $\Omega \subset \mathbb{R}^n$ be open. Let $f : \overline{\Omega} \times \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ be continuous and such that $f(x, s, p) \geq f_0 > 0$ for some constant f_0 and for every $(x, s, p) \in \overline{\Omega} \times \mathbb{R} \times \mathbb{R}^n$. Let $\varphi \in C^2(\overline{\Omega})$ (or $C^2_{piec}(\overline{\Omega})$) satisfy*

$$\left|\det D^2\varphi(x)\right| < f(x, \varphi(x), D\varphi(x)), \quad x \in \overline{\Omega}.$$

Then there exists a function $u \in W^{2,\infty}(\Omega)$ such that

$$\begin{cases} \left|\det D^2u(x)\right| = f(x, u(x), Du(x)), & a.e.\ x \in \Omega, \\ u = \varphi,\ Du = D\varphi, & on\ \partial\Omega. \end{cases} \tag{1.5}$$

The above Theorems 1.1 and 1.2 are consequences of the general results proved in this paper (see Theorem 6.1 and, in the case independent of lower order terms, Theorem 3.1). When the differential problems above are independent of lower order terms, the first order case has been established in [**18**], [**19**], [**20**], [**21**] when $n = 2$ and, with the same proof, in [**24**] for the general case. When $n = 3$, $a_i \equiv 1$ and $\varphi \equiv 0$, the result can be found in Cellina-Perrotta [**12**]; see also Celada-Perrotta [**10**].

Motivated by an application to nonlinear elasticity, more precisely in the study of a problem of *potential wells*, S. Muller and V. Sverak [**44**], [**45**] recently obtained very interesting attainment results, that in particular cases can be compared with ours, at least in a model case of first order vectorial problem $m \geq 1$, $N = 1$, without x dependence and without lower order terms, by using Gromov's method of *convex integration* (see M. Gromov [**32**] and D. Spring [**47**]). Some related results obtained by mean of Gromov method are also due to P. Celada and S. Perrotta [**10**].

The use of *viscosity solutions* to solve this kind of problems is classical and much older, although it seems related essentially with the scalar case. The bibliography in this subject is very wide; there are several excellent books and articles in this field and we can mention here only a few of them: M. Bardi-I. Capuzzo Dolcetta [**3**], G. Barles [**4**], S.H. Benton [**6**], I. Capuzzo Dolcetta-L.C. Evans [**8**], I. Capuzzo Dolcetta-P.L. Lions [**9**], M.G. Crandall-L.C. Evans-P.L. Lions [**13**], M.G. Crandall-H. Ishii-P.L. Lions [**14**], M.G. Crandall-P.L. Lions [**15**], A. Douglis [**27**], W.H. Fleming-H.M. Soner [**28**], H. Frankowska [**30**], E. Hopf [**34**], H. Ishii [**35**], S.N. Kruzkov [**38**], P.D. Lax [**39**], P.L. Lions [**40**] and A.I. Subbotin [**48**].

2. Notations

We introduce some notations and definitions to handle the higher order case. Our presentation is slightly different from those of [**2**], [**33**], [**46**]. We start first with some notations, with the aim to write in a simple way the matrix $D^N u$ of all partial derivatives of order N of a map $u : \mathbb{R}^n \to \mathbb{R}^m$.

Let $N, n, m \geq 1$ be integers. We denote by $\mathbb{R}_s^{m\times n^N}$ the set of matrices

$$A = \left(A^i_{j_1\dots j_N}\right)^{1\leq i\leq m}_{1\leq j_1,\dots,j_N\leq n} \in \mathbb{R}^{m\times n^N}$$

such that for every permutation σ of $\{j_1, ..., j_N\}$ we have

$$A^i_{\sigma(j_1\dots j_N)} = A^i_{j_1\dots j_N}.$$

When $N = 1$ we have $\mathbb{R}_s^{m\times n} = \mathbb{R}^{m\times n}$, while if $m = 1$ and $N = 2$ we get $\mathbb{R}_s^{n^2} = \mathbb{R}_s^{n\times n}$, i.e. the usual set of symmetric matrices.

As already stated in the introduction, for $u : \mathbb{R}^n \to \mathbb{R}^m$ we write

$$D^N u = \left(\frac{\partial^N u^i}{\partial x_{j_1}\dots\partial x_{j_N}}\right)^{1\leq i\leq m}_{1\leq j_1,\dots,j_N\leq n} \in \mathbb{R}_s^{m\times n^N}.$$

We also use the notation

$$D^{[N]}u = \left(u, Du, \dots, D^N u\right) \in \mathbb{R}^m \times \mathbb{R}^{m\times n} \times \dots \times \mathbb{R}_s^{m\times n^N},$$

which stands for the matrix of all partial derivatives of u up to the order N. We shall write

$$\mathbb{R}_s^{m\times M} = \mathbb{R}^m \times \mathbb{R}^{m\times n} \times \mathbb{R}_s^{m\times n^2} \times ... \times \mathbb{R}_s^{m\times n^{(N-1)}},$$

where

$$M = 1 + n + ... + n^{(N-1)} = \frac{n^N - 1}{n - 1};$$

hence

$$D^{[N]}u = \left(D^{[N-1]}u, D^N u\right) \in \mathbb{R}_s^{m\times M} \times \mathbb{R}_s^{m\times n^N}.$$

Given $\alpha \in \mathbb{R}^n$, we denote by $\alpha^{\otimes_N} = \alpha \otimes \alpha ... \otimes \alpha$ (N times); it is a matrix in $\mathbb{R}_s^{n^N}$. Therefore a generic matrix of rank one in $\mathbb{R}_s^{m\times n^N}$ has the form

$$\beta \otimes \alpha^{\otimes_N} = (\beta_i \alpha_{j_1} ... \alpha_{j_N})^{1\leq i\leq m}_{1\leq j_1,\dots,j_N\leq n},$$

where $\beta \in \mathbb{R}^m$ and $\alpha \in \mathbb{R}^n$.

We now give the definitions of quasiconvexity and of rank one convexity in the higher order case. Let $f : \mathbb{R}_s^{m\times n^N} \to \mathbb{R}$ be a continuous function. We say that f is *quasiconvex* if

$$\int_\Omega f\left(\xi + D^N u\,(x)\right) dx \geq f\,(\xi)\,\mathrm{meas}\,\Omega\,,$$

for every $\xi \in \mathbb{R}_s^{m\times n^N}$ and every $u \in W_0^{N,\infty}(\Omega; \mathbb{R}^m)$. We say that f is *rank one convex* if the function of one real variable

$$F\,(t) = f\left(\xi + tw \otimes v^{\otimes_N}\right)$$

is convex in $t \in \mathbb{R}$ for every $\xi \in \mathbb{R}_s^{m\times n^N}$, $w \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$. It has been established by N.G. Meyers [**42**] (c.f. also [**2**]) that quasiconvexity implies rank one

convexity. In the case $N = 2$ it is proved in [**33**] that if the function is quasiconvex or rank one convex then it is automatically locally Lipschitz (as well as for $N = 1$).

Semicontinuity results for quasiconvex integrands have been studied by C.B. Morrey [**43**], who introduced the concept of quasiconvexity, and then by N.G. Meyers [**42**], E. Acerbi and N. Fusco [**1**], P. Marcellini [**41**] and many others since then. In particular, in the higher order case we refer to N.G. Meyers [**42**], N. Fusco [**31**] and M. Guidorzi and L. Poggiolini [**33**].

In a similar way as we defined the different notions of convexity for functions we may define these notions for sets. In particular we say that a set $K \subset \mathbb{R}^{m\times n^N}_s$ is *rank one convex* if for every $t \in [0,1]$ and every $A, B \in K$ with $\operatorname{rank}(A-B) \le 1$ then $tA + (1-t)B \in K$. Moreover, for a set $E \subset \mathbb{R}^{m\times n^N}_s$ we define

$$\operatorname{Rco} E = \left\{ \begin{array}{c} \xi \in \mathbb{R}^{m\times n} : f(\xi) \le 0, \quad \forall f : \mathbb{R}^{m\times n^N}_s \to \overline{\mathbb{R}} = \mathbb{R}\cup\{+\infty\}, \\ f\,|_E = 0\,, \ \ f \text{ rank one convex} \end{array} \right\},$$

called the *rank one convex hull* of E, and

$$\overline{\operatorname{Qco} E} = \left\{ \begin{array}{l} \xi \in \mathbb{R}^{m\times n^N}_s : \ \ f(\xi) \le 0, \ \ \forall f : \mathbb{R}^{m\times n^N}_s \to \mathbb{R} \\ f\,|_E = 0\,, \quad f \ \text{ quasiconvex and continuous} \end{array} \right\},$$

called the (closure of the) *quasiconvex hull* of E.

3. Statement of Some Model Results

Among some other existence results presented in Sections 5, 6, four model theorems, consequence of the theory developed in the next sections, are stated below. Note that the boundary condition is to be interpreted as

$$u - \varphi \in W^{N,\infty}_0(\Omega;\mathbb{R}^m).$$

The first result that we state in this section and that we will prove in this paper is the following Theorem 3.1. It has applications in particular in *singular values problems* (see [**20**], [**21**], [**22**], [**10**] or [**23**]).

THEOREM 3.1. *Let $\Omega \subset \mathbb{R}^n$ be open. Let $F_i : \mathbb{R}^{m\times n^N}_s \to \mathbb{R}$, $i = 1, 2, ..., I$, be quasiconvex, locally Lipschitz and positively homogeneous of degree $\alpha_i > 0$. Let $a_i > 0$ and*

$$E = \left\{ \xi \in \mathbb{R}^{m\times n^N}_s \ : \ F_i(\xi) = a_i, \ \ i = 1, 2, ..., I \right\}.$$

Assume that $\operatorname{Rco} E$ *is compact and satisfies*

$$\operatorname{Rco} E = \left\{ \xi \in \mathbb{R}^{m\times n^N}_s \ : \ F_i(\xi) \le a_i, \ \ i = 1, 2, ..., I \right\}.$$

Let $\varphi \in C^N_{piec}(\overline{\Omega};\mathbb{R}^m)$ be such that

$$F_i(D^N\varphi(x)) < a_i, \qquad \text{a.e. } x \in \Omega, \ i = 1, ..., I.$$

Then there exists (a dense set of) $u \in W^{N,\infty}(\Omega;\mathbb{R}^m)$ such that

$$\left\{ \begin{array}{ll} F_i(D^N u(x)) = a_i, & \text{a.e. } x \in \Omega, \ i = 1, ..., I \\ D^\alpha u(x) = D^\alpha \varphi(x), & x \in \partial\Omega, \quad \alpha = 0, ..., N-1. \end{array} \right.$$

THEOREM 3.2. *Let $\Omega \subset \mathbb{R}^n$ be open. Let $F_i : \mathbb{R}_s^{m\times n^N} \to \mathbb{R}$, $i = 1, 2, ..., I$, be convex and let*

$$E = \left\{\xi \in \mathbb{R}_s^{m\times n^N} : \ F_i(\xi) = 0, \quad i = 1, 2, ..., I\right\}.$$

Assume that $\operatorname{Rco} E$ *is compact and satisfies*

$$\operatorname{Rco} E = \left\{\xi \in \mathbb{R}_s^{m\times n^N} : \ F_i(\xi) \leq 0, \quad i = 1, 2, ..., I\right\}.$$

Let $\varphi \in C^N_{piec}(\overline{\Omega}; \mathbb{R}^m)$ satisfy

$$F_i(D^N\varphi(x)) < 0, \quad \text{a.e. } x \in \Omega, \ \ i = 1, ..., I$$

or $\varphi \in W^{N,\infty}(\Omega; \mathbb{R}^m)$ satisfy

$$F_i(D^N\varphi(x)) \leq -\theta, \quad \text{a.e. } x \in \Omega, \ i = 1, ..., I$$

for a certain $\theta > 0$. Then there exists (a dense set of) $u \in W^{N,\infty}(\Omega; \mathbb{R}^m)$ such that

$$\begin{cases} F_i(D^N u(x)) = 0, \quad \text{a.e. } x \in \Omega, \ \ i = 1, ..., I \\ D^\alpha u(x) = D^\alpha\varphi(x), \quad x \in \partial\Omega, \ \ \alpha = 0, ..., N-1. \end{cases}$$

The generalization of the previous results to the case with explicit dependence on lower order terms is not as simple, from the technical point of view, as it could seem at a first glance. However, in particular (see Section 6 for more general results) we can obtain the following two theorems.

THEOREM 3.3. *Let $\Omega \subset \mathbb{R}^n$ be open. Let $F_i : \overline{\Omega} \times \mathbb{R}_s^{m\times M} \times \mathbb{R}_s^{m\times n^N} \to \mathbb{R}$, $F_i = F_i(x, s, \xi)$, $i = 1, ..., I$, be continuous with respect to $(x, s) \in \overline{\Omega} \times \mathbb{R}_s^{m\times M}$ and quasiconvex, locally Lipschitz and positively homogeneous of degree $\alpha_i > 0$ with respect to the last variable $\xi \in \mathbb{R}_s^{m\times n^N}$.*

Let $a_i : \overline{\Omega} \times \mathbb{R}_s^{m\times M} \to \mathbb{R}$, $i = 1, ..., I$, be continuous and satisfy for a certain $a_0 > 0$

$$a_i(x, s) \geq a_0 > 0, \ \ i = 1, ..., I, \ \ \forall (x, s) \in \overline{\Omega} \times \mathbb{R}_s^{m\times M}.$$

Assume that, for every $(x, s) \in \overline{\Omega} \times \mathbb{R}_s^{m\times M}$

$$\begin{aligned} &\operatorname{Rco}\left\{\xi \in \mathbb{R}_s^{m\times n^N} : \ F_i(x, s, \xi) = a_i(x, s), \ \ i = 1, ..., I\right\} \\ &= \left\{\xi \in \mathbb{R}_s^{m\times n^N} : \ F_i(x, s, \xi) \leq a_i(x, s), \ \ i = 1, ..., I\right\} \end{aligned}$$

and is bounded in $\mathbb{R}_s^{m\times n^N}$ uniformly with respect to (x, s) in a bounded set of $\overline{\Omega} \times \mathbb{R}_s^{m\times M}$. If $\varphi \in C^N_{piec}(\overline{\Omega}; \mathbb{R}^m)$ satisfies

$$F_i(x, D^{[N-1]}\varphi(x), D^N\varphi(x)) < a_i\left(x, D^{[N-1]}\varphi(x)\right), \ \ \text{a.e. } x \in \Omega, \ \ i = 1, ..., I,$$

then there exists (a dense set of) $u \in W^{N,\infty}(\Omega; \mathbb{R}^m)$ such that

$$\begin{cases} F_i(x, D^{[N-1]}u(x), D^N u(x)) = a_i\left(x, D^{[N-1]}u(x)\right), \ \ \text{a.e. } x \in \Omega, \ \ i = 1, ..., I \\ D^\alpha u(x) = D^\alpha\varphi(x), \ \ x \in \partial\Omega, \ \ \alpha = 0, ..., N-1. \end{cases}$$

THEOREM 3.4. *Let $\Omega \subset \mathbb{R}^n$ be open. Let $F_i : \overline{\Omega} \times \mathbb{R}_s^{m\times M} \times \mathbb{R}_s^{m\times n^N} \to \mathbb{R}$, $F_i = F_i(x,s,\xi)$, $i = 1,...,I$, be continuous with respect to $(x,s) \in \overline{\Omega} \times \mathbb{R}_s^{m\times M}$ and convex with respect to the last variable $\xi \in \mathbb{R}_s^{m\times n^N}$. Assume that, for every $(x,s) \in \overline{\Omega} \times \mathbb{R}_s^{m\times M}$*

$$\begin{aligned}&\mathrm{Rco}\left\{\xi \in \mathbb{R}_s^{m\times n^N} : \ F_i(x,s,\xi) = 0, \ \ i = 1,...,I\right\}\\ &= \left\{\xi \in \mathbb{R}_s^{m\times n^N} : \ F_i(x,s,\xi) \le 0, \ \ i = 1,...,I\right\}\end{aligned}$$

and is bounded in $\mathbb{R}_s^{m\times n^N}$ uniformly with respect to (x,s) in a bounded set of $\overline{\Omega} \times \mathbb{R}_s^{m\times M}$. Assume also that for every $(x,s) \in \overline{\Omega} \times \mathbb{R}_s^{m\times M}$ there exists $\xi_0 = \xi_0(x,s) \in \mathbb{R}_s^{m\times n^N}$ such that

$$F_i(x,s,\xi_0) < 0, \quad i = 1,...,I.$$

Let $\varphi \in C_{piec}^N(\overline{\Omega};\mathbb{R}^m)$ satisfy

$$F_i(x, D^{[N-1]}\varphi(x), D^N\varphi(x)) < 0, \quad \text{a.e. } x \in \Omega, \ i = 1,...,I,$$

or $\varphi \in W^{N,\infty}(\Omega;\mathbb{R}^m)$ be such that

$$F_i(x, D^{[N-1]}\varphi(x), D^N\varphi(x)) \le -\theta, \quad \text{a.e. } x \in \Omega, \ i = 1,...,I,$$

for a certain $\theta > 0$. Then there exists (a dense set of) $u \in W^{N,\infty}(\Omega;\mathbb{R}^m)$ such that

$$\begin{cases} F_i(x, D^{[N-1]}u(x), D^Nu(x)) = 0, & \text{a.e. } x \in \Omega, \ i = 1,...,I \\ D^\alpha u(x) = D^\alpha\varphi(x), & x \in \partial\Omega, \ \alpha = 0,...,N-1. \end{cases}$$

Proofs of Theorems 3.1, 3.2, 3.3, 3.4 will be given in Section 5.

4. Problems without Lower Order Terms

4.1. Weakly Extreme Sets and the Relaxation Property. We start with the definition of *weakly extreme set* E_{ext} of a given set $E \subset \mathbb{R}_s^{m\times n^N}$.

DEFINITION 4.1 (Weakly extreme set). *A set E_{ext} is said to be weakly extreme for a subset E of $\mathbb{R}_s^{m\times n^N}$ if $E_{\mathrm{ext}} \subset E$ and if for every $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that, for every $u, u_\nu \in W^{N,\infty}(\Omega;\mathbb{R}^m)$ satisfying*

$$\begin{cases} u_\nu \overset{*}{\rightharpoonup} u \quad \text{in } W^{N,\infty}(\Omega;\mathbb{R}^m) \\ D^N u_\nu(x) \in \overline{\mathrm{Qco}\, E}, \quad \text{a.e in } \Omega, \end{cases} \tag{4.1}$$

then the following implication holds for ν sufficiently large

$$\int_\Omega \mathrm{dist}\left(D^N u(x); E_{\mathrm{ext}}\right) dx \le \delta \ \Rightarrow \int_\Omega \mathrm{dist}\left(D^N u_\nu(x); E_{\mathrm{ext}}\right) dx \le \varepsilon. \tag{4.2}$$

In some cases we can choose in the previous Definition 4.1 the set E_{ext} equal to E; by the following result we give sufficient conditions in order to make this choice.

THEOREM 4.1. *Let $\Omega \subset \mathbb{R}^n$ be open and bounded. Assume $E \subset \mathbb{R}_s^{m\times n^N}$ is compact and that there exist F_i, $i = 1,2,...,I$, quasiconvex and locally Lipschitz functions such that*

$$E = \left\{\xi \in \mathbb{R}_s^{m\times n^N} : \ F_i(\xi) = 0, \ \ i = 1,2,...,I\right\}. \tag{4.3}$$

Then the set E_{ext} in Definition 4.1 can be chosen equal to E.

PROOF. Under assumption (4.3) we will prove that for every $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that, for every $u, u_\nu \in W^{N,\infty}(\Omega;\mathbb{R}^m)$ satisfying (4.1), for ν sufficiently large the following implication holds

$$\int_\Omega \operatorname{dist}\left(D^N u(x); E\right) dx \leq \delta \;\Rightarrow\; \int_\Omega \operatorname{dist}\left(D^N u_\nu(x); E\right) dx \leq \varepsilon. \tag{4.4}$$

Step 1 : We first observe that, by definition of $\overline{\operatorname{Qco} E}$ and because of the weak* convergence, from (4.1) we also have

$$D^N u(x) \in \overline{\operatorname{Qco} E}, \quad \text{a.e in } \Omega.$$

We next fix the constants. Since $\overline{\operatorname{Qco} E}$ is compact we can find $\beta > 0$ such that

$$\eta \in \overline{\operatorname{Qco} E} \;\Rightarrow\; \operatorname{dist}(\eta; E) \leq \beta. \tag{4.5}$$

We also see that, by definition of $\overline{\operatorname{Qco} E}$ (since $F_i = 0$ on E), we have $F_i(\eta) \leq 0$ for every $\eta \in \overline{\operatorname{Qco} E}$ and since F_i is Lipschitz on $\overline{\operatorname{Qco} E}$ we can find $\alpha > 0$ such that

$$0 \leq -F_i(\eta) \leq \alpha \cdot \operatorname{dist}(\eta; E), \quad \forall \eta \in \overline{\operatorname{Qco} E}, \;\; \forall\, i = 1, 2, ..., I. \tag{4.6}$$

We finally observe that by continuity of the distance function and of the F_i we can find for every $\sigma > 0$, $\rho = \rho(\sigma) > 0$ such that for every $\eta \in \overline{\operatorname{Qco} E}$

$$\left.\begin{array}{c} 0 \leq -F_i(\eta) < \rho \\ \forall\, i = 1, 2, ..., I \end{array}\right\} \quad \Rightarrow \quad \operatorname{dist}(\eta; E) < \sigma; \tag{4.7}$$

in fact the choice

$$\rho(\sigma) = \min_{\substack{\operatorname{dist}(\eta; E) \geq \sigma \\ \eta \in \overline{\operatorname{Qco} E}}} \max_{i=1,2,...,I} \{-F_i(\eta)\}$$

leads immediately to (4.7).

Step 2 : We now fix $\varepsilon > 0$ and we choose

$$\sigma = \frac{\varepsilon}{2 \operatorname{meas} \Omega}, \quad \delta(\varepsilon) = \frac{\varepsilon \rho(\sigma)}{4 I \alpha \beta}.$$

With this choice we will prove that, for ν sufficiently large, (4.4) holds.

In fact, by weak lower semicontinuity, by (4.6) and by the fact that $D^N u_\nu(x) \in \overline{\operatorname{Qco} E}$, we get

$$\begin{aligned} 0 \;&\geq\; \liminf_{\nu\to\infty} \int_\Omega F_i\left(D^N u_\nu(x)\right) dx \geq \int_\Omega F_i\left(D^N u(x)\right) dx \\ &\geq\; -\alpha \int_\Omega \operatorname{dist}\left(D^N u(x); E\right) dx \geq -\alpha\delta, \quad \forall\, i = 1, 2, ..., I. \end{aligned}$$

Therefore, for ν large enough, we have

$$\int_\Omega F_i\left(D^N u_\nu(x)\right) dx \geq -2\alpha\delta = -\frac{\varepsilon\rho}{2I\beta}, \quad \forall\, i = 1, 2, ..., I. \tag{4.8}$$

Setting

$$\Omega_{i,\nu} = \left\{x \in \Omega :\; -F_i\left(D^N u_\nu(x)\right) \geq \rho\right\},$$

we get

$$\rho \cdot \operatorname{meas} \Omega_{i,\nu} \leq -\int_\Omega F_i\left(D^N u_\nu(x)\right) dx \leq \frac{\varepsilon\rho}{2I\beta}$$

and hence

$$\operatorname{meas}\Omega_{i,\nu} \leq \frac{\varepsilon}{2I\beta}\,. \tag{4.9}$$

Letting $\Omega_\nu = \cup_{i=1}^{I}\Omega_{i,\nu}$, we find

$$\operatorname{meas}\Omega_{\nu} \leq \frac{\varepsilon}{2\beta}\,. \tag{4.10}$$

We now use (4.7) on $\Omega\diagdown\Omega_\nu$ to obtain

$$\operatorname{dist}\left(D^N u_\nu(x)\,;E\right) \leq \sigma = \frac{\varepsilon}{2\operatorname{meas}\Omega}\,, \quad \text{a.e. } x \in \Omega\diagdown\Omega_\nu$$

and hence, by using (4.5),

$$\begin{aligned}\int_\Omega \operatorname{dist}\left(D^N u_\nu(x)\,;E\right)dx &= \int_{\Omega\diagdown\Omega_\nu} \operatorname{dist}\left(D^N u_\nu(x)\,;E\right)dx \\ &\quad + \int_{\Omega_\nu} \operatorname{dist}\left(D^N u_\nu(x)\,;E\right)dx \\ &\leq \frac{\varepsilon\operatorname{meas}(\Omega\diagdown\Omega_\nu)}{2\operatorname{meas}\Omega} + \beta\operatorname{meas}\Omega_\nu \leq \varepsilon\,,\end{aligned}$$

which is the claimed result. ∎

We now state the main hypothesis that should satisfy $\overline{\operatorname{Qco} E}$ or rather a subset K of it; it says that for every boundary datum which is a polynomial of degree at most N we can find an approximate solution of our problem that remains almost everywhere in $\operatorname{int}\overline{\operatorname{Qco} E}$.

Definition 4.2 (Relaxation property). *Let E be a set of $\mathbb{R}^{m\times n^N}_s$. A subset $K \subset \overline{\operatorname{Qco} E}$ is said to have the relaxation property with respect to E if the following condition is satisfied: for every $\xi \in \operatorname{int} K$, for every bounded open set Ω of $\mathbb{R}^n$, there exists a sequence $u_\nu \in C^N_{piec}\left(\overline{\Omega};\mathbb{R}^m\right)$ such that (defining $u_\xi(x)$ by $D^N u_\xi(x) = \xi$)*

$$\begin{cases} D^\alpha u_\nu(x) = D^\alpha u_\xi(x) & \text{on } \partial\Omega, \quad \alpha = 0,...,N-1 \\ u_\nu \overset{*}{\rightharpoonup} u_\xi & \text{in } W^{N,\infty}(\Omega;\mathbb{R}^m) \\ D^N u_\nu(x) \in E \cup \operatorname{int} K, & \text{a.e in } \Omega \\ \int_\Omega \operatorname{dist}\left(D^N u_\nu(x)\,;E\right)dx \to 0 & \text{as } \nu \to \infty. \end{cases}$$

Remark 4.1. *(i) Below we will give some cases where the relaxation property can be established.*

(ii) What is interesting about defining the relaxation property for K and not for $\overline{\operatorname{Qco} E}$ (in some sense the relaxation property could be taken as the definition of the quasiconvex hull) is that we do not need to compute this last quantity and we can choose for example $K = \operatorname{Rco} E$ (or even a subset), our approximation property (c.f. Definition 4.3) then ensures that K has the relaxation property (c.f. also [**21**] *or Lemma 2.2 and Lemma 3.4 of* [**22**]*).*

4.2. An Abstract Result. The main result of this section is

Theorem 4.2. *Let $\Omega \subset \mathbb{R}^n$ be open and let $E \subset \mathbb{R}^{m\times n^N}_s$ be compact. Let us assume that $K \subset \mathbb{R}^{m\times n^N}_s$ has the relaxation property with respect to E_{ext}. Let $\varphi \in C^N_{piec}\left(\overline{\Omega};\mathbb{R}^m\right)$ be such that*

$$D^N\varphi(x) \in E_{\text{ext}} \cup \operatorname{int} K, \quad \text{a.e in } \Omega.$$

Then there exists (a dense set of) $u \in W^{N,\infty}(\Omega;\mathbb{R}^m)$ *such that*

$$\begin{cases} D^N u(x) \in E_{\text{ext}}, \quad a.e.\ x \in \Omega \\ D^\alpha u(x) = D^\alpha \varphi(x) \quad on\ \partial\Omega, \quad \alpha = 0, ..., N-1. \end{cases} \tag{4.11}$$

REMARK 4.2. *(i) Since* $E_{\text{ext}} \subset E$, *in particular* $u \in W^{N,\infty}(\Omega;\mathbb{R}^m)$ *and it satisfies*

$$\begin{cases} D^N u(x) \in E, \quad a.e.\ x \in \Omega \\ D^\alpha u(x) = D^\alpha \varphi(x) \quad on\ \partial\Omega, \quad \alpha = 0, ..., N-1. \end{cases} \tag{4.12}$$

Of course conclusion (4.11) is sharper than (4.12).

(ii) The boundary condition is to be interpreted as

$$u - \varphi \in W_0^{N,\infty}(\Omega;\mathbb{R}^m).$$

(iii) The conclusion of the theorem is in fact more precise. The solution found is in the C^{N-1} *closure of the set*

$$\left\{ \begin{array}{c} u \in C^N_{piec}\left(\overline{\Omega};\mathbb{R}^m\right) : D^\alpha u(x) = D^\alpha \varphi(x) \quad on\ \partial\Omega, \quad \alpha = 0, ..., N-1 \\ and \quad D^N u(x) \in E_{\text{ext}} \cup \operatorname{int} K, \quad a.e\ in\ \Omega \end{array} \right\}.$$

PROOF. Step 1 : We observe first that $\Omega \subset \mathbb{R}^n$ can be assumed bounded, without loss of generality. We then let V be the closure in the $C^{N-1}\left(\overline{\Omega};\mathbb{R}^m\right)$ norm of

$$\left\{ \begin{array}{c} u \in C^N_{piec}\left(\overline{\Omega};\mathbb{R}^m\right) : \ D^\alpha u(x) = D^\alpha \varphi(x) \quad \text{on } \partial\Omega, \quad \alpha = 0, ..., N-1 \\ \text{and} \quad D^N u(x) \in E_{\text{ext}} \cup \operatorname{int} K, \quad \text{a.e in } \Omega \end{array} \right\}.$$

Note that $\varphi \in V$, that V is a complete metric space when endowed with the C^{N-1} norm and that (by weak lower semicontinuity and since $E_{\text{ext}} \cup K \subset \overline{\text{Qco}\, E_{\text{ext}}}$)

$$V \subset \left\{ u \in \varphi + W_0^{N,\infty}(\Omega;\mathbb{R}^m) : \ D^N u(x) \in \overline{\text{Qco}\, E_{\text{ext}}}, \quad \text{a.e in } \Omega \right\}$$

Step 2 : Let for $u \in V$

$$I(u) = \inf_{\{u_\nu\}} \liminf_{u_\nu \overset{*}{\rightharpoonup} u,\ u_\nu \in V} \left[- \int_\Omega \operatorname{dist}\left(D^N u_\nu(x); E_{\text{ext}}\right) dx \right].$$

Observe that I is lower semicontinuous on V, i.e. more precisely

$$\liminf_{u_\nu \overset{*}{\rightharpoonup} u,\ u_\nu \in V} I(u_\nu) \geq I(u), \quad \forall u \in V. \tag{4.13}$$

Since $I \leq 0$ on V, by taking $u_\nu \equiv u$ in the definition of I we see that

$$u \in V,\ I(u) = 0 \ \Rightarrow \ D^N u(x) \in E_{\text{ext}}, \text{ a.e. in } \Omega. \tag{4.14}$$

The above implication admits a converse (see (4.17) below). Indeed, by definition of I, for every $u \in V$ and for every $\varepsilon > 0$ we can find $u_\nu \in V$ (and hence in particular $D^N u_\nu(x) \in \overline{\text{Qco}\, E_{\text{ext}}}$), $u_\nu \overset{*}{\rightharpoonup} u$ such that

$$I(u) \geq -\varepsilon - \int_\Omega \operatorname{dist}\left(D^N u_\nu(x); E_{\text{ext}}\right) dx. \tag{4.15}$$

By the Definition 4.1 of E_{ext}, there exists $\delta = \delta(\varepsilon) > 0$ such that the following implication holds for ν sufficiently large

$$\int_\Omega \operatorname{dist}\left(D^N u(x); E_{\text{ext}}\right) dx \leq \delta \quad \Rightarrow \quad \int_\Omega \operatorname{dist}\left(D^N u_\nu(x); E_{\text{ext}}\right) dx \leq \varepsilon. \tag{4.16}$$

Therefore, by combining (4.15), (4.16) we get that, for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\int_{\Omega} \operatorname{dist}\left(D^N u\left(x\right); E_{\text{ext}}\right) dx \leq \delta \quad \Rightarrow \quad I\left(u\right) \geq -2\varepsilon. \tag{4.17}$$

Step 3 : Let

$$V^k = \left\{u \in V : I\left(u\right) > -1/k\right\}.$$

The set V^k is open in V (c.f. (4.13)). Furthermore it is dense in V. This follows from the relaxation property and we will prove this fact below. If this property has been established we deduce from Baire category theorem that $\cap V^k \subset \left\{u \in V : \ I\left(u\right) = 0\right\}$ is dense in V. Thus the result by (4.14).

It remains to prove that for any $u \in V$ and any $\varepsilon > 0$ sufficiently small we can find $u_\varepsilon \in V^k$ so that

$$\left\|u_\varepsilon - u\right\|_{N-1,\infty} \leq \varepsilon.$$

We will prove this property under the further assumption that $u \in C^N_{piec}$ and

$$D^N u\left(x\right) \in E_{\text{ext}} \cup \operatorname{int} K, \text{ a.e in } \Omega.$$

The general case will follow by definition of V. By working on each piece where $u \in C^N$ we can assume, without loss of generality, that $u \in C^N\left(\overline{\Omega}; \mathbb{R}^m\right)$ and $D^N u\left(x\right) \in E_{\text{ext}} \cup \operatorname{int} K$. We introduce the notations

$$\Omega_0 = \left\{x \in \Omega : D^N u\left(x\right) \in E_{\text{ext}}\right\}, \quad \Omega_1 = \Omega - \Omega_0.$$

It is clear, by continuity, that Ω_0 is closed and hence Ω_1 is open.

Before proceeding further we fix the constants. We let k be an integer and choose $0 < \varepsilon < 1/\left(2k\right)$, $\delta = \delta\left(\varepsilon\right)$ be as in Definition 4.1 and $\beta > 0$ be such that ($K \subset \overline{\text{Qco}\, E}$ being compact)

$$\xi \in K \ \Rightarrow \ \operatorname{dist}\left(\xi; E_{\text{ext}}\right) \leq \beta.$$

By approximation (c.f. the Appendix of [**23**]) we can find $u_s \in C^N\left(\overline{\Omega}_1; \mathbb{R}^m\right)$, an integer $I = I\left(\varepsilon\right)$ and $\Omega_{s,i} \subset \Omega_1$, $1 \leq i \leq I$, disjoint open sets such that

$$\begin{cases} u_s \equiv u, \ \text{ near } \partial\Omega_1 \\ \left\|u_s - u\right\|_{N,\infty} \leq \frac{\varepsilon}{2} \\ D^N u_s\left(x\right) \in \operatorname{int} K, \ \text{ a.e. } x \in \Omega_1 \\ \operatorname{meas}\left(\Omega_1 - \cup_{i=1}^I \Omega_{s,i}\right) \leq \frac{\delta}{2\beta} \\ D^N u_s\left(x\right) = \xi_{s,i} = \ \text{constant}, \ \ x \in \Omega_{s,i}. \end{cases}$$

We next use the relaxation property (since $\xi_{s,i} \in \operatorname{int} K$) to find $u_{i,\nu} \in C^N_{piec}\left(\overline{\Omega}_{s,i}; \mathbb{R}^m\right)$ satisfying

$$\begin{cases} D^\alpha u_{i,\nu} = D^\alpha u_s, \text{ on } \partial\Omega_{s,i}\,, \ \ \alpha = 0, ..., N-1 \\ \left\|u_s - u_{i,\nu}\right\|_{N-1,\infty} \leq \frac{\varepsilon}{2}\,, \ \text{ in } \Omega_{s,i} \\ D^N u_{i,\nu}\left(x\right) \in E_{\text{ext}} \cup \operatorname{int} K \\ \int_{\Omega_{s,i}} \operatorname{dist}\left(D^N u_{i,\nu}\left(x\right); E_{\text{ext}}\right) dx \leq \frac{\delta}{2} \cdot \frac{\operatorname{meas}(\Omega_{s,i})}{\operatorname{meas}(\Omega_1)}\,. \end{cases}$$

Now we can define

$$u_\varepsilon\left(x\right) = \begin{cases} u\left(x\right) & \text{if } x \in \Omega_0 \\ u_s\left(x\right) & \text{if } x \in \Omega_1 - \cup_{i=1}^I \Omega_{s,i} \\ u_{i,\nu}\left(x\right) & \text{if } x \in \Omega_{s,i}\,. \end{cases}$$

Observe that $u_\varepsilon \in C^N_{piec}\left(\overline{\Omega};\mathbb{R}^m\right)$,

$$\begin{cases} D^\alpha u_\varepsilon = D^\alpha u, \quad \text{on } \partial\Omega, \quad \alpha = 0, ..., N-1 \\ \|u_\varepsilon - u\|_{N-1,\infty} \leq \varepsilon, \quad \text{in } \Omega \\ D^N u_\varepsilon(x) \in E_{\text{ext}} \cup \text{int}\, K, \quad \text{a.e. } x \in \Omega \end{cases}$$

and that

$$\begin{aligned} \int_\Omega \text{dist}\left(D^N u_\varepsilon(x); E_{\text{ext}}\right) dx &= \int_{\Omega_0} \text{dist}\left(D^N u_\varepsilon(x); E_{\text{ext}}\right) dx \\ &\quad + \int_{\Omega_1} \text{dist}\left(D^N u_\varepsilon(x); E_{\text{ext}}\right) dx \\ &= \int_{\Omega_1} \text{dist}\left(D^N u_\varepsilon(x); E_{\text{ext}}\right) dx \\ &= \int_{\Omega_1 - \cup\Omega_{s,i}} \text{dist}\left(D^N u_\varepsilon(x); E_{\text{ext}}\right) dx \\ &\quad + \sum_{i=1}^{I} \int_{\Omega_{s,i}} \text{dist}\left(D^N u_\varepsilon(x); E_{\text{ext}}\right) dx \\ &\leq \frac{\delta}{2} + \frac{\delta}{2} \leq \delta. \end{aligned}$$

Hence combining (4.17) and the above inequality we get

$$I(u_\varepsilon) \geq -2\varepsilon > -\frac{1}{k},$$

which implies that $u_\varepsilon \in V^k$. The claimed density has therefore been established and the proof is thus complete. ∎

Direct consequence of Theorems 4.1 and 4.2 is the following

COROLLARY 4.3. *Let $\Omega \subset \mathbb{R}^n$ be open. Let $E \subset \mathbb{R}^{m\times n^N}_s$ be compact and satisfy (4.3), i.e. there exist F_i, $i = 1, 2, ..., I$, quasiconvex and locally Lipschitz functions such that*

$$E = \left\{\xi \in \mathbb{R}^{m\times n^N}_s : \ F_i(\xi) = 0, \quad i = 1, 2, ..., I\right\}.$$

Assume $K \subset \overline{\text{Qco}\, E} \subset \mathbb{R}^{m\times n^N}_s$ has the relaxation property with respect to E. Let $\varphi \in C^N_{piec}\left(\overline{\Omega};\mathbb{R}^m\right)$ be such that

$$D^N\varphi(x) \in E \cup \text{int}\, K, \quad \text{a.e in } \Omega;$$

then there exists (a dense set of) $u \in W^{N,\infty}(\Omega;\mathbb{R}^m)$ such that

$$\begin{cases} D^N u(x) \in E, \quad \text{a.e. } x \in \Omega \\ D^\alpha u(x) = D^\alpha\varphi(x) \quad \text{on } \partial\Omega, \quad \alpha = 0, ..., N-1, \end{cases}$$

or equivalently $u \in W^{N,\infty}(\Omega;\mathbb{R}^m)$ satisfies the differential problem

$$\begin{cases} F_i(D^N u(x)) = 0, \quad \text{a.e. } x \in \Omega, \quad i = 1, 2, ..., I \\ D^\alpha u(x) = D^\alpha\varphi(x) \quad \text{on } \partial\Omega, \quad \alpha = 0, ..., N-1. \end{cases}$$

4.3. The Key Approximation Lemma. The following lemma will be useful to establish the relaxation property.

LEMMA 4.4. *Let $\Omega \subset \mathbb{R}^n$ be an open set with finite measure. Let $t \in [0,1]$ and $A, B \in \mathbb{R}_s^{m\times n^N}$ with rank $\{A-B\} = 1$. Let φ be such that*

$$D^N\varphi(x) = tA + (1-t)B, \quad \forall x \in \overline{\Omega}.$$

Then, for every $\varepsilon > 0$, there exist $u \in C^N_{piec}\left(\overline{\Omega}; \mathbb{R}^m\right)$ and disjoint open sets $\Omega_A, \Omega_B \subset \Omega$, so that

$$\begin{cases} \left|\text{meas } \Omega_A - t \text{ meas } \Omega\right|, \ \left|\text{meas } \Omega_B - (1-t)\,\text{meas } \Omega\right| \leq \varepsilon \\ u \equiv \varphi \text{ near } \partial\Omega \\ \|u-\varphi\|_{N-1,\infty} \leq \varepsilon \\ D^N u(x) = \begin{cases} A & \text{in } \Omega_A \\ B & \text{in } \Omega_B \end{cases} \\ \text{dist}\left(D^N u(x), \text{co}\,\{A,B\}\right) \leq \varepsilon \quad \text{a.e. in } \Omega. \end{cases}$$

REMARK 4.3. *(i) By $\text{co}\,\{A,B\} = [A,B]$ we mean the closed segment joining A to B.*

(ii) It is interesting to note that when $n = 1$ the construction is exact, i.e. $\overline{\Omega} = \overline{\Omega}_A \cup \overline{\Omega}_B$.

PROOF. We divide the proof into two steps (unfortunately, since the notations are not easy, every time we will point out the case $n, m \geq 1$ and $N = 1$ as well as the case $m = 1$ and $N = 2$).

Step 1: Let us first assume that the matrix has the form

$$A - B = \alpha \otimes e_1^{\otimes_N}$$

where $e_1 = (1, 0, ..., 0) \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}^m$. For example when $N = 1$ we have

$$A - B = \begin{pmatrix} \alpha_1 & 0 & ... & 0 \\ \alpha_2 & 0 & ... & 0 \\ ... & ... & ... & ... \\ \alpha_m & 0 & ... & 0 \end{pmatrix} \in \mathbb{R}^{m\times n},$$

or when $m = 1$ and $N = 2$ we can write

$$A - B = \begin{pmatrix} \alpha & 0 & ... & 0 \\ 0 & 0 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 0 \end{pmatrix} \in \mathbb{R}_s^{n\times n}.$$

We can express Ω as union of cubes with faces parallel to the coordinate axes and a set of small measure. Then, by posing $u \equiv \varphi$ on the set of small measure, and by dilatations and translations, we can reduce ourselves to work with Ω equal to the unit cube.

Let Ω_ε be a set compactly contained in Ω and let $\eta \in C_0^\infty(\Omega)$ and $L > 0$ be such that

$$\begin{cases} \text{meas}\,(\Omega - \Omega_\varepsilon) \leq \varepsilon \\ 0 \leq \eta(x) \leq 1, \ \ \forall x \in \Omega \\ \eta(x) = 1, \ \ \forall x \in \Omega_\varepsilon \\ \left|D^k\eta(x)\right| \leq \frac{L}{\varepsilon^k}, \ \ \forall x \in \Omega - \Omega_\varepsilon \text{ and } \forall k = 1, ..., N. \end{cases} \tag{4.18}$$

Let us define a function $v : [0,1] \to \mathbb{R}^m$ in the following way: given $\delta > 0$, divide the interval $(0,1)$ into two finite unions I, J of disjoint open subintervals such that

$$\begin{cases} \overline{I} \cup \overline{J} = [0,1], \;\; I \cap J = \emptyset \\ \text{meas } I = t, \; \text{meas } J = 1-t \\ v^{(N)}(x_1) = \begin{cases} (1-t)\,\alpha & \text{if } x_1 \in I \\ -t\,\alpha & \text{if } x_1 \in J \end{cases} \\ \left| v^{(k)}(x_1) \right| \leq \delta, \;\; \forall x_1 \in (0,1), \;\; \forall k = 0,1,...,N-1. \end{cases}$$

In particular $v^{(N)}(x_1)$ can assume the two values $(1-t)\,\alpha$ and $-t\,\alpha$, and at the same time $v^{(k)}(x_1)$ can be small for every k up to order $N-1$, i.e. in absolute value less than or equal to δ, since 0 is a convex combination of the two values, with coefficients t and $1-t$.

We define u as a convex combination of $v + \varphi$ (by abuse of notations now the function $v : \mathbb{R}^n \to \mathbb{R}^m$ depends explicitly only on the first variable) and φ in the following way

$$u = \eta(v + \varphi) + (1-\eta)\varphi = \eta v + \varphi.$$

Choosing $\delta > 0$ sufficiently small (of the order ε^N), we find that u satisfies the conclusions of the lemma, with

$$\Omega_A = \{x \in \Omega_\varepsilon : \; x_1 \in I\}, \qquad \Omega_B = \{x \in \Omega_\varepsilon : \; x_1 \in J\}.$$

In fact $u \equiv \varphi$ near $\partial\Omega$ and for $k = 0,1,...,N-1$ we have, for every $x \in \Omega$,

$$\left|D^k u - D^k \varphi\right| \leq \sum_{l=0}^{k} \left|D^l \eta\right| \left|D^{k-l} v\right| \leq \sum_{l=0}^{k} \frac{L}{\varepsilon^l} \delta$$

and hence $\|u - \varphi\|_{N-1,\infty} \leq \varepsilon$. Since in Ω_ε we have $\eta \equiv 1$ we deduce that

$$D^N u = D^N v + D^N \varphi = D^N v + tA + (1-t)B = \begin{cases} A & \text{in } \Omega_A \\ B & \text{in } \Omega_B \end{cases}.$$

Finally it remains to show that

$$\text{dist}\left(D^N u(x), \text{co}\,\{A, B\}\right) \leq \varepsilon \qquad \text{a.e. in } \Omega.$$

We have that

$$D^N u = \eta D^N v + D^N \varphi + R\left(D^1\eta, ..., D^N\eta, D^0 v, ..., D^{N-1} v\right)$$

where (choosing δ smaller if necessary)

$$\left|R\left(D^1\eta, ..., D^N\eta, D^0 v, ..., D^{N-1} v\right)\right| \leq \gamma \sum_{l=1}^{N} \left|D^l \eta\right| \left|D^{N-l} v\right| \leq \varepsilon.$$

In the case $n, m \geq 1$ and $N = 1$ we have $R(D\eta, v) = v \otimes D\eta$, while when $m = 1$ and $N = 2$

$$R\left(D^1\eta, D^2\eta, D^0 v, D^1 v\right) = D^1 v \otimes D^1\eta + D^1\eta \otimes D^1 v + v\, D^2\eta.$$

Since both $D^N v + D^N \varphi$ ($= A$ or B) and $D^N \varphi = tA + (1-t)B$ belong to $\text{co}\,\{A, B\}$ we obtain that

$$\eta D^N v + D^N \varphi = \eta\left(D^N v + D^N \varphi\right) + (1-\eta)\, D^N \varphi \in \text{co}\,\{A, B\};$$

since the remaining term is arbitrarily small we deduce the result i.e.

$$\text{dist}\left(D^N u; \text{co}\,\{A, B\}\right) \leq \varepsilon.$$

Step 2: Let us assume now that $A-B$ is any matrix of rank one of $\mathbb{R}_s^{m\times n^N}$ and therefore it can be written as $A-B=\alpha\otimes v^{\otimes_N}$, i.e.

$$(A-B)^i_{j_1\dots j_N}=\alpha_i v_{j_1}\dots v_{j_N}$$

for a certain $\alpha\in\mathbb{R}^m$ and $v\in\mathbb{R}^n$ (v not necessarily e_1 as in Step 1). Replacing α by $|v|^N\,\alpha$ we can assume that $|v|=1$. We can then find $R=(r_{ij})\in SO(n)\subset\mathbb{R}^{n\times n}$ (i.e. a rotation) so that $v=e_1R$, and hence $e_1=vR^t$. We then set $\widetilde{\Omega}=R^t\Omega$ and for $1\le i\le m$, $1\le j_1,\dots,j_N\le n$ we let

$$\begin{aligned}\widetilde{A}^i_{j_1\dots j_N} &= \sum_{k_1,\dots,k_N=1}^{n} A^i_{k_1\dots k_N} r_{j_1k_1}\dots r_{j_Nk_N}\\ \widetilde{B}^i_{j_1\dots j_N} &= \sum_{k_1,\dots,k_N=1}^{n} B^i_{k_1\dots k_N} r_{j_1k_1}\dots r_{j_Nk_N}.\end{aligned}$$

For example if $n,m\ge 1$ and $N=1$ we have

$$\widetilde{A}^i_j=\sum_{k=1}^{n}A^i_k r_{jk}\ ,\quad \widetilde{B}^i_j=\sum_{k=1}^{n}B^i_k r_{jk}\ ,\ \text{ i.e. }\ \widetilde{A}=AR^t\ \text{ and }\ \widetilde{B}=BR^t.$$

While if $m=1$ and $N=2$ we have

$$\widetilde{A}_{j_1j_2}=\sum_{k_1,k_2=1}^{n}A_{k_1k_2}r_{j_1k_1}r_{j_2k_2}\ ,\ \widetilde{B}_{j_1j_2}=\sum_{k_1,k_2=1}^{n}B_{k_1k_2}r_{j_1k_1}r_{j_2k_2}\ ,$$

i.e. we get $\widetilde{A}=RAR^t$ and $\widetilde{B}=RBR^t$. We observe that by construction that

$$\widetilde{A}-\widetilde{B}=\alpha\otimes e_1^{\otimes_N}.$$

Indeed, since $e_1=vR^t$, we have

$$\begin{aligned}\left(\widetilde{A}-\widetilde{B}\right)^i_{j_1\dots j_N} &= \sum_{k_1,\dots,k_N=1}^{n}\alpha_i v_{k_1}\dots v_{k_N}r_{j_1k_1}\dots r_{j_Nk_N}\\ &= \alpha_i\sum_{k_1,\dots,k_N=1}^{n}(v_{k_1}r_{j_1k_1})\dots(v_{k_N}r_{j_Nk_N})\\ &= \alpha_i\,(e_1)_{j_1}\dots(e_1)_{j_N}.\end{aligned}$$

We can therefore apply Step 1 to $\widetilde{\Omega}$ and to $\widetilde{\varphi}(y)=\varphi(Ry)$ and find $\widetilde{\Omega}_{\widetilde{A}}$, $\widetilde{\Omega}_{\widetilde{B}}$ and $\widetilde{u}\in C^N_{piec}(\overline{\widetilde{\Omega}};\mathbb{R}^m)$ with the claimed properties. By setting

$$\begin{cases} u(x)=\widetilde{u}(R^tx), & x\in\Omega\\ \Omega_A=R\widetilde{\Omega}_{\widetilde{A}}, & \Omega_B=R\widetilde{\Omega}_{\widetilde{B}}\end{cases}$$

we get the result. ∎

4.4. Sufficient Conditions for the Relaxation Property. We will now give some conditions that can ensure the relaxation property (c.f. Definition 4.2), which is the main condition to prove existence of solutions. We give three types of results arranged by increasing order of difficulty. The first one will apply to the case of one quasiconvex function (c.f. Theorem 6.2).

THEOREM 4.5. *Let $E \subset \mathbb{R}^{m\times n^N}_s$ be closed and such that $\partial \operatorname{Rco} E \subset E$ and $\operatorname{Rco} E$ is bounded in at least one direction of rank one, then $\operatorname{Rco} E$ has the relaxation property with respect to E.*

REMARK 4.4. *By $\operatorname{Rco} E$ is bounded in at least one direction of rank one, we mean that there exists $\eta \in \mathbb{R}^{m\times n^N}_s$ with rank $\{\eta\} = 1$ such that, for every $\xi \in \operatorname{int} \operatorname{Rco} E$, there exist $t_1 < 0 < t_2$ with $\xi + t_1\eta$, $\xi + t_2\eta \in \partial \operatorname{Rco} E \subset E$.*

PROOF. The proof is elementary. Let $\xi \in \operatorname{int} \operatorname{Rco} E$ then by boundedness we can find $t_1 < 0 < t_2$ such that

$$\begin{cases} \xi_t = \xi + t\eta \in \operatorname{int} \operatorname{Rco} E, & \forall t \in (t_1, t_2) \\ \xi_{t_1}, \xi_{t_2} \in \partial \operatorname{Rco} E \subset E. \end{cases}$$

The approximation lemma (c.f. Lemma 4.4) (with $A = \xi_{t_1+\varepsilon}$ and $B = \xi_{t_2-\varepsilon}$ for ε small enough and $\xi = \frac{t_2-\varepsilon}{t_2-t_1-2\varepsilon}A + \frac{-(t_1+\varepsilon)}{t_2-t_1-2\varepsilon}B$) leads immediately to the result. ∎

We now consider the more difficult case (which corresponds to the case of systems or the case of one non quasiconvex equation) where $\partial \operatorname{Rco} E \nsubseteq E$. To handle this case we need to assume more structure on $\operatorname{Rco} E$ or more generally on a subset $K(E)$ of E (note that, to emphasize the dependence of K on E we will use the notation $K = K(E)$.

DEFINITION 4.3 (Approximation property). *Let $E \subset K(E) \subset \mathbb{R}^{m\times n^N}_s$. The sets E and $K(E)$ are said to have the approximation property if for every $\delta > 0$, there exist closed sets E_δ and $K(E_\delta)$ such that*

(i) $E_\delta \subset K(E_\delta) \subset \operatorname{int} K(E)$ for every $\delta > 0$

(ii) $\operatorname{dist}(\eta; E) < \delta$ for every $\eta \in E_\delta$

(iii) for every $\xi \in \operatorname{int} K(E)$ there exists $\delta = \delta(\xi) > 0$ such that $\xi \in K(E_\delta)$.

REMARK 4.5. *The above definition is similar to the so called in-approximation of convex integration (c.f. S. Muller and V. Sverak* [**44**], [**45**]*).*

THEOREM 4.6. *Let $E \subset \mathbb{R}^{m\times n^N}_s$ be compact and $\operatorname{Rco} E$ has the approximation property with $K(E_\delta) = \operatorname{Rco} E_\delta$; then $\operatorname{Rco} E$ has the relaxation property with respect to E.*

PROOF. The proof is a direct consequence of the next theorem. Since if $\xi \in \operatorname{Rco} E_\delta$ for some δ, we can find an integer $I = I(\xi)$ such that $\xi \in R_I \operatorname{co} E_\delta$. We have therefore with the notations of the following theorem

$$\begin{aligned} E^i_\delta &= R_i \operatorname{co} E_\delta, \ i = 1, ..., I \\ K(E_\delta) &= \operatorname{Rco} E_\delta \end{aligned}$$

and all the hypotheses of that theorem are satisfied by definition of $R_i \operatorname{co} E_\delta$. ∎

In some applications, such as the one on singular values, one may want to work not with the whole of $\operatorname{Rco} E$ but rather on a subset $K(E)$ which is not necessary rank one convex. The next theorem provides some answers to these types of problems.

THEOREM 4.7. *Let $E \subset K(E) \subset \mathbb{R}^{m\times n^N}_s$ be compact and have the approximation property. Let $\delta > 0$ and assume that for every $\xi \in K(E_\delta)$ (as in Definition 4.3) there exist an integer $I = I(\xi)$ and E^i_δ, $i = 0, 1, ..., I$, such that*

(i) $E^0_\delta = E_\delta \subset E^1_\delta \subset ... \subset E^I_\delta \subset K(E_\delta)$

(ii) $\xi \in E_\delta^I$

(iii) there exist $\xi_1, \xi_2 \in E_\delta^{I-1}$ *with* $\operatorname{rank}[\xi_1 - \xi_2] \leq 1$ *such that* $\xi \in [\xi_1, \xi_2] \subset K(E_\delta)$

(iv) for every $\eta \in E_\delta^i$, $i = 1, ..., I-1$, *there exist* $\eta_1, \eta_2 \in E_\delta^{i-1}$, $\operatorname{rank}[\eta_1 - \eta_2] \leq 1$, *such that* $\eta \in [\eta_1, \eta_2] \subset K(E_\delta)$.

Then $K(E)$ *has the relaxation property with respect to* E.

REMARK 4.6. *The four conditions presented in the above theorem capture the essential features of* $\operatorname{Rco} E$ *but are more general and thus more flexible. Examples of application to singular values will be given in the book* [**23**] *(see also* [**21**], [**24**]*).*

PROOF. Let $\varepsilon > 0$, we wish to show that for every $\xi \in \operatorname{int} K(E)$ and $\Omega \subset \mathbb{R}^n$ a bounded open set, we can find $u_\nu \in C_{piec}^N\left(\overline{\Omega}; \mathbb{R}^m\right)$ such that (defining φ by $D^N\varphi(x) = \xi$)

$$\begin{cases} D^\alpha u_\nu(x) = D^\alpha \varphi(x) \quad \text{on } \partial\Omega, \quad \alpha = 0, ..., N-1 \\ u_\nu \overset{*}{\rightharpoonup} \varphi \quad \text{in } W^{N,\infty} \\ D^N u_\nu(x) \in E \cup \operatorname{int} K(E), \quad \text{a.e in } \Omega \\ \int_\Omega \operatorname{dist}\left(D^N u_\nu(x); E\right) dx \leq \varepsilon \text{ as } \nu \to \infty. \end{cases}$$

Since $K(E)$ has the approximation property and is compact, we can find, for every $\delta > 0$ sufficiently small, a closed set E_δ, with $\xi \in K(E_\delta)$ and $K(E_\delta)$ is compactly contained in $\operatorname{int} K(E)$. It will be sufficient to find a C_{piec}^Nvector valued function u and an open set $\widetilde{\Omega} \subset \Omega$ so that

$$\begin{cases} \operatorname{meas}(\Omega - \widetilde{\Omega}) \leq \varepsilon \\ u \equiv \varphi \quad \text{near } \partial\Omega \\ \|u - \varphi\|_{N-1,\infty} \leq \varepsilon \\ \operatorname{dist}\left(D^N u(x); E_\delta\right) \leq \varepsilon, \quad \text{a.e. } x \in \widetilde{\Omega} \\ \operatorname{dist}\left(D^N u(x); K(E_\delta)\right) \leq \varepsilon, \quad \text{a.e. } x \in \Omega. \end{cases}$$

The fact that $K(E_\delta) \subset\subset \operatorname{int} K(E)$ and the last inequality imply that $D^N u(x) \in \operatorname{int} K(E)$ for ε sufficiently small. Finally since E_δ is close to E for δ sufficiently small we will have indeed obtained the theorem.

By hypothesis $\xi \in E_\delta^I$ for a certain I. We proceed by induction on I.

Step 1: We start with $I = 1$. We can therefore write

$$\xi = \lambda A + (1-\lambda)B, \ \operatorname{rank}\{A - B\} = 1, \quad A, B \in E_\delta.$$

We then use the approximation lemma (c.f. Lemma 4.4) to get the claimed result by setting $\widetilde{\Omega} = \Omega_A \cup \Omega_B$ and since $\operatorname{co}\{A, B\} = [A; B] \subset K(E_\delta)$ and hence for ε sufficiently small

$$\operatorname{dist}\left(D^N u(x); K(E_\delta)\right) \leq \varepsilon.$$

Step 2: For $I > 1$ we consider $\xi \in E_\delta^I$. Therefore there exist $A, B \in \mathbb{R}_s^{m \times n^N}$ such that

$$\begin{cases} \xi = \lambda A + (1-\lambda)B, \ \operatorname{rank}\{A - B\} = 1 \\ A, B \in E_\delta^{I-1}. \end{cases}$$

We then apply the approximation lemma (c.f. Lemma 4.4) and find that there exist a C^N_{piec} vector valued function v and Ω_A, Ω_B disjoint open sets such that

$$\begin{cases} \text{meas}\left(\Omega - (\Omega_A \cup \Omega_B)\right) \leq \varepsilon/2 \\ v \equiv \varphi \quad \text{near } \partial\Omega \\ \|v - \varphi\|_{N-1,\infty} \leq \varepsilon/2 \\ D^N v(x) = \begin{cases} A \text{ in } \Omega_A \\ B \text{ in } \Omega_B \end{cases} \\ \text{dist}\left(D^N v(x); K\left(E_\delta\right)\right) \leq \varepsilon, \quad \text{in } \Omega. \end{cases}$$

We now use the hypothesis of induction on Ω_A, Ω_B and A, B. We then can find $\widetilde{\Omega}_A, \widetilde{\Omega}_B$ $v_A \in C^N_{piec}$ in Ω_A, $v_B \in C^N_{piec}$ in Ω_B satisfying

$$\begin{cases} \text{meas}(\Omega_A - \widetilde{\Omega}_A), \ \text{meas}(\Omega_B - \widetilde{\Omega}_B) \leq \varepsilon/2 \\ v_A \equiv v \text{ near } \partial\Omega_A, \ v_B \equiv v \text{ near } \partial\Omega_B \\ \|v_A - v\|_{N-1,\infty} \leq \varepsilon/2 \text{ in } \overline{\Omega}_A, \ \|v_B - v\|_{N-1,\infty} \leq \varepsilon/2 \text{ in } \overline{\Omega}_B \\ \text{dist}(D^N v_A; E_\delta) \leq \varepsilon, \text{ a.e. in } \widetilde{\Omega}_A, \ \text{dist}(D^N v_B; E_\delta) \leq \varepsilon, \text{ a.e. in } \widetilde{\Omega}_B \\ \text{dist}(D^N v_A; K\left(E_\delta\right)) \leq \varepsilon, \text{ a.e. in } \Omega_A, \ \text{dist}(D^N v_B; K\left(E_\delta\right)) \leq \varepsilon, \text{ a.e. in } \Omega_B. \end{cases}$$

Letting $\widetilde{\Omega} = \widetilde{\Omega}_A \cup \widetilde{\Omega}_B$ and

$$u(x) = \begin{cases} v(x) & \text{in } \ \Omega - (\Omega_A \cup \Omega_B) \\ v_A(x) & \text{in } \ \Omega_A \\ v_B(x) & \text{in } \ \Omega_B \end{cases}$$

we have indeed obtained the result. ∎

5. Proofs of the Model Results

We start with the following

THEOREM 5.1. *Let $\Omega \subset \mathbb{R}^n$ be open. Let $F_i : \mathbb{R}^{m \times n^N}_s \to \mathbb{R}$, $i = 1, 2, ..., I$, be quasiconvex and locally Lipschitz functions and let*

$$E = \left\{ \xi \in \mathbb{R}^{m \times n^N}_s : F_i(\xi) = 0, \ \ i = 1, 2, ..., I \right\}.$$

Assume that $\text{Rco}\, E$ *is compact and strongly star shaped with respect to a fixed* $\xi_0 \in \text{int}\, \text{Rco}\, E$ *(i.e. for every* $\xi \in \text{Rco}\, E$ *and every* $t \in (0, 1]$ *then* $t\xi_0 + (1-t)\,\xi \in \text{int}\, \text{Rco}\, E$*). Let* $\varphi \in C^N_{piec}(\overline{\Omega}; \mathbb{R}^m)$ *satisfy*

$$D^N \varphi\,(x) \in E \cup \text{int}\, \text{Rco}\, E, \ \ a.e. \ x \in \Omega.$$

Then there exists (a dense set of) $u \in W^{N,\infty}(\Omega; \mathbb{R}^m)$ *such that*

$$\begin{cases} F_i(D^N u(x)) = 0, \quad a.e. \ x \in \Omega, \quad i = 1, ..., I \\ D^\alpha u(x) = D^\alpha \varphi(x), \quad x \in \partial\Omega, \quad \alpha = 0, ..., N-1. \end{cases}$$

PROOF. The result is a consequence of Corollary 4.3 when we set $K = K\,(E) = \text{Rco}\, E$. The only hypothesis that remains to be proved is that $\text{Rco}\, E$ has the relaxation property with respect to E. This will follow from Theorem 4.6 and from the fact that $\text{Rco}\, E$ is strongly star shaped.

We thus let for $\delta \in (0, 1]$

$$E_\delta = \delta\xi_0 + (1 - \delta)\, E$$

and observe that (by induction)

$$K(E_\delta) = \operatorname{Rco} E_\delta = \delta\xi_0 + (1-\delta)\operatorname{Rco} E \subset \operatorname{int}\operatorname{Rco} E, \quad \forall \delta \in (0,1].$$

Therefore E and $\operatorname{Rco} E$ have the approximation property (c.f. Definition 4.3) and hence Theorem 4.6 applies. ∎

Theorem 5.1 has as direct consequences Theorem 3.1 of Section 3.

PROOF. (Of Theorem 3.1) In order to apply Theorem 5.1 the only thing to be checked is that $\operatorname{Rco} E$ is strongly star shaped with respect to $0 \in \mathbb{R}^{m\times n^N}_s$. This is elementary. Indeed $0 \in \operatorname{int}\operatorname{Rco} E$, since $F_i(0) = 0$ for every $i = 1, ..., I$. Moreover we have for every $\xi \in \operatorname{Rco} E$ and $t \in (0,1]$

$$F_i((1-t)\,\xi) = (1-t)^{\alpha_i} F_i(\xi) \leq (1-t)^{\alpha_i} a_i < a_i, \quad i = 1, 2, ..., I;$$

thus the claimed result. ∎

COROLLARY 5.2. *Let $\Omega \subset \mathbb{R}^n$ be open. Let $F_i : \mathbb{R}^{m\times n^N}_s \to \mathbb{R}$, $i = 1, 2, ..., I$, be quasiconvex and locally Lipschitz functions and let*

$$E = \left\{\xi \in \mathbb{R}^{m\times n^N}_s \,:\, F_i(\xi) = 0, \quad i = 1, 2, ..., I\right\}.$$

Assume that $\operatorname{Rco} E$ is compact and $\operatorname{Rco} E = \operatorname{co} E$. Let $\varphi \in C^N_{piec}(\overline{\Omega}; \mathbb{R}^m)$ verify

$$D^N\varphi(x) \in E \cup \operatorname{int}\operatorname{Rco} E, \quad \text{a.e. } x \in \Omega$$

or $\varphi \in W^{N,\infty}(\Omega; \mathbb{R}^m)$ satisfy

$$D^N\varphi(x) \text{ compactly contained in } \operatorname{int}\operatorname{Rco} E, \quad \text{a.e. } x \in \Omega.$$

Then there exists (a dense set of) $u \in W^{N,\infty}(\Omega; \mathbb{R}^m)$ such that

$$\begin{cases} F_i(D^N u(x)) = 0, & \text{a.e. } x \in \Omega,\ i = 1, ..., I \\ D^\alpha u(x) = D^\alpha \varphi(x), & x \in \partial\Omega, \quad \alpha = 0, ..., N-1. \end{cases}$$

PROOF. We start by considering the case where $\varphi \in C^N_{piec}(\overline{\Omega}; \mathbb{R}^m)$. Assume that $\operatorname{int}\operatorname{Rco} E \neq \emptyset$, otherwise the corollary is trivial. So to apply Theorem 5.1 it is only necessary to show that $\operatorname{Rco} E$ is strongly star shaped with respect to any $\xi_0 \in \operatorname{int}\operatorname{Rco} E$. This is however a trivial property of convex sets.

The case $\varphi \in W^{N,\infty}(\Omega; \mathbb{R}^m)$ is deduced from the preceding one by applying an approximation Theorem that can be found in the appendix of [**23**]; this result allows to replace the boundary condition φ by $\psi \in C^N_{piec}(\overline{\Omega}; \mathbb{R}^m)$ with

$$D^N\psi(x) \in \operatorname{int}\operatorname{Rco} E, \quad \text{a.e. } x \in \Omega.$$

The corollary then follows. ∎

Theorem 3.2 is a direct consequence of the preceding Corollary 5.2. We now state an extension of Theorem 5.1, that can be used when dealing with the problem of *potential wells* (see [**21**], [**44**], [**45**]).

THEOREM 5.3. *Let $\Omega \subset \mathbb{R}^n$ be open. Let $F^\delta_i : \mathbb{R}^{m\times n^N}_s \to \mathbb{R}$, $i = 1, ..., I$, be quasiconvex, locally Lipschitz and continuous with respect to $\delta \in [0, \delta_0]$, for some $\delta_0 > 0$. Assume that*

$$\begin{aligned} (i) \quad & \operatorname{Rco}\left\{\xi \in \mathbb{R}^{m\times n^N}_s \,:\, F^\delta_i(\xi) = 0,\ i = 1, ..., I\right\} \\ & = \left\{\xi \in \mathbb{R}^{m\times n^N}_s \,:\, F^\delta_i(\xi) \leq 0,\ i = 1, ..., I\right\}, \quad \forall \delta \in [0, \delta_0] \end{aligned}$$

and it is compact;

$$(ii)\quad \left\{\xi \in \mathbb{R}^{m\times n^N}_s :\ F^\delta_i(\xi)=0,\ i=1,...,I\right\} \subset \left\{\xi \in \mathbb{R}^{m\times n^N}_s :\ F^0_i(\xi)<0,\ i=1,...,I\right\},\quad \forall\, 0<\delta\le\delta_0.$$

If $\varphi \in C^N_{piec}(\overline{\Omega};\mathbb{R}^m)$ *satisfies*

$$F^0_i(D^N\varphi(x))<0, \qquad \forall x\in\Omega,\ \ i=1,...,I,$$

then there exists (a dense set of) $u\in W^{N,\infty}(\Omega;\mathbb{R}^m)$ *such that*

$$\begin{cases} F^0_i(D^N u(x))=0, & a.e.\ x\in\Omega,\ \ i=1,...,I\\ D^\alpha u(x)=D^\alpha\varphi(x), & x\in\partial\Omega,\ \ \alpha=0,...,N-1.\end{cases}$$

Moreover u *is in the* C^{N-1} *closure of*

$$\left\{ \begin{array}{c} u\in C^N_{piec}\left(\overline{\Omega};\mathbb{R}^m\right): D^\alpha u=D^\alpha\varphi,\ \text{on } \partial\Omega,\ \alpha=0,...,N-1,\\ \text{and } F^0_i(D^N u(x))<0,\quad a.e \text{ in } \Omega \end{array} \right\}.$$

Proof. The proof is a direct consequence of the abstract results of the preceding section. Indeed let

$$\begin{aligned} E &= \left\{\xi \in \mathbb{R}^{m\times n^N}_s :\ F^0_i(\xi)=0,\ \ i=1,...,I\right\}\\ E_\delta &= \left\{\xi \in \mathbb{R}^{m\times n^N}_s :\ F^\delta_i(\xi)=0,\ \ i=1,...,I\right\}\\ K(E) &= \operatorname{Rco} E=\left\{\xi \in \mathbb{R}^{m\times n^N}_s :\ F^0_i(\xi)\le 0,\ \ i=1,...,I\right\}\\ K(E_\delta) &= \operatorname{Rco} E_\delta=\left\{\xi \in \mathbb{R}^{m\times n^N}_s :\ F^\delta_i(\xi)\le 0,\ \ i=1,...,I\right\}. \end{aligned}$$

Note that, since

$$\left\{\xi \in \mathbb{R}^{m\times n^N}_s :\ F^0_i(\xi)<0,\ \ i=1,...,I\right\}$$

is rank one convex and open, we have

$$\begin{aligned} &\operatorname{Rco}\left\{\xi \in \mathbb{R}^{m\times n^N}_s : F^0_i(\xi)<0,\ i=1,...,I\right\}\\ =\ &\left\{\xi \in \mathbb{R}^{m\times n^N}_s : F^0_i(\xi)<0,\ i=1,...,I\right\}\\ \subset\ &\operatorname{int}\left\{\xi \in \mathbb{R}^{m\times n^N}_s : F^0_i(\xi)\le 0,\ i=1,...,I\right\}=\operatorname{int}\operatorname{Rco}E. \end{aligned}$$

Applying therefore hypotheses (i) and (ii) we deduce that

$$\begin{aligned} K(E_\delta) &= \operatorname{Rco}E_\delta\subset\left\{\xi \in \mathbb{R}^{m\times n^N}_s : F^0_i(\xi)<0,\ i=1,...,I\right\}\\ &\subset \operatorname{int}\operatorname{Rco}E=\operatorname{int}K(E). \end{aligned}$$

Then $\operatorname{Rco}E$ has the relaxation property (since it has by construction the approximation property and therefore, by applying Theorem 4.6, it has necessarily the claimed property). Thus our result follows from Theorem 4.2. ∎

Now we turn our attention to the case with dependence on lower order terms. The first result that we will prove is Theorem 3.3 of Section 3.

PROOF. (Of Theorem 3.3) The claim follows from Theorem 6.1 that is stated below in Section 6; we only need to construct functions F_i^δ that satisfy all the hypotheses of this theorem. We therefore let for $\delta \in [0,1)$

$$F_i^\delta(x,s,\xi) = F_i(x,s,\xi) - (1-\delta)^{\alpha_i} a_i(x,s).$$

The first obvious claim is that

$$F_i^{\delta'}(x,s,\xi) < F_i^\delta(x,s,\xi), \quad \text{whenever } 0 \le \delta' < \delta < 1$$

which implies (ii) of Theorem 6.1. Therefore the only hypothesis that remains to be checked is that

$$\begin{aligned} &\operatorname{Rco}\left\{\xi \in \mathbb{R}_s^{m\times n^N} : \ F_i^\delta(x,s,\xi) = 0, \ \ i=1,...,I\right\} \\ = \ &\left\{\xi \in \mathbb{R}_s^{m\times n^N} : \ F_i^\delta(x,s,\xi) \le 0, \ \ i=1,...,I\right\}, \ \ \forall \delta \in [0,1). \end{aligned}$$

For $\delta = 0$ this is the assumption of the present theorem. Since (x,s) act only as parameters, we will drop below the dependence on these variables. We let

$$\begin{aligned} E_\delta &= \left\{\xi \in \mathbb{R}_s^{m\times n^N} : \ F_i^\delta(\xi) = 0, \ \ i=1,...,I\right\} \\ &= \left\{\xi \in \mathbb{R}_s^{m\times n^N} : \ F_i(\xi) = (1-\delta)^{\alpha_i} a_i, \ \ i=1,...,I\right\} \end{aligned}$$

and

$$E = \left\{\xi \in \mathbb{R}_s^{m\times n^N} : \ F_i(\xi) = a_i, \ i=1,...,I\right\}.$$

Observe that by homogeneity we have $E_\delta = (1-\delta)E$ and hence (by induction)

$$\begin{aligned} \operatorname{Rco} E_\delta &= (1-\delta)\operatorname{Rco} E \\ &= (1-\delta)\left\{\xi \in \mathbb{R}_s^{m\times n^N} : \ F_i(\xi) \le a_i, \ \ i=1,...,I\right\} \\ &= \left\{\xi \in \mathbb{R}_s^{m\times n^N} : \ F_i^\delta(\xi) \le 0, \ \ i=1,...,I\right\} \end{aligned}$$

which is (i) of Theorem 6.1. ∎

It remains to prove Theorem 3.4 of Section 3.

PROOF. (Of Theorem 3.4) We start by noticing that the case $\varphi \in W^{N,\infty}(\Omega;\mathbb{R}^m)$ follows from the case $\varphi \in C^N_{piec}(\overline{\Omega};\mathbb{R}^m)$ via the use of an approximation result of function in $W^{N,\infty}(\Omega;\mathbb{R}^m)$ by mean of piecewise smooth functions. This approximation result is proved in the appendix of the book [**23**]. We then proceed as in the proof of the above Theorem 3.3. We wish to find F_i^δ satisfying the hypotheses of Theorem 6.1 below. We let

$$F_i^\delta(x,s,\xi) = F_i(x,s,\frac{\xi}{1-\delta} - \frac{\delta}{1-\delta}\xi_0).$$

Since (x,s) are only parameters, we will drop below the dependence on these variables. We let

$$E = \left\{\xi \in \mathbb{R}_s^{m\times n^N} : \ F_i(\xi) = 0, \ i=1,...,I\right\}$$

and for $\delta \in (0,1)$ we have $E_\delta = \delta\xi_0 + (1-\delta)\, E$. Note that as above

$$\begin{aligned} \operatorname{Rco} E_\delta &= \delta\xi_0 + (1-\delta) \operatorname{Rco} E \\ &= \left\{\eta \in \mathbb{R}^{m\times n^N}_s : \eta = \delta\xi_0 + (1-\delta)\,\xi \text{ with } F_i(\xi) \le 0,\ i = 1,...,I\right\} \\ &= \left\{\eta \in \mathbb{R}^{m\times n^N}_s : F_i^\delta(\eta) \le 0,\ i = 1,...,I\right\} \end{aligned}$$

which is (i) of Theorem 6.1. Therefore the only thing that remains to be checked is that

$$\begin{aligned} E_\delta &= \left\{\xi \in \mathbb{R}^{m\times n^N}_s : F_i^\delta(\xi) = 0,\ i = 1,...,I\right\} \\ &\subset \left\{\xi \in \mathbb{R}^{m\times n^N}_s : F_i^{\delta'}(\xi) < 0,\ i = 1,...,I\right\},\ \forall\, 0 \le \delta' < \delta < 1. \end{aligned}$$

So let

$$t = \frac{1-\delta}{1-\delta'} \in (0,1)$$

and $\xi \in E_\delta$. Bearing in mind that $F_i(\xi_0) < 0$ and the convexity of F_i we have for every $i = 1,...,I$

$$\begin{aligned} 0 &= F_i^\delta(\xi) = tF_i^\delta(\xi) > tF_i^\delta(\xi) + (1-t)\, F_i(\xi_0) \\ &\ge tF_i\Big(\frac{\xi}{1-\delta} - \frac{\delta}{1-\delta}\xi_0\Big) + (1-t)\, F_i(\xi_0) \\ &\ge F_i\Big(\frac{t\xi}{1-\delta} + \left(1 - t - \frac{t\delta}{1-\delta}\right)\xi_0\Big) = F_i^{\delta'}(\xi) \end{aligned}$$

thus the result. ∎

6. Other Differential Problems with Lower Order Terms

The previous results generalize to the case with dependence on lower order terms. More precisely, a generalization of Theorem 5.3 to the case of explicit dependence on lower order terms is the following.

THEOREM 6.1 (Attainment for general systems). *Let $\Omega \subset \mathbb{R}^n$ be open. Let $F_i^\delta : \overline{\Omega} \times \mathbb{R}^{m\times M}_s \times \mathbb{R}^{m\times n^N}_s \to \mathbb{R}$, $F_i^\delta = F_i^\delta(x,s,\xi)$, $i = 1,...,I$, be quasiconvex and locally Lipschitz with respect to $\xi \in \mathbb{R}^{m\times n^N}_s$ and continuous with respect to $(x,s) \in \overline{\Omega} \times \mathbb{R}^{m\times M}_s$ and with respect to $\delta \in [0,\delta_0]$, for some $\delta_0 > 0$. Assume that, for every $(x,s) \in \overline{\Omega} \times \mathbb{R}^{m\times M}_s$,*

$$\begin{aligned} (i)\quad & \operatorname{Rco}\left\{\xi \in \mathbb{R}^{m\times n^N}_s : F_i^\delta(x,s,\xi) = 0,\ i = 1,...,I\right\} \\ & = \left\{\xi \in \mathbb{R}^{m\times n^N}_s : F_i^\delta(x,s,\xi) \le 0,\ i = 1,...,I\right\},\ \forall \delta \in [0,\delta_0] \end{aligned}$$

and it is bounded in $\mathbb{R}^{m\times n^N}_s$ uniformly with respect to (x,s) in a bounded set of $\overline{\Omega} \times \mathbb{R}^{m\times M}_s$;

$$\begin{aligned} (ii)\quad & \left\{\xi \in \mathbb{R}^{m\times n^N}_s : F_i^\delta(x,s,\xi) = 0,\ i = 1,...,I\right\} \\ & \subset \left\{\xi \in \mathbb{R}^{m\times n^N}_s : F_i^{\delta'}(x,s,\xi) < 0,\ i = 1,...,I\right\},\ \forall\, 0 \le \delta' < \delta \le \delta_0. \end{aligned}$$

If $\varphi \in C^N_{piec}(\overline{\Omega};\mathbb{R}^m)$ satisfies

$$F_i^0(x, D^{[N-1]}\varphi(x), D^N\varphi(x)) < 0, \quad \text{a.e. } x \in \Omega,\ i = 1,...,I,$$

then there exists (a dense set of) $u \in W^{N,\infty}(\Omega;\mathbb{R}^m)$ *such that*

$$\begin{cases} F_i^0(x, D^{[N-1]}u(x), D^N u(x)) = 0, & a.e.\ x \in \Omega,\ i = 1, ..., I \\ D^\alpha u(x) = D^\alpha \varphi(x), & x \in \partial\Omega,\ \alpha = 0, ..., N-1. \end{cases}$$

The proof of Theorem 6.1 is not simple from the technical point of view. However it follows the lines of a similar proof given by the authors in [**22**] for the second order case $N = 2$ (see also L. Poggiolini [**46**]). For this reason, and since we have here a limited room, we will not give in this paper the details of the proof of Theorem 6.1, but we refer to the book [**23**]. If $I = 1$ we obtain a simpler result as follows.

COROLLARY 6.2. *Let* $\Omega \subset \mathbb{R}^n$ *be open. Let* $F : \overline{\Omega} \times \mathbb{R}_s^{m\times M} \times \mathbb{R}_s^{m\times n^N} \to \mathbb{R}$ *be continuous and quasiconvex and locally Lipschitz in the last variable. Assume that* $\left\{\xi \in \mathbb{R}_s^{m\times n^N} : F(x, s, \xi) \le 0\right\}$ *is bounded in* $\mathbb{R}_s^{m\times n^N}$ *uniformly with respect to* (x, s) *in a bounded set of* $\overline{\Omega} \times \mathbb{R}_s^{m\times M}$. *If* $\varphi \in C^N_{piec}(\overline{\Omega};\mathbb{R}^m)$ *is such that*

$$F\left(x, D^{[N-1]}\varphi(x), D^N\varphi(x)\right) \le 0, \quad \forall x \in \Omega,$$

then there exists (a dense set of) $u \in W^{N,\infty}(\Omega;\mathbb{R}^m)$ *such that*

$$\begin{cases} F\left(x, D^{[N-1]}u(x), D^N u(x)\right) = 0, & a.e.\ x \in \Omega \\ D^\alpha u(x) = D^\alpha \varphi(x),\ x \in \partial\Omega, & \alpha = 0, ..., N-1. \end{cases}$$

PROOF. The proof follows from Theorem 6.1. Indeed let for $\delta \ge 0$

$$\begin{aligned} F^\delta(x, s, \xi) &= F(x, s, \xi) + \delta\,, \\ E_\delta &= \left\{\xi \in \mathbb{R}_s^{m\times n^N} : F^\delta(x, s, \xi) = 0\right\}. \end{aligned}$$

We then have $\mathrm{Rco}\, E_\delta = \left\{\xi \in \mathbb{R}_s^{m\times n^N} : F^\delta(x, s, \xi) \le 0\right\}$ and therefore all the hypotheses of the previous theorem are satisfied. The fact that the compatibility on the boundary datum is not given by a strict inequality is in this case acceptable, since we can choose $u = \varphi$ on the set where $F\left(x, D^{[N-1]}\varphi(x), D^N\varphi(x)\right) = 0$. ∎

REMARK 6.1. *(i) The required compactness can be weakened and it is sufficient to assume that there exists* $\eta \in \mathbb{R}_s^{m\times n^N}$ *with rank* $\{\eta\} = 1$ *such that* $F(x, s, \xi + t\eta) \to +\infty$ *as* $|t| \to \infty$, *for every* $(x, s, \xi) \in \Omega \times \mathbb{R}_s^{m\times M} \times \mathbb{R}_s^{m\times n^N}$.

(ii) Observe that the vectorial problem can here be obtained from the scalar one since $\varphi \in C^N_{piec}(\overline{\Omega};\mathbb{R}^m)$. *Indeed choosing the* $(m-1)$ *first components of* u *equal to the* $(m-1)$ *first components of* φ, *we would reduce the problem to a scalar one. If in addition* $N = 1$ *i.e. the first order case, the problem is then reduced to a convex scalar problem (since quasiconvexity of* F *implies convexity with respect to the last vector of* Du*).*

We give here an example of applications of the above theorems. The example is a scalar problem which is an Nth order version of the *eikonal equation.*

COROLLARY 6.3. *Let* $\Omega \subset \mathbb{R}^n$ *be open. Let* $a : \overline{\Omega} \times \mathbb{R}_s^M \to \mathbb{R}_+$ *be continuous and* $\varphi \in C^N_{piec}(\overline{\Omega})$ *satisfy*

$$\left|D^N\varphi(x)\right| \le a\left(x, D^{[N-1]}\varphi(x)\right), \quad a.e.\ x \in \Omega,$$

or $\varphi \in W^{N,\infty}(\Omega)$ *such that*

$$\left|D^N\varphi(x)\right| \leq a\left(x, D^{[N-1]}\varphi(x)\right) - \theta, \quad a.e.\ x \in \Omega,$$

for a certain $\theta > 0$. *Then there exists (a dense set of)* $u \in W^{N,\infty}(\Omega)$ *verifying*

$$\begin{cases} \left|D^N u(x)\right| = a\left(x, D^{[N-1]}u(x)\right), & a.e.\ x \in \Omega \\ D^\alpha u(x) = D^\alpha \varphi(x),\ x \in \partial\Omega, & \alpha = 0, ..., N-1. \end{cases}$$

References

[1] Acerbi E. and Fusco N. : *Semicontinuity problems in the calculus of variations*; Archive for Rational Mechanics and Analysis 86 (1984), 125-145.

[2] Ball J.M., Curie J.C. and Olver P.J. : *Null Lagrangians, weak continuity and variational problems of arbitrary order*; Journal of Functional Analysis, 41 (1981), 135-174.

[3] Bardi M. and Capuzzo Dolcetta I. : *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*; Birkhäuser (1997).

[4] Barles G. : *Solutions de viscosité des équations de Hamilton-Jacobi*; Mathématiques et Applications 17, Springer, Berlin (1994).

[5] Bellman R. : *Introduction to matrix analysis*; Mac Graw-Hill, New York (1960).

[6] Benton S.H. : *The Hamilton-Jacobi equation. A global approach*; Academic Press, New York (1977).

[7] Bressan A. and Flores F. : *On total differential inclusions*; Rend. Sem. Mat. Univ. Padova, 92 (1994), 9-16.

[8] Capuzzo Dolcetta I. and Evans L.C. : *Optimal switching for ordinary differential equations*; SIAM J. Optim. Control 22 (1988), 1133-1148.

[9] Capuzzo Dolcetta I. and Lions P.L. : *Viscosity solutions of Hamilton-Jacobi equations and state constraint problem*; Trans. Amer. Math. Soc. 318 (1990), 643-683.

[10] Celada P. and Perrotta S. : *Functions with prescribed singular values of the gradient*; preprint.

[11] Cellina A. : *On the differential inclusion* $x' \in \{-1, 1\}$; Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. 69 (1980), 1-6.

[12] Cellina A. and Perrotta S. : *On a problem of potential wells*; J. Convex Analysis 2 (1995), 103-115.

[13] Crandall M.G., Evans L.C. and Lions P.L. : *Some properties of viscosity solutions of Hamilton-Jacobi equations*; Trans. Amer. Math. Soc. 282 (1984), 487-502.

[14] Crandall M.G., Ishii H. and Lions P.L. : *User's guide to viscosity solutions of second order partial differential equations*; Bull. Amer. Math. Soc. 27 (1992), 1-67.

[15] Crandall M.G. and Lions P.L. : *Viscosity solutions of Hamilton-Jacobi equations*; Trans. Amer. Math. Soc. 277 (1983), 1-42.

[16] Dacorogna B. : *Direct methods in the calculus of variations*; Springer Verlag (1989).

[17] Dacorogna B. and Marcellini P. : *Existence of minimizers for non quasiconvex integrals*; Archive for Rational Mechanics and Analysis 131 (1995), 359-399.

[18] Dacorogna B. and Marcellini P. : *Théorème d'existence dans le cas scalaire et vectoriel pour les équations de Hamilton-Jacobi*; C.R. Acad. Sci. Paris 322, Série I, (1996), 237-240.

[19] Dacorogna B. and Marcellini P. : *Sur le problème de Cauchy-Dirichlet pour les systèmes d'équations non linéaires du premier ordre*; C.R. Acad. Sci. Paris 323, Série I, (1996), 599-602.

[20] Dacorogna B. and Marcellini P. : *General existence theorems for Hamilton-Jacobi equations in the scalar and vectorial case*; Acta Mathematica 178 (1997), 1-37.

[21] Dacorogna B. and Marcellini P. : *Cauchy-Dirichlet problem for first order nonlinear systems*; Journal of Functional Analysis 152 (1998), 404-446.

[22] Dacorogna B. and Marcellini P. : *Implicit second order partial differential equations*; Annali Scuola Normale Superiore di Pisa, 1998, to appear.

[23] Dacorogna B. and Marcellini P. : *Implicit partial differential equations*; Birkhäuser (1999), to appear.

[24] Dacorogna B. and Tanteri C. : *On the different convex hulls of sets involving singular values*; Proc. Royal Soc. Edinburgh, to appear.

[25] De Blasi F.S. and Pianigiani G. : *Non convex valued differential inclusions in Banach spaces*; J. Math. Anal. Appl. 157 (1991), 469-494.

[26] De Blasi F.S. and Pianigiani G. : *On the Dirichlet problem for Hamilton-Jacobi equations. A Baire category approach*; preprint (1997).
[27] Douglis A. : *The continuous dependence of generalized solutions of nonlinear partial differential equations upon initial data*; Comm. Pure Appl. Math. 14 (1961), 267-284.
[28] Fleming W.H. and Soner H.M. : *Controlled Markov processes and viscosity solutions; Applications of Mathematics*, Springer, Berlin (1993).
[29] Fonseca I. and Tartar L. : *The gradient theory of phase transition for systems with two potential wells*; Proc. Royal Soc. Edinburgh 111A (1989), 89-102.
[30] Frankowska H. : *Hamilton-Jacobi equations : viscosity solutions and generalized gradients*; J. Math. Anal. 141 (1989), 21-26.
[31] Fusco N. : *Quasi-convessità e semicontinuità per integrali multipli di ordine superiore*; Ricerche di Matematica, 29 (1980), 307-323.
[32] Gromov M. : *Partial differential relations*; Springer, Berlin (1986).
[33] Guidorzi M. and Poggiolini L. : *Lower semicontinuity for quasiconvex integrals of higher order*; NoDEA, to appear.
[34] Hopf E. : *Generalized solutions of nonlinear equations of first order*; J. Math. Mech. 14 (1965), 951-974.
[35] Ishii H. : *Perron's method for monotone systems of second order elliptic pdes*; Differential Integral Equations 5 (1992), 1-24.
[36] Kinderlehrer D. and Pedregal P. : *Remarks about the analysis of gradient Young measures*; edited by M. Miranda, Pitman Research Notes in Math. Longman 262 (1992), 125-150.
[37] Kohn R.V. and Strang G. : *Optimal design and relaxation of variational problems I, II, III*; Comm. Pure Appl. Math. 39 (1986), 113-137, 139-182, 353-377.
[38] Kruzkov S.N. : *Generalized solutions of Hamilton-Jacobi equation of eikonal type*; USSR Sbornik 27 (1975), 406-446.
[39] Lax P.D. : *Hyperbolic systems of conservation laws II*; Comm. Pure Appl. Math. 10 (1957), 537-566.
[40] Lions P.L. : *Generalized solutions of Hamilton-Jacobi equations*; Research Notes in Math. 69, Pitman, London (1982).
[41] Marcellini P. : *Approximation of quasiconvex functions and lower semicontinuity of multiple integrals*; Manuscripta Math. 51 (1985), 1-28.
[42] Meyers N.G. : *Quasiconvexity and the semicontinuity of multiple integrals*; Trans. Amer. Math. Soc., 119 (1965), 125-149.
[43] Morrey C.B. : *Multiple integrals in the calculus of variations*; Springer, Berlin (1966).
[44] Müller S. and Sverak V. : *Attainment results for the two-well problem by convex integration*; edited by J. Jost, International Press (1996), 239-251.
[45] Müller S. and Sverak V. : *Unexpected solutions of first and second order partial differential equations*; Proceedings of the International Congress of Mathematicians, Berlin 1998, Documents mathematica, vol II (1998), 691-702.
[46] Poggiolini L. : *Almost everywhere solutions of partial differential equations and systems of any order*; SIAM Journal on Mathematical Analysis, to appear.
[47] Spring D. : *Convex integration theory; Birkhäuser*, Basel (1998).
[48] Subbotin A.I. : *Generalized solutions of first order partial differential equations: the dynamical optimization perspective*; Birkhäuser, Boston (1995).
[49] Sychev M. A. : *Characterization of homogeneous scalar variational problems solvable for all boundary data*; Preprint CMU Pittsburgh, 97-203, October 1997.
[50] Zagatti S. : *Minimization of functionals of the gradient by Baire's theorem*; Preprint SISSA, 71/97, May 1997.
[51] Zhang K. : *On various semiconvex hulls in the calculus of variations*; Journal of Calculus of Variations and PDEs, 6 (1998), 143-160.

Departement de Mathematiques, Ecole Polytechnique Federale de Lausanne, 1015 Lausanne, Switzerland

E-mail address: Bernard.Dacorogna@epfl.ch

Dipartimento di Matematica "U. Dini", Universita' di Firenze, Viale Morgagni 67/A, 50134 Firenze, Italy

E-mail address: marcell@udini.math.unifi.it

Contemporary Mathematics
Volume **238**, 1999
B 0-8218-1196-7-03542-8

Genuinely Nonlinear Hyperbolic Systems of Two Conservation Laws

Constantine M. Dafermos

Abstract. This is an expository paper discussing the regularity and large time behavior of admissible BV solutions of genuinely nonlinear, strictly hyperbolic systems of two conservation laws. The approach will be via the theory of generalized characteristics.

1. Introduction

As is well-known, the theory of the Cauchy problem for nonlinear hyperbolic systems of conservation laws has to surmount numerous obstacles. The multi-space dimensional case is still terra incognita. Considerable progress has been made in one-space dimension, but the theory has yet to achieve definitive status. The source of the difficulty lies in that first derivatives of solutions starting out from even smooth and "small" initial values eventually blow up, triggering the development of jump discontinuities which propagate as shock waves. Thus, at best, only weak solutions may exist in the large. On the other hand, uniqueness generally fails within the class of weak solutions, so that extraneous "entropy" conditions have to be imposed in order to single out the admissible solution.

For strictly hyperbolic systems and initial data of small total variation, the random scheme [11;16] as well as front tracking algorithms [1;18] have successfully been employed for constructing admissible weak solutions in the class BV of functions of bounded variation. Furthermore, it has been established [3] that these solutions are L^1-stable, at least when the system is genuinely nonlinear. When the total variation of the initial data is large, even the L^∞ norm may blow up in finite time [13], so that the existence of even weak solutions is problematic.

Conditions are more favorable for genuinely nonlinear systems endowed with a coordinate system of Riemann invariants, in particular, systems of two conservation laws. In that case, the coupling between distinct characteristic families is weaker and, as a result of the spreading of rarefaction waves, even solutions starting out from initial values with unbounded total variation instantaneously acquire bounded variation. This remarkable property was first derived in the pioneering memoir [12], by appealing to the notion of approximate characteristics, within the framework of

1991 *Mathematics Subject Classification.* Primary: 35L65.
The author was supported by grants from the National Science Foundation and the Office of Naval Research.

the random choice scheme. The same technique was subsequently employed by several authors in order to study the local structure and the large time behavior of solutions.

The aim of this expository paper is to provide an outline of a comprehensive theory of genuinely nonlinear, strictly hyperbolic systems of two conservation laws, developed from a different standpoint: One is to consider, at the outset, an admissible solution in an appropriate function class and derive its properties, without regard to any particular method of construction. The principal tool in the investigation will be the theory of generalized characteristics [5]. The results will include bounds on the total variation of the trace of the solution along any space-like curve, as well as a description of local structure and large time asymptotics of solutions under initial data in L^1, of compact support, or periodic. The above shall be reported here without demonstration; detailed proofs will be presented in Chapter XII of the forthcoming book [7] by the author. Proofs obtained under stronger a priori restrictions on the function class of solutions, have appeared in [5;6;19].

The author thanks the Editors, Gui-Qiang Chen and Emmanuele DiBenedetto, for giving him the opportunity to announce these results here.

2. BV Solutions and Generalized Characteristics

Consider a genuinely nonlinear, strictly hyperbolic system of two conservation laws:

$$\partial_t U(x,t) + \partial_x F(U(x,t)) = 0. \tag{2.1}$$

Thus U takes values in $\mathbb{R}^2$ and F is a given, smooth map from $\mathbb{R}^2$ to $\mathbb{R}^2$ such that, for any $U \in \mathbb{R}^2$, the Jacobian matrix $DF(U)$ has real distinct eigenvalues $\lambda(U) < \mu(U)$ associated with linearly independent eigenvectors $R(U), S(U)$, which satisfy

$$D\lambda(U)R(U) < 0\ , \qquad D\mu(U)S(U) > 0\,. \tag{2.2}$$

The system is endowed with a coordinate system of Riemann invariants (z,w), normalized by

$$DzR = 1\ ,\ \ DzS = 0\,,\ \ DwR = 0\,,\ \ \ DwS = 1. \tag{2.3}$$

By taking composition with the inverse of the local diffeomorphism $U \mapsto (z,w)$, one may realize functions of U as functions of (z,w); for economy in notation, the same symbol shall be employed to denote both representations.

It is assumed, further, that the system has the Glimm-Lax interaction property, namely the collision of any two shocks of the same family produces a shock of that family together with a rarefaction wave of the opposite family. This condition is here expressed by

$$S^T D^2 z S > 0\ , \qquad R^T D^2 w R > 0\,. \tag{2.4}$$

The normalization (2.3) in conjunction with the direction of the inequalities (2.2) and (2.4) imply that z increases across admissible weak 1-shocks and 2-shocks while w decreases across admissible weak 1-shocks and 2-shocks.

We now assume that $U(x,t)$ is a weak solution of (2.1) on $(-\infty,\infty)\times[0,\infty)$, which is a bounded measurable function of class $BV_{\rm loc}$. In particular, $(-\infty,\infty)\times[0,\infty) = \mathcal{C}\bigcup\mathcal{J}\bigcup\mathcal{I}$, where $\mathcal{C}$ is the set of points of approximate continuity of U, $\mathcal{J}$ is the shock set of U and $\mathcal{I}$ is the set of irregular points of U. The one-dimensional

Hausdorff measure of $\mathcal{I}$ is zero. $\mathcal{J}$ is essentially covered by the countable union of C^1 arcs. With any $(\overline{x},\overline{t}) \in \mathcal{J}$ are associated one-sided approximate limits $U^{\pm}$ and a tangent line of slope (shock speed) s, which are related through the Rankine-Hugoniot jump condition

$$F(U^+) - F(U^-) = s[U^+ - U^-]\,. \tag{2.5}$$

We will be assuming that the solution U has sufficiently small oscillation, in which case (2.5) implies that s must be close to either λ or μ. This allows us to classify each shock point as belonging to the first or the second characteristic family. We then assume that the solution satisfies the Lax E-condition, namely

$$\lambda(U^+) < s < \lambda(U^-)\ , \tag{2.6}_1$$

$$\mu(U^+) < s < \mu(U^-)\ , \tag{2.6}_2$$

for 1-shocks or 2-shocks, respectively.

For convenience, we normalize U by requiring that it assumes at any point $(\overline{x},\overline{t}) \in \mathcal{C}$ the approximate value U^0 at that point, $U(\overline{x},\overline{t}) = U^0$. Furthermore, we extend $U^{\pm}$ from $\mathcal{J}$ to $\mathcal{J}\bigcup\mathcal{C}$ by setting $U^+ = U^- = U^0$ at any $(\overline{x},\overline{t}) \in \mathcal{C}$.

Characteristics of the first or second family, associated with a classical, Lipschitz continuous, solution U of (2.1) are integral curves of the ordinary differential equations

$$\frac{dx}{dt} = \lambda(U(x,t))\ , \tag{2.7}_1$$

or

$$\frac{dx}{dt} = \mu(U(x,t)). \tag{2.7}_2$$

Extending this notion, we associate characteristics with weak solutions in the function class discussed above by adopting the same definition, except that now (2.7) have to be interpreted as generalized ordinary differential equations, in the sense of Filippov [10]:

Definition 2.1. A *generalized characteristic* of the first or second family on the time interval $[\sigma,\tau] \subset [0,\infty)$, associated with the weak solution U, is a Lipschitz curve $\xi : [\sigma,\tau] \to (-\infty,\infty)$ which satisfies the differential inclusion

$$\dot{\xi} \in [\lambda(U^+)\,,\ \lambda(U^-)]\ , \tag{2.8}_1$$

or

$$\dot{\xi} \in [\mu(U^+)\,,\ \mu(U^-)]\ , \tag{2.8}_2$$

almost everywhere on $[\sigma,\tau]$.

In particular, shocks of either family are generalized characteristics of that family.

By standard theory of differential inclusions, through any fixed point $(\overline{x},\overline{t}) \in (-\infty,\infty)\times[0,\infty)$ pass two (not necessarily distinct) generalized characteristics of each family, associated with U and defined on $[0,\infty)$, namely the *minimal* $\xi_-(\cdot)$ and the *maximal* $\xi_+(\cdot)$, with $\xi_-(t) \le \xi_+(t)$ for $t \in [0,\infty)$. The funnel-shaped region confined between the graphs of ξ_- and ξ_+ comprises the set of points (x,t) that may be connected to $(\overline{x},\overline{t})$ by a generalized characteristic of that family. The extremal

backward characteristics will be playing a pivotal role throughout the paper. Their first important property is that they propagate with classical characteristic speed:

Theorem 2.1. *Let $\xi(\cdot)$ denote any of the four extremal backward characteristics emanating from some point $(\overline{x},\overline{t})$ of the upper half-plane. Then $(\xi(t),t)\in\mathcal{C}$ for almost all $t\in[0,\overline{t}]$. In particular, for almost all $t\in[0,\overline{t}]$,*

$$\dot{\xi}=\lambda(U^{\pm})\ , \tag{2.9}_1$$

if ξ is a 1-characteristic, or

$$\dot{\xi}=\mu(U^{\pm})\ , \tag{2.9}_2$$

if ξ is a 2-characteristic.

The extremal backward characteristics mark the paths of signals travelling with extremal speed and may thus be employed in order to characterize space-like curves:

Definition 2.2. A Lipschitz curve, with graph $\mathcal{A}$ embedded in the upper half-plane, is called *space-like* relative to U when every point $(\overline{x},\overline{t})\in\mathcal{A}$ has the following property: The set $\{(x,t):0\le t<\overline{t},\ \zeta(t)<x<\xi(t)\}$ of points confined between the maximal backward 2-characteristic ζ and the minimal backward 1-characteristic ξ, emanating from $(\overline{x},\overline{t})$, has empty intersection with $\mathcal{A}$.

Clearly, any generalized characteristic, of either family, associated with U, is space-like relative to U. Similarly, all time-lines, t=constant, are space-like.

We now impose the following *structural condition* on our solution U: The traces of the Riemann invariants (z,w) along any space-like curve are functions of (locally) bounded variation.

The justification for the above assumption shall be provided, a posteriori, in Section 3, where such bounds on the variation will indeed be established. In fact, only part of the assumption is necessary in the analysis: The condition need only be tested for special space-like curves, namely generalized characteristics and time-lines. It should also be noted that, as shown in [4], any solution satisfying the structural condition must necessarily coincide with the solution with the same initial data constructed by either the random choice method or the front tracking algorithm.

In consequence of the structural condition, one-sided limits $U(x\pm,t)$ exist for all $-\infty<x<\infty,\ t>0$, and $(x,t)\in\mathcal{C}$ implies $U(x-,t)=U(x+,t)=U(x,t)$ while $(x,t)\in\mathcal{J}$ implies $U(x-,t)=U^-,U(x+,t)=U^+$.

If U were a classical, Lipschitz continuous, solution of (2.1), then the trace of z along any 1-characteristic and the trace of w along any 2-characteristic would be constant. On the other hand, if U were a piecewise smooth admissible solution, then the trace of z along classical 1-characteristics and the trace of w along classical 2-characteristics would be step functions, with jumps at the points where the characteristic crosses shocks of the opposite family. Moreover, by classical theory, the sign of the jump would be fixed and the strength of the jump would be of cubic order in the strength of the crossed shock. It is interesting that the above essentially hold even in the context of weak solutions, for the extremal backward characteristics:

Theorem 2.2. *Let ξ be the minimal (or maximal) backward 1-characteristic (or 2-characteristic) emanating from any fixed point $(\overline{x},\overline{t})$ of the upper half-plane. Set*

$$\overline{z}(t) = z(\xi(t)-,t)\ ,\quad \overline{w}(t) = w(\xi(t)+,t)\ ,\quad 0\le t\le \overline{t}. \tag{2.10}$$

Then $\overline{z}(\cdot)$ (or $\overline{w}(\cdot)$) is a nonincreasing saltus function whose variation is concentrated in the set of jump points of $\overline{w}(\cdot)$ (or $\overline{z}(\cdot)$). Furthermore, if $\tau\in(0,\overline{t})$ is any point of jump discontinuity of $\overline{z}(\cdot)$ (or $\overline{w}(\cdot)$), then

$$\overline{z}(\tau-)-\overline{z}(\tau+)\le a[\overline{w}(\tau+)-\overline{w}(\tau)]^3\ , \tag{2.11}_1$$

or

$$\overline{w}(\tau-)-\overline{w}(\tau+)\le a[\overline{z}(\tau+)-\overline{z}(\tau)]^3\ , \tag{2.11}_2$$

where a is a constant depending solely on F.

The proof of the above theorem is based on estimates induced by entropy inequalities and is quite lengthy. It is given in [7, Ch. XII]. For earlier proofs, requiring more restrictive structural conditions on U, see [5;6].

Since we are operating in the realm of solutions with small oscillation, (2.11) imply that z and w are nearly constant along the extremal backward characteristics of the corresponding family. From the perspective of the present approach, it is this property that induces the decoupling of the two characteristic families and thereby all the distinctive properties of solutions of our system that will be presented in following sections.

3. Bounds on the Variation

A priori bounds are reported here on admissible weak solutions U of (2.1) of class $BV_{\rm loc}$, with small oscillation, which satisfy the structural condition laid down in Section 2. They are similar to the estimates derived in [12], in the context of the random choice scheme. The proofs are found in [7, Ch. XII] or, under somewhat stronger assumptions on U, in [19].

The solution is conveniently monitored through its Riemann invariants (z,w). The oscillation is controlled by a small positive constant δ:

$$|z(x,t)|+|w(x,t)|<2\delta\ ,\quad -\infty<x<\infty\ ,\quad 0<t<\infty\ . \tag{3.1}$$

The first set of estimates depends on the initial data. We assume

$$\sup_{(-\infty,\infty)}|z(\cdot,0)|+\sup_{(-\infty,\infty)}|w(\cdot,0)|\le\delta\ , \tag{3.2}$$

$$TV_{(-\infty,\infty)}z(\cdot,0)+TV_{(-\infty,\infty)}w(\cdot,0)<b\delta^{-1}\ , \tag{3.3}$$

where b is a fixed, small constant, independent of δ. Thus, there is a tradeoff, allowing for arbitrarily large total variation at the expense of keeping the oscillation sufficiently small.

Theorem 3.1. *Consider any space-like curve $t=t^*(x)$, $x_\ell\le x\le x_r$, in the upper half-plane, along which the trace of (z,w) is denoted by (z^*,w^*). Then*

$$\begin{aligned} TV_{[x_\ell,x_r]}z^*(\cdot)\le{}& TV_{[\xi_\ell(0),\xi_r(0)]}z(\cdot,0)\\ &+c\delta^2\{TV_{[\zeta_\ell(0),\xi_r(0)]}z(\cdot,0)+TV_{[\zeta_\ell(0),\xi_r(0)]}w(\cdot,0)\}\ , \end{aligned} \tag{3.4}_1$$

$$TV_{[x_\ell,x_r]}w^*(\cdot)\le TV_{[\zeta_\ell(0),\zeta_r(0)]}w(\cdot,0) \tag{3.4}_2$$

$$+c\delta^2\{TV_{[\zeta_\ell(0),\xi_r(0)]}z(\cdot,0)+TV_{[\zeta_\ell(0),\xi_r(0)]}w(\cdot,0)\},$$

where $\xi_\ell(\cdot),\xi_r(\cdot)$ *are the minimal backward* 1*-characteristics and* $\zeta_\ell(\cdot),\zeta_r(\cdot)$ *are the maximal backward* 2*-characteristics emanating from the end-points* (x_ℓ,t_ℓ) *and* (x_r,t_r) *of the graph of* $t^*(\cdot)$.

The estimates (3.4) reflect the fact that z and w are nearly constant along minimal backward 1-characteristics and maximal backward 2-characteristics, respectively. Indeed, we notice that the effect of the coupling of the two characteristic families is $O(\delta^2)$.

Since generalized characteristics are space-like curves, one may combine Theorems 2.2 and 3.1 to deduce the following corollary:

Theorem 3.2. *For any point* (x,t) *of the upper half-plane:*

$$(3.5)_1 \qquad \sup_{(-\infty,\infty)} z(\cdot,0)\geq z(x,t)\geq \inf_{(-\infty,\infty)} z(\cdot,0)-cb\delta\ ,$$

$$(3.5)_2 \qquad \sup_{(-\infty,\infty)} w(\cdot,0)\geq w(x,t)\geq \inf_{(-\infty,\infty)} w(\cdot,0)-cb\delta\,.$$

Thus, on account of (3.2) and by selecting b sufficiently small, we secure a posteriori that the solution will satisfy (3.1).

Due to the spreading of rarefaction waves, solutions acquire instantaneously bounded variation, independently of the initial data. This is reflected in the following proposition, which applies to any solution with small oscillation (3.1), without any assumptions on the initial data:

Theorem 3.3. *For any* $-\infty<x<y<\infty$ *and* $t>0$,

$$(3.6) \qquad TV_{[x,y]}z(\cdot,t)+TV_{[x,y]}w(\cdot,t)\leq \beta\frac{y-x}{t}+\gamma\delta\ ,$$

where β *and* γ *are constants that may depend on* F *but are independent of the initial data.*

The oscillation of the solution is also controlled by just the oscillation, and not the variation, of the initial data:

Theorem 3.4. *There is a positive constant* κ*, depending solely on* F*, such that solutions generated by initial data with small oscillation*

$$(3.7) \qquad |z(x,0)|+|w(x,0)|<\kappa\delta^2\ ,\qquad -\infty<x<\infty\ ,$$

but unrestricted, possibly infinite, total variation, satisfy (3.1).

4. Regularity of Solutions

The invariance of the system (2.1) under uniform stretching of the space-time variables suggests that, in the vicinity of any fixed point $(\overline{x},\overline{t})$ of the upper half-plane, the solution U should behave like a self-similar solution relative to that point: In the most general situation, shocks and/or centered compression waves converge and collide at $(\overline{x},\overline{t})$ to produce a jump discontinuity which is then resolved into an outgoing fan of shocks and/or rarefaction waves, corresponding to the solution of a Riemann problem. Indeed, such behavior has been established in [9], for solutions constructed by the random choice scheme. See also [2]. Similar results will be reported here for our solution U, which satisfies the conditions laid down in Section 2. The proofs, found in [7, Ch. XII], rely heavily on Theorem 2.2.

With any fixed point $(\overline{x},\overline{t})$ of the upper half-plane, we associate eight generalized characteristics emanating from it, namely, four backward, $\xi_-,\xi_+,\zeta_-,\zeta_+$, and four forward, $\phi_-,\phi_+,\psi_-,\psi_+$, determined as follows: ξ_- is the minimal backward 1-characteristic, ξ_+ is the maximal backward 1-characteristic, ζ_- is the minimal backward 2-characteristic, ζ_+ is the maximal backward 2-characteristic, ϕ_+ is the maximal forward 1-characteristic, and ψ_- is the minimal forward 2-characteristic. For $t>\overline{t}$, $\phi_-(t)$ is identified by the property that the minimal backward 1-characteristic ξ emanating from any point (x,t) is intercepted by the $\overline{t}$-time line at $\xi(\overline{t})$, with $\xi(\overline{t})<\overline{x}$ if $x<\phi_-(t)$ and $\xi(\overline{t})\geq\overline{x}$ if $x\geq\phi_-(t)$. Similarly, $\phi_+(t)$ is identified by the property that the maximal backward 2-characteristic ζ emanating from any point (x,t) is intercepted by the $\overline{t}$-time line at $\zeta(\overline{t})$, with $\zeta(\overline{t})>\overline{x}$ if $x>\psi_+(t)$ and $\zeta(\overline{t})\leq\overline{x}$ if $x\leq\psi_+(t)$. Of course, the above eight characteristics are not necessarily distinct: we may have coincidence of ξ_- with ξ_+,ζ_- with ζ_+,ϕ_- with ϕ_+, and/or ψ_- with ψ_+.

The characteristics $\xi_-,\xi_+,\zeta_-,\zeta_+,\phi_-,\phi_+,\psi_-$ and ψ_+ border regions

$$\mathcal{S}_W=\{(x,t):x<\overline{x},\ \zeta_-^{-1}=(x)<t<\phi_-^{-1}(x)\}\ , \tag{4.1}$$

$$\mathcal{S}_E=\{(x,t):x>\overline{x},\ =\xi_+^{-1}(x)<t<\psi_+^{-1}(x)\}\ , \tag{4.2}$$

$$\mathcal{S}_N=\{(x,t):t>\overline{t},\ \phi_+(t)<x<\psi_-(t)\}\ , \tag{4.3}$$

$$\mathcal{S}_S=\{(x,t):t<\overline{t},\ \zeta_+(t)<x<\xi_-(t)\}. \tag{4.4}$$

Theorem 4.1. *The solution U, with Riemann invariants (z,w), has the following properties, at any fixed point $(\overline{x},\overline{t})$ of the upper half-plane:*

(a) *As (x,t) tends to $(\overline{x},\overline{t})$ through any one of the four regions $\mathcal{S}_W,\mathcal{S}_E,\mathcal{S}_N$ or $\mathcal{S}_S$ $(z(x,t),w(x,t))$ tend to respective limits $(z_W,w_W),(z_E,w_E),(z_N,w_N)$ or (z_S,w_S). In particular $z_W=z(\overline{x}-,\overline{t}),w_W=w(\overline{x}-,\overline{t})$, $z_E=z(\overline{x}+,\overline{t}),w_E=w(\overline{x}+,\overline{t})$.*

(b)$_1$ *If $p_\ell(\cdot)$ and $p_r(\cdot)$ are any two backward 1-characteristics emanating from $(\overline{x},\overline{t})$, with $\xi_-(t)\leq p_\ell(t)<p_r(t)\leq\xi_+(t)$, for $t<\overline{t}$, then*

$$\begin{aligned}z_S=\lim_{t\uparrow\overline{t}}z(\xi_-(t)\pm,t)\leq\lim_{t\uparrow\overline{t}}z(p_\ell(t)-,t)\leq\lim_{t\uparrow\overline{t}}z(p_\ell(t)+,t)\\ \leq\lim_{t\uparrow\overline{t}}z(p_r(t)-,t)\leq\lim_{t\uparrow\overline{t}}z(p_r(t)+,t)\leq\lim_{t\uparrow\overline{t}}z(\xi_+(t)\pm,t)=z_E\ ,\end{aligned} \tag{4.5}$$

$$\begin{aligned}w_S=\lim_{t\uparrow\overline{t}}w(\xi_-(t)\pm,t)\geq\lim_{t\uparrow\overline{t}}w(p_\ell(t)-,t)\geq\lim_{t\uparrow\overline{t}}w(p_\ell(t)+,t)\\ \geq\lim_{t\uparrow\overline{t}}w(p_r(t)-,t)\geq\lim_{t\uparrow\overline{t}}w(p_r(t)+,t)\geq\lim_{t\uparrow\overline{t}}w(\xi_+(t)\pm,t)=w_E.\end{aligned} \tag{4.6}$$

(b)$_2$ *If $q_\ell(\cdot)$ and $q_r(\cdot)$ are any two backward 2-characteristics emanating from $(\overline{x},\overline{t})$, with $\zeta_-(t)\leq q_\ell(t)<q_r(t)\leq\zeta_+(t)$, for $t<\overline{t}$, then*

$$\begin{aligned}w_W=\lim_{t\uparrow\overline{t}}w(\zeta_-(t)\pm,t)\geq\lim_{t\uparrow\overline{t}}w(q_\ell(t)-,t)\geq\lim_{t\uparrow\overline{t}}w(q_\ell(t)+,t)\\ \geq\lim_{t\uparrow\overline{t}}w(q_r(t)-,t)\geq\lim_{t\uparrow\overline{t}}w(q_r(t)+,t)\geq\lim_{t\uparrow\overline{t}}w(\zeta_+(t)\pm,t)=w_S\ ,\end{aligned} \tag{4.5}$$

$$\begin{aligned}z_W=\lim_{t\uparrow\overline{t}}z(\zeta_-(t)\pm,t)\leq\lim_{t\uparrow\overline{t}}z(q_\ell(t)-,t)\leq\lim_{t\uparrow\overline{t}}z(q_\ell(t)+,t)\\ \leq\lim_{t\uparrow\overline{t}}z(q_r(t)-,t)\leq\lim_{t\uparrow\overline{t}}z(q_r(t)+,t)\leq\lim_{t\uparrow\overline{t}}z(\zeta_+(t)\pm,t)=z_S.\end{aligned} \tag{4.6}$$

(c)$_1$ *If* $\phi_-(t) = \phi_+(t)$, *for* $\overline{t} < t < \overline{t} + s$, *then* $z_W \le z_N, w_W \ge w_N$. *On the other hand, if* $\phi_-(t) < \phi_+(t)$, *for* $\overline{t} < t < \overline{t} + s$, *then, as* (x,t) *tends to* $(\overline{x},\overline{t})$ *through the region* $\{(x,t) : t > \overline{t},\ \phi_-(t) < x < \phi_+(t)\}$, $w(x,t)$ *tends to* w_W. *Furthermore, if* $p_\ell(\cdot)$ *and* $p_r(\cdot)$ *are any two forward* 1*-characteristics issuing from* $(\overline{x},\overline{t})$, *with* $\phi_-(t) \le p_\ell(t) \le p_r(t) \le \phi_+(t)$, *for* $\overline{t} < t < \overline{t} + s$, *then*

$$(4.7)_1 \qquad z_W = \lim_{t\downarrow\overline{t}} z(\phi_-(t)\pm,t) \ge \lim_{t\downarrow\overline{t}} z(p_\ell(t)-,t) = \lim_{t\downarrow\overline{t}} z(p_\ell(t)+,t)$$

$$\ge \lim_{t\downarrow\overline{t}} z(p_r(t)-,t) = \lim_{t\downarrow\overline{t}} z(p_r(t)+,t) \ge \lim_{t\downarrow\overline{t}} z(\phi_+(t)\pm,t) = z_N.$$

(c)$_2$ *If* $\psi_-(t) = \psi_+(t)$, *for* $\overline{t} < t < \overline{t} + s$, *then* $w_N \ge w_E,\ z_N \le z_E$. *On the other hand, if* $\psi_-(t) < \psi_+(t)$, *for* $\overline{t} < t < \overline{t} + s$, *then, as* (x,t) *tends to* $(\overline{x},\overline{t})$ *through the region* $\{(x,t) : t > \overline{t},\ \psi_-(t) < x < \psi_+(t)\}$, $z(x,t)$ *tends to* z_E. *Furthermore, if* $q_\ell(\cdot)$ *and* $q_r(\cdot)$ *are any two forward* 2*-characteristics issuing from* $(\overline{x},\overline{t})$, *with* $\psi_-(t) \le q_\ell(t) \le q_r(t) \le \psi_+(t)$, *for* $\overline{t} < t < \overline{t} + s$, *then*

$$(4.7)_2 \qquad w_N = \lim_{t\downarrow\overline{t}} w(\psi_-(t)\pm,t) \le \lim_{t\downarrow\overline{t}} w(q_\ell(t)-,t) = \lim_{t\downarrow\overline{t}} w(q_\ell(t)+,t)$$

$$\le \lim_{t\downarrow\overline{t}} w(q_r(t)-,t) = \lim_{t\downarrow\overline{t}} w(q_r(t)+,t) \le \lim_{t\downarrow\overline{t}} w(\psi_+(t)\pm,t) = w_E.$$

Statements (b)$_1$ and (b)$_2$ regulate the incoming waves, allowing for any combination of admissible shocks and focussing compression waves. Statements (c)$_1$ and (c)$_2$ characterize the outgoing wave fan. In particular, (c)$_1$ implies that the state (z_W, w_W), on the left, may be joined with the state (z_N, w_N), on the right, by a 1-rarefaction wave or admissible 1-shock; while (c)$_2$ imples that the state (z_N, w_N), on the left, may be joined with the state (z_E, w_E), on the right, by a 2-rarefaction wave or admissible 2-shock. Thus, the outgoing wave fan is locally approximated by the solution of the Riemann problem with end-states $(z(\overline{x}-,\overline{t}), w(\overline{x}-,\overline{t}))$ and $(z(\overline{x}+,\overline{t}), w(\overline{x}+,\overline{t}))$.

In Section 2 we noted that membership in $BV_{\rm loc}$ endows the solution U with certain regularity. This is now improved, in consequence of Theorem 4.1: Any point $(\overline{x},\overline{t}) \in \mathcal{C}$ of approximate continuity is actually a point of continuity, characterized by the property that the four states $(z_W, w_W), (z_E, w_E), (z_N, w_N)$ and (z_S, w_S) coincide. Similarly, any point $(\overline{x},\overline{t}) \in \mathcal{J}$ of the shock set is a point of jump discontinuity, characterized by either $(z_W, w_W) = (z_S, w_S) \ne (z_E, w_E) = (z_N, w_N)$, for 1-shocks, or $(z_W, w_W) = (z_N, w_N) \ne (z_E, w_E) = (z_S, w_S)$, for 2-shocks. Finally, $\mathcal{I}$ comprises all points $(\overline{x},\overline{t})$ for which at least three of the four states $(z_W, w_W), (z_E, w_E), (z_N, w_N)$ and (z_S, w_S) are distinct. It can be shown that $\mathcal{I}$ is at most countable.

The focussing of characteristics, induced by genuine nonlinearity, is responsible for the demise of Lipschitz continuity and the generation of shocks. However, this same pattern, viewed in reverse time, has the opposite effect of lowering the Lipschitz constant of the solution. This "schizophrenic" role of genuine nonlinearity is reflected in the following

Theorem 4.2. *Assume the set* $\mathcal{C}$ *of points of continuity of the solution* U *has nonempty interior* $\mathcal{C}^0$. *Then* U *is locally Lipschitz continuous on* $\mathcal{C}^0$.

5. Initial Data in L^1

Genuine nonlinearity gives rise to a host of dissipative mechanisms that affect the large time behavior of solutions. The following proposition reports $O(t^{-\frac{1}{2}})$ decay when the initial data are summable. The proof is given in [7, Ch. XII].

Theorem 5.1. *When* $(z(\cdot,0), w(\cdot,0)) \in L^1(-\infty,\infty)$, *then, as* $t \to \infty$,

$$(z(\cdot,t), w(\cdot,t)) = O(t^{-\frac{1}{2}}) \ , \tag{5.1}$$

uniformly in x *on* $(-\infty,\infty)$.

6. Initial Data with Compact Support

Here we discuss the large time behavior of solutions with initial data $(z(\cdot,0), w(\cdot,0))$ that vanish outside a bounded interval $[-\ell,\ell]$. In the first place, by virtue of Theorem 5.1, the Riemann invariants decay at the rate $O(t^{-\frac{1}{2}})$. As a result, the two characteristic families decouple asymptotically, and each one develops a N-wave profile of width $O(t^{\frac{1}{2}})$ and strength $O(t^{-\frac{1}{2}})$, which propagates into the rest state at characteristic speed. This asymptotic portrait was established in [8;15;17], for solutions constructed by the random choice scheme. For BV solutions satisfying the structural condition, the study of the spreading of generalized characteristics leads to the following, sharp result, whose proof is given in [7, Ch. XII]:

Theorem 6.1. *Employing the notation introduced in Section 4, consider the special forward characteristics* $\phi_-(\cdot), \psi_-(\cdot)$ *issuing from* $(-\ell,0)$ *and* $\phi_+(\cdot), \psi_+(\cdot)$ *issuing from* $(\ell,0)$. *Then*

(a) *For* t *large,* ϕ_-, ψ_-, ϕ_+ *and* ψ_+ *propagate according to*

$$\phi_-(t) = \lambda(0,0)t - (p_-t)^{\frac{1}{2}} + O(1) \ , \tag{6.1}_1$$

$$\psi_+(t) = \mu(0,0)t + (q_+t)^{\frac{1}{2}} + O(1) \ , \tag{6.1}_2$$

$$\phi_+(t) = \lambda(0,0)t + (p_+t)^{\frac{1}{2}} + O(t^{\frac{1}{4}}) \ , \tag{6.2}_1$$

$$\psi_-(t) = \mu(0,0)t - (q_-t)^{\frac{1}{2}} + O(t^{\frac{1}{4}}) \ , \tag{6.2}_2$$

where p_-, p_+, q_- *and* q_+ *are nonnegative constants.*

(b) *For* $t > 0$ *and either* $x < \phi_-(t)$ *or* $x > \psi_+(t)$,

$$z(x,t) = 0 \ , \qquad w(x,t) = 0. \tag{6.3}$$

(c) *For* t *large,*

$$TV_{[\phi_-(t),\psi_+(t)]} z(\cdot,t) + TV_{[\phi_-(t),\psi_+(t)]} w(\cdot,t) = O(t^{-\frac{1}{2}}). \tag{6.4}$$

(d) *For* t *large and* $\phi_-(t) < x < \phi_+(t)$,

$$\lambda(z(x,t),0) = \frac{x}{t} + O(\frac{1}{t}) \ , \tag{6.5}_1$$

while for $\psi_-(t) < x < \psi_+(t)$,

$$\mu(0,w(x,t)) = \frac{x}{t} + O(\frac{1}{t}). \tag{6.5}_2$$

(e) *For t large and $x > \phi_+(t)$, if $p_+ > 0$ then*

$$0 \le -z(x,t) \le c[x - \lambda(0,0)t]^{-\frac{3}{2}}, \tag{6.6}_1$$

while for $x < \psi_-(t)$, if $q_- > 0$ then

$$0 \le -w(x,t) \le c[\mu(0,0)t - x]^{-\frac{3}{2}}. \tag{6.6}_2$$

Thus, in the wake of nondegenerate N-waves the Riemann invariants decay at the rate $O(t^{-\frac{3}{4}})$. In cones properly contained in the wake, the decay is even faster, $O(t^{-\frac{3}{2}})$.

The pointwise decay estimates of Theorem 6.1 induce the following asymptotic behavior of solutions in $L^1(-\infty,\infty)$:

Theorem 6.2. *Assume $p_+ > 0$ and $q_- > 0$. Then, as $t \to \infty$,*

$$\|U(x,t) - M(x,t;p_-,p_+)R(0,0) - N(x,t;q_-,q_+)S(0,0)\|_{L^1(-\infty,\infty)} = O(t^{-\frac{1}{4}}), \tag{6.7}$$

where M and N denote the N-wave profiles:

$$M(x,t;p_-,p_+) = \begin{cases} \dfrac{x - \lambda(0,0)t}{\lambda_z(0,0)t} \ , \ \text{for } -(p_-t)^{\frac{1}{2}} \le x - \lambda(0,0)t \le (p_+t)^{\frac{1}{2}} \\ \\ 0 \qquad \text{otherwise} \ , \end{cases} \tag{6.8}_1$$

$$N(x,t;q_-,q_+) = \begin{cases} \dfrac{x - \mu(0,0)t}{\mu_w(0,0)t} \ , \ \text{for } -(q_-t)^{\frac{1}{2}} \le x - \mu(0,0)t \le (q_+t)^{\frac{1}{2}} \\ \\ 0 \qquad \text{otherwise.} \end{cases} \tag{6.8}_2$$

7. Periodic Solutions

The study of genuinely nonlinear systems of two conservation laws will be completed with a discussion of the large time behavior of solutions that are periodic,

$$U(x+\ell,t) = U(x,t) \ , \quad -\infty < x < \infty \ , \quad t > 0 \ , \tag{7.1}$$

and have zero mean:

$$\int_y^{y+\ell} U(x,t)dx = 0 \ , \qquad -\infty < y < \infty \ , \quad t > 0. \tag{7.2}$$

The confinement of waves resulting from periodicity induces active interactions and cancellation. As a result, the total variation per period decays at the rate $O(t^{-1})$:

Theorem 7.1. *For any $x \in (-\infty,\infty)$,*

$$TV_{[x,x+\ell]} z(\cdot,t) + TV_{[x,x+\ell]} w(\cdot,t) \le \frac{\beta\ell}{t}. \tag{7.3}$$

The proof of (7.3), originally given in [12], is an immediate corollary of (3.6) and periodicity.

An important feature of periodic solutions is the existence of divides. A *divide* of the first (or second) characteristic family associated with U, is a curve $\chi : [0,\infty) \to (-\infty,\infty)$ with the property that for any $t \in [0,\infty)$ the restriction of χ to the interval $[0,t]$ coincides with the minimal (or maximal) backward characteristic of the first (or second) family emanating from the point $(\chi(t),t)$. It can be shown

[6] that in the case of our periodic solution at least one (and probably generically just one) divide of each family originates from any interval of the x-axis of period length ℓ. Of course, the ℓ-translate of any divide is necessarily a divide.

As shown in [6;7, Ch. XII], an interesting mechanism is at work here which decouples the two characteristic families, as $t \to \infty$, and induces the formation of saw-toothed shaped profiles, familiar in the case of scalar conservation laws [14], of strength $O(t^{-1})$:

Theorem 7.2. *The upper half-plane is partitioned by divides of the first (or second) family along which z (or w) decays at the rate $O(t^{-2})$. Let $\chi_-(\cdot)$ and $\chi_+(\cdot)$ be any two adjacent divides of the first (or second) family, with $\chi_-(t) < \chi_+(t)$. Then $\chi_+(t) - \chi_-(t)$ approaches a constant at the rate $O(t^{-1})$, as $t \to \infty$. Furthermore, between χ_- and χ_+ lies a characteristic ψ, of the first (or second) family, such that, as $t \to \infty$,*

$$\psi(t) = \frac{1}{2}[\chi_-(t) + \chi_+(t)] + o(1) \ , \tag{7.4}$$

$$\lambda_z(0,0)z(x,t) = \begin{cases} \dfrac{x - \chi_-(t)}{t} + o(\dfrac{1}{t}) \ , & \chi_-(t) < x < \psi(t) \ , \\ \dfrac{x - \chi_+(t)}{t} + o(\dfrac{1}{t}) \ , & \psi(t) < x < \chi_+(t) \ , \end{cases} \tag{7.5}$$

or

$$\mu_w(0,0)w(x,t) = \begin{cases} \dfrac{x - \chi_-(t)}{t} + o(\dfrac{1}{t}) \ , & \chi_-(t) < x < \psi(t) \ , \\ \dfrac{x - \chi_+(t)}{t} + o(\dfrac{1}{t}) \ , & \psi(t) < x < \chi_+(t). \end{cases} \tag{7.5}$$

References

1. A. Bressan, *Global solutions of systems of conservation laws by wave-front tracking.* J. Math. Anal. Appl. **170** (1992), 414-432.
2. A. Bressan and P. LeFloch, *Structural stability and regularity of entropy solutions to hyperbolic systems of conservation laws.* Indiana U. Math. J. (to appear).
3. A. Bressan, T.-P. Liu and T. Yang, *L^1 stability estimates for $n \times n$ conservation laws.* Arch. Rational Mech. Anal. (to appear).
4. A. Bressan and M. Lewicka, *A uniqueness condition for hyperbolic systems of conservation laws* (preprint).
5. C. Dafermos, *Generalized characteristics in hyperbolic systems of conservation laws.* Arch. Rational Mech. Anal. **107** (1989), 127-155.
6. C. Dafermos, *Large time behavior of solutions of hyperbolic systems of conservation laws with periodic initial data.* J. Diff. Eqs. **121** (1995), 183-202.
7. C. Dafermos, *Hyperbolic Systems of Conservation Laws in Continuum Physics.* (To be published by Springer-Verlag).
8. R. DiPerna, *Decay and asymptotic behavior of solutions to nonlinear hyperbolic conservation laws.* Indiana U. Math. J. **24** (1975), 1047-1071.

9. R. DiPerna, *Singularities of solutions of nonlinear hyperbolic systems of conservation laws*. Arch. Rational Mech. Anal. **60** (1975), 75-100.
10. A. Filippov, *Differential equations with discontinuous right-hand side*. Mat. Sb. (N.S.) **51** (1960), 99-128.
11. J. Glimm, *Solutions in the large for nonlinear hyperbolic systems of equations*. Comm. Pure Appl. Math. **18** (1965), 697-715.
12. J. Glimm and P. Lax, *Decay of solutions of nonlinear hyperbolic conservation laws*. Memoirs AMS **101** (1970).
13. K. Jenssen, *Blowup for systems of conservation laws*. (Preprint).
14. P. Lax, *Hyperbolic systems of conservation laws* II. Comm. Pure Appl. Math. **10** (1957), 537-566.
15. T.-P. Liu, *Decay to N-waves of solutions of general systems of nonlinear hyperbolic conservation laws*. Comm. Pure Appl. Math. **30** (1977), 585-610.
16. T.-P. Liu, *The deterministic version of Glimm scheme*. Comm. Math. Phys. **57** (1977), 135-148.
17. T.-P. Liu, *Pointwise convergence to N-waves for solutions of hyperbolic conservation laws*. Inst. Mittag-Leffler Report No. 4 (1986).
18. N. Risebro, *A front-tracking alternative to the random choice method*. Proceeding AMS **117** (1993), 1125-1139.
19. K. Trivisa, *A priori estimates in hyperbolic systems of conservation laws via generalized characteristics*. Comm. PDE **22** (1997), 235-267.

Division of Applied Mathematics, Brown University, Providence, RI 02912
dafermos@cfm.brown.edu

Contemporary Mathematics
Volume **238**, 1999
B 0-8218-1196-7-03543-X

Milne Problem for Strong Force Scaling

Irene M. Gamba

Abstract

High field kinetic semiconductor equations with a linear collision operator are considered under strong force scaling corresponding to a strong non-equilibrium regime. Boundary and interface layers are studied and a kinetic half-space problem for a slab geometry is stated and solved analytically for negative constant fields.

The solution of this problem is necessary in order to produce numerical implementations of strong-weak forcing decomposition as already implemented in the kinetic linking of Boltzmann-Stokes equation for the linking of kinetic-fluid interfaces of gas flow.

Introduction

In this lecture I will summarize recent work in collaboration with Axel Klar on kinetic high field models and their associated macroscopic models and transition regimes.

These models have been considered in [**CG1**], [**CG2**], [**TT**], [**FV**], [**Pp2**], based on scalings taken form the range of parameters as obtained in the computational experiments in [**BW**] and recently in [**CGJ**].

However, up to now, no analysis of the kinetic boundary layer problem to find the correct boundary conditions for the fluid approximation has been performed. Such an analysis is also required, if one wants to solve the matching problem for kinetic and macroscopic equations. Here an interface regions between the two equations has to be considered. The matching problem has to be solved, for example, for domain decomposition approaches solving simultaneously kinetic and macroscopic equations in different regions of the computational domain.

Boundary and interface regions are described by a transition layer where a stationary kinetic equation is solved as in [**C1**]. For semiconductor models, see, e.g. [**Pp1**], [**Ya**] [**Kl**].

We assume this layer to have slab symmetry, that is, the particle distribution is constant on surfaces parallel to the interface. (This is generically the case whenever the curvature of the interface is small compared to the reciprocal of the mean free path). Hence, the space coordinate reduces to x, the distance to the boundary or interface. After scaling it like $\frac{x}{\varepsilon}$, where ε is the order magnitude of the mean

1991 *Mathematics Subject Classification.* Primary 35Q35, 35B30, 82C40.

free path, one has to solve a kinetic half-space problem. Existence and uniqueness results for this half-space problem have been given in [**GK**] for a relaxation model and negative field. However they can be easily extended to a fully linear collision operator as pointed out to us by Naoufel Ben Abdallah [**BA**].

In fact, one is not really interested in the full solution of the half-space problem: the only objects of interest to obtain boundary or matching conditions are the asymptotic states and the outgoing distribution.

In fact, in [**GK**] we have described a numerical procedure which computes just those quantities by using a Chapman-Enskog type expansion to approximate the solution. The method is seen to converge very fast numerically. For approaches to the numerical solution of the standard half-space problem in gas dynamics and semiconductor equations we refer to [**AS**],[**Co**],[**GA**], [**ST**] and for a mathematical investigation to [**AC**], [**CGS**] and [**GMP**].

The High Field Semiconductor Equations

We consider the semi-classical Boltzmann equation for an electron gas in a semiconductor in the parabolic band approximation in nondimensionalized form with the high field (strong force) scaling:

$$\eta\partial_t f + v\cdot\nabla_z f - \frac{\eta}{\varepsilon}E(z,t)\cdot\nabla_v f = \frac{1}{\varepsilon}Q(f)$$

with $z, v \in \mathbb{R}^3$. The collision operator reads

$$Q(f) = \int s(v,v')[M(v)f(v') - M(v')f(v)]\,dv' = Q^+(f) - Q^-(f),$$

where

$$0 < s_0 \le s(v,v') \le s_1 < +\infty \quad \text{and} \quad s(v,v') = s(v',v).$$

Here we denoted by M the centered, reduced Maxwellian $M = (2\pi)^{-\frac{3}{2}}\exp(-\frac{v^2}{2})$, $E = E(z,t) = -\nabla_z\Phi$ denotes the opposite to the electric field, which is determined by a Poisson equation for the potential Φ:

$$\nabla_z\cdot(\nabla_z\Phi) = \gamma\left(\frac{1}{\eta^d}\int_{\mathbb{R}^3} f dv - C(z)\right).$$

The function $C(z)$ denotes the ion background. The parameters η, γ are dimensionless and of order $O(1)$. The scaled mean free path ε is of order $O(\varepsilon) << 1$. This is the high field scaling, see [**CG1**].

To obtain the boundary or interface layer equations we fix a point $\hat{z}$ on the boundary and re-scale as usual the space coordinate in the layer normal to the boundary with the mean free path ε, introducing the new coordinate x orthogonal to the boundary:

$$x = \frac{(z-\hat{z})\cdot n}{\varepsilon}.$$

Here, n denotes the normal to the boundary or interface. This yields the new coordinates $(x,\hat{z})$ instead of z in the layer. To $O(1)$ one obtains from the rescaled transport equation for a bounded field E at $\hat{z}$:

$$v\cdot n\partial_x\varphi - \eta E\cdot\nabla_v\varphi = Q(\varphi)$$

where $x \in [0,\infty)$ and $E = E(x=0,\hat{z},t)$ does not depend on x. This problem has to be supplied with the in going function at the boundary, i.e. at $x=0$: We have to prescribe $\varphi(0,v), v\cdot n>0$.

To simplify the problem, we assume from now on that the z_1-coordinate points in the direction of the normal, that $E=(E_1,0,0)$ and that $\eta=1$. Then the above reduces to the following one dimensional problem

$$v_1\partial_x\varphi - E_1\partial_{v_1}\varphi = Q(\varphi)$$

with $x\in[0,\infty), v_1\in\mathbb{R}$, $\varphi=\varphi(x,v_1)$. M is now the one-dimensional Maxwellian $M(v)=(2\pi)^{-\frac{1}{2}}\exp(-\frac{v^2}{2}), v\in\mathbb{R}$ and we have used the definition $<f>:=\int\limits_{\mathbb{R}} f(v)dv$.

One observes that

$$\partial_x < v\varphi >= 0,$$

which means $<v\varphi>$ is constant in x.

We shall pose the half-space problem for strong forcing and sketch the proof presented in [**GK**].

The Milne Problem for Strong Negative Field E

For given $-E$ a positive constant, let $P=P(E,v)$ be the unique distribution that solves the problem

$$E\frac{\partial P}{\partial v} = Q(P)\ , \qquad \int P\,dv = 1\ . \tag{1}$$

We shall call P the space homogeneous stationary solution. In fact, P is the leading term of the the renormalized distribution function obtained by the Chapman-Enskog expansion under a strong force scaling, given the higher order term to a distribution corresponding to strong non-equilibrium states.

The solvability of (1) in L^∞ can be found in Trugman and Taylor [TT] for the relaxation type operator in one dimension, has also been discussed in Frosali, Van der Mee and Paveri Fontana [FV] and Poupaud [P1] for the general linear collision operators in higher dimensions (in L^1).

In the case of the relaxation operator with relaxation parameter τ (that is when $s(v,v')=constant=\tau^{-1}$), the distribution solution P satisfies the following explicit formula for $u=-\tau E$ given by

$$P_u = \frac{1}{u}\exp\left(-\frac{\lambda}{u}\right) erf\left(\frac{\lambda}{\sqrt{2\theta}}\right), \tag{2}$$

with $u>0$, $\lambda = v-\frac{2\theta}{u}$ and $erf(x)=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{x} e^{-t^2}dt$.

For $u<0$, the solution is given by $P_u(v)=P_{-u}(-v)$.

In this case P satisfies

$$< vP >= u \quad \text{and} \quad < v^2P >= 1+2u^2.$$

Clearly, P yields distributions that are small perturbations of strong non-equilibrium states.

Hence, in the general linear case, we consider the following problem

$$\begin{cases} v\dfrac{\partial f}{\partial x} + E\dfrac{\partial f}{\partial v} = Q(f) \\ f(0,v)=k(v) \quad \text{on } v>0, \quad 0\le k(v)\le KP(v). \end{cases} \tag{3}$$

The following theorem shows that the solution of problem (3) is unique and relaxes to a multiple of P when the space variable tends to infinity.

THEOREM. *(***Strong forcing Milne problem***) If $-E > 0$, then problem (3) has a* **unique** *solution φ with $0 \le \varphi \le KP(E,v)$, K a positive constant, $P(E,v)$ the space-homogeneous solution of the stationary equation (1). Moreover,*

$$\lim_{x\to\infty} \varphi(x,v) = \lambda_\infty\, P(v), \qquad \textit{with } 0 \le \lambda_\infty \le K\ . \tag{4}$$

The proof of this theorem, contained in [**GK**] for the relaxation case, requires several intermediate steps. The extension to the linear case follows the same strategy as in the relaxation case.

1. The first step consists in making a construction of a solution for the half space problem (3). This is done by the construction of minimal and maximal solutions that control any possible solution of (3) by a constant factor of the homogeneous solution P.
2. The second step is to study the asymptotic behavior of the solutions as the space variable x goes to infinite, i.e the limiting behavior in (4). This step requires several estimates:
   - The control the gain operator (or the average of f in the relaxation case) by a factor μ_0 of the gain operator acting on P. The multiplicative factor μ_0 depends on λ given by the quotient of the first moments of f by P.
   - The construction of a decreasing sequence μ_k and an increasing unbounded one x_k, such that an upper estimate for f by a factor μ_k of P is obtained whenever the gain operator acting on f is bounded above by a factor μ_k of the gain operator acting on P, in a set depending on the characteristic surfaces of the equation (3) that passes through x_k, that is

$$Q^+(f) \le \mu_k Q^+(P) \ \text{ for } \ (x,v) \in D_k = \left\{(x,v),\ x \ge x_k,\ v \le \sqrt{2E(x-x_k)}\right\},$$

     then $f \le \mu_k P$ on D_k. In addition the sequence μ_K is telescoping and bounded below by λ.
   - From the two previous estimates, as k goes to infinity, $\mu_k \to \lambda$, $x_k \to \infty$ and (4) holds with $\lambda_\infty = \lambda$.

3. The third and final step consists into proving uniqueness in within the class of functions that satisfy the data and the homogeneous behavior at infinity.

References

[AS] K. Aoki and Y. Sone, *Gas Flows Around the Condensed Phase with Strong Evaporation or Condensation*, Advances in Kinetic Theory and Continuum Mechanics, Proceedings of a Symposium in Honor of H. Cabannes (1991).

[AC] M.D. Arthur and C. Cercignani, *Nonexistence of a Steady Rarefied Supersonic Flow in a Half Space*, ZAMP, Vol 31, (1980).

[BW2] H.U. Barenger and J.W. Wilkins, *Ballistic structure in the electron distribution function of small semiconducting structures: General features and specific trends*, Physical Review B **36** No 3 (1987), 1487–1502.

[BA] N. Ben Abdallah, *Personal communication* (1998).

[C1] C. Cercignani, *"Half-space problems in the Kinetic Theory of Gases"*, *Trends in Applications of Pure Mathematics to Mechanics*, Edited by E. Kröner and K. Kirchgässner, Springer, Berlin (1986).

[CG1] C. Cercignani, I.M. Gamba, and C.D. Levermore, *High field approximations to a Boltzmann-Poisson system boundary conditions in a semiconductor*, Applied Math Letters, Vol 10 (4), pp.111-118, (1997).

[CG2] C. Cercignani, I.M. Gamba, and C.L. Levermore, *A High Field Approximation to a Boltzmann-Poisson System in Bounded Domains*, TICAM report 97–20, submitted for publication (1998).
http : *//www.ma.utexas.edu/mp_arc/e/*98 − 69.*ps*.

[CGJ] C. Cercignani, I.M. Gamba, J. Jerome, and C.W. Shu, *Device Benchmark Comparisons via Kinetic, Hydrodynamic and High-Field Models,* to appear in CMAME (1999).

[Co] F. Coron, *Computation of the Asymptotic States for Linear Halfspace Problems*, TTSP, 19 (1990).

[CGS] F. Coron, F. Golse and C. Sulem, *a classification of well posed Kinetic problems,* Communications in Pure and Applied Mathematics, vol. 41, 4, pp.409–436, (1988).

[FV] G. Frosali, C.V.M. van der Mee and S.L. Paveri Fontana, *Conditions for run–away phenomena in the kinetic theory of particle swarms*, J. Math. Physics, **30**(5), 1177–1186 (1989).

[GK] I.M. Gamba and A. Klar, *The Milne problem for high field kinetic semiconductor equations*, preprint, submitted for publication (1998).

[GA] F. Golse and A. Klar, *A Numerical Methods for Computing Asymptotic States and Outgoing Distributions for Kinetic Linear Half Space Problems* J. Statist. Phys. **80** no. 5–6, 1033–1061 (1995).

[GMP] W. Greenberg, C. van der Mee and V. Protopopescu, *Boundary Value Problems in Abstract Kinetic Theory*, Birkhäuser, (1987).

[Kl] A. Klar, *A Numerical Method for Kinetic Semiconductor Equations in the Drift Diffusion Limit*, to appear in SIAM J. Sci. Comp. (1998).

[Pp1] F. Poupaud, *Diffusion Approximation of the Linear Semiconductor Equation*, J. Asympt. Anal., **4:**293 (1991).

[Pp2] F. Poupaud, *Runaway Phenomena and Fluid Approximation Under High Fields in Semiconductor Kinetic Theory* ZAMM. Z.angew. Math. Mech. **72** 8, 359-372, (1992).

[ST] C.E. Siewert and J.R. Thomas, *Strong Evaporation into a Half Space II*, Z. Angew. Math. Physik, **33:**202, (1982).

[TT] S. A. Trugman and A. J. Taylor, *Analytic solution of the Boltzmann equation with applications to electron transport in inhomogeneous semiconductors. Phys. Rev. B* **33**, 5575–5584 (1986).

[Ya] A. Yamnahakki, *Second order boundary conditions of Drift Diffusion Equations of Semiconductors*, Math. Meth. Appl. Sci., **5:**429, (1995).

Department of Mathematics, The University of Texas at Austin, Austin, TX 78712-1082

Contemporary Mathematics
Volume **238**, 1999
B 0-8218-1196-7-03544-1

Simple Front Tracking

James Glimm, John W. Grove, X. L. Li, and N. Zhao

Abstract. A new and simplified front tracking algorithm has been developed as an aspect of the extension of this algorithm to three dimensions. Here we emphasize two main results: (1) a simplified description of the microtopology of the interface, based on interface crossings with cell block edges, and (2) an improved algorithm for the interaction of a tracked contact discontinuity with an untracked shock wave. For the latter question, we focus on the post interaction jump at the contact, which is a purely 1D issue. Comparisons to other methods, including the level set method, are included.

1. Introduction

Fluid interface instabilities and chaotic multiphase mixing define the microphysics of multiphase flow. Some of the best computations of these instabilities, in the regime of strongly accelerated flows, have been obtained using the front tracking method [**17, 3, 36**]. Front tracking employs a fundamental approach to the numerical modeling of a fluid interface through the use of a numerically defined interface which plays an explicit role in the algorithm. At the opposite extreme, numerical algorithms based on modern finite differences alone and a very fine grid, made feasible by adaptive mesh refinement (AMR), have also produced high quality simulations in these flow regimes [**21**]. The problem of fluid instabilities remains difficult numerically, especially when the practical requirements of multiscale simulations in three dimensions with flow fields not aligned with a rectangular grid are considered.

Section 2 is a review of the essential features that distinguish front tracking from other numerical methods, while section 3 explains simplifications and improvements recently introduced into this method. An illustrative three dimensional simulation shows a dynamic example of interface bifurcation. Section 4 compares the front tracking and the level set method Osher, Sethian and others [**33, 34, 31**]. In

1991 *Mathematics Subject Classification.* Primary: 76N15, 76T05, 76M20.

Supported by the Applied Mathematics Subprogram of the U.S. Department of Energy DE-FG02-90ER25084, the Department of Energy Office of Inertial Fusion, the Army Research Office, grant DAAG559810313 and the National Science Foundation, grant DMS-9732876.

Supported by the U.S. Department of Energy.

Supported by the Applied Mathematics Subprogram of the U.S. Department of Energy DE-FG02-90ER25084.

Section 5 we present numerical results for the interaction of a shock wave with a contact discontinuity.

2. The Front Tracking Algorithm

2.1. Modularity and Data Structures. Front tracking, as implemented in the code *FronTier*, makes extensive use of modern programming concepts, including data structures and modular organization. We indicate the use of data structures for description of the interface, as illustration. For more details, see [**12**].

The interface library describes the geometry and topology of piecewise smooth manifolds with piecewise smooth boundaries, embedded in R^3. Boundary and coboundary operators, to map from a manifold to its boundary, and to the manifolds which it bounds, are included in this library. The library forms a base class for the remainder *FronTier*. We begin with a description of the main data structures (whose names are in capital letters) and their interrelationships. At a continuum level, an INTERFACE [**14**] is a collection of non-intersecting geometric objects, NODEs, CURVEs, and SURFACEs, that correspond to zero, one, or two dimensional manifolds respectively. Both CURVEs and SURFACEs are oriented manifolds. NODEs correspond to boundary points of CURVEs, while CURVEs correspond to the boundaries of SURFACEs. We designate as COMPONENT, some labeling scheme, *i.e.* equivalence class, for the connected components in R^3 produced by the SURFACEs. Several connected components may share a common component label and constitute a single COMPONENT.

The discretized version of the INTERFACE has the same structure, with a piecewise linear description built from simplices of the appropriate dimensions. The CURVEs are composed of BONDs. Each BOND is a pair of POINTs, and (conceptually) the straight line segment joining them. SURFACEs are discretized in terms of TRIANGLEs.

In Figure 1, we illustrate the geometric data structures used to represent three dimensional interfaces in *FronTier*.

2.2. The Time Step Algorithm. The solution of systems of conservation laws, of the form

$$\mathbf{U}_t + \nabla \cdot \mathbf{F}(\mathbf{U}) = \mathbf{G}(\mathbf{U})$$

are supported by the framework discussed here.

2.2.1. *Interior States: Codimension 0.* The propagation in time of interior states uses a one dimensional regular grid stencil, for a sweep along each coordinate direction, and a choice of finite difference operators for this stencil, such as the higher order Godunov method, the Lax-Wendroff scheme, *etc.* Special care is needed only when the stencil is cut by a front; in this case there are missing state values, as the finite difference operator is expected to receive states from a single component only. In this sense, the method takes the idea of weak derivatives seriously, and will never compute a finite difference across a tracked front.

The missing points of the stencil, in the case of a front cutting through the stencil, are filled in as ghost cells, with the state values obtained by extrapolation from nearby front states of the same component. Thus the state values are double valued near the front, with the left-component states extending by extrapolation for a small distance into the right component, and *vice versa*. The use of ghost cell states was introduced into front tracking in [**13**]. With the ghost states thus defined, the interior solver follows a conventional finite difference algorithm. The

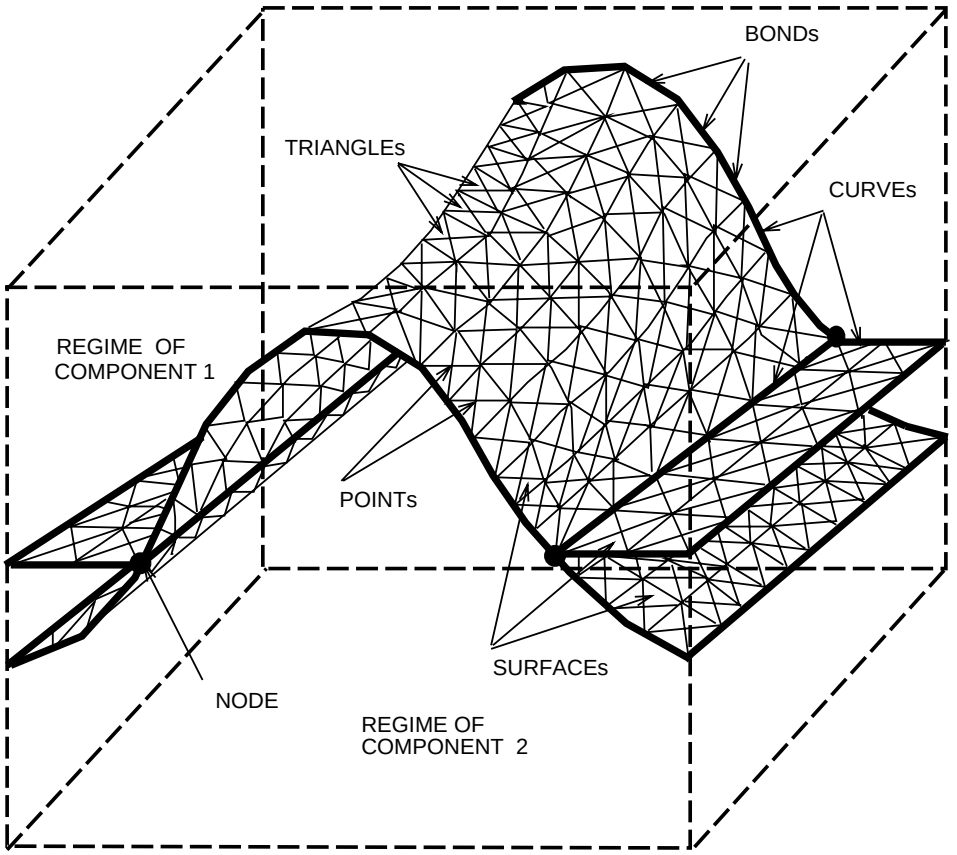

FIGURE 1. An illustration of the three dimensional geometric data structures used in *FronTier*.

stencil for a state on the right side of the interface is composed solely of right component states, and similarly for the stencil for a state on the left side of the interface.

2.2.2. *Regular Front States: Codimension 1.* The propagation of front states and positions is performed in a single step. Operator splitting, in a rotated coordinate system, allows separate propagation steps in directions normal to and tangent to the front. First consider the normal propagation step. The analysis reduces to the integration of a differential equation in one space dimension (the normal direction), and thus is largely independent of spatial dimension.

The leading order term in the propagation of a discontinuity, in the direction normal to the front, is given by the solution of a Riemann problem. This is a one dimensional Cauchy problem, with idealized initial conditions consisting of a single jump discontinuity. The solution will, in general, contain a number of waves. Of these, one is identified with the discontinuity being tracked. The Riemann solution gives the wave speed and states immediately ahead of and behind the advancing front. This speed defines the new interface position, and the states the updated flow states at the propagated points, and thus the lowest order version of the normal propagation algorithm.

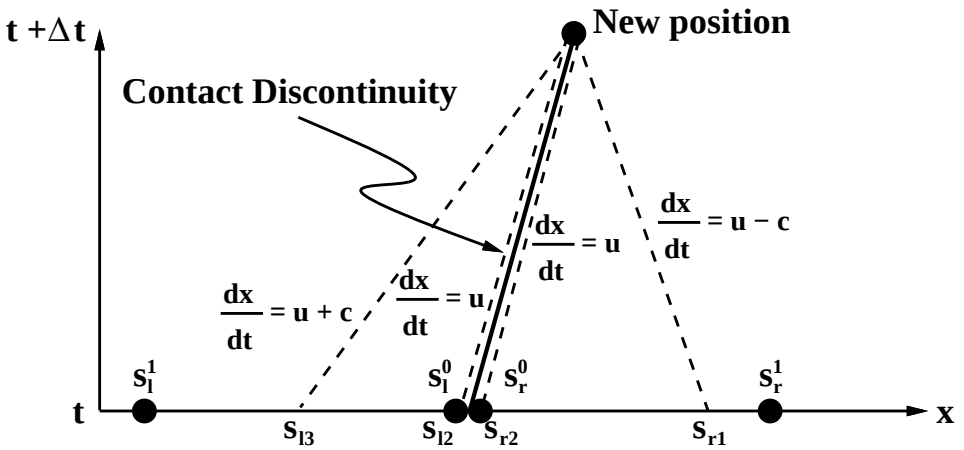

FIGURE 2. A schematic picture of the data used for the normal propagation of a contact discontinuity. The front data at the old time step provides a Riemann solution, that is corrected by interior data, using the method of characteristics.

Corrections are needed, to couple the interior variation of the solution states to the front propagation. For this purpose, a *generalized* Riemann problem is solved. By this we mean a Cauchy problem having a single jump discontinuity in the initial data. However, the initial data on each side of the jump discontinuity, rather than being constant, is now allowed a separate variation, linear in the distance from the front point, on each side of the front. The linear approximation to the nearby interior solution states is constructed by moving in mesh increments Δs away from the point being propagated in the direction of the normal on either side of the interface. The resulting points for solution evaluation, called s_l^i on the left side of the front in Figure 2 and s_r^i on the right, are not, in general, regular grid points, *i.e.*, a cell center corresponding to the rectangular lattice on which the interior states are distributed. The solution at such points must thus be constructed by interpolation using the regular grid points (cell centers) and the front states, but using only states from the same component, *i.e.*, coming from the same side of the interface. The states s_{l3}, s_{l2}, s_{r2}, and s_{r1} in Figure 2 at the feet of the backward characteristics are in turn interpolated from the states s_l^i and s_r^i. The solution of the nonlocal Riemann problem is constructed as a finite difference correction to the previously discussed (local) Riemann problem, using the method of characteristics. Because of the use of finite differences to solve the characteristic equations, the algorithm as specified here (see [**4**]) is suitable for simulations in which the flow variation on either side of the tracked fronts is small relative to the mesh width Δs. One purpose of the present paper is to remove this restriction, so that tracking of (strong) shock waves is not required, and thus so that the complexity of wave interactions solved in three dimensions is reduced. Since a modification of the normal sweep is proposed in §3.2 to allow for strong untracked shock waves interacting with a tracked contact discontinuity, further details are omitted here.

Curvature dependent corrections to the normal propagation are contained implicitly in the tangential sweep. The tangential propagation step modifies the interface states but not the points. The tangential motion of the interface is a reparameterization of the interface, and does not contribute to its dynamics. Thus the tangential motion of the interface is arbitrary, and as a convention, the reparameterization is taken to be the identity.

Separate finite difference steps are carried out for the states on each side of the interface. The splitting into normal and tangential directions is locally orthogonal, and for this reason no explicit source terms are introduced into the difference equations by the splitting. While this seems paradoxical, since *e.g.* radial expanding flow must decrease as the wave front expands, the decay mechanism is found not in an explicit source term, but in the divergence of the velocity field, as seen by the tangential finite difference stencil, after the states are projected onto the tangent plane to the surface.

2.3. Summary. Front tracking offers sharp resolution of fronts, interfaces, or other solution discontinuities. In contrast to other algorithms with features of this nature:

- Front tracking uses Riemann solvers to enforce jump discontinuities and proper jump relations in solution variables.
- Front tracking applies finite difference operations only to states on the same side of the front.
- The front tracking algorithm is applicable to complex physics, in principle to arbitrary systems of conservation laws.

Differences specific to the comparison to the level set method will be noted in §4.

3. Simple Front Tracking

3.1. A Simplified Geometrical Description of Fronts. Domain decomposition parallel computing requires the decomposition and reassembly of interface fragments during a communication phase of the time step operator. The need for robustness of the reassembly places a premium on operations that are purely local to individual processors. In reference [**11**] we described a new method to resolve the changing topology of the tracked fronts by the reconstruction of the interface based on the micro-topology within each rectangular grid block. In this method, the reconstructed interface is uniquely defined by its intersections with cell block edges.

The algorithm as currently implemented assumes a two fluid model. Suppose that a region is occupied by two immiscible fluids. For simplicity let us label the fluids by colors, say black and white. We wish to reconstruct the material interface separating the two fluids using some subset of the geometrical information associated with the "true" material interface. This subset will consist of the crossings of the original interface with the cell edges associated with a specified three dimensional lattice superimposed over the given region.

The reconstruction is based on the following three hypotheses:

1. At most two fluid components intersect any individual cell in the computational lattice.
2. Each cell edge has at most one interface crossing.

3. The cell corners and edges that lie on the same side of the reconstructed interface form a connected set.

The first hypothesis is clearly true for the two fluid flows considered here. More generally the basic algorithm is still valid for multi-fluid flows provided this condition holds locally in each computational cell. The second hypothesis says that the reconstructed interface will cross each cell edge at most once, while the third implies that if two corners of a cell lie in the same fluid component, then the entire edge connecting those corners also remains in the same fluid. These latter two assumptions rule out oscillations in the reconstructed interface at length scales below that of the lattice grid size.

The reconstruction algorithm is based on two steps. The topological reconstruction of an interface segment internal to a single grid block, and the continuity of the interface between adjacent grid blocks. The latter condition is enforced by the obvious geometric observation that interface crossings along specified lattice edge are the same for all cells that contain that edge.

The topological reconstruction of an interface inside a cell is based on the elementary reconstruction of the interface using the component labels (colors) for the cell corners. In general, given two colors, there are $2^8 = 256$ different possible colorings of the vertices of the graph of a cube. We can subdivide this set of colorings into isomorphy classes based on the subgraphs of the cube generated by vertices of the same color. Since we can always remap the cube by transposing the vertex colors (relabeling white to black and black to white), we need only consider the cases where the black vertices appear four or fewer times.

There are fourteen distinct cases.

1. No black vertices.
2. One black vertex.
3. Two black vertices subdivided into three cases:
   (a) Two connected black vertices.
   (b) Two disconnected black vertices sharing a common face.
   (c) Two disconnected black vertices with no common face.
4. Three black vertices subdivided into three cases:
   (a) Three connected black vertices.
   (b) Two connected and one disconnected black vertices.
   (c) Three disconnected black vertices.
5. Four black vertices subdivided into six cases:
   (a) Four connected black vertices sharing a common face of the cube.
   (b) Four connected black vertices whose subgraph has a single vertex of degree three.
   (c) Four connected black vertices whose subgraph has vertices of degree one or two (*i.e.*, the graph is a broken line).
   (d) Three connected and one disconnected black vertices.
   (e) Two disconnected black edges.
   (f) Four disconnected black vertices.

It is trivial to verify that any two coloring of the graph of a cube must correspond to one of these distinct cases, see Figure 3. A block interface is defined as an interface segment formed by connecting the grid line crossings that occur along edges where the coloring changes. For zero black vertices the corresponding interface is empty. In each of the remaining cases we form the interface by

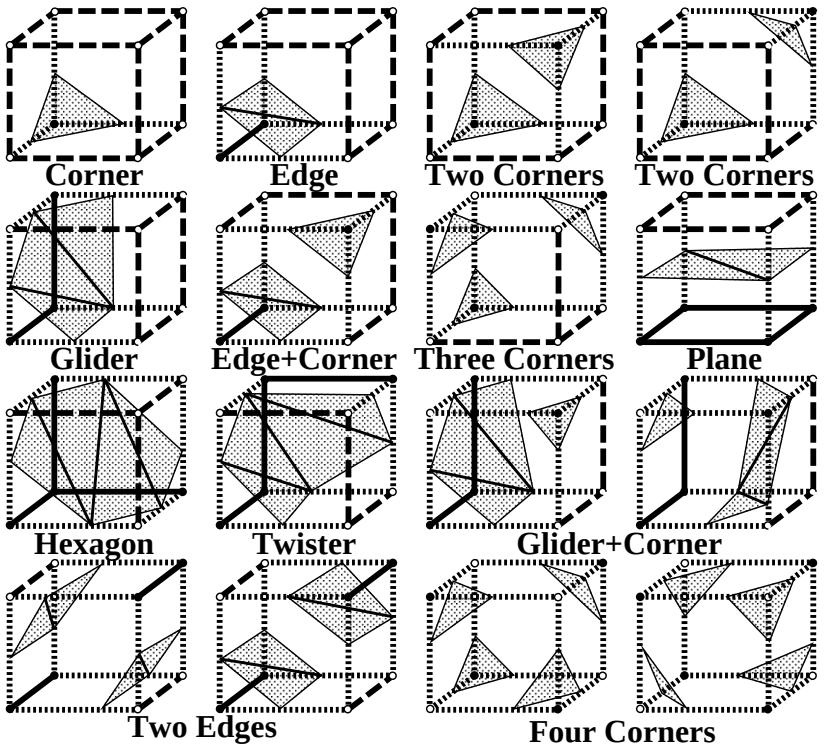

FIGURE 3. Grid based interface reconstruction. For a two-fluid interface within each grid cell, there are $2^8 = 256$ possible configurations for the crossings of the cell edge by the interface. Through elementary operations of rotation, reflection, and separation, these cases can be reduced to the sixteen cases shown above.

constructing a polygon whose sides consist of edge crossings with a common face. These cases are illustrated in Figure 3. We see there are six distinct topological surface elements (corner, edge, glider, plane, hexagon, and twister) and a total of sixteen (not counting the trivial no crossings case) cases in all that are composites of these six elementary surface elements. Note that the cases, glider+corner, two edges, and four corners do not uniquely determine the interface (unlike the previous cases). This is because there are two ways to select the interfaces, those separating out the black or those separating out the white vertices. For example in the glider+corner case we could construct the glider using the three connected black vertices or the three connected white vertices. At the level of this algorithm this choice is arbitrary since either construction will yield a globally consistent interface. In practice one might use additional topological information from the original (*i.e.*, unreconstructed) interface to select the choice that best fits that interface.

The power of the new interface description to untangle interfaces and to automate the dynamic pinchoff or change of interface topology is illustrated in Figure 4. In this figure, we show an instance of interface pinchoff and reconstruction using the grid based interface description in a three dimensional simulation. The numerical

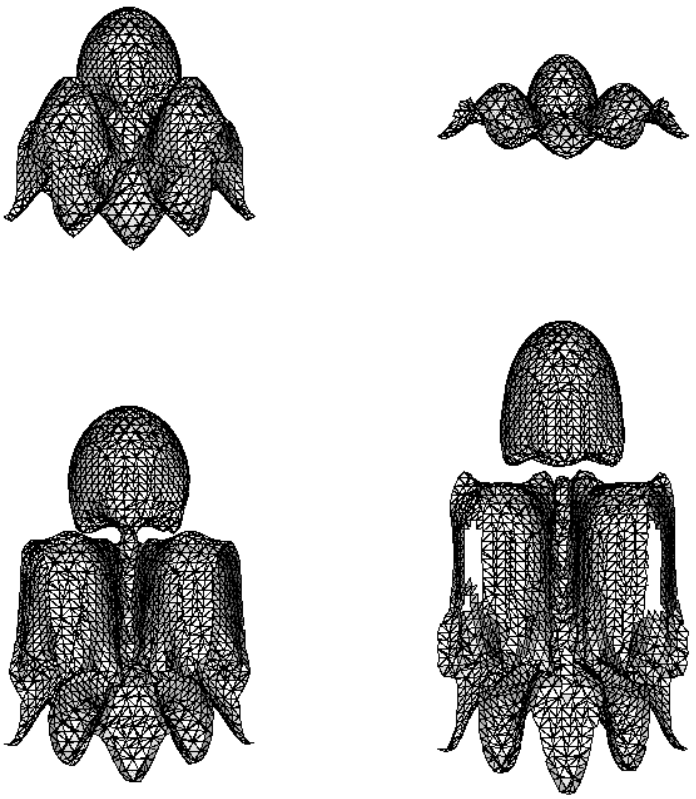

FIGURE 4. A simulation of Rayleigh-Taylor instability using *FronTier*. The initial configuration has four modes of unequal amplitude that evolve at different speeds. In the final frame the lead bubble has pinched off from the main body of light fluid. This bifurcation in the flow is an important part of the dynamics of this instability, and is handled correctly by *FronTier*.

handling of the pinch-off consists of three steps. First the fluid interface is propagated as described in §2. After the propagation, the interface may become tangled. This will be manifested by an inconsistency between the regular grid components and the components at the interface-grid intersections. We walk through the grid cell edges to check the consistency of the components. Unphysical interface-grid intersections, when detected, are removed, leaving only fully consistent interface crossings. Finally, we reconstruct the interface within each grid block as described above using the new crossings. Pinchoff and other topological bifurcations in the interface follow automatically from use of the grid based description with no additional programming complexity.

This grid based interface reconstruction algorithm is comparable in simplicity to the level set method as a geometric interface description.

3.2. Simplified Shock-Contact Interaction. The refraction of a shock wave by a material interface or contact discontinuity is an important physical interaction that has a surprisingly rich and complex structure. An illustration of the variety

of possible refraction behaviors can be seen in the shock tube experiments of Jahn, Henderson, and others, (see for example [**24, 6, 7, 8**]), and numerical computations [**2, 10, 15, 18, 16, 19**].

Basically, the one dimensional interaction of a shock wave with a material interface consists of the acceleration of the material interface due to the differential transfer of momentum from the shock to particles on either side of the interface, and the refraction of the shock front into a set of reflected and transmitted waves. The solution to this interaction is found by solving the Riemann problem with data given by the flow on the upstream side of the material interface and the flow behind the shock wave. It can be shown that the transmitted wave must always be a shock, while the reflected wave may be either a shock or rarefaction wave. If both fronts are tracked, then this Riemann problem data is explicit in the numerical representation of the flow, and the resolution of the interaction is essentially exact.

The situation for two dimensional oblique refractions is considerably more complex. If the angle of incidence between the two waves is not too large, then the local behavior of the interaction can be computed using shock polar analysis [**5**]. The refraction consists of a locally steady state flow with the upstream waves (*i.e.*, the incoming shock and material interface) providing Riemann data for the outgoing downstream waves (the transmitted and reflected waves and the accelerated material interface). In this case the refraction is said to be regular [**15**]. Simulations of locally (in space and time) regular refractions using front tracking have been made with considerable success in a variety of simulations [**15, 17, 23, 22**]. As in the one dimensional case, the application of this algorithm requires the tracking of both the incoming shock and the material interface so that the states about these waves can be used in the shock polar analysis. As the angle of incidence between the two waves is increased the refraction becomes unsteady in space and time and the shock polar equations have no solution. In some cases one can extrapolate the shock polar solutions beyond their strict domain of validity and still obtain reasonable estimates of the flow states [**20, 32, 35**], or else apply a node scattering analysis to interpret the wave behavior into the irregular refraction regime [**18**]. However the virtually infinite set of outcomes for irregular refractions places a practical limitation on the ability of explicit tracking to handle this interaction. In practice, this means that simulations of shock refractions, as occur say in Richtmyer-Meshkov instability, are limited to interactions where the initial perturbations of the material interface are not too large. Fortunately this has not proven to be too onerous a restriction for the simulations conducted previously. The situation in three dimensional flows is even more complex and the need to represent the dynamic local geometry of the interacting waves makes explicit tracking of both the shock front and the material interface virtually impossible. Thus there is a need to formulate versions of the tracking algorithms that do not require explicit tracking of both the incoming shock and material interface.

In general the material interface needs to be tracked to eliminate numerical diffusion across that wave front, so we are primarily interested in formulating a modification of the point propagate algorithm that allows for the interaction of a captured (*i.e.*, untracked) shock with a tracked material interface. The basic problem is this, the point propagate algorithm as described in §2 uses the method of characteristics to integrate the incoming waves into the material interface. In the event that a captured shock front is nearby the interface, the flow gradients near

that wave are not small, and the integration of the characteristic differential equations on either side of the interface produce substantial errors. Furthermore these errors produce spurious wave modes that move with the local fluid velocity. Since this is the same velocity as the material interface, these error waves tend to lock into interface and do not dissipate. In reality, errors of this type occur in all finite difference methods, and these method rely on the numerical viscosity inherent in their schemes to diffuse these error modes away from the fluid fronts, which in most cases are captured waves themselves. Paradoxically, it is precisely the attempt to control numerical diffusion at the material interface that reduces the numerical viscosity near the front and prevents these spurious wave modes from dissipating away from the interface. This problem is compounded since subsequent wave refractions through the material interface are affected by the previously established errors.

The modified point propagation algorithm is based on the analogy between one dimensional front tracking and Lagrangian hydrodynamics. Indeed in one space dimension front tracking can be interpreted as a locally Lagrangian update to the states at the fronts. In multiple space dimensions this same analogy applies to the normal propagation sweep to the interface points. As described in §2, the data for the normal sweep is obtained by sampling the flow state at points p_l^i, p_r^i distributed from the interface in equal spatial increments Δs along the normal direction to the point being propagated. We denote the states at these points by s_l^i and s_r^i for $0 \le i < N$ where N is the number of stencil points on either side of the interface. Thus s_l^0 and s_r^0 are the states at the unpropagated front, and the s_l^i and s_r^i are the states at locations a distance $i\Delta s$ on the left or right hand side of the interface respectively. Next, to each state and position in the propagation stencil we associate a corresponding slope to be used in interpolating the flow in a one dimensional interval centered at that point. Let us denote these slopes by δs_l^i and δs_r^i respectively. For $i = 0$ and $i = N - 1$ the interpolation intervals are of width $\frac{\Delta s}{2}$ while for $1 \le i < N-1$ the intervals are of length Δs. The original algorithm is included in this formulation if we choose $N = 2$ and compute the slopes so that the total interpolation profile between the two points on the same side of the interface is linear with endpoints given by the states at those points. The modification of the point propagate algorithm is to use the van Leer limiter to compute the slopes. For example on the right hand side we choose

$$\delta s_r^i = sgn \min(|\frac{s_r^{i+1} - s_r^{i-1}}{2}|, |s_r^{i+1} - s_r^i|, |s_r^i - s_r^{i-1}|) \tag{1}$$

where sgn is the common algebraic sign of the three differences if all agree in sign, zero otherwise. At the endpoints (in particular at the interface) we simply copy the slopes from the adjacent interior stencil point (*i.e.*, $\delta s_r^0 = \delta s_r^1$). In practice we choose $N = 3$ so that we only compute one slope for all three states, and this is the value centered about a point a distance of one Δs from the interface.

The basic algorithm then proceeds as before. We use the method of characteristics to trace back along the incoming sound waves to the interface. States at the feet of these characteristics are computed using the data states and computed slopes. If the flow gradient is not large and the flow profile is monotone this is a second order interpolation and agrees to this order with the original algorithm. In the case of a strong wave near the interface, the slopes are limited to zero, which provides the additional numerical dissipation needed to allow the entropy waves

to escape from the material interface where they can diffuse into the surrounding region.

4. Comparison of Front Tracking and Level Sets

4.1. Level Sets. We constructed a level set code [**26, 27, 28, 29**] and obtained the first level set simulations of accelerated fluid instabilities in three dimensions. The code was validated by comparison to previous simulations and to perturbative analytic solutions [**37, 38**]. Two changes in the level set algorithm were needed for successful validation, each improving upon the original algorithm [**30**]. Mass diffusion across the interface in the algorithm of [**30**] grows as $t^{1/2}$. This effect is undesirable, and was cured [**29**] with artificial compression. The strength of an artificial compression parameter controls the resulting narrowness of the diffused interfacial density discontinuity. Our point of view was only to control the long time growth of this diffused band, while not trying to limit it too sharply, say below about five mesh cells in width. When used in this mode, we found artificial compression to be satisfactory in performance. Other workers have reported growth of secondary instabilities along the interface resulting from the use of artificial compression. These secondary instabilities have been attributed to a larger artificial compression parameter, used in an attempt to gain a narrower diffused interface.

Inconsistency between the diffused mass and the sharp change in the equation of state (EOS) at the location of the level zero surface, $\phi = 0$, of the level set function ϕ is a second problem for [**30**]. Due to the gradual change in density and the sharp change in the equation of state, at most one additional thermodynamic quantity can have one sided continuity at the interface, and all the rest will generically have spikes, or standing waves located at the interface. As discovered and documented in [**12**], standing pressure waves at the interface result from differencing in conserved variables. The usual, if not totally satisfactory solution, to this problem is to use a mixed material EOS descriptive of atomically mixed fluids in varying proportions of mixture. In this case, the level set is totally decoupled from the flow and has the role of graphics post processing. We followed a special case of this approach, by restricting our studies to fluids with identical EOS, in which case the density discontinuity results from a jump in temperature.

A proposed cure to the level set EOS problem is to difference in nonconserved variables, such as pressure [**25**]. A more fundamental solution was recently proposed [**9**]. In cells that are near the front, in the sense of not having a full stencil of states on one side of it, ghost cell states are constructed to complete the stencil and to allow a standard finite difference operator update.

To begin the ghost cell state construction, entropy is extrapolated from the nearest state on the proper side of the front. From these extrapolated entropies and from the existing pressures at the grid locations, the equation of state reconstructs a density. The velocities are decomposed into components v_n and v_t, normal and tangential to the interface. Then v_n comes from the existing fluid grid value, while v_t is extrapolated. The extrapolated/real velocities, the pseudo densities, and the real pressures at the grid cell define a "ghost state" used to complete the stencil and to allow update of the near front states with insufficient data on their own side of the front.

This use of extrapolated values to complete missing stencil states for the update of near front states is similar to the treatment of near front states used for

many years in *FronTier* [**13**] and brings these two methods closer algorithmically. Here, the missing states are filled in completely by extrapolation, so that the stencil is composed of states taken from a single component. The states used for extrapolation are associated with the front position itself, rather than the cell centered states. Communication between the states on the two sides of the interface occurs through Riemann solvers based on these states located at the front.

4.2. Riemann Solvers for Front Propagation. Front tracking [**13, 4**] differs from ordinary finite differences through three features, that are supported through an algorithm applicable to a general system of conservation laws:

1. A data structure to support the definition of a sharp interface;
2. Special algorithms to compute updated finite differences for cells whose regular finite difference stencil crosses the front;
3. Riemann solvers to enforce correct propagation velocities and jump discontinuities at the front.

A fundamental difference between the front tracking method and the original formulation of the level set method [**30, 9**] is in the representation of the velocity of the moving front. Using a one dimensional formalism for simplicity, the Euler equations gives the motion of a front as a solution to the equation:

$$\frac{dx}{dt} = v_{\text{front}}, \tag{2}$$

where the front velocity v_{front} is determined by the flow state near the front and the Rankine-Hugoniot conditions across the front. Front tracking treats this equation literally and computes the discrete motion of the front in terms of the coupling between the front motion and the flow on either side of the front. One point of importance is that this method only requires that the velocity field be defined at the front and does not seek to extrapolate v_{front} into a space-time region adjacent to the front. A further property is that the discrete representation of v_{front} allows for the jump in the derivative of v_{front} across the front. In contrast, the level set method as formulated in [**30, 9**], first seeks an extrapolated front velocity field $v_{\text{grid}}(x,t)$ in a neighborhood of the front that agrees with v_{front} at the interface. It then seeks to update the interface position by integrating the equation:

$$\phi_t + v_{\text{grid}} \cdot \nabla\phi = 0 \tag{3}$$

where $\phi(x,t)$ is the level set function whose zero set $\phi = 0$ provides the space-time position of the front. Since this construction requires that v_{grid} be a smooth space-time field, a first order error is introduced at the front since the "true" interface velocity must in general have a jump in its derivative across the wave. A second source of truncation error occurs due to the practical requirement that v_{grid} be constructed at a single time level. This allows for the explicit integration of (3) but restricts the time accuracy of the discrete solution to first order in time.

This error is particularly significant when the flow field near the front is strongly nonlinear, as in the case of a shock refracting through a material interface. For a strong shock interacting with a contact discontinuity, the velocity gradient variation at the edge of the shock is $O(\Delta x^{-1})$. This truncation error in the level set propagation was identified by Adalsteinson and Sethian [**1**], who proposed to eliminate this error by using pure front velocities in the propagation of the level set. The level set function is constructed initially to be the distance to the interface. The given front velocity v_{front} is extended to a velocity v_{ext} defined in a band around the front,

so that the variation in the velocity v_{ext} occurs only on level sets of the level set function ϕ. Thus v_{ext} is constructed as a solution of the equation $\nabla v_{\text{ext}} \cdot \nabla \phi = 0$ at every time step. The front is advanced through the level set function ϕ using:

$$\phi_t + v_{\text{ext}} \cdot \nabla \phi = 0 \ . \tag{4}$$

Using (4), the entire set of level curves moves with rigid spacing under the extended front velocity v_{ext}, rather than with a spatially variable grid velocity v_{grid}. The velocity extension method makes no reference to v_{grid}, and it can thus be used in cases when only a front velocity is defined. This is the basis of Sethian's fast marching algorithm.

The truncation error of (4) appears to be identical to the front tracking propagation of (2). The method (4) of front propagation is algorithmically a close approximation to the front tracking use of a local velocity v_{front} defined at the front. These two algorithms utilize identical data, namely v_{front}. In front tracking, only the level set $\phi = 0$, *i.e.*, only the front itself is involved, and there is no need to construct an extended front velocity. From this point of view, front tracking can be thought of as an ultra narrow band level set method.

5. Numerical Experiments

5.1. One Dimensional Simulations. Numerical simulation shows that the L_∞ errors associated with the interaction of an captured strong shock and a tracked or level set contact are $O(1)$, and increase with shock strength. These errors are concentrated in a region of width $O(\Delta x)$ near the contact, so that the L_1 error is first order. These errors appear to have an origin similar to that of shock wall heating, which are common to many numerical methods.

For strong shocks, the L_1 density errors at the contact are an order of magnitude larger for the level set contacts than they are for tracked contacts. This difference appears to be due to the use of Riemann solvers in front tracking to describe nonlinear wave interactions as opposed to the finite differencing used in the level set method. In fact the level set density errors for strong shocks are comparable to those for the artificial compression method. See Table 1 and Figure 5. Point value density errors at the contact are given in Table 2.

For the numerical experiments, illustrated in Tables 1 and 2 and Figure 5, the fluid ahead of the incident shock was at rest, $v_a = 0$, and at an ambient pressure of $p_a = 1$. The ahead-shock densities to the left and right of the interface were $d_l = 1$ and $d_r = 5$ respectively. The computational domain is $[0, 6]$ and both fluids were taken as perfect gases with $\gamma = 1.4$. Simulations were conducted for various incident shock strengths as measured by the shock Mach number. Initially, the contact interface is located at $x = 3$. The shock travels to the left from the point $x = 3.5$. The computations used a total of 240 grid points for the spatial mesh. The numerical results are compared with the level set entropy extrapolation method of Fedkiw and Osher [**9**] and the artificial compression method [**29**].

5.2. Other Comparisons. A five way comparison (theory, experiment, and three simulation codes) was conducted [**21, 22**] for the single mode Richtmyer Meshkov instability of a shock refracting through a sinusoidally perturbed interface separating two fluids. The agreement was impressive. The three codes agreed in their computations of the instability growth rate and their solutions were within

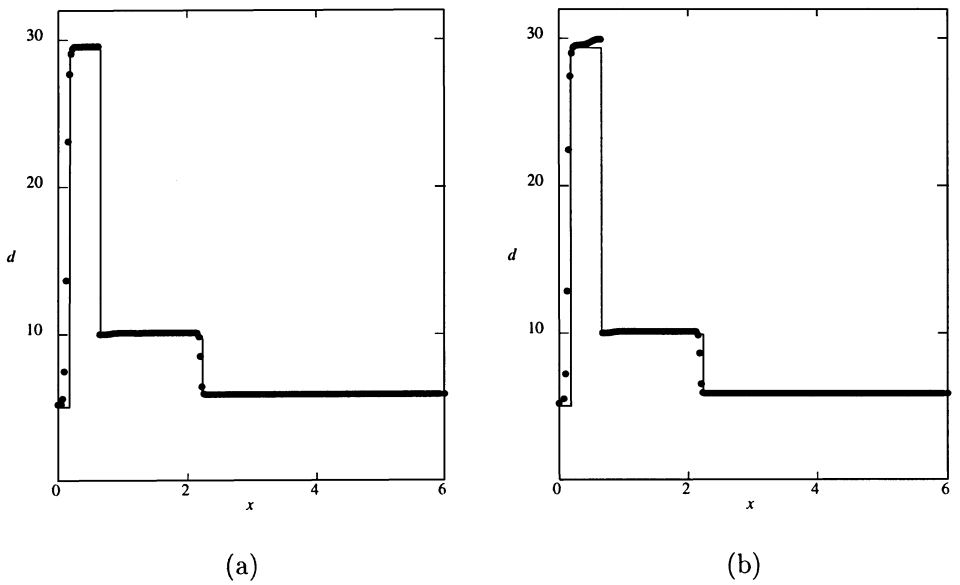

FIGURE 5. Density plot for the solution of an untracked Mach 10 shock interacting with a contact. The contact is tracked in the left frame and represented by a level set in the right. In each frame, the post interaction transmitted shock is on the left, the contact in the center, and the reflected shock wave is on the right. Also shown is the exact solution (solid line in each figure).

Mach Number	*FronTier*	Level Set	TVD/AC
2	9.0238×10^{-3}	1.7165×10^{-2}	7.3113×10^{-2}
5	1.3620×10^{-2}	1.7065×10^{-2}	3.1773×10^{-1}
10	1.8129×10^{-2}	4.9970×10^{-2}	3.7688×10^{-1}
50	1.9120×10^{-2}	2.9411×10^{-1}	4.1944×10^{-1}

TABLE 1. Comparison of L_1 Density Errors

Mach number	*FronTier* Error	Level Set Error
$M = 2$	-0.130781	-0.242735
$M = 5$	-0.063719	-0.027313
$M = 10$	-0.025553	0.430499
$M = 50$	0.071407	1.76907

TABLE 2. Comparison of Density Errors at Contact.

the experimental error bars as was the theory [**37, 38**], based on a low order perturbation expansion in powers of the interface amplitude, Padé resummation, and matched asymptotic expansions. *FronTier* and two higher order Godunov codes, the adaptive mesh refinement code, RAGE, and the PPM based code, PROMETHEUS, were compared. At the level of differences among the three computations, *FronTier*

and RAGE were quite similar while PROMETHEUS showed high frequency secondary instabilities at late time on the interface. The general opinion is that the high order interfacial oscillations shown by PROMETHEUS are numerical in origin and do not correspond to real surface instabilities. *FronTier* used a factor of two less resolution per linear dimension (a factor of eight in space-time mesh cells) than did RAGE, and a factor of three less than did PROMETHEUS (a factor of 27 in space-time mesh cells).

The authors have conducted a series of comparisons of an artificial compression code with *FronTier*, for both Rayleigh Taylor and Richtmyer Meshkov problems, with good results. Comparisons to theory [**37, 38**] were also included.

References

[1] D. Adalsteinson and J. A. Sethian. The fast construction of extension velocities in level set methods. *J. Comp. Phys.*, 1998. submitted.

[2] G. Ben-Dor and I. Glass. Domains and boundaries of non-stationary oblique shock-wave reflexions: 2. monatomic gas. *J. Fluid Mech.*, 96(4):735–756, 1980.

[3] Y. Chen, Y. Deng, J. Glimm, G. Li, D. H. Sharp, and Q. Zhang. A renormalization group scaling analysis for compressible two-phase flow. *Phys. Fluids A*, 5(11):2929–2937, 1993.

[4] I-L. Chern, J. Glimm, O. McBryan, B. Plohr, and S. Yaniv. Front tracking for gas dynamics. *J. Comput. Phys.*, 62:83–110, 1986.

[5] R. Courant and K. Friedrichs. *Supersonic Flow and Shock Waves*. Springer-Verlag, New York, 1976.

[6] A. Abd el Fattah and L. Henderson. Shock waves at a fast-slow gas interface. *J. Fluid Mech.*, 86:15–32, 1978.

[7] A. Abd el Fattah and L. Henderson. Shock waves at a slow-fast gas interface. *J. Fluid Mech.*, 89:79–95, 1978.

[8] A. Abd el Fattah, L. Henderson, and A. Lozzi. Precursor shock waves at a slow-fast gas interface. *J. Fluid Mech.*, 76:157–176, 1976.

[9] R. P. Fedkiw, T. Aslam, B. Merriman, and S. Osher. A non-oscillatory Eulerian approach to interfaces in multimaterial flows (the ghost fluid method). *J. Comp. Phys.*, 1998. Submitted.

[10] H. Glaz, P. Colella, I. I. Glass, and R. L. Deschambault. A numerical and experimental study of pseudostationary oblique shock wave reflections with experimental comparisons. *Proc. Royal Soc. London A*, 398:117–140, 1985.

[11] J. Glimm, J. Grove, X. L. Li, and D. C. Tan. Robust computational algorithms for dynamic interface tracking in three dimensions. *SIAM J. Sci. Comp.*, (SUNYSB-AMS-98-03). Submitted.

[12] J. Glimm, J. W. Grove, X.-L. Li, K.-M. Shyue, Q. Zhang, and Y. Zeng. Three dimensional front tracking. *SIAM J. Sci. Comp.*, 19:703–727, 1998.

[13] J. Glimm, D. Marchesin, and O. McBryan. Subgrid resolution of fluid discontinuities II. *J. Comput. Phys.*, 37:336–354, 1980.

[14] J. Glimm and O. McBryan. A computational model for interfaces. *Adv. Appl. Math.*, 6:422–435, 1985.

[15] J. Grove. The interaction of shock waves with fluid interfaces. *Adv. Appl. Math.*, 10:201–227, 1989.

[16] J. Grove. Irregular shock refractions at a material interface. In S. Schmidt, R. Dick, J. Forbes, and D. Tasker, editors, *Shock Compression of Condensed Matter 1991*, pages 241–244. North-Holland, 1992.

[17] J. Grove, R. Holmes, D. H. Sharp, Y. Yang, and Q. Zhang. Quantitative theory of Richtmyer-Meshkov instability. *Phys. Rev. Lett.*, 71(21):3473–3476, 1993.

[18] J. Grove and R. Menikoff. The anomalous reflection of a shock wave at a material interface. *J. Fluid Mech.*, 219:313–336, 1990.

[19] J. W. Grove. Applications of front tracking to the simulation of shock refractions and unstable mixing. *J. Appl. Num. Math.*, 14:213–237, 1994.

[20] J. F. Hawley and N. Zabusky. Vortex paradigm for shock accelerated density stratified interfaces. *Phys. Rev. L.*, 63:1241, 1989.

[21] R. L. Holmes, G. Dimonte, B. Fryxell, M. Gittings, J. W. Grove, M. Schneider, D. H. Sharp, A. Velikovich, R. P. Weaver, and Q. Zhang. Single mode Richtmyer-Meshkov instability growth: Experiment, simulation, and theory. In G. Jourdan and L. Houas, editors, *Proceedings of the 6th International Workshop on the Physics of Compressible Turbulent Mixing*, pages 197–202. Imprimerie Caractère, Marseille, France, 1997.

[22] R. L. Holmes, B. Fryxell, M. Gittings, J. W. Grove, G. Dimonte, M. Schneider, D. H. Sharp, A. Velikovich, R. P. Weaver, and Q. Zhang. Richtmyer-meshkov instability growth: Experiment, simulation, and theory. Technical Report LA-UR-97-2606, Los Alamos National Laboratory, Los Alamos, NM, 1997. too appear in J. Fluid Mech.

[23] R. L. Holmes, J. W. Grove, and D. H. Sharp. Numerical investigation of Richtmyer-Meshkov instability using front tracking. *J. Fluid Mech.*, 301:51–64, 1995.

[24] R. G. Jahn. The refraction of shock waves at a gaseous interface. *J. Fluid Mech.*, 1:457–489, 1956.

[25] S. Karni. Hybrid multifluid algorithms. *SIAM J. Sci. Comput.*, 17:1019–1039, 1996.

[26] X.-L. Li and J. Glimm. A numerical study of Richtmyer-Meshkov instability in three dimensions. In H. Kubota and S. Aso, editors, *Proceedings of the Second Asia CFD Conference, Tokyo*, volume 1, pages 87–92. Japan Society of Computational Fluid Dynamics, 1996.

[27] X.-L. Li, J. W. Grove, and Q. Zhang. Parallel computation of three dimensional Rayleigh-Taylor instability in compressible fluids through the front tracking method and level set methods. In *Proceedings of the 4th International Workshop on the Physics of Compressible Turbulent Mixing*. Cambridge University Press, Cambridge, 1993.

[28] X. L. Li and B. X. Jin. Parallel computation of Rayleigh-Taylor instability through high resolution scheme for contact discontinuity. In *Proceedings of the First Asia Conference on Computational Fluid Dynamics*, volume 2, pages 811–817. HKUST Press, 1995.

[29] X.-L. Li, B. X. Jin, and J. Glimm. Numerical study for the three dimensional Rayleigh-Taylor instability using the TVD/AC scheme and parallel computation. *J. Comp. Phys.*, 126:343–355, 1996.

[30] R. Mulder, S. Osher, and J. A. Sethian. Computing interface motion in compressible gas dynamics. *J. Comp. Phys.*, 100:209–228, 1992.

[31] S. Osher and J. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi equations. *Jour. Comp. Phys*, 79:12–49, 1988.

[32] R. Samtaney and N. J. Zabusky. Circulation deposition on shock-accelerated planar and curved density-stratified interfaces: Models and scaling laws. *J. Fluid. Mech.*, 269:45–78, 1994.

[33] J. A. Sethian. Numerical algorithms for propagating interfaces: Hamilton-Jacobi equations and conservation laws. *J. Differential Geometry*, 31:131–161, 1990.

[34] J. A. Sethian. *Level Set Methods*. Cambridge University Press, 1996.

[35] X. L. Yang, I-L. Chern, N. J. Zabusky, R. Samtaney, and J. F. Hawley. Vorticity generation and evolution in shock-accelerated density-stratified interfaces. *Phys. Fluids A*, 4(7):1531–1540, 1992.

[36] Q. Zhang and M. J. Graham. A numerical study of Richtmyer-Meshkov instability driven by cylindrical shocks. *Phys. Fluids*, 10:974–992, 1998.

[37] Q. Zhang and S. Sohn. Nonlinear theory of unstable fluid mixing driven by shock waves. *Phys. Fluids*, 9:1106–1124, 1997.

[38] Q. Zhang and S. Sohn. Quantitative theory of Richtmyer-Meshkov instability in three dimensions. *ZAMP*, 1998. To appear.

Department of Applied Mathematics and Statistics, University at Stony Brook, Stony Brook, NY 11794-3600
E-mail address: glimm@ams.sunysb.edu

Hydrodynamics Methods Group, Applied Theoretical and Computational Physics Division, Los Alamos National Laboratory, Los Alamos, NM 87545
E-mail address: jgrove@lanl.gov

Department of Applied Mathematics and Statistics, University at Stony Brook, Stony Brook, NY 11794-3600
E-mail address: linli@ams.sunysb.edu

Department of Aerodynamics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, P.R.China

Contemporary Mathematics
Volume **238**, 1999
B 0-8218-1196-7-03545-3

Formation of Singularities in Relativistic Fluid Dynamics and in Spherically Symmetric Plasma Dynamics

Yan Guo and A. Shadi Tahvildar-Zadeh

1. Introduction

Quasilinear hyperbolic systems have a special place in the theory of partial differential equations since most of the PDEs arising in continuum physics are of this form. Well-known examples are the Euler equations for a perfect compressible fluid, the equations of elastodynamics for a perfect elastic solid, and equations describing a variety of field-matter interactions, such as magnetohydrodynamics etc. It is well-known that for all these systems the Cauchy problem is well-posed, i.e., it has a unique classical solution in a small neighborhood (in space-time) of the hypersurface on which the initial data are given.

On the other hand, it is not expected that these systems will have a global-in-time regular solution, because shock discontinuities are expected to form at some point, at least as long as the initial data are not very small. In more than one space dimension, there are no general theorems to that effect however, mainly because in higher dimensions, the method of characteristics, which is a powerful tool in one dimension for the study of hyperbolic systems, becomes intractable.

In 1985 T. C. Sideris published a remarkable paper on the formation of singularities in three-dimensional compressible fluids [**13**], proving that the classical solution to Euler equations has to break down in finite time. His proof was based on studying certain averaged quantities formed out of the solution, showing that they satisfy differential inequalities whose solutions have finite life-span. Such a technique was already employed by Glassey [**3**] in the case of a nonlinear Schrödinger equation. The idea is that by using averaged quantities one is able to avoid local analysis of solutions. The same technique was subsequently used to prove other formation of singularity theorems: for a compressible fluid body surrounded by vacuum in the nonrelativistic [**7**] and relativistic [**10**] cases, for the spherically symmetric Euler-Poisson equations in the attractive [**6**] and repulsive [**8**] cases, for

1991 *Mathematics Subject Classification.* Primary 76Y05, 35Q05.

The first author was supported in part by a National Science Foundation Postdoctoral Fellowship and the National Science Foundation grant DMS-9623253.

The second author was supported in part by the National Science Foundation grant DMS-9704430.

Both authors are Alfred P. Sloan research fellows.

magnetohydrodynamics [**9**], and for elastodynamics [**14**]. In this paper we present two more such "Siderian" blowup theorems: one in relativistic fluid mechanics, and the other in plasma dynamics.

2. Relativistic Fluid Dynamics

Let (M, g) be the Minkowski spacetime, with (x^μ), $\mu = 0, \dots, 3$ the global coordinate system on M in which $g_{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1)$. We will use the standard convention that Greek indices run from 0 to 3, while Latin ones run from 1 to 3. Indices are raised and lowered using the metric tensor g, and all up-and-down repeated indices are summed over the range. We also denote $t = x^0$ and $\mathbf{x} = (x^1, x^2, x^3)$. In the following, we adopt the notation and terminology of [**1**] and quote from it some of the basic facts regarding relativistic dynamics:

The energy tensor for a relativistic perfect fluid is

$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu + pg^{\mu\nu}. \tag{2.1}$$

In this formula,

1. $u = (u^0, \mathbf{u})$ is the four-velocity field of the fluid, a unit future-directed timelike vectorfield on M, so that $g(u, u) = -1$ and hence
$$u^0 = \sqrt{1 + |\mathbf{u}|^2}.$$
We note that here, unlike the nonrelativistic case considered by Sideris, *all* components of the energy tensor are quadratic in the velocity.
2. $\rho \geq 0$ is the *proper energy density* of the fluid, the eigenvalue of T corresponding to the eigenvector u. It is a function of the (nonnegative) thermodynamic variables n, the *number density* and s, the *entropy per particle*. The particular dependence of ρ on these variables is given by the *equation of state*
$$\rho = \rho(n, s). \tag{2.2}$$
3. $p \geq 0$ is the fluid pressure, defined by
$$p = n\frac{\partial \rho}{\partial n} - \rho. \tag{2.3}$$
Basic assumptions on the equation of state of a perfect fluid are
$$\frac{\partial \rho}{\partial n} > 0, \quad \frac{\partial p}{\partial n} > 0, \quad \frac{\partial \rho}{\partial s} \geq 0 \text{ and } = 0 \text{ iff } s = 0. \tag{2.4}$$
In particular, these insure that η, the speed of sound in the fluid, is always real:
$$\eta^2 := \left(\frac{dp}{d\rho}\right)_s.$$
In addition, the energy tensor (2.1) must satisfy the positivity condition, which implies that we must have
$$p \leq \rho. \tag{2.5}$$

A typical example of an equation of state is that of a polytropic gas. A perfect fluid is called *polytropic* if the equation of state is of the form

$$\rho = n + \frac{A(s)}{\gamma - 1}n^\gamma, \tag{2.6}$$

where $1 < \gamma < 2$ and A is a positive increasing function of s (The speed of light is equal to one). This implies that $p = An^\gamma$ and thus the sound speed $\eta(n, s)$ is determined as follows:

$$\eta^2 = \left(\frac{dp}{d\rho}\right)_s = \frac{\partial p/\partial n}{\partial \rho/\partial n} = \frac{\gamma(\gamma-1)An^{\gamma-1}}{\gamma - 1 + \gamma An^{\gamma-1}}.$$

In particular, the sound speed is increasing with density and is bounded above by $\sqrt{\gamma - 1}$.

The equations of motion for a relativistic perfect fluid are:

$$\partial_\nu T^{\mu\nu} = 0. \tag{2.7}$$

Moreover, $n = n(x)$ satisfies the *continuity equation*

$$\partial_\nu (nu^\nu) = 0. \tag{2.8}$$

Given an equation of state (2.2), the system of equations (2.7-2.8) provides 5 equations for the 5 unknowns $n(x)$, $s(x)$ and $\mathbf{u}(x)$. The component of (2.7) in the direction of u is

$$u^\nu \partial_\nu \rho + (\rho + p)\partial_\nu u^\nu = 0. \tag{2.9}$$

As long as the solution is C^1, this is equivalent to the *adiabatic condition*

$$u^\nu \partial_\nu s = 0. \tag{2.10}$$

The component of (2.7) in the direction orthogonal to u is

$$(\rho + p)u^\nu \partial_\nu u^\mu + h^{\mu\nu}\partial_\nu p = 0, \tag{2.11}$$

where

$$h_{\mu\nu} := g_{\mu\nu} + u_\mu u_\nu$$

is the projection tensor onto the orthogonal complement of $u(x)$ in T_xM.

Thus the system of equations for a relativistic fluid can be written as follows:

$$\begin{cases} \partial_\nu (nu^\nu) &= 0, \\ (\rho + p)u^\nu \partial_\nu u^\mu + h^{\mu\nu}\partial_\nu p &= 0, \\ u^\nu \partial_\nu s &= 0. \end{cases} \tag{2.12}$$

The *Cauchy problem* for a relativistic fluid consists of specifying the values of n, s and $\mathbf{u}$ on a spacelike hypersurface Σ_0 of M,

$$n|_{\Sigma_0} = n_0, \quad s|_{\Sigma_0} = s_0, \quad \mathbf{u}|_{\Sigma_0} = \mathbf{u}_0, \tag{2.13}$$

and finding a solution $(n, \mathbf{u}, s)$ to (2.12,2.13) in a neighborhood of Σ_0 in M. In particular, let $\Sigma_0 = \mathbb{R}^3 \times \{0\}$ be the hyperplane $t = 0$ in M and suppose the initial data (2.13) correspond to a smooth compactly supported perturbation of a quiet fluid filling the space, i.e., assume

$$\begin{array}{l} n_0,\ s_0 \text{ and } \mathbf{u}_0 \text{ are smooth functions on } \mathbb{R}^3 \text{ and there are positive} \\ \text{constants } R_0,\ \bar{n} \text{ and } \bar{s} \text{ such that outside the ball } B_{R_0}(0) \text{ we have} \\ n_0 = \bar{n},\ s_0 = \bar{s}, \text{ and } \mathbf{u}_0 = 0. \end{array} \tag{2.14}$$

Let $\bar{\eta} = \eta(\bar{n}, \bar{s})$ be the sound speed in the background quiet state. We then have

PROPOSITION 2.1. *Any C^1 solution of* (2.12,2.13,2.14) *will satisfy*

$$n = \bar{n}, \quad s = \bar{s}, \quad \mathbf{u} = 0,$$

outside the ball $B_{R(t)}(0)$ where $R(t) = R_0 + \bar{\eta}t$.

PROOF. It is enough to check that the system (2.12) can be written in symmetric hyperbolic form:

$$A^\mu_{ij}(U)\partial_\mu U^j = 0 \text{ where } A^\mu_{ij} = A^\mu_{ji} \text{ and } A^0_{ij} \text{ is positive definite.} \tag{2.15}$$

This can be accomplished for example by using p instead of n as an unknown. By (2.4), we can think of n as a function of p and and s and thus of ρ as a function of p and s. By (2.3) it is then easy to see that (2.12) is equivalent to the following system for the unknowns $U = (p, u, s)$:

$$\left\{ \begin{array}{rcl} \dfrac{1}{(\rho+p)\eta} u^\nu \partial_\nu p + \eta \partial_\nu u^\nu &=& 0 \\ \eta h^{\mu\nu}\partial_\nu p + (\rho+p)\eta u^\nu \partial_\nu u^\mu &=& 0 \\ u^\nu \partial_\nu s &=& 0. \end{array} \right. \tag{2.16}$$

Let $\bar{U} = (\bar{p}, 1, 0, 0, 0, \bar{s})$ denote the constant background solution to (2.16). Let $\bar{\zeta} := (\bar{\rho} + \bar{p})\bar{\eta} > 0$. We then have that the differential operator $P = \bar{A}^\mu \partial_\mu$ corresponding to the linearization of (2.15) at $\bar{U}$ is symmetric hyperbolic, with

$$\bar{A}^0 = A^0(\bar{U}) = \mathrm{diag}(\frac{1}{\bar{\zeta}}, \bar{\zeta}, \bar{\zeta}, \bar{\zeta}, \bar{\zeta}, 1), \qquad \bar{A}^i = A^i(\bar{U}) = \begin{pmatrix} 0 & 0 & \bar{\eta}\mathbf{e}_i^T & 0 \\ 0 & 0 & 0 & 0 \\ \bar{\eta}\mathbf{e}_i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Once we have this, we can use energy estimates, as in [**12**], to conclude the desired domain of dependence statement. □

We now prove that for large enough initial data, the solution to (2.12,2.13,2.14) cannot remain C^1 for all $t > 0$. Such a result was announced in [**10**], but the unpublished proof contained an error which invalidated the argument [**11**].

First of all, a scaling analysis shows that without loss of generality we can set $R_0 = 1$. Let

$$B_t = \{x \in M \mid x^0 = t,\ |\mathbf{x}| \leq R(t) = 1 + \bar{\eta} t\}$$

denote the time t slice of the range of influence of the data, and let

$$Q(t) := \int_{\mathbb{R}^3} g_{ij} x^i T^{0j} = \int \mathbf{x} \cdot \mathbf{u} u^0 (\rho + p) \tag{2.17}$$

be the total radial momentum of the fluid at time t. We then have

$$\begin{aligned} Q'(t) &= \int g_{ij} x^i \partial_0 T^{0j} = -\int g_{ij} x^i \partial_k T^{kj} \\ &= \int g_{ij}(T^{ij} - \bar{T}^{ij}) = \int (\rho + p)|\mathbf{u}|^2 + 3(p - \bar{p}). \end{aligned} \tag{2.18}$$

Let

$$E = \int_{\mathbb{R}^3} T^{00} - \bar{T}^{00} = \int (\rho + p)|\mathbf{u}|^2 + \rho - \bar{\rho} \tag{2.19}$$

be the total energy of the perturbation. By (2.7) it is a conserved quantity, $E = E_0$. Our goal is to use E to obtain a differential inequality for Q that would lead to blowup.

We are going to make two assumptions on the equation of state of the fluid, which are quite natural from a physical point of view. First we note that, as mentioned before, we can use the pressure p as a thermodynamic variable in place

of n. The equation of state of the fluid then has the form $\rho = \rho(p, s)$. The two assumptions are:

(A1) $\rho(p, s)$ is a non-increasing function of s, for each p.
(A2) $\eta(p, s)$ is a non-decreasing function of p, for each s.

These two assumptions are in particular satisfied for a polytropic equation of state (2.6). In order to see that, we observe that $n = (p/A(s))^{1/\gamma}$, and from there we get

$$\rho(p, s) = \frac{1}{\gamma - 1} p + \frac{1}{A^{1/\gamma}(s)} p^{1/\gamma}.$$

It is then clear that **(A1)** holds. Moreover

$$\eta^2(p, s) = \frac{\gamma(\gamma - 1) A^{1/\gamma}(s) p^{(\gamma-1)/\gamma}}{\gamma - 1 + \gamma A^{1/\gamma}(s) p^{(\gamma-1)/\gamma}}$$

shows that **(A2)** is satisfied.

We also make the following assumptions on the initial data:

(D1) $\bar{\eta} < \frac{1}{3}$.
(D2) $E > 0$.
(D3) $s_0(\mathbf{x}) \geq \bar{s}$ for all $\mathbf{x} \in B_0$.

By (2.10), the entropy s is constant along the flow lines, and thus **(D3)** implies that $s(x) \geq \bar{s}$ for $x \in B_t$. By **(A1)** and **(A2)** we then have

$$\begin{aligned} \rho - \bar{\rho} &= \rho(p, s) - \rho(\bar{p}, \bar{s}) = \rho(p, s) - \rho(p, \bar{s}) + \rho(p, \bar{s}) - \rho(\bar{p}, \bar{s}) \leq \rho(p, \bar{s}) - \rho(\bar{p}, \bar{s}) \\ &= \int_{\bar{p}}^{p} \frac{\partial \rho}{\partial p}(p', \bar{s})\, dp' = \int_{\bar{p}}^{p} \frac{1}{\eta^2(p', \bar{s})}\, dp' \leq \frac{1}{\bar{\eta}^2}(p - \bar{p}). \end{aligned}$$

By (2.18) and (2.19) we then obtain

$$Q'(t) \geq 3\bar{\eta}^2 E + (1 - 3\bar{\eta}^2) \int (\rho + p)|\mathbf{u}|^2,$$

which implies, by virtue of **(D1)** and **(D2)** that

$$Q'(t) \geq (1 - 3\bar{\eta}^2) \int (\rho + p)|\mathbf{u}|^2 > 0.$$

In particular $Q(t) > 0$ if $Q(0) > 0$.

On the other hand, we can always estimate $Q(t)$ from above, using (2.5):

$$\begin{aligned} Q^2(t) &\leq \left(\int (\rho + p)|\mathbf{u}|^2 \right) R^2(t) \left(\int_{B_t} (\rho + p)(|\mathbf{u}|^2 + 1) \right) \\ &\leq 2 \left(\int (\rho + p)|\mathbf{u}|^2 \right) R^2(t) \left(\int_{B_t} (\rho + p)|\mathbf{u}|^2 + \rho - \bar{\rho} + \bar{\rho} \right) \\ &\leq \frac{2}{1 - 3\bar{\eta}^2} Q'(t) R^2(t) [E + \frac{4\pi}{3} \bar{\rho} R^3(t)]. \end{aligned}$$

Integrating this differential inequality and changing the integration variable to $r = R(t)$, we obtain

$$\frac{1}{Q(t)} \leq \frac{1}{Q(0)} - \frac{1 - 3\bar{\eta}^2}{2\bar{\eta}} \int_1^{R(t)} \frac{dr}{Er^2 + \frac{4\pi}{3}\bar{\rho} r^5},$$

which contradicts the positivity of Q for all time provided the initial data satisfies the following final assumption:

(D4) $Q(0) > \dfrac{2\bar\eta}{1-3\bar\eta^2}\left(\displaystyle\int_1^\infty \frac{dr}{Er^2+\frac{4\pi}{3}\bar\rho r^5}\right)^{-1}$.

The contradiction implies that there exists a certain $T^* < \infty$ by which time a C^1 solution has to have broken down. In particular, the domain of dependence may break down at an earlier time, perhaps because a shock discontinuity forms. We have thus proved

THEOREM 2.2. *Suppose that the equation of state of a fluid satisfies* **(A1)** *and* **(A2)**. *Then the Cauchy problem* (2.12,2.13,2.14) *with initial data satisfying* **(D1–D4)** *cannot have a global-in-time C^1 solution.*

REMARK 2.3. It is easy to obtain a simpler, sufficient condition for blowup: Let

$$f(y) := \left(\int_1^\infty \frac{dr}{r^2(r^3+y)}\right)^{-1}.$$

By **(D4)** we thus need

$$Q(0) > \frac{2\bar\eta}{1-3\bar\eta^2}\frac{4\pi}{3}\bar\rho\, f(\frac{E}{\frac{4\pi}{3}\bar\rho}). \tag{2.20}$$

It is easy to see that $f(0)=4$, $f'(0)=16/7$ and that f is a concave function of y, so that $f(y) < \frac{16}{7}y + 4$. It is therefore enough to have

$$Q(0) > \frac{32\bar\eta}{7(1-3\bar\eta^2)}(E + \frac{7\pi}{3}\bar\rho). \tag{2.21}$$

We note that unlike the nonrelativistic case, the lower bound for the initial radial momentum in **(D4)** or (2.21) depends on the initial energy, and thus on the initial velocity. Since Q is of the same order of magnitude as E, it is worthwhile to show that there *exist* data sets satisfying these largeness conditions. In fact, (2.21) can be satisfied for $\bar n$ small enough. All that is needed is $\partial\rho/\partial n > 0$ at $n = 0$. We illustrate this in the following by considering the polytropic case.

Let us consider a fluid with a polytropic equation of state (2.6), and consider initial data of the following form

$$n_0(\mathbf{x}) = \bar n\psi(r), \qquad \mathbf{u}_0(\mathbf{x}) = \frac{\mathbf{x}}{r}\phi(r), \qquad s_0(\mathbf{x}) = \bar s + \phi(r), \tag{2.22}$$

where $r = |\mathbf{x}|$. ϕ and ψ are smooth, positive functions on $[0,\infty)$ such that

$$\phi(r) \equiv 0 \text{ for } r \geq 1, \qquad \phi(0) = 0, \tag{2.23}$$

and

$$\psi(r) \equiv 1 \text{ for } r \geq 1, \quad \int_0^1 (\psi(r)-1)r^2dr = 0. \tag{2.24}$$

We then compute

$$\begin{aligned} E &= \int_{B_0} (\rho_0 + p_0)|\mathbf{u}_0|^2 + \rho_0 - \bar\rho \\ &= 4\pi\bar n\int_0^1 \left\{\psi\phi^2 + \frac{1}{\gamma-1}\bar n^{\gamma-1}[A(s)\psi^\gamma(\gamma\phi^2+1) - A(\bar s)]\right\} r^2dr, \end{aligned}$$

and thus $E > 0$ by **(D3)** and (2.24). Now,

$$Q(0) = 4\pi\bar n\int_0^1 \phi\sqrt{1+\phi^2}(\psi + A\frac{\gamma}{\gamma-1}\bar n^{\gamma-1}\psi^\gamma)r^3dr.$$

Dividing (2.21) by $4\pi\bar{n}$, all we need is that the following inequality be satisfied for $\bar{n}$ small enough:

(2.25)
$$\int_0^1 \psi\phi\sqrt{1+\phi^2}r^3dr + O(\bar{n}^{\gamma-1}) > \frac{32}{7(1-3\bar{\eta}^2)}\bar{\eta}\left\{\int_0^1 \psi\phi^2 r^2 dr + \frac{7}{12} + O(\bar{n}^{\gamma-1})\right\}.$$

This is clearly true since $\bar{\eta} \to 0$ as $\bar{n} \to 0$. We have thus shown

PROPOSITION 2.4. *Let ϕ and ψ be two smooth, positive functions on $[0,\infty)$ satisfying* (2.23,2.24)*. Then there exists $\bar{n} > 0$ small enough (depending on ϕ, ψ and γ) such that the initial data set $(n_0, \mathbf{u}_0, s_0)$ of the form* (2.22) *satisfy the conditions* **(D1–D4)***, and thus lead to a blowup for* (2.12,2.13,2.14)*.*

3. Euler-Maxwell with Constant Background Charge

A simple two-fluid model to describe plasma dynamics is the so called Euler-Maxwell system, where a compressible electron fluid interacts with a constant ion background. Let $n(t,\mathbf{x})$, $s(t,\mathbf{x})$ and $\mathbf{v}(t,\mathbf{x})$ be the average electron density, entropy, and velocity, let $\bar{n}$ be the constant ion density, and let $\mathbf{E}(t,\mathbf{x})$ and $\mathbf{B}(t,\mathbf{x})$ be the electric and magnetic fields. Let c = speed of light in vacuum, e = the charge of an electron, and m = the mass of an electron. The Euler-Maxwell system (see [**5**, pp. 490–491]) then takes the form:

$$(3.1)\quad \begin{cases} \partial_t n + \partial_i(nv^i) = 0 \\ \partial_t \Pi^i + \partial_j T^{ij} = \dfrac{e\bar{n}}{m}E^i \\ \partial_t s + v^i\partial_i s = 0 \end{cases} \qquad \begin{aligned} \partial_t B^i + c(\nabla\times\mathbf{E})^i &= 0 \\ \partial_t E^i - c(\nabla\times\mathbf{B})^i &= -4\pi e n v^i, \end{aligned}$$

together with the constraint equations

$$(3.2)\qquad \partial_i E^i = 4\pi e(n-\bar{n}), \qquad \partial_i B^i = 0.$$

In the above, Π is the momentum vector,

$$\Pi = n\mathbf{v} + \frac{1}{4\pi mc}(\mathbf{E}\times\mathbf{B}),$$

and T is the stress tensor, which can be decomposed into material and electromagnetic parts: $T = T_M + T_E$, with

$$\begin{aligned} T_M^{ij} &= nv^iv^j + \frac{1}{m}p\delta^{ij}, \\ T_E^{ij} &= \frac{1}{4\pi m}[\frac{1}{2}(|\mathbf{E}|^2+|\mathbf{B}|^2)\delta^{ij} - E^iE^j - B^iB^j]. \end{aligned}$$

p is the electron pressure, which is modeled by a polytropic law $p(n,s) = A(s)n^\gamma$, where $\gamma > 1$ and A is a positive increasing function.

The system (3.1) being hyperbolic, we once again have the domain of dependence property. However, this time the largest characteristic speed in the background will be c, the speed of light. We recall that in Sideris's original argument [**13**], the largeness condition on the initial data implied that the initial velocity had to be supersonic at some point, relative to the sound speed in the background. An analogous result in the Euler-Maxwell case would thus require that the initial velocity be superluminar at some point, which is absurd. However, we note that if the data is spherically symmetric, so will be the solution, and thus there will be no electromagnetic waves, and the largest characteristic speed will once again be the sound speed, so a Siderian blowup theorem is possible in the spherically symmetric

case. Moreover, since spherical symmetry implies that the flow is irrotational, such a blowup result is complementary to the recent construction [**4**] of global smooth irrotational solutions with small amplitude for the above system. We note that a blowup result in the spherically symmetric, isentropic case with no background charge has been obtained [**2**] using Riemann invariants.

REMARK 3.1. Under the assumption of spherical symmetry, the Euler-Maxwell system reduces to what is often referred to as the spherically symmetric Euler-Poisson system (with repulsive force). We note the important distinction between this, and the general Euler-Poisson system obtained by taking the Newtonian limit $c \to \infty$ in (3.1). The latter is not a hyperbolic system, and does not have finite propagation speeds.

We have the following theorem:

THEOREM 3.2. *Let ν_0, σ_0 and u_0 be smooth functions on $\mathbb{R}^+$ satisfying*

$$u_0(r) \equiv \sigma_0(r) \equiv \nu_0(r) \equiv 0 \text{ for } r \geq 1, \qquad u_0(0) = 0, \qquad \sigma_0(r) \geq 0,$$

and the neutrality condition

$$\int_0^1 \nu_0(r) r^2 dr = 0. \tag{3.3}$$

Let $\bar{s} \geq 0$ be fixed. Then

(a) *There exists $T > 0$ and functions $\nu, \sigma, u, E \in C^1([0,T) \times \mathbb{R}^+)$ such that*

$$\begin{array}{l} \nu(0,r) = \nu_0(r) \\ \sigma(0,r) = \sigma_0(r) \\ u(0,r) = u_0(r) \end{array} \qquad E(0,r) = \frac{4\pi e}{r^2} \int_0^r \nu_0(r') r'^2 dr'.$$

and such that the Euler-Maxwell system (3.1) *has a unique solution of the form:*

$$\begin{array}{rclcrcl} n(t,\mathbf{x}) & = & \bar{n} + \nu(t,r) & & \mathbf{E}(t,\mathbf{x}) & = & E(t,r)\dfrac{\mathbf{x}}{r} \\ s(t,\mathbf{x}) & = & \bar{s} + \sigma(t,r) & & & & \\ \mathbf{v}(t,\mathbf{x}) & = & u(t,r)\dfrac{\mathbf{x}}{r} & & \mathbf{B}(t,\mathbf{x}) & \equiv & 0, \end{array} \tag{3.4}$$

where $r = |\mathbf{x}|$.

(b) *For $t \in [0,T)$, $(n, s, \mathbf{v}, \mathbf{E})$ satisfy the reduced Euler-Maxwell system:*

$$\left\{ \begin{array}{rcl} \partial_t n + \partial_i (n v^i) & = & 0 \\ \partial_t s + v^i \partial_i s & = & 0 \\ \partial_t (n v^i) + \partial_j T^{ij} & = & \dfrac{e\bar{n}}{m} E^i \\ \partial_t E^i + 4\pi e n v^i & = & 0, \end{array} \right. \tag{3.5}$$

where

$$T^{ij} = n v^i v^j + \frac{1}{m} p \delta^{ij} + \frac{1}{4\pi m} \left(\frac{1}{2} |\mathbf{E}|^2 \delta^{ij} - E^i E^j \right),$$

together with the constraint Poisson equation:

$$\partial_i E^i = 4\pi e (n - \bar{n}).$$

(c) *Let* $\bar{\eta} = \sqrt{\gamma A(\bar{s})\bar{n}^{\gamma-1}}$ *be the sound speed in the background,* $R(t) := 1 + \bar{\eta}t$, *and let*

$$D_T := \{(t, \mathbf{x}) \mid 0 \le t < T,\ |\mathbf{x}| \ge R(t)\}.$$

Then we have $(n, s, \mathbf{v}, \mathbf{E}) \equiv (\bar{n}, \bar{s}, 0, 0)$ *on* D_T.

(d) *For any fixed* $\nu_0(r)$ *which satisfies* (3.3), *there exists* $u_0(r)$ *sufficiently large, such that the life-span of the* C^1 *solution* (3.4) *is finite.*

PROOF. (a) The Euler-Maxwell system (3.1) can be written as a positive, symmetric hyperbolic system, and therefore has a unique, local C^1 solution with $n > 0$ provided its initial data are sufficiently smooth. Notice that the initial data are spherically symmetric. Because of the rotational covariance properties of the Euler-Maxwell system and the uniqueness of the local solution, the solution remains spherically symmetric and (a) follows.

(b) follows since $B^i \equiv 0$.

(c) Notice that from the Poisson equation at $t = 0$,

$$E(0, r) = \frac{4\pi e}{r^2}\int_0^r \nu_0(r) r^2 dr \equiv 0 \text{ for } r \ge 1$$

by the neutrality assumption (3.3). Now the reduced Euler-Maxwell system (3.5) is still a hyperbolic system, and we can deduce (c) via the Proposition in [**12**].

(d) Let

$$Q(t) := \frac{1}{4\pi}\int_{\mathbb{R}^3} x \cdot \Pi = \int_0^\infty rnu\, r^2 dr.$$

A direct computation yields:

$$Q'(t) = \int_0^\infty \left\{ nu^2 + \frac{3}{m}(p - \bar{p}) + \frac{1}{8\pi m}E^2 \right\} r^2 dr + \frac{e\bar{n}}{m}\int_0^\infty rE\ r^2 dr,$$

where $\bar{p} = p(\bar{n}, \bar{s})$. Meanwhile, by the first and fourth equations in (3.5),

$$\int_0^\infty rE(t, r) r^2 dr = \int_0^\infty rE(0, r) r^2 dr - 4\pi e \int_0^t Q(t') dt'.$$

Integrating by parts, we notice that

$$\int_0^\infty rE(0, r) r^2 dr = -4\pi e \int_0^\infty \nu_0(r) r^4 dr.$$

We now define $y(t) := \displaystyle\int_0^t Q(t') dt'$ and obtain

$$y''(t) + \omega^2 y(t) = G(t), \tag{3.6}$$

where $\omega^2 = \dfrac{4\pi e^2 \bar{n}}{m}$ is the *plasma frequency*, and

$$G(t) := -\omega^2 \int_0^\infty \nu_0(r) r^4 dr + \int_0^\infty \left\{ nu^2 + \frac{3}{m}(p - \bar{p}) + \frac{1}{8\pi m}E^2 \right\} r^2 dr.$$

Therefore, from solving the ODE (3.6) for $y(t)$, we have

$$y''(t) = -\omega y'(0) \sin \omega t + G(t) - \omega \int_0^t \sin \omega(t - \tau) G(\tau) d\tau. \tag{3.7}$$

We recall the conserved quantities energy:

$$\mathcal{E} = \int_0^\infty \left\{ \frac{1}{2} n u^2 + \frac{1}{m(\gamma-1)} (A(s) n^\gamma - A(\bar{s}) \bar{n}^\gamma) + \frac{1}{8\pi m} E^2 \right\} r^2 dr,$$

and mass

$$M = \frac{1}{4\pi} \int_{\mathbb{R}^3} (n - \bar{n}) = \int_0^\infty \nu(t, r) r^2 dr.$$

From the neutrality condition (3.3) we have $M \equiv 0$. Also, $s(0, \mathbf{x}) \geq \bar{s}$ since $\sigma_0 \geq 0$ by assumption. By the adiabatic condition (the second equation in (3.5)) entropy is constant along flow lines, and thus $s(t, \mathbf{x}) \geq \bar{s}$ for $t < T$. Since A is an increasing function,

$$\int A(s) n^\gamma - A(\bar{s}) \bar{n}^\gamma \geq A(\bar{s}) \int n^\gamma - \bar{n}^\gamma \geq \bar{\eta}^2 \int n - \bar{n} = 0.$$

Hence we have

$$\alpha \mathcal{E} \leq G(t) + \omega^2 \int_0^\infty \nu_0(r) r^4 dr \leq \beta \mathcal{E},$$

with $\alpha = \min\{1, 3(\gamma-1)\}$, $\beta = \max\{2, 3(\gamma-1)\}$. But for large enough $u_0(r)$, $\int_0^\infty \nu_0(r) r^4 dr$ is dominated by $\mathcal{E}(0)$. Hence, we have

$$\frac{\alpha}{2} \mathcal{E} \leq G(t) \leq 2\beta \mathcal{E}$$

for sufficiently large $u_0(r)$. Moreover, we have

$$Q^2(t) \leq R^2(t) \int_0^\infty n u^2 \int_0^{R(t)} n \leq C R^5(t) \bar{n} \mathcal{E}. \tag{3.8}$$

C will henceforth denote a generic numerical constant. By choosing $u_0(r)$ large such that $\mathcal{E}(t) = \mathcal{E}(0) \geq 1$, we have

$$|y'(0)| = |Q(0)| \leq C \sqrt{\bar{n}} \mathcal{E}.$$

Thus from (3.7), there exists $T_0 = T_0(\gamma, \bar{n}, \omega) > 0$ such that for $0 \leq t \leq T_0$,

$$Q'(t) \geq C \mathcal{E}.$$

Together with (3.8), we deduce for $0 \leq t \leq T_0$,

$$Q'(t) \geq \frac{C}{R^5(t) \bar{n}} Q^2(t).$$

Integrating over $[0, T_0]$ we obtain

$$\frac{1}{Q(0)} - \frac{1}{Q(T_0)} \geq \frac{C}{\bar{n}\bar{\eta}} [1 - \frac{1}{(1 + \bar{\eta} T_0)^4}]. \tag{3.9}$$

We can then choose $u_0(r)$ sufficiently large, so that $Q(0)$ is so large to contradict (3.9).

□

Acknowledgments

We would like to thank S. Cordier and the anonymous referee for their helpful comments, and D. Christodoulou for his lucid presentation of the results of Sideris, which provided us with with a starting point for this project.

References

1. D. Christodoulou, *Self-gravitating relativistic fluids: A two-phase model*, Arch. Rational Mech. Anal. **130** (1995), 343–400.
2. S. Engelberg, *Formation of singularities in the Euler and Euler-Poisson equations*, Physica D. **98** (1996), 67–74.
3. R. T. Glassey, *On the blowing up of solutions to the Cauchy problem for nonlinear Schrödinger equations*, J. Math. Phys. **18** (1977), no. 9, 1794–1797.
4. Y. Guo, *Smooth irrotational flows in the large to the Euler-Poisson system in* $\mathbb{R}^{3+1}$, Comm. Math Phys. (to appear).
5. J. D. Jackson, *Classical electrodynamics*, 2nd ed., Wiley, New York, 1975.
6. T. Makino and B. Perthame, *Sur les solutions à symétrie sphérique de l'equation d'Euler-Poisson pour l'evolution d'etoiles gazeuses*, Japan J. Appl. Math. **7** (1990), 165–170.
7. T. Makino, S. Ukai, and S. Kawashima, *Sur la solution à support compact de l'equation d'Euler compressible*, Japan J. Appl. Math. **3** (1986), 249–257.
8. B. Perthame, *Non-existence of global solutions to Euler-Poisson equations for repulsive forces*, Japan J. Appl. Math. **7** (1990), 363–367.
9. M. A. Rammaha, *On the formation of singularities in magnetohydrodynamic waves*, J. Math. Anal. Appl. **188** (1994), 940–955.
10. A. Rendall, *The initial value problem for self-gravitating fluid bodies*, Mathematical physics, X (Leipzig, 1991) (Berlin), Springer, Berlin, 1992, pp. 470–474.
11. ———, personal communication, 1998.
12. T. C. Sideris, *Formation of singularities in solutions to nonlinear hyperbolic equations*, Arch. Rat. Mech. Anal. **86** (1984), 369–381.
13. ———, *Formation of singularities in three-dimensional compressible fluids*, Comm. Math. Phys. **101** (1985), 475–485.
14. A. S. Tahvildar-Zadeh, *Relativistic and nonrelativistic elastodynamics with small shear strains*, Ann. Inst. H. Poincaré Phys. Théor. **69** (1998), no. 3.

Division of Applied Mathematics, Brown University, Providence RI 02912-1932
E-mail address: guoy@cfm.brown.edu

Department of Mathematics, Princeton University, Princeton NJ 08544
E-mail address: shadi@math.princeton.edu

Contemporary Mathematics
Volume **238**, 1999
B 0-8218-1196-7-03546-5

Asymptotic Stability of Plane Diffusion Waves for the 2-D Quasilinear Wave Equation

Corrado Lattanzio and Pierangelo Marcati

Abstract. In this paper we consider the asymptotic stability of the solutions to the nonlinear damped wave equation in 2-D of space. In particular we deal with initial data which are small perturbation (in Sobolev norms) of a self-similar plane diffusive profile which solve a related parabolic equation. The results are achieved by using the classical energy method and in addition we provide polynomial rates of convergences.

1. Introduction

The present paper is part of a general program of understanding the connections between nonlinear nonhomogeneous hyperbolic systems and nonlinear parabolic equations. Concerning these problems, there are several points of view which can be pieced together in order to have a good comprehension of the underlying dynamics. Here we are concerned with the large time behavior of the solutions to the following nonlinear wave equation with a frictional damping term

$$\begin{aligned} &w_{tt} - \operatorname{div}\left[\vartheta\left(|\nabla w|\right)\nabla w\right] + \alpha w_t = 0, \\ &t \geq 0, \quad (x,y) \in \mathbb{R}^2, \end{aligned} \tag{1.1}$$

where $\alpha > 0$, $w = w(x,y,t) \in \mathbb{R}$ and $\vartheta(\lambda) > 0$ is a smooth nonlinear function such that $\sigma(\lambda) = \vartheta(\lambda)\lambda$ satisfies $\sigma'(0) > 0$, $\sigma''(\lambda)\lambda > 0$ for any $\lambda \neq 0$. As usual, we denote

$$w_x(x,y,t) = v(x,y,t), \quad w_y(x,y,t) = m(x,y,t), \quad w_t(x,y,t) = u(x,y,t),$$

then the equation (1.1) can be reformulated as the following nonlinear hyperbolic system

$$\begin{cases} v_t - u_x = 0 \\ m_t - u_y = 0 \\ u_t - \operatorname{div}\left[\vartheta\left(|p|\right)p\right] = -\alpha u, \end{cases} \tag{1.2}$$

where $p = (v,m)$.

1991 *Mathematics Subject Classification.* Primary 35L65, 35L70.

Key words and phrases. Quasilinear wave equation, asymptotic stability, diffusion waves.

Both the authors are partially supported by TMR-Network "Hyperbolic Systems of Conservation Laws", contract nº ERB – FMRX – CT96 – 0033.

We consider solutions of (1.2) which are small perturbations in H^s of a plane diffusion wave obtained from the caloric self-similar solution to the 1-D parabolic system related with (1.2)

$$\begin{cases} \overline{v}_t - \overline{u}_x = 0 \\ \alpha \overline{u} = \left[\vartheta\left(\overline{v}\right)\overline{v}\right]_x, \end{cases} \tag{1.3}$$

with limiting conditions

$$\overline{v}(\pm\infty, t) = v_\pm, \quad \overline{u}(\pm\infty, t) = 0. \tag{1.4}$$

We wish to prove that these perturbed solutions to the full system (1.2) converge asymptotically, in higher order energy norms, to a related solution of the parabolic equation (1.3). Actually, this large-time dynamic is somehow decoupled into a typical 1-D phenomenon (see [**HL92, HL93, Nis96**]) and a more genuine 2-D convergence.

This type of analysis has been initiated by the previously mentioned papers of Hsiao and Liu [**HL92, HL93**] and later continued by Nishihara [**Nis96**] in one space dimension. All of these papers are based completely on the use of the classical energy techniques and they provide stability and polynomial decay rates. Recently, in [**MM**], it has been proved a related result concerning the initial-boundary value problem.

The asymptotic study for weak solutions of hyperbolic systems with damping has been carried out in [**MM90, MMS88**], by introducing an appropriate parabolic-type scaling and then by studying the related relaxation problem via the theory of compensated compactness. The general 2×2 case is treated in [**MR**], together with some multi-D results. Recently, the convergence obtained in [**MR**] in the general 2×2 case, which can be viewed as a convergence "in the mean", has been improved in [**LR97**] to an almost pointwise convergence, by following an idea of [**SX97**] for the p-system with linear damping. Related results for semiconductors hydrodynamic models have been obtained in [**MN95, Nat96, LM, Lat, JR**].

In the present paper, we prove that a 2-D perturbation of the plane wave $\overline{v}(x,t)$ converges as $t \uparrow +\infty$ with polynomial rates to the 1-D solution of [**HL92, Nis96**]. In particular, let us denote by $\widetilde{v}(x,t)$ this solution, our analysis is based on the splitting between the 1-D component and the 2-D component of the initial perturbation $\psi(x,y) = v(x,y,0) - \widetilde{v}(x,0)$, by using the condition

$$\int_{-\infty}^{+\infty} \psi(x,y)dx \equiv 0.$$

This zero-mean condition is necessary to avoid interactions between the one dimensional and the two dimensional dynamics, which could destabilize the convergence process.

In the next section, we will recall some properties of the self-similar solution of the parabolic equation [**HL92**] and of the solution of the 1-D hyperbolic problem [**HL92, Nis96**], which will be useful in the proof of the decay estimates.

The section 3 is devoted to prove the energy estimate which will show the convergence of the 2-D perturbation of $\widetilde{v}(x,t)$ as $t \uparrow +\infty$, thanks to the results of [**HL92, Nis96**], the convergence of the solutions of (1.2) toward the self-similar solutions of the parabolic system (1.3).

2. Statement of the Problem and Main Results

In this section we recall the main results regarding the 1-D problem [**HL92, PVD97, Nis96**]. Let us consider the nonlinear diffusion equation

$$f_t = -\frac{1}{\alpha}(\vartheta(f)f)_{xx}, \tag{2.1}$$

with the following conditions at $\pm\infty$

$$f(\pm\infty, t) = v_\pm, \quad v_+ > v_- > 0. \tag{2.2}$$

The problem (2.1)-(2.2) has a caloric self-similar solution

$$\overline{v}(x,t) = \varphi\left(\frac{x}{\sqrt{1+t}}\right), \quad \varphi(\pm\infty) = v_\pm, \ \varphi(\xi) > 0. \tag{2.3}$$

This solution verifies the inequalities [**HL92**]

$$\sum_{k=1}^{3}\left|\frac{d^k}{d\xi^k}\varphi(\xi)\right| + |\varphi(\xi) - v_+|_{\xi>0} + |\varphi(\xi) - v_-|_{\xi<0} \le C|v_+ - v_-|e^{-c\alpha\xi^2}, \tag{2.4}$$

and the pointwise decay estimates for all the derivatives of $\overline{v}$ can be easily obtained by differentiating (2.3) in x and t

$$\overline{v}_x = \frac{\varphi'(\xi)}{\sqrt{1+t}}, \qquad \overline{v}_t = -\frac{\xi\varphi'(\xi)}{2(1+t)}. \tag{2.5}$$

Then, let us consider a solution $(\widetilde{u}, \widetilde{v})$ of the 1-D system

$$\begin{cases} \widetilde{v}_t - \widetilde{u}_x = 0 \\ \widetilde{u}_t - \left[\vartheta\left(\widetilde{v}\right)\widetilde{v}\right]_x = -\alpha\widetilde{u}, \end{cases} \tag{2.6}$$

which verifies

$$\widetilde{v}(\pm\infty, 0) = v_\pm, \quad \widetilde{u}(\pm\infty, 0) = u_\pm. \tag{2.7}$$

Thus, it is known [**HL92, Nis96**] that the shift x_0 and the correctors $\widehat{u}$ and $\widehat{v}$ have the following expressions

$$\begin{aligned} x_0 &= \frac{u_+ - u_-}{\alpha(v_+ - v_-)} + \frac{1}{v_+ - v_-}\int_{-\infty}^{+\infty}\left(\widetilde{v}(x,0) - \overline{v}(x,0)\right)dx, \\ \widehat{u}(x,t) &= e^{-\alpha t}\left[u_+ + (u_+ - u_-)\int_{-\infty}^{x} m_0(\xi)d\xi\right], \\ \widehat{v} &= \frac{u_+ - u_-}{-\alpha}e^{-\alpha t}m_0(x), \end{aligned}$$

where m_0 is a nonnegative test function such that $\displaystyle\int_{-\infty}^{+\infty} m_0(x)dx = 1$. Let us denote

$$\begin{aligned} \widetilde{V}(x,t) &= \int_{-\infty}^{x}\left(\widetilde{v}(\xi,t) - \overline{v}(\xi + x_0, t) - \widehat{v}(\xi,t)\right)d\xi, \\ \widetilde{z}(x,t) &= \widetilde{u}(x,t) - \widehat{u}(x,t) - \overline{u}(x + x_0, t). \end{aligned}$$

With this notation, the problem (2.6)-(2.7) becomes

$$
\begin{cases}
\widetilde{V}_t - \widetilde{z} = 0 \\
\widetilde{z}_t - \left[\vartheta(\widetilde{V}_x + \overline{v} + \widehat{v})(\widetilde{V}_x + \overline{v} + \widehat{v}) - \vartheta(\overline{v})\overline{v}\right]_x + \alpha\widetilde{z} = -\overline{u}_t = -\dfrac{1}{\alpha}\left[\vartheta(\overline{v})\overline{v}\right]_x \\
\widetilde{V}(x,0) = \widetilde{V}_0(x) \\
\widetilde{z}(x,0) = \widetilde{z}_0(x) \\
\widetilde{V}_0(\pm\infty) = \widetilde{z}_0(\pm\infty) = 0.
\end{cases} \tag{2.8}
$$

Hence, the following theorem holds [**HL92, Nis96**].

THEOREM 2.1. *Suppose* $\delta = |v_+ - v_-| + |u_+ - u_-|$ *and* $\|\widetilde{V}_0\|_3 + \|\widetilde{z}_0\|_2$ *are sufficiently small. Then there exists a unique global solution* $(\widetilde{V}(x,t), \widetilde{z}(x,t))$ *to (2.8) which satisfies*

$$
\widetilde{V} \in W^{i,\infty}([0,+\infty); H^i), \quad i = 0, \dots, 3,
$$

and moreover

$$
\begin{aligned}
&\sum_{k=0}^{3}(1+t)^k\|\partial_x^k\widetilde{V}(\cdot,t)\|^2 + \sum_{k=0}^{2}(1+t)^{k+2}\|\partial_x^k\widetilde{z}(\cdot,t)\|^2 \\
&\quad + \int_0^t\left[\sum_{j=1}^{3}(1+\tau)^{j-1}\|\partial_x^j\widetilde{V}(\cdot,\tau)\|^2 + \sum_{j=0}^{2}(1+\tau)^{j+1}\|\partial_x^j\widetilde{z}(\cdot,\tau)\|^2\right]d\tau \\
&\quad \le C(\|\widetilde{V}_0\|_3^2 + \|\widetilde{z}_0\|_2^2 + \delta).
\end{aligned} \tag{2.9}
$$

REMARK 2.2. By using a recursive procedure, it is possible to improve the previous result when the initial data are more regular. In particular, the estimate (2.9) can be achieved for a larger k. For our purposes, we will assume $\|\widetilde{V}_0\|_8 + \|\widetilde{z}_0\|_7 \le \delta$ small enough to have

$$
\begin{aligned}
&\sum_{k=0}^{8}(1+t)^k\|\partial_x^k\widetilde{V}(\cdot,t)\|^2 + \sum_{k=0}^{7}(1+t)^{k+2}\|\partial_x^k\widetilde{z}(\cdot,t)\|^2 \\
&\quad + \int_0^t\left[\sum_{j=1}^{8}(1+\tau)^{j-1}\|\partial_x^j\widetilde{V}(\cdot,\tau)\|^2 + \sum_{j=0}^{7}(1+\tau)^{j+1}\|\partial_x^j\widetilde{z}(\cdot,\tau)\|^2\right]d\tau \\
&\quad \le C(\|\widetilde{V}_0\|_8^2 + \|\widetilde{z}_0\|_7^2 + \delta) \le C\delta.
\end{aligned} \tag{2.10}
$$

REMARK 2.3. We know that the solution $\overline{v}$ of (2.1)-(2.2) is positive. Due to (2.10) and due to the expression of the corrector $\widehat{v}$, the difference $\widetilde{v} - \overline{v}$ is $O(\delta)$. Therefore, for δ sufficiently small, we have $\widetilde{v} > 0$.

Now we analyze the 2-D perturbation of this 1-D solution. Let us consider the following 2-D system given by the wave equation (1.1)

$$
\begin{cases}
v_t - u_x = 0 \\
m_t - u_y = 0 \\
u_t - \operatorname{div}\left[\vartheta\left(|p|\right)p\right] = -\alpha u,
\end{cases} \tag{2.11}
$$

where $p = (v, m)$. We choose the initial data $v(x, y, 0)$ and $\widetilde{v}(x, 0)$ so that

$$\int_{-\infty}^{+\infty} (v(x, y, 0) - \widetilde{v}(x, 0))\, dx = 0 \tag{2.12}$$

and we assume the following limiting conditions

$$\begin{array}{ll} v(\pm\infty, y, t) = v_{\pm}, & v(x, \pm\infty, t) = \widetilde{v}(x, t), \\ m(\pm\infty, y, t) = 0, & m(x, \pm\infty, t) = 0, \\ u(\pm\infty, y, t) = u_{\pm} e^{-\alpha t}, & u(x, \pm\infty, t) = \widetilde{u}(x, t). \end{array} \tag{2.13}$$

REMARK 2.4. The condition (2.12) implies in particular that the new perturbation due to the difference $v(x, y, 0) - \widetilde{v}(x, 0)$ does not affect the shift of the final plane wave. Therefore, the asymptotic profile of our 2-D solution is selected by the 1-D solution $\widetilde{v}$. This phenomenon provides a big advantage since it allows us to consider directly the convergence of the 2-D perturbation of $\widetilde{v}(x, t)$. Once we know this kind of convergence, we can simply make use of the estimate (2.10) to show the asymptotic behavior of the 2-D solution.

As in the 1-D case, we introduce a new set of variables

$$\begin{aligned} V(x, y, t) &= \int_{-\infty}^{x} (v(\xi, y, t) - \widetilde{v}(\xi, t))\, d\xi \\ M(x, y, t) &= \int_{-\infty}^{y} m(x, \eta, t) d\eta \\ z(x, y, t) &= u(x, y, t) - \widetilde{u}(x, t), \end{aligned}$$

and the problem (2.11)-(2.13) can be rewritten as follows

$$\left\{ \begin{array}{l} V_t = z \\ M(x, y, t) = V(x, y, t) + \displaystyle\int_0^t \widetilde{u}(x, s) ds + M(x, y, 0) - V(x, y, 0) \\ z_t - \operatorname{div} \left[\vartheta \left(\left| \begin{array}{c} V_x + \widetilde{v} \\ V_y \end{array} \right| \right) \left(\begin{array}{c} V_x + \widetilde{v} \\ V_y \end{array} \right) + \vartheta(\widetilde{v}) \left(\begin{array}{c} \widetilde{v} \\ 0 \end{array} \right) \right] + \alpha z = 0, \end{array} \right. \tag{2.14}$$

with the limiting conditions

$$V(\pm\infty, y, 0) = 0, \quad V(x, \pm\infty, 0) = 0, \quad z(\pm\infty, y, 0) = 0, \quad z(x, \pm\infty, 0) = 0. \tag{2.15}$$

Now we can state our main theorem. We recall that

$$\|f\| = \left(\int \int |f(x, y)|^2 dx dy \right)^{\frac{1}{2}}$$

denotes the classical $L^2(\mathbb{R}^2)$ norm and the Sobolev norm is given by

$$\|f\|_s = \left(\sum_{j=0}^{s} \int \int |D^j f(x, y)|^2 dx dy \right)^{\frac{1}{2}},$$

where D^j is any differential operator of the form $\dfrac{\partial^{j_1}}{\partial x^{j_1}} \dfrac{\partial^{j_2}}{\partial y^{j_2}}$ with $j_1 + j_2 = j$.

THEOREM 2.5. *Suppose δ and $\|V_0\|_7 + \|z_0\|_6$ are sufficiently small. Then there exists a unique global solution $(V(x, y, t), z(x, y, t))$ to (2.14)-(2.15) which satisfies*

$$V \in W^{i,\infty}([0, +\infty); H^{7-i}), \quad i = 0, \ldots, 7,$$

and moreover

$$\sum_{k=0}^{7}(1+t)^k\|D^kV(\cdot,t)\|^2+\sum_{k=0}^{6}(1+t)^{k+2}\|D^kz(\cdot,t)\|^2$$
$$+\int_0^t\left[\sum_{j=1}^{7}(1+\tau)^{j-1}\|D^jV(\cdot,\tau)\|^2+\sum_{j=0}^{6}(1+\tau)^{j+1}\|D^jz(\cdot,\tau)\|^2\right]d\tau$$
$$=O(1)\left(\|V_0\|_7^2+\|z_0\|_6^2+\delta\right). \tag{2.16}$$

Remark 2.6. In view of the Sobolev embeddings, the estimate (2.16) of theorem 2.5 and the estimate (2.10) of remark 2.2 imply that the C^4 norm of $V(x,y,t)$ and $\widetilde{V}(x,t)$ decays in time with polynomial rates. Therefore, the same kind of C^4 convergence holds also for the quantity

$$\mathcal{V}(x,y,t)=\int_{-\infty}^{x}\left(v(\xi,y,t)-\overline{v}(\xi+x_0,t)-\widehat{v}(\xi,t)\right)d\xi.$$

Hence, the full 2-D solution converges toward the plane wave with the same rates established in [**HL92, Nis96**] for the 1-D problem.

3. The Proof of the Main Theorem

In this section we deal with the proof of the theorem 2.5, namely, the proof of the asymptotic behavior (2.16). We achieve this result by using energy methods, together with a continuation principle. As usual in this framework, we start with an a priori assumption

$$N(T)=\sup_{0<t<T}\left\{\sum_{k=0}^{7}(1+t)^k\left[\|\partial_x^kV(\cdot,t)\|^2+\|\partial_y^kV(\cdot,t)\|^2\right]\right.$$
$$\left.+\sum_{k=0}^{6}(1+t)^{k+2}\left[\|\partial_x^kz(\cdot,t)\|^2+\|\partial_y^kz(\cdot,t)\|^2\right]\right\}\le\varepsilon. \tag{3.1}$$

Let us use the following notations

$$\vartheta=\vartheta\left(\left|\begin{array}{c}V_x+\widetilde{v}\\V_y\end{array}\right|\right),\quad\widetilde{\vartheta}=\vartheta(\widetilde{v}).$$

In the next lemma, we establish some useful properties of the nonlinear function ϑ.

Lemma 3.1. *Let ϑ be a smooth function such that $\sigma(\lambda)=\vartheta(\lambda)\lambda$ satisfies $\sigma'(0)>0$ and $\sigma''(\lambda)\lambda>0$ for any $\lambda\neq0$. Then $\vartheta'(\lambda)\lambda>0$ for any $\lambda\neq0$.*

The following lemma concerns the bound of the H^1 norm of the solution V.

Lemma 3.2. *Suppose ε, δ and $\|V_0\|_7^2+\|z_0\|_6^2$ are sufficiently small. Then*

$$\|V(t)\|_1^2+\|z(t)\|^2+\int_0^t\left(\|V_x(\tau)\|^2+\|V_y(\tau)\|^2+\|z(\tau)\|^2\right)d\tau$$
$$=O(1)\left(\|V_0\|_1^2+\|z_0\|^2+\delta\right). \tag{3.2}$$

PROOF. The system (2.14) can be rewritten as a hyperbolic equation for the function V

$$V_{tt} - \operatorname{div}\left[\vartheta \begin{pmatrix} V_x + \widetilde{v} \\ V_y \end{pmatrix} - \widetilde{\vartheta} \begin{pmatrix} \widetilde{v} \\ 0 \end{pmatrix}\right] + \alpha V_t = 0, \tag{3.3}$$

which can be linearized as follows

$$V_{tt} - \operatorname{div}\left[\widetilde{\vartheta} \begin{pmatrix} V_x \\ V_y \end{pmatrix}\right] + \alpha V_t = \operatorname{div}\left[(\vartheta - \widetilde{\vartheta}) \begin{pmatrix} V_x + \widetilde{v} \\ V_y \end{pmatrix}\right] = F. \tag{3.4}$$

Multiplying (3.4) for $V_t + \lambda V$ and integrating on $dxdy$ one has

$$\begin{aligned} &\frac{1}{2}\frac{d}{dt}\int\int \left[V_t^2 + 2\lambda V V_t + \widetilde{\vartheta}(V_x^2 + V_y^2) + \alpha\lambda V^2\right] dxdy \\ &+ \int\int \left[(\alpha - \lambda)V_t^2 + \lambda\widetilde{\vartheta}(V_x^2 + V_y^2)\right] dxdy \\ &= \frac{1}{2}\int\int \widetilde{\vartheta}_t(V_x^2 + V_y^2)dxdy + \int\int (V_t + \lambda V)F dxdy. \end{aligned} \tag{3.5}$$

The left hand side of the above relation clearly gives the quantity we have to estimate, once we control the product VV_t in terms of V^2 and V_t^2, which is possible by choosing an appropriate small value for the constant λ. Therefore, we have to estimate the right hand side of (3.5) to conclude the proof. We start by investigating the term $\int\int zF dxdy$. To this end, we introduce the function $H : \mathbb{R}^2 \longrightarrow \mathbb{R}$ defined by

$$H(|P|) = \int_{\widetilde{v}}^{|P|} s\vartheta(s)ds, \quad P \in \mathbb{R}^2.$$

With the notation

$$G = \begin{pmatrix} V_x + \widetilde{v} \\ V_y \end{pmatrix},$$

this term becomes

$$\begin{aligned} \int\int zF dxdy &= -\int\int \begin{pmatrix} V_x \\ V_y \end{pmatrix}_t \cdot \left[\vartheta \begin{pmatrix} V_x + \widetilde{v} \\ V_y \end{pmatrix} - \widetilde{\vartheta} \begin{pmatrix} \widetilde{v} \\ 0 \end{pmatrix}\right. \\ &\quad \left. - \widetilde{\vartheta} \begin{pmatrix} V_x \\ V_y \end{pmatrix}\right] dxdy \\ &= -\frac{d}{dt}\int\int \left[H(|G|) - \widetilde{\vartheta}\widetilde{v}V_x - \frac{1}{2}\widetilde{\vartheta}(V_x^2 + V_y^2)\right] dxdy \\ &\quad + \int\int \begin{pmatrix} \widetilde{v} \\ 0 \end{pmatrix}_t \cdot \left[\vartheta \begin{pmatrix} V_x + \widetilde{v} \\ V_y \end{pmatrix} - \widetilde{\vartheta} \begin{pmatrix} \widetilde{v} \\ 0 \end{pmatrix}\right] dxdy \\ &\quad - \int\int V_x(\widetilde{\vartheta}_t\widetilde{v} - \widetilde{v}_t\widetilde{\vartheta})dxdy - \frac{1}{2}\int\int \widetilde{\vartheta}_t(V_x^2 + V_y^2)dxdy \\ &= -\frac{d}{dt}\int\int \left[H(|G|) - \widetilde{\vartheta}\widetilde{v}V_x - \frac{1}{2}\widetilde{\vartheta}(V_x^2 + V_y^2)\right] dxdy \\ &\quad - \frac{1}{2}\int\int \widetilde{\vartheta}_t(V_x^2 + V_y^2)dxdy + \int\int \begin{pmatrix} \widetilde{v} \\ 0 \end{pmatrix}_t \cdot \left[\vartheta \begin{pmatrix} V_x + \widetilde{v} \\ V_y \end{pmatrix}\right. \\ &\quad \left. - \widetilde{\vartheta} \begin{pmatrix} \widetilde{v} \\ 0 \end{pmatrix} - \widetilde{\vartheta}' \begin{pmatrix} \widetilde{v} \\ 0 \end{pmatrix} V_x - \widetilde{\vartheta} \begin{pmatrix} V_x \\ V_y \end{pmatrix}\right] dxdy. \end{aligned}$$

With the above equality, we can rewrite the right hand side of (3.5) as follows

$$
\begin{aligned}
&-\lambda\int\int(\vartheta-\widetilde{\vartheta})\begin{pmatrix}V_x\\V_y\end{pmatrix}\cdot\begin{pmatrix}V_x+\widetilde{v}\\V_y\end{pmatrix}dxdy\\
&-\frac{d}{dt}\int\int\left[H(|G|)-\widetilde{\vartheta}\widetilde{v}V_x-\frac{1}{2}\widetilde{\vartheta}(V_x^2+V_y^2)\right]dxdy\\
&+\int\int\begin{pmatrix}\widetilde{v}\\0\end{pmatrix}_t\cdot\left[\vartheta\begin{pmatrix}V_x+\widetilde{v}\\V_y\end{pmatrix}-\widetilde{\vartheta}\begin{pmatrix}\widetilde{v}\\0\end{pmatrix}-\widetilde{\vartheta}'\begin{pmatrix}\widetilde{v}\\0\end{pmatrix}V_x\right.\\
&\left.-\widetilde{\vartheta}\begin{pmatrix}V_x\\V_y\end{pmatrix}\right]dxdy=I_1+I_2+I_3. \qquad (3.6)
\end{aligned}
$$

Thus, we have to estimate the terms $I_1+I_2+I_3$ in (3.6). Since

$$\vartheta-\widetilde{\vartheta}=\widetilde{\vartheta}'V_x+O\left(V_x^2+V_y^2\right),$$

combining the result of lemma 3.1 and the a priori assumption (3.1), we get

$$
\begin{aligned}
I_1&=-\lambda O(1)\int\int\widetilde{\vartheta}'\widetilde{v}V_x^2dxdy+\lambda\varepsilon O(1)\int\int(V_x^2+V_y^2)dxdy\\
&\leq\lambda\varepsilon O(1)\int\int(V_x^2+V_y^2)dxdy.
\end{aligned}
$$

The Taylor expansion of the functions $H(|P|)$ and $\vartheta(|P|)P$ yields

$$H(|G|)-\widetilde{\vartheta}\widetilde{v}V_x-\frac{1}{2}\widetilde{\vartheta}(V_x^2+V_y^2)=\frac{1}{2}\widetilde{\vartheta}'\widetilde{v}V_x^2+O\left(V_x^3+V_y^3\right) \qquad (3.7)$$

and

$$\vartheta\begin{pmatrix}V_x+\widetilde{v}\\V_y\end{pmatrix}-\widetilde{\vartheta}\begin{pmatrix}\widetilde{v}\\0\end{pmatrix}-\widetilde{\vartheta}'\begin{pmatrix}\widetilde{v}\\0\end{pmatrix}V_x-\widetilde{\vartheta}\begin{pmatrix}V_x\\V_y\end{pmatrix}=O\left(V_x^2+V_y^2\right).$$

Hence, in view of (2.4), (2.5), (2.10) and (3.1), the previous relation implies

$$|I_3|=O(1)\int\int|\widetilde{v}_t|\left(|V_x|^2+|V_y^2|\right)dxdy=\frac{O(1)\delta}{(1+t)^2}.$$

Finally, integrating (3.7) in dt and using again lemma 3.1 and (3.1), we get

$$
\begin{aligned}
\int_0^t I_2ds&=\frac{1}{2}\int\int\widetilde{\vartheta}'\widetilde{v}V_x^2dxdy+\varepsilon O(1)\int\int\left(V_x^2+V_y^2\right)dxdy+O(1)\|V_0\|_1\\
&\leq\varepsilon O(1)\int\int\left(V_x^2+V_y^2\right)dxdy+O(1)\|V_0\|_1.
\end{aligned}
$$

Therefore, integrating (3.5) in dt, for δ, ε and λ small enough, it follows (3.2). □

The previous lemma gives a bound of the H^1 norm of V, without any decay property.We can improve the estimate (3.2) by showing the first convergence result for the functions V and z. The proof of such property is based essentially on the decays of the 1-D solution $\widetilde{v}$ contained in (2.4), (2.5) and (2.10).

LEMMA 3.3. *Suppose ε, δ and $\|V_0\|_7^2+\|z_0\|_6^2$ are sufficiently small. Then*

$$
\begin{aligned}
(1+t)\left(\|V_x(t)\|^2+\|V_y(t)\|^2+\|z(t)\|^2\right)+\int_0^t(1+\tau)\|z(\tau)\|^2d\tau\\
=O(1)\left(\|V_0\|_1^2+\|z_0\|^2+\delta\right).
\end{aligned}
$$

PROOF. We multiply the linearized equation (3.4) by $(1+t)V_t$. Therefore, after integrating in $dxdy$ we obtain

$$\begin{aligned}&\frac{1}{2}\frac{d}{dt}\int\int(1+t)\left[V_t^2+\widetilde{\vartheta}(V_x^2+V_y^2)\right]dxdy+\int\int\alpha(1+t)V_t^2dxdy\\&=\frac{1}{2}\int\int(1+t)\widetilde{\vartheta}_t(V_x^2+V_y^2)dxdy+\frac{1}{2}\int\int\left[V_t^2+\widetilde{\vartheta}(V_x^2+V_y^2)\right]dxdy\\&+\int\int(1+t)V_tFdxdy=I_1+I_2+I_3.\end{aligned}\tag{3.8}$$

The results of lemma 3.2 yield

$$I_2=O(1)\left(\|V_0\|_1^2+\|z_0\|^2+\delta\right).$$

With the previous notations, we have

$$\begin{aligned}I_3=&-\frac{d}{dt}\int\int(1+t)\left[H(|G|)-\widetilde{\vartheta}\widetilde{v}V_x-\frac{1}{2}\widetilde{\vartheta}(V_x^2+V_y^2)\right]dxdy\\&+\int\int\left[H(|G|)-\widetilde{\vartheta}\widetilde{v}V_x-\frac{1}{2}\widetilde{\vartheta}(V_x^2+V_y^2)\right]dxdy\\&-\frac{1}{2}(1+t)\int\int\widetilde{\vartheta}_t(V_x^2+V_y^2)dxdy+\int\int(1+t)\begin{pmatrix}\widetilde{v}\\0\end{pmatrix}_t\cdot\left[\vartheta\begin{pmatrix}V_x+\widetilde{v}\\V_y\end{pmatrix}\right.\\&\left.-\widetilde{\vartheta}\begin{pmatrix}\widetilde{v}\\0\end{pmatrix}-\widetilde{\vartheta}'\begin{pmatrix}\widetilde{v}\\0\end{pmatrix}V_x-\widetilde{\vartheta}\begin{pmatrix}V_x\\V_y\end{pmatrix}\right]dxdy.\end{aligned}$$

Therefore,

$$\begin{aligned}I_1+I_3=&-\frac{d}{dt}\int\int(1+t)\left[H(|G|)-\widetilde{\vartheta}\widetilde{v}V_x-\frac{1}{2}\widetilde{\vartheta}(V_x^2+V_y^2)\right]dxdy\\&+\int\int\left[H(|G|)-\widetilde{\vartheta}\widetilde{v}V_x-\frac{1}{2}\widetilde{\vartheta}(V_x^2+V_y^2)\right]dxdy\\&+\int\int(1+t)\begin{pmatrix}\widetilde{v}\\0\end{pmatrix}_t\cdot\left[\vartheta\begin{pmatrix}V_x+\widetilde{v}\\V_y\end{pmatrix}-\widetilde{\vartheta}\begin{pmatrix}\widetilde{v}\\0\end{pmatrix}\right.\\&\left.-\widetilde{\vartheta}'\begin{pmatrix}\widetilde{v}\\0\end{pmatrix}V_x-\widetilde{\vartheta}\begin{pmatrix}V_x\\V_y\end{pmatrix}\right]dxdy=J_1+J_2+J_3.\end{aligned}$$

Moreover, proceeding as in the proof of lemma 3.2, we have

$$\int_0^t J_1ds\le O(1)\varepsilon\int\int(1+t)\left(V_x^2+V_y^2\right)dxdy+O(1)\|V_0\|_1;$$

$$J_2\le O(1)\varepsilon\int\int\left(V_x^2+V_y^2\right)dxdy.$$

Thus, (3.2) implies

$$\int_0^t J_2ds=O(1)\left(\|V_0\|_1^2+\|z_0\|^2+\delta\right).$$

Finally, the last term can be bounded by using again (3.2) and the decay of $\widetilde{v}_t$

$$\begin{aligned}J_3&=O(1)\int\int(1+t)|\widetilde{v}_t|\left(V_x^2+V_y^2\right)dxdy=O(1)\int\int\left(V_x^2+V_y^2\right)dxdy\\&=O(1)\left(\|V_0\|_1^2+\|z_0\|^2+\delta\right).\end{aligned}$$

As before, we conclude the proof integrating (3.8) in dt and choosing δ, ε and λ small enough. □

Now we turn to the study of the estimates for the higher derivatives of V and z. In the next lemma, we prove the first H^2 result, regarding essentially the x and y derivatives of V and z.

LEMMA 3.4. *Suppose ε, δ and $\|V_0\|_7^2 + \|z_0\|_6^2$ are sufficiently small. Then*

$$\begin{aligned}
&(1+t)^2(\|V_{xx}(t)\|^2 + \|V_{xy}(t)\|^2 + \|V_{yy}(t)\|^2 + \|z_x(t)\|^2 + \|z_y(t)\|^2)\\
&+ \int_0^t (1+\tau)\left[\|V_{xx}(\tau)\|^2 + \|V_{xy}(\tau)\|^2 + \|V_{yy}(\tau)\|^2\right] d\tau\\
&+ \int_0^t (1+\tau)^2\left[\|z_x(\tau)\|^2 + \|z_y(\tau)\|^2\right] d\tau\\
&= O(1)(\|V_0\|_2^2 + \|z_0\|_1^2 + \delta).
\end{aligned}$$

PROOF. We start by differentiating the linearized equation (3.4) in x and y in order to have

$$\mathcal{Z}_{tt} - \operatorname{div}\left[\widetilde{\vartheta}\begin{pmatrix}\mathcal{Z}_x\\ \mathcal{Z}_y\end{pmatrix}\right] + \alpha\mathcal{Z}_t = F_x + \operatorname{div}\left[\widetilde{\vartheta}_x\begin{pmatrix}\mathcal{Z}\\ \mathcal{W}\end{pmatrix}\right] \tag{3.9}$$

and

$$\mathcal{W}_{tt} - \operatorname{div}\left[\widetilde{\vartheta}\begin{pmatrix}\mathcal{W}_x\\ \mathcal{W}_y\end{pmatrix}\right] + \alpha\mathcal{W}_t = F_y, \tag{3.10}$$

where $\mathcal{Z} = V_x$ and $\mathcal{W} = V_y$. We multiply (3.9) for $\mathcal{Z}_t$ and we integrate on $dxdy$ and we obtain

$$\begin{aligned}
&\frac{1}{2}\frac{d}{dt}\int\int\left[\mathcal{Z}_t^2 + \widetilde{\vartheta}\left(\mathcal{Z}_x^2 + \mathcal{Z}_y^2\right)\right]dxdy + \alpha\int\int\mathcal{Z}_t^2 dxdy\\
&= \frac{1}{2}\int\int\widetilde{\vartheta}_t\left(\mathcal{Z}_x^2 + \mathcal{Z}_y^2\right)dxdy + \int\int\mathcal{Z}_t F_x dxdy\\
&+ \int\int\mathcal{Z}_t\operatorname{div}\left[\widetilde{\vartheta}_x\begin{pmatrix}\mathcal{Z}\\ \mathcal{W}\end{pmatrix}\right]dxdy = I_1 + I_2 + I_3.
\end{aligned} \tag{3.11}$$

Due to (2.4), (2.5) and (2.10), the first term is estimated as follows

$$|I_1| = O(1)\delta\int\int\left(\mathcal{Z}_x^2 + \mathcal{Z}_y^2\right)dxdy.$$

Moreover, the last term can be bounded by using also the Young inequality

$$\begin{aligned}
I_3 &= \int\int\left[\mathcal{Z}_t\left(\widetilde{\vartheta}_{xx}V_x + \widetilde{\vartheta}_x\mathcal{Z}_x + \widetilde{\vartheta}_x\mathcal{W}_y\right)\right]dxdy \le E_\alpha\int\int\mathcal{Z}_t^2 dxdy\\
&+ O(1)(1+t)^{-2}\int\int V_x^2 dxdy + O(1)\delta\int\int\left(\mathcal{Z}_x^2 + \mathcal{W}_y^2\right)dxdy,
\end{aligned}$$

where E_α is a small positive constant (depending only on α) which will be chosen afterwards. Now, let us consider the second term in the right-hand-side of (3.11).

Integration by parts yields

$$\begin{aligned} I_2 &= -\int\int \begin{pmatrix} V_{xxt} \\ v_{xyt} \end{pmatrix} \cdot \left((\vartheta - \widetilde{\vartheta}) \begin{pmatrix} V_x + \widetilde{v} \\ V_y \end{pmatrix} \right)_x dxdy \\ &= -\int\int V_{xxt} \left[(\vartheta - \widetilde{\vartheta})_x (V_x - \widetilde{v}) + (\vartheta - \widetilde{\vartheta})(V_{xx} - \widetilde{v}_x) \right] dxdy \\ &\quad - \int\int V_{xyt} \left[(\vartheta - \widetilde{\vartheta})_x V_y + (\vartheta - \widetilde{\vartheta}) V_{xy} \right] dxdy \\ &= J_1 + J_2 + J_3 + J_4. \end{aligned}$$

Developing the x-derivative of $\vartheta - \widetilde{\vartheta}$, J_1 becomes

$$\begin{aligned} J_1 = -\int\int V_{xxt} \Bigg[&\frac{\vartheta'(V_x - \widetilde{v})}{|G|} V_{xx} + \frac{\vartheta' V_y}{|G|} V_{xy} \\ &+ \left(\frac{\vartheta'(V_x - \widetilde{v})}{|G|} - \widetilde{\vartheta}' \right) \widetilde{v}_x \Bigg] (V_x - \widetilde{v}) dxdy, \end{aligned}$$

where G represent again the vector $\begin{pmatrix} V_x + \widetilde{v} \\ V_y \end{pmatrix}$. We examine the terms in J_1 one by one.

$$\begin{aligned} -\int\int V_{xxt} \frac{\vartheta'(V_x - \widetilde{v})^2}{|G|} V_{xx} dxdy = &-\frac{1}{2} \frac{d}{dt} \int\int \frac{\vartheta'(V_x - \widetilde{v})^2}{|G|} \mathcal{Z}_x^2 dxdy \\ &+ O(1)(\varepsilon + \delta) \int\int \mathcal{Z}_x^2 dxdy. \end{aligned}$$

We emphasize that the total derivative with respect to t in the above relation has the "right" sign for ε small enough, because

$$\left. \frac{\vartheta'(V_x - \widetilde{v})^2}{|G|} \right|_{V_x = V_y = 0} = \widetilde{\vartheta}' \widetilde{v} > 0,$$

thanks to lemma 3.1. The last term in J_1 can be treated as follows

$$\begin{aligned} &-\int\int V_{xxt} (V_x - \widetilde{v}) \left(\frac{\vartheta'(V_x - \widetilde{v})}{|G|} - \widetilde{\vartheta}' \right) \widetilde{v}_x dxdy \\ &= \int\int V_{xt} (V_x - \widetilde{v})_x \left(\frac{\vartheta'(V_x - \widetilde{v})}{|G|} - \widetilde{\vartheta}' \right) \widetilde{v}_x dxdy \\ &\quad + \int\int V_{xt} (V_x - \widetilde{v}) \left(\frac{\vartheta'(V_x - \widetilde{v})}{|G|} - \widetilde{\vartheta}' \right) \widetilde{v}_{xx} dxdy \\ &\quad + \int\int V_{xt} (V_x - \widetilde{v}) \left(\frac{\vartheta'(V_x - \widetilde{v})}{|G|} - \widetilde{\vartheta}' \right)_x \widetilde{v}_x dxdy. \end{aligned} \tag{3.12}$$

Since

$$\frac{\vartheta'(V_x - \widetilde{v})}{|G|} - \widetilde{\vartheta}' = O\left(|V_x| + |V_y| \right),$$

in view of (2.4), (2.5) and (2.10) and using the Young inequality, the first two terms in (3.12) are bounded by

$$\begin{aligned} O(1) \int\int |\mathcal{Z}_t| |V_x| \left(|\widetilde{v}_{xx}| + |\widetilde{v}_x| |V_{xx} + \widetilde{v}_x| \right) dxdy \le E_\alpha \int\int \mathcal{Z}_t^2 dxdy \\ + O(1)(1+t)^{-2} \int\int V_x^2 dxdy. \end{aligned}$$

Evaluating the x-derivative in the last part of (3.12) we prove that this quantity is controlled by

$$
\begin{aligned}
&O(1)\int\int |\mathcal{Z}_t||\widetilde{v}_x|\left[|V_{xx}|+|\widetilde{v}_x|\left(|V_x|+|V_y|\right)+|V_y||V_{xy}|+|V_{xx}|\left(|V_x|+|V_y|\right)\right]dxdy\\
&\leq E_\alpha\int\int \mathcal{Z}_t^2 dxdy+O(1)(1+t)^{-2}\int\int \left(V_x^2+V_y^2\right)dxdy\\
&+O(1)\delta\int\int \left(\mathcal{Z}_x^2+\mathcal{Z}_y^2\right)dxdy.
\end{aligned}
$$

The remaining term of J_1

$$
-\int\int V_{xxt}\frac{\vartheta'(V_x+\widetilde{v})}{|G|}V_yV_{xy}dxdy \tag{3.13}
$$

can not be bounded for the moment: it will become a part of a total derivative with respect to t. Let us turn now on

$$
J_2=\int\int V_{xxt}(\vartheta-\widetilde{\vartheta})V_{xx}dxdy-\int\int V_{xxt}(\vartheta-\widetilde{\vartheta})\widetilde{v}_xdxdy.
$$

An integration by part in the last term gives

$$
\begin{aligned}
&\int\int \mathcal{Z}_t\left[(\vartheta-\widetilde{\vartheta})_x\widetilde{v}_x+(\vartheta-\widetilde{\vartheta})\widetilde{v}_{xx}\right]dxdy\\
&=O(1)\int\int |\mathcal{Z}_t|\left[|\widetilde{v}_x||V_{xx}|+|\widetilde{v}_x||V_y||V_{xy}|+|\widetilde{v}_x|^2\left(|V_x|+|V_y|\right)\right.\\
&\left.+\ |\widetilde{v}_x||\widetilde{v}_{xx}|\left(|V_x|+|V_y|\right)\right]dxdy\\
&\leq E_\alpha\int\int \mathcal{Z}_t^2dxdy+O(1)(1+t)^{-2}\int\int \left(V_x^2+V_y^2\right)dxdy\\
&+O(1)\delta\int\int \left(\mathcal{Z}_x^2+\mathcal{Z}_y^2\right)dxdy,
\end{aligned}
$$

by using also $\vartheta-\widetilde{\vartheta}=O\left(|V_x|+|V_y|\right)$. Moreover, the first term is equal to

$$
\begin{aligned}
&-\frac{1}{2}\frac{d}{dt}\int\int(\vartheta-\widetilde{\vartheta})\mathcal{Z}_x^2dxdy+\frac{1}{2}\int\int(\vartheta-\widetilde{\vartheta})_t\mathcal{Z}_x^2dxdy\\
&=-\frac{1}{2}\frac{d}{dt}\int\int(\vartheta-\widetilde{\vartheta})\mathcal{Z}_x^2dxdy+O(1)(\varepsilon+\delta)\int\int\mathcal{Z}_x^2dxdy.
\end{aligned}
$$

Proceeding in the same way, we bound J_4

$$
\begin{aligned}
J_4&=-\frac{1}{2}\frac{d}{dt}\int\int(\vartheta-\widetilde{\vartheta})\mathcal{Z}_y^2dxdy+\frac{1}{2}\int\int(\vartheta-\widetilde{\vartheta})_t\mathcal{Z}_y^2dxdy\\
&=-\frac{1}{2}\frac{d}{dt}\int\int(\vartheta-\widetilde{\vartheta})\mathcal{Z}_y^2dxdy+O(1)(\varepsilon+\delta)\int\int\mathcal{Z}_y^2dxdy.
\end{aligned}
$$

Finally, the last term is

$$
\begin{aligned}
J_3=-\int\int V_{xyt}&\left[\frac{\vartheta'(V_x-\widetilde{v})}{|G|}V_{xx}+\frac{\vartheta'V_y}{|G|}V_{xy}\right.\\
&\left.+\left(\frac{\vartheta'(V_x-\widetilde{v})}{|G|}-\widetilde{\vartheta}'\right)\widetilde{v}_x\right]V_ydxdy.
\end{aligned}
$$

As before, the term

$$-\int\int V_{xyt}\frac{\vartheta'(V_x-\widetilde{v})}{|G|}V_{xx}V_y dxdy \tag{3.14}$$

will be considered later. By using arguments similar to the previous ones, we get

$$\begin{aligned}
-\int\int V_{xyt}\frac{\vartheta' V_y^2}{|G|}V_{xy}dxdy &= -\frac{1}{2}\frac{d}{dt}\int\int \frac{\vartheta' V_y^2}{|G|}\mathcal{Z}_y^2 dxdy \\
&\quad + O(1)(\varepsilon+\delta)\int\int \mathcal{Z}_y^2 dxdy \\
&\quad -\int\int V_{xyt}V_y\widetilde{v}_x\left(\frac{\vartheta'(V_x+\widetilde{v})}{|G|}-\widetilde{\vartheta}'\right)dxdy \\
&= -\int\int V_{xt}V_{yy}\widetilde{v}_x\left(\frac{\vartheta'(V_x+\widetilde{v})}{|G|}-\widetilde{\vartheta}'\right)dxdy \\
&\quad -\int\int V_{xt}V_y\widetilde{v}_x\left(\frac{\vartheta'(V_x+\widetilde{v})}{|G|}-\widetilde{\vartheta}'\right)_y dxdy.
\end{aligned} \tag{3.15}$$

As before, the first term of (3.15) is bounded by

$$\begin{aligned}
O(1)\int\int |\mathcal{Z}_t||\widetilde{v}_x||\mathcal{W}_y|\left(|V_x|+|V_y|\right)dxdy &\le E_\alpha\int\int \mathcal{Z}_t^2 dxdy \\
&\quad + O(1)\delta\int\int \mathcal{W}_y^2 dxdy,
\end{aligned}$$

while the second is studied by developing the y-derivative

$$\begin{aligned}
&\int\int V_{xt}V_y\widetilde{v}_x\left(\frac{\vartheta'(V_x+\widetilde{v})}{|G|}-\widetilde{\vartheta}'\right)_y dxdy \\
&= O(1)\int\int |\mathcal{Z}_t||\widetilde{v}_x||V_y|\left(|\mathcal{Z}_y|+|\mathcal{W}_y||V_y|\right)dxdy \\
&\le E_\alpha\int\int \mathcal{Z}_t^2 dxdy + O(1)\delta\int\int\left(\mathcal{Z}_y^2+\mathcal{W}_y^2\right)dxdy.
\end{aligned}$$

Grouping together (3.13) and (3.14) we get

$$\begin{aligned}
&-\int\int \frac{\vartheta'(V_x-\widetilde{v})}{|G|}V_y\left(V_{xxt}V_{xy}+V_{xyt}V_{xx}\right)dxdy \\
&= -\frac{d}{dt}\int\int \frac{\vartheta'(V_x-\widetilde{v})}{|G|}V_y\mathcal{Z}_x\mathcal{Z}_y dxdy + O(1)(\varepsilon+\delta)\int\int\left(\mathcal{Z}_x^2+\mathcal{Z}_y^2\right)dxdy.
\end{aligned}$$

Therefore, the relation (3.11) becomes

$$\begin{aligned}
&\frac{1}{2}\frac{d}{dt}\int\int\left[\mathcal{Z}_t^2+\widetilde{\vartheta}\left(\mathcal{Z}_x^2+\mathcal{Z}_y^2\right)\right]dxdy+\alpha\int\int\mathcal{Z}_t^2 dxdy \\
&\le E_\alpha\int\int\mathcal{Z}_t^2 dxdy + O(1)(\varepsilon+\delta)\int\int\left(\mathcal{Z}_x^2+\mathcal{Z}_y^2+\mathcal{W}_y^2\right)dxdy \\
&\quad -\frac{1}{2}\frac{d}{dt}\int\int\left[\frac{\vartheta'(V_x-\widetilde{v})^2}{|G|}\mathcal{Z}_x^2+(\vartheta-\widetilde{\vartheta})(\mathcal{Z}_x^2+\mathcal{Z}_y^2)+\frac{\vartheta' V_y^2}{|G|}\mathcal{Z}_y^2\right. \\
&\quad \left.+2\frac{\vartheta'(V_x-\widetilde{v})}{|G|}V_y\mathcal{Z}_x\mathcal{Z}_y\right]dxdy + O(1)(1+t)^{-2}\int\int\left(V_x^2+V_y^2\right)dxdy.
\end{aligned} \tag{3.16}$$

We pass now to the estimates regarding the quantity $\mathcal{W} = V_y$. Multiplying (3.10) by $\mathcal{W}_t$ and integrating by part one has

$$\begin{aligned}&\frac{1}{2}\frac{d}{dt}\int\int\left[\mathcal{W}_t^2+\widetilde{\vartheta}\left(\mathcal{Z}_y^2+\mathcal{W}_y^2\right)\right]dxdy+\alpha\int\int\mathcal{W}_t^2dxdy\\&=\frac{1}{2}\int\int\widetilde{\vartheta}\left(\mathcal{Z}_y^2+\mathcal{W}_y^2\right)dxdy+\int\int\mathcal{W}_tF_ydxdy=I_1+I_2.\end{aligned}\tag{3.17}$$

As we did in the previous estimate, the first term is easily bounded in the following way

$$|I_1|=O(1)\delta\int\int\left(\mathcal{Z}_y^2+\mathcal{W}_y^2\right)dxdy.$$

Moreover,

$$\begin{aligned}I_2&=-\int\int\begin{pmatrix}V_{xyt}\\v_{yyt}\end{pmatrix}\cdot\left((\vartheta-\widetilde{\vartheta})\begin{pmatrix}V_x+\widetilde{v}\\V_y\end{pmatrix}\right)_y dxdy\\&=-\int\int V_{xyt}\left[\vartheta_y(V_x-\widetilde{v})+(\vartheta-\widetilde{\vartheta})V_{xy}\right]dxdy\\&\quad-\int\int V_{yyt}\left[\vartheta_yV_y+(\vartheta-\widetilde{\vartheta})V_{yy}\right]dxdy\\&=J_1+J_2+J_3+J_4.\end{aligned}$$

Hence,

$$\begin{aligned}J_1&=-\int\int V_{xyt}(V_x+\widetilde{v})\left[\frac{\vartheta'(V_x+\widetilde{v})}{|G|}V_{xy}+\frac{\vartheta'V_y}{|G|}V_{yy}\right]dxdy\\&=-\frac{1}{2}\frac{d}{dt}\int\int\mathcal{Z}_y^2\frac{\vartheta'(V_x+\widetilde{v})^2}{|G|}dxdy+O(1)(\varepsilon+\delta)\int\int\mathcal{Z}_y^2dxdy\\&\quad-\int\int\frac{\vartheta'(V_x+\widetilde{v})}{|G|}V_yV_{xyt}V_{yy}dxdy,\end{aligned}$$

where, as before, the first term has the "right" sign, while the last term will be studied in the sequel. The terms J_2 and J_4 are similar to those we considered above

$$\begin{aligned}J_2&=-\frac{1}{2}\frac{d}{dt}\int\int(\vartheta-\widetilde{\vartheta})\mathcal{Z}_y^2dxdy+O(1)(\varepsilon+\delta)\int\int\mathcal{Z}_y^2dxdy;\\J_4&=-\frac{1}{2}\frac{d}{dt}\int\int(\vartheta-\widetilde{\vartheta})\mathcal{W}_y^2dxdy+O(1)(\varepsilon+\delta)\int\int\mathcal{W}_y^2dxdy.\end{aligned}$$

Evaluating ϑ_y in J_3, we get

$$\begin{aligned}J_3&=-\int\int V_{yyt}V_y\left[\frac{\vartheta'(V_x+\widetilde{v})}{|G|}V_{xy}+\frac{\vartheta'V_y}{|G|}V_{yy}\right]dxdy\\&=-\frac{1}{2}\frac{d}{dt}\int\int\mathcal{W}_y^2\frac{\vartheta'V_y^2}{|G|}dxdy+O(1)(\varepsilon+\delta)\int\int\mathcal{W}_y^2dxdy\\&\quad-\int\int\frac{\vartheta'(V_x+\widetilde{v})}{|G|}V_yV_{yyt}V_{xy}dxdy.\end{aligned}$$

Finally,

$$
\begin{aligned}
&-\int\int \frac{\vartheta'(V_x-\widetilde{v})}{|G|}V_y\left(V_{xyt}V_{yy}+V_{yyt}V_{xy}\right)dxdy\\
&=-\frac{d}{dt}\int\int \frac{\vartheta'(V_x-\widetilde{v})}{|G|}V_y\mathcal{Z}_y\mathcal{W}_y dxdy+O(1)(\varepsilon+\delta)\int\int\left(\mathcal{Z}_y^2+\mathcal{W}_y^2\right)dxdy.
\end{aligned}
$$

Thus, (3.17) can be rewritten as follows

$$
\begin{aligned}
&\frac{1}{2}\frac{d}{dt}\int\int\left[\mathcal{W}_t^2+\widetilde{\vartheta}\left(\mathcal{Z}_y^2+\mathcal{W}_y^2\right)\right]dxdy+\alpha\int\int\mathcal{W}_t^2dxdy\\
&=O(1)(\varepsilon+\delta)\int\int\left(\mathcal{Z}_y^2+\mathcal{W}_y^2\right)dxdy\\
&-\frac{1}{2}\frac{d}{dt}\int\int\left[\frac{\vartheta'(V_x-\widetilde{v})^2}{|G|}\mathcal{Z}_y^2+(\vartheta-\widetilde{\vartheta})(\mathcal{Z}_y^2+\mathcal{W}_y^2)+\frac{\vartheta'V_y^2}{|G|}\mathcal{W}_y^2\right.\\
&\left.+2\frac{\vartheta'(V_x-\widetilde{v})}{|G|}V_y\mathcal{Z}_y\mathcal{W}_y\right]dxdy. \qquad (3.18)
\end{aligned}
$$

Now we multiply (3.9) for $\lambda\mathcal{Z}$, where, as in lemma 3.2, λ is a small, nonnegative constant which will be chosen at the end. Integration in $dxdy$ yields

$$
\begin{aligned}
&\frac{1}{2}\frac{d}{dt}\int\int\left[\lambda\alpha\mathcal{Z}^2+2\lambda\mathcal{Z}\mathcal{Z}_t\right]dxdy+\lambda\int\int\widetilde{\vartheta}\left(\mathcal{Z}_x^2+\mathcal{Z}_y^2\right)dxdy\\
&-\lambda\int\int\mathcal{Z}_t^2dxdy=\lambda\int\int\mathcal{Z}\,\mathrm{div}\left[\widetilde{\vartheta}_x\begin{pmatrix}\mathcal{Z}\\ \mathcal{W}\end{pmatrix}\right]dxdy\\
&+\lambda\int\int\mathcal{Z}F_xdxdy=I_1+I_2. \qquad (3.19)
\end{aligned}
$$

Young inequality implies

$$
\begin{aligned}
I_2&=-\lambda\int\int\mathcal{Z}_x\widetilde{\vartheta}_xV_xdxdy+\lambda\int\int\mathcal{W}_y\widetilde{\vartheta}_xV_xdxdy\\
&\le\lambda E_{\widetilde{\vartheta}}\int\int\left(\mathcal{Z}_x^2+\mathcal{W}_y^2\right)dxdy+\lambda O(1)(1+t)^{-1}\int\int V_x^2dxdy,
\end{aligned}
$$

where $E_{\widetilde{\vartheta}}$ is a small, positive constant, depending only on the (positive) quantity

$$
\min\left\{\vartheta(v):v\in[-\|\widetilde{v}\|_\infty,+\|\widetilde{v}\|_\infty]\right\},
$$

which will be chosen afterwards. Moreover,

$$
\begin{aligned}
I_2&=-\lambda\int\int\begin{pmatrix}V_{xx}\\ v_{xy}\end{pmatrix}\cdot\left((\vartheta-\widetilde{\vartheta})\begin{pmatrix}V_x+\widetilde{v}\\ V_y\end{pmatrix}\right)_x dxdy\\
&=-\lambda\int\int V_{xx}\left[(\vartheta-\widetilde{\vartheta})_x(V_x-\widetilde{v})+(\vartheta-\widetilde{\vartheta})(V_{xx}-\widetilde{v}_x)\right]dxdy\\
&-\lambda\int\int V_{xy}\left[(\vartheta-\widetilde{\vartheta})_xV_y+(\vartheta-\widetilde{\vartheta})V_{xy}\right]dxdy\\
&=J_1+J_2+J_3+J_4.
\end{aligned}
$$

As in the previous calculations,

$$\begin{aligned}
J_1 &= -\lambda \int\int V_{xx} \left[\frac{\vartheta'(V_x - \widetilde{v})}{|G|} V_{xx} + \frac{\vartheta' V_y}{|G|} V_{xy} \right. \\
&\quad \left. + \left(\frac{\vartheta'(V_x - \widetilde{v})}{|G|} - \widetilde{\vartheta}' \right) \widetilde{v}_x \right] (V_x - \widetilde{v}) dxdy \\
&\leq \lambda O(1)\varepsilon \int\int \left(\mathcal{Z}_x^2 + \mathcal{Z}_y^2 \right) dxdy + \lambda O(1) \int\int |\mathcal{Z}_x| |\widetilde{v}_x| \left(|V_x| + |V_y| \right) dxdy \\
&\leq \lambda O(1)\varepsilon \int\int \left(\mathcal{Z}_x^2 + \mathcal{Z}_y^2 \right) dxdy + \lambda E_{\widetilde{\vartheta}} \int\int \mathcal{Z}_x^2 dxdy \\
&\quad + \lambda O(1)(1+t)^{-1} \int\int \left(V_x^2 + V_y^2 \right) dxdy,
\end{aligned}$$

since the first term in J_1,

$$-\lambda \int\int \frac{\vartheta'(V_x + \widetilde{v})^2}{|G|} V_{xx}^2 dxdy,$$

is negative for ε sufficiently small, as we pointed out previously. Moreover,

$$\begin{aligned}
J_2 &= -\lambda O(1)\varepsilon \int\int \mathcal{Z}_x^2 dxdy + \lambda E_{\widetilde{\vartheta}} \int\int \mathcal{Z}_x^2 dxdy \\
&\quad + \lambda O(1)(1+t)^{-1} \int\int \left(V_x^2 + V_y^2 \right) dxdy; \\
J_4 &= \lambda O(1) \int\int \mathcal{Z}_y^2 dxdy.
\end{aligned}$$

Finally,

$$\begin{aligned}
J_3 &= -\lambda \int\int V_{xy} \left[\frac{\vartheta'(V_x - \widetilde{v})}{|G|} V_{xx} + \frac{\vartheta' V_y}{|G|} V_{xy} \right. \\
&\quad \left. + \left(\frac{\vartheta'(V_x - \widetilde{v})}{|G|} - \widetilde{\vartheta}' \right) \widetilde{v}_x \right] V_y dxdy \\
&\leq \lambda O(1)\varepsilon \int\int \left(\mathcal{Z}_x^2 + \mathcal{Z}_y^2 \right) dxdy + \lambda O(1)\varepsilon \int\int \mathcal{Z}_y^2 dxdy \\
&\quad + \lambda O(1) \int\int |\mathcal{Z}_y| |\widetilde{v}_x| |V_y| \left(|V_x| + |V_y| \right) dxdy \\
&\leq \lambda O(1)\varepsilon \int\int \left(\mathcal{Z}_x^2 + \mathcal{Z}_y^2 \right) dxdy + \lambda E_{\widetilde{\vartheta}} \int\int \mathcal{Z}_x^2 dxdy \\
&\quad + \lambda O(1)(1+t)^{-1} \int\int \left(V_x^2 + V_y^2 \right) dxdy.
\end{aligned}$$

Thus, the relation (3.19) becomes

$$\begin{aligned}
&\frac{1}{2} \frac{d}{dt} \int\int \left[\lambda \alpha \mathcal{Z}^2 + 2\lambda \mathcal{Z} \mathcal{Z}_t \right] dxdy + \lambda \int\int \widetilde{\vartheta} \left(\mathcal{Z}_x^2 + \mathcal{Z}_y^2 \right) dxdy \\
&\quad - \lambda \int\int \mathcal{Z}_t^2 dxdy \leq \lambda E_{\widetilde{\vartheta}} \int\int \left(\mathcal{Z}_x^2 + \mathcal{Z}_y^2 + \mathcal{W}_y^2 \right) dxdy \\
&\quad + \lambda \varepsilon O(1) \int\int \left(\mathcal{Z}_x^2 + \mathcal{Z}_y^2 \right) dxdy \\
&\quad + \lambda O(1)(1+t)^{-1} \int\int \left(V_x^2 + V_y^2 \right) dxdy. \qquad (3.20)
\end{aligned}$$

A similar estimate can be achieved for the quantity $\mathcal{W}$, by multiplying (3.10) by $\lambda\mathcal{W}$. Therefore, proceeding as before, we end up to a relation of the form

$$\begin{aligned}&\frac{1}{2}\frac{d}{dt}\int\int\left[\lambda\alpha\mathcal{W}^2+2\lambda\mathcal{W}\mathcal{W}_t\right]dxdy+\lambda\int\int\widetilde{\vartheta}\left(\mathcal{Z}_y^2+\mathcal{W}_y^2\right)dxdy\\&-\lambda\int\int\mathcal{W}_t^2dxdy\leq\lambda O(1)\varepsilon\int\int\left(\mathcal{Z}_y^2+\mathcal{W}_y^2\right)dxdy.\end{aligned}\tag{3.21}$$

Summing the estimates (3.16), (3.18), (3.20) and (3.21) we get

$$\begin{aligned}&\frac{1}{2}\frac{d}{dt}\int\int\left[\mathcal{Z}_t^2+\widetilde{\vartheta}\left(\mathcal{Z}_x^2+\mathcal{Z}_y^2\right)\right]dxdy+(\alpha-\lambda)\int\int\mathcal{Z}_t^2dxdy\\&+\frac{1}{2}\frac{d}{dt}\int\int\left[\mathcal{W}_t^2+\widetilde{\vartheta}\left(\mathcal{Z}_y^2+\mathcal{W}_y^2\right)\right]dxdy+(\alpha-\lambda)\int\int\mathcal{W}_t^2dxdy\\&+\frac{1}{2}\frac{d}{dt}\int\int\left[\lambda\alpha\mathcal{Z}^2+2\lambda\mathcal{Z}\mathcal{Z}_t\right]dxdy+\lambda\int\int\widetilde{\vartheta}\left(\mathcal{Z}_x^2+\mathcal{Z}_y^2\right)dxdy\\&+\frac{1}{2}\frac{d}{dt}\int\int\left[\lambda\alpha\mathcal{W}^2+2\lambda\mathcal{W}\mathcal{W}_t\right]dxdy+\lambda\int\int\widetilde{\vartheta}\left(\mathcal{Z}_y^2+\mathcal{W}_y^2\right)dxdy\\&\leq E_\alpha\int\int\mathcal{Z}_t^2dxdy+O(1)(\varepsilon+\delta)\int\int\left(\mathcal{Z}_x^2+\mathcal{Z}_y^2+\mathcal{W}_y^2\right)dxdy\\&-\frac{1}{2}\frac{d}{dt}\int\int\left[\frac{\vartheta'(V_x-\widetilde{v})^2}{|G|}\mathcal{Z}_x^2+(\vartheta-\widetilde{\vartheta})(\mathcal{Z}_x^2+\mathcal{Z}_y^2)+\frac{\vartheta'V_y^2}{|G|}\mathcal{Z}_y^2\right.\\&\left.+2\frac{\vartheta'(V_x-\widetilde{v})}{|G|}V_y\mathcal{Z}_x\mathcal{Z}_y\right]dxdy+O(1)(1+t)^{-2}\int\int\left(V_x^2+V_y^2\right)dxdy\\&+O(1)(\varepsilon+\delta)\int\int\left(\mathcal{Z}_y^2+\mathcal{W}_y^2\right)dxdy\\&-\frac{1}{2}\frac{d}{dt}\int\int\left[\frac{\vartheta'(V_x-\widetilde{v})^2}{|G|}\mathcal{Z}_y^2+(\vartheta-\widetilde{\vartheta})(\mathcal{Z}_y^2+\mathcal{W}_y^2)+\frac{\vartheta'V_y^2}{|G|}\mathcal{W}_y^2\right.\\&\left.+2\frac{\vartheta'(V_x-\widetilde{v})}{|G|}V_y\mathcal{Z}_y\mathcal{W}_y\right]dxdy+\lambda E_{\widetilde{\vartheta}}\int\int\left(\mathcal{Z}_x^2+\mathcal{Z}_y^2+\mathcal{W}_y^2\right)dxdy\\&+\lambda\varepsilon O(1)\int\int\left(\mathcal{Z}_x^2+\mathcal{Z}_y^2\right)dxdy+\lambda O(1)(1+t)^{-1}\int\int\left(V_x^2+V_y^2\right)dxdy\\&+\lambda O(1)\varepsilon\int\int\left(\mathcal{Z}_y^2+\mathcal{W}_y^2\right)dxdy.\end{aligned}$$

At this point, we choose λ, ε, δ, E_α, $E_{\widetilde{\vartheta}}$ sufficiently small and we control the products $\mathcal{Z}\mathcal{Z}_t$ and $\mathcal{W}\mathcal{W}_t$ in order to have

$$\begin{aligned}&\frac{d}{dt}\int\int\left[\mathcal{Z}_t^2+\mathcal{W}_t^2+\widetilde{\vartheta}\left(\mathcal{Z}_x^2+\mathcal{Z}_y^2+\mathcal{W}_y^2\right)+\lambda(\mathcal{Z}+\mathcal{W})\right]dxdy\\&+\int\int\left[\mathcal{Z}_t^2+\mathcal{W}_t^2+\lambda\widetilde{\vartheta}\left(\mathcal{Z}_x^2+\mathcal{Z}_y^2+\mathcal{W}_y^2\right)\right]dxdy\\&\leq-\frac{1}{2}\frac{d}{dt}\int\int\left[\frac{\vartheta'(V_x-\widetilde{v})^2}{|G|}\mathcal{Z}_x^2+(\vartheta-\widetilde{\vartheta})(\mathcal{Z}_x^2+\mathcal{Z}_y^2)+\frac{\vartheta'V_y^2}{|G|}\mathcal{Z}_y^2\right.\end{aligned}$$

$$
\begin{aligned}
&+2\frac{\vartheta'(V_x-\widetilde{v})}{|G|}V_y\mathcal{Z}_x\mathcal{Z}_y\Big]\,dxdy+O(1)(1+t)^{-2}\int\int\left(V_x^2+V_y^2\right)dxdy\\
&-\frac{1}{2}\frac{d}{dt}\int\int\Bigg[\frac{\vartheta'(V_x-\widetilde{v})^2}{|G|}\mathcal{Z}_y^2+(\vartheta-\widetilde{\vartheta})(\mathcal{Z}_y^2+\mathcal{W}_y^2)+\frac{\vartheta'V_y^2}{|G|}\mathcal{W}_y^2\\
&+2\frac{\vartheta'(V_x-\widetilde{v})}{|G|}V_y\mathcal{Z}_y\mathcal{W}_y\Big]\,dxdy\\
&+\lambda O(1)(1+t)^{-1}\int\int\left(V_x^2+V_y^2\right)dxdy. \qquad (3.22)
\end{aligned}
$$

Hence, we integrate (3.22) in dt and, using the relations

$$
\begin{aligned}
&|\vartheta-\widetilde{\vartheta}|=O(1)\varepsilon\\
&|V_y|=O(1)\varepsilon\\
&\frac{\vartheta'(V_x-\widetilde{v})^2}{|G|}>0\quad\text{for }\varepsilon<<1,
\end{aligned}
$$

we get the first H^2 estimate

$$
\begin{aligned}
&\|V_{xx}(t)\|^2+\|V_{xy}(t)\|^2+\|V_{yy}(t)\|^2+\|z_x(t)\|^2+\|z_y(t)\|^2\\
&+\int_0^t\left(\|V_{xx}(\tau)\|^2+\|V_{xy}(\tau)\|^2+\|V_{yy}(\tau)\|^2+\|z_x(\tau)\|^2+\|z_y(\tau)\|^2\right)d\tau\\
&=O(1)\left(\|V_0\|_2^2+\|z_0\|_1^2+\delta\right).
\end{aligned}
$$

Moreover, we first multiply (3.22) by $(1+t)$ and then we integrate in dt to get (using the relations and the estimate above)

$$
\begin{aligned}
&(1+t)\left[\|V_{xx}(t)\|^2+\|V_{xy}(t)\|^2+\|V_{yy}(t)\|^2+\|z_x(t)\|^2+\|z_y(t)\|^2\right]\\
&+\int_0^t(1+\tau)\left[\|V_{xx}(\tau)\|^2+\|V_{xy}(\tau)\|^2+\|V_{yy}(\tau)\|^2+\|z_x(\tau)\|^2+\|z_y(\tau)\|^2\right]d\tau\\
&=O(1)\left(\|V_0\|_2^2+\|z_0\|_1^2+\delta\right).
\end{aligned}
$$

Finally, we consider (3.22) for $\lambda=0$ and we multiply it by $(1+t)^2$. Integrating the relation obtained in dt and using all the relations above, we end up with the last estimate we need to conclude the proof. □

The differentiation of the equation (3.4) with respect to t gives a better asymptotic result, contained in the following lemma. This phenomenon follows from the fact that the t-derivatives of $\overline{v}$ (and hence of $\widetilde{v}$) have better asymptotic decays than the x-derivatives of $\overline{v}$.

LEMMA 3.5. *Suppose ε, δ and $\|V_0\|_7^2+\|z_0\|_6^2$ are sufficiently small. Then*

$$
\begin{aligned}
&(1+t)^2\|z(t)\|^2+(1+t)^3(\|z_t(t)\|^2+\|z_x(t)\|^2+\|z_y(t)\|^2)\\
&+\int_0^t\left[(1+\tau)^2(\|z_x(\tau)\|^2+\|z_y(\tau)\|^2)+(1+\tau)^3\|z_t(\tau)\|^2\right]d\tau\\
&=O(1)\left(\|V_0\|_2^2+\|z_0\|_1^2+\delta\right).
\end{aligned}
$$

The proof of this lemma follows step by step the proof of lemma 3.4 and it is omitted. Finally, iterating the procedure, it is possible to prove the following lemmas.

LEMMA 3.6. *Suppose ε, δ and $\|V_0\|_7^2 + \|z_0\|_6^2$ are sufficiently small. Then, for any $k \le 6$,*

$$\begin{aligned}&(1+t)^{k+1}\|D^{k+1}V(t)\|^2 + (1+t)^{k+1}\|D^k z(t)\|^2\\ &+\int_0^t (1+\tau)^k\|D^{k+1}V(\tau)\|^2 d\tau + \int_0^t (1+\tau)^{k+1}\|D^k z(\tau)\|^2 d\tau\\ &= O(1)\left(\|V_0\|_{k+1}^2 + \|z_0\|_k^2 + \delta\right).\end{aligned}$$

LEMMA 3.7. *Suppose ε, δ and $\|V_0\|_7^2 + \|z_0\|_6^2$ are sufficiently small. Then, for any $k \le 6$,*

$$\begin{aligned}&(1+t)^{k+2}\|D^k z(t)\|^2 + (1+t)^{k+2}\|D^{k-1} z_t(t)\|^2\\ &+\int_0^t (1+\tau)^{k+1}\|D^k z(\tau)\|^2 d\tau + \int_0^t (1+\tau)^{k+2}\|D^{k-1} z_t(\tau)\|^2 d\tau\\ &= O(1)\left(\|V_0\|_{k+1}^2 + \|z_0\|_k^2 + \delta\right).\end{aligned}$$

REMARK 3.8. Since the nonlinear function ϑ depends on V_x and V_y, in order to compute the energy estimates, we have to bound the H^4 norm of V, so we bound, by Sobolev embedding (in 2-D), the L^∞ norm of V_x and V_y. However, the trilinear terms which appears in the energy estimates to achieve the H^4 bounds are of the form

$$D^\alpha V D^\beta V D^\gamma V,$$

with $|\alpha| + |\beta| + |\gamma| \le 10$. Therefore, since in all the terms of the H^4 estimate we must have α, β, $\gamma \le 4$, there are terms with the property α, β, $\gamma \ge 3$. Therefore, the H^4 bounds are not enough to close the estimate and we need at least H^6 to control third derivatives in L^∞. With a simple argument, we can prove that the H^7 norm is enough to close the proof. Indeed, in the H^7 case, the trilinear terms are of the form

$$D^\alpha V D^\beta V D^\gamma V,$$

with $|\alpha| + |\beta| + |\gamma| \le 16$. Thus, the terms with the maximum number of derivatives can be reduced, by integration by parts, in one of the two following forms:

$$\partial_t D^7 V D^7 V D^1 V$$
$$D^7 V D^\alpha V D^\beta V,$$

with $|\alpha| + |\beta| = 9$. The first term can be written as a total derivative with respect to t and it is controlled by the energy (using the smallness of $|D^1 V| = O(1)\varepsilon$). The second one is no longer trilinear, since now either α or β is less or equal to 4 and hence either $D^\alpha V$ or $D^\beta V$ is controlled in L^∞.

References

[HL92] L. Hsiao and T.-P. Liu, *Convergence to nonlinear diffusion waves for solutions of a system of hyperbolic conservation laws with damping*, Comm. Math. Physics **143** (1992), 599–605.

[HL93] L. Hsiao and T.-P. Liu, *Nonlinear diffusive phenomena of nonlinear hyperbolic systems*, Chin. Ann. Math. Ser. B **14** (1993), 465–480.

[JR] S. Junca and M. Rascle, *Relaxation of the isothermal Euler-Poisson system to the drift-diffusion equations*, Quart. Appl. Math., to appear.

[Lat] C. Lattanzio, *On the* 3*-D Bipolar Isentropic Euler-Poisson Model for Semiconductors and the Drift-Diffusion Limit*, Math. Models Methods Appl. Sci., to appear.

[LM] C. Lattanzio and P. Marcati, *The relaxation to the drift-diffusion system for the* 3*-D isentropic Euler-Poisson model for semiconductors*, Discrete Contin. Dynam. Systems, to appear.

[LR97] C. Lattanzio and B. Rubino, *Limiting Behavior for Hyperbolic Systems of Conservation Laws with Damping*, Tech. Report 159, Dipartimento di Matematica Pura ed Applicata, Università di L'Aquila, 1997.

[MM] P. Marcati and M. Mei, *Convergence to nonlinear diffusion waves for solutions of the initial boundary value problems in the hyperbolic conservation laws with damping*, Quart. Appl. Math., to appear.

[MM90] P. Marcati and A. Milani, *The one-dimensional Darcy's law as the limit of a compressible Euler flow*, J. Differential Equations **13** (1990), 129–147.

[MMS88] P. Marcati, A. Milani, and P. Secchi, *Singular convergence of weak solutions for a quasilinear nonhomogeneous hyperbolic system*, Manuscripta Math. **60** (1988), 49–69.

[MN95] P. Marcati and R. Natalini, *Weak solutions to a hydrodynamic model for semiconductors and relaxation to the drift-diffusion equation*, Arch. Rational Mech. Anal. **129** (1995), 129–145.

[MR] P. Marcati and B. Rubino, *Hyperbolic to Parabolic Relaxation Theory for Quasilinear First Order Systems*, J. Differential Equations, to appear.

[Nat96] R. Natalini, *The bipolar hydrodynamic model for semiconductors and the drift-diffusion equations*, J. Math. Anal. Appl. **198** (1996), 262–281.

[Nis96] K. Nishihara, *Convergence Rates to Nonlinear Diffusion Waves for Solutions of Systems of Hyperbolic Conservation Laws with Damping*, J. Differential Equations **131** (1996), 171–188.

[PVD97] L.A. Peletier and C.T. Van Duyn, *A class of similarity solutions of the nonlinear diffusion equation*, J. Nonlinear Analysis: TMA **1** (1997), 223–233.

[SX97] D. Serre and L. Xsiao, *Asymptotic behaviour of large weak entropy solutions of the damped p-system*, J. Partial Diff. Eqs. **10** (1997), 355–368.

Dipartimento di Matematica Pura ed Applicata, Università degli Studi di L'Aquila, via Vetoio, loc. Coppito – 67010 L'Aquila, Italy
E-mail address: corrado@univaq.it

Dipartimento di Matematica Pura ed Applicata, Università degli Studi di L'Aquila, via Vetoio, loc. Coppito – 67010 L'Aquila, Italy
E-mail address: marcati@univaq.it

Contemporary Mathematics
Volume **238**, 1999
B 0-8218-1196-7-03547-7

L_1 Stability for Systems of Hyperbolic Conservation Laws

Tai-Ping Liu and Tong Yang

Abstract. In this paper, we summarize our results on constructing a nonlinear functional which is equivalent to L_1 distance between two weak solutions to systems of hyperbolic conservation laws and non-increasing in time. The weak solutions are constructed by Glimm scheme through the wave tracing methed. Therefore, such an explicit functional depending only on the two wave patterns of the solutions yields directly the uniqueness of solutions by Glimm scheme and reveals the effects of nonlinear interaction and coupling on the L_1 topology.

1. Introduction

Consider the Cauchy problem for a system of conservation laws,

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = 0, \quad t \geq 0, \ -\infty < x < \infty, \tag{1.1}$$

$$u(x,0) = u_0(x), \qquad -\infty < x < \infty, \tag{1.2}$$

where u and $f(u)$ are n-vectors. We assume that the system is strictly hyperbolic, i.e. the matrix $\partial f(u)/\partial u$ has real and distinct eigenvalues $\lambda_1(u) < \lambda_2(u) < \cdots < \lambda_n(u)$ for all u under consideration, with the corresponding right eigenvectors $r_i(u)$, $i = 1, 2, \ldots, n$. Each characteristic field is assumed to be either linearly degenerate or genuinely nonlinear, [13], i.e. $r_i(u) \cdot \nabla \lambda_i(u) \equiv 0$ or $r_i(u) \cdot \nabla \lambda_i(u) \neq 0$, $i = 1, 2, \ldots, n$.

The purpose of our research is to construct a nonlinear functional $H(t) = H[u(\cdot,t), v(\cdot,t)]$, which is equivalent to $\|u - v\|_{L_1}$ of two weak solutions $u(x,t)$ and $v(x,t)$ to (1.1) and (1.2) and is non-increasing in time. It also depends explicitly on the wave patterns of these two solutions. In general, the functional $H[u(\cdot,t), v(\cdot,t)]$ consists three parts: the first part is the product of the Glimm's functional and $L(t)$ representing the L_1 distance between $u(x,t)$ and $v(x,t)$, which reveals the interaction effects of nonlinear waves on the L_1 topology; the second part is $Q_d(t)$, which registers the effect of nonlinear coupling of waves in different families on $\|(u-v)(x,t)\|_{L_1(x)}$, making use of the strict hyperbolicity of the system; the third part,

1991 *Mathematics Subject Classification.* 35L40, 35L65, 35L67.

The research of the first author was supported in part by NSF Grant DMS-9623025. The research of second author was supported in part by the RGC Competitive Earmarked Research Grant 9040394.

which is called the generalised entropy functional, captures the genuine nonlinearity of the characteristic fields. If the distance between two states in the phase plane is measured by rarefaction wave curves instead of Hugoniot wave curves, there is one more main component in the nonlinear functional $H[u(\cdot,t),v(\cdot,t)]$ denoted by $L_h(t)$. This functional is needed because the generalised entropy functional can be used only to control the difference of the third orders of shocks pertaining to different solutions instead of shock waves strengths to the cubic power. The summation of $L(t)$ and $L_h(t)$ shows that some of the effect of a shock on the L_1 distance is conservative.

The weak solutions in consideration are constructed by Glimm scheme through the wave tracing method, [11, 16]. Therefore the nonlinear functional $H[u(\cdot,t),$ $v(\cdot,t)]$ immediately yields the uniqueness of weak solutions obtained by Glimm scheme. Furthermore, measuring the L_1 distance between the solutions by the nonlinear functional $H[u(\cdot,t),v(\cdot,t)]$ is robust and therefore it does not require any particular approximation scheme. In [5], this functional is defined for weak solutions obtained by the wave front tracking method, [1, 3]. In fact, our analysis would be applied to any approximate scheme based on the characterisitc method, c.f. [7, 10].

There has been much progress on the well-posedness problem. In [2], the problem when the two solutions are infinitesimally close is studied by making uses of the fact that the topology of the shock waves are close in this case. This analysis is used to study the continuous dependence of the solutions on its initial data for 2×2 systems in [1] and for $n\times n$ systems in [3]. By homotopically deforming one solution to the other to construct a Riemann semigroup, this line of approach requires the monitoring of the changes of the topology of shocks in the approximate solutions. Hence the nonlinear functional thus defined depends not only on the two wave patterns of $u(x,t)$ and $v(x,t)$.

Glimm's nonlinear functional can be defined for any BV entropy solutions and was proved to be non-increasing in time for piecewise smooth solutions [5]. It will be interesting to define the above nonlinear functional $H[u(\cdot,t),v(\cdot,t)]$ for general BV entropy solutions and show that it is non-increasing in time. Also it will be interesting to study the more general case without the assumption of genuine nonlinearity.

Some general uniqueness formulations have been formulated by various authors, [3, 4] and references therein. For attempts on the uniqueness based on the L_2-norm, see [8, 9, 15, 17, 24]. For comments on non-contractiveness in the L_1-norm, see [27].

2. Glimm's Functional and Wave Tracing

Glimm scheme uses Riemann solutions as building blocks and consists of constructing Riemann data by using a random sequence. At each time step, a nonlinear functional corresponding to the interaction potential is used to control the increase in the new waves strengths after interaction. The Riemann problem for (1.1) with initial value

$$u(x,0)=\begin{cases} u_l, & x<0,\\ u_r, & x>0,\end{cases} \tag{2.1}$$

has the scattering property which represents the large time behaviour of the solution. The rarefaction curve is the integral curve $R_i(u_0)$ of $r_i(u)$ through a given

state u_0, and the Hugoniot curves through a state u_0 are

$$H(u_0) \equiv \{u \ : \ u - u_0 = \sigma(u_0, u)(f(u) - f(u_0))\} \tag{2.2}$$

for some scalar $\sigma = \sigma(u_0, u)$. The following basic theorem is from [13].

THEOREM 2.1. *Suppose that system* (1.1) *is strictly hyperbolic. Then, in a small neighborhood of a state* u_0*, the Hugoniot curves consists of* n *curves* $H_i(u_0)$*,* $i = 1, 2, \ldots, n$*, with the following properties:*
(i) *The Hugoniot curve* $H_i(u_0)$ *and the rarefaction curve* $R_i(u_0)$ *have second order contact at* $u = u_0$*,*
(ii) *The shock speed* $\sigma(u_0, u)$*,* $u \in H_i(u_0)$ *satisfies:*

$$\sigma(u_0, u) = \frac{1}{2}(\lambda_i(u) + \lambda_i(u_0)) + O(1)|u - u_0|^2. \tag{2.3}$$

(iii) *For a genuinely nonlinear field,* (u_0, u)*,* $u \in R_i^+(u_0)$*, is a rarefaction wave; and* $(u_-, u_+) = (u_0, u)$*,* $u \in H_i^-(u_0)$*, is a shock wave satisfying the following Lax entropy condition*

$$\lambda_i(u_-) > s > \lambda_i(u_+). \tag{2.4}$$

(iv) *For a linearly degenerate field* $H_i(u_0) = R_i(u_0)$ *and* (u_0, u)*,* $u \in R_i(u_0)$*, forms a contact discontinuity with speed* s*:*

$$s = \lambda_i(u) = \lambda_i(u_0). \tag{2.5}$$

We construct the wave curves $W_i(u_0)$ as follows:

$$W_i(u_0) \equiv \begin{cases} R_i^+(u_0) \ U \ H_i^-(u_0), & i\text{-th characteristic g.nl.;} \\ R_i(u_0) = H_i(u_0), & i\text{-th characteristic l.dg.} \end{cases} \tag{2.6}$$

where for each genuinely nonlinear field, we let $\lambda_i(u) < \lambda_i(u_0)$ for the states u on $H_i^-(u_0)$ and $R_i^-(u_0)$ and $\lambda_i(u) \geq \lambda_i(u_0)$ for u on $H_i^+(u_0)$ and $R_i^+(u_0)$. Thus (u_0, u) forms an elementary i-wave when $u \in W_i(u_0)$. These waves are the building blocks for the solution of the Riemann problem.

By using strict hyperbolicity and the inverse function theorem, the Riemann problem can be solved in the class of elementary waves [13]:

THEOREM 2.2. *Suppose that each characteristic field is either genuinely nonlinear or linearly degenerate. Then the Riemann problem for* (1.1) *has a unique solution in the class of elementary waves provided that the states are in a small neighborhood of a given state.*

The Glimm scheme is a finite difference scheme involving a random sequence a_i, $i = 0, 1, \ldots$, $0 < a_i < 1$. Let $r = \triangle x$, $s = \triangle t$ be the mesh sizes satisfying the (C-F-L) condition

$$\frac{r}{s} > \ 2|\lambda_i(u)|, \qquad 1 \leq i \leq n, \tag{2.7}$$

for all states u under consideration. The approximate solutions $u(x, t) = u_r(x, t)$ depends on the random sequence $\{a_k\}$ and is defined inductively in time as follows:

$$u(x, 0) = u_0((h + a_0)r), \qquad hr < x < (h + 1)r, \tag{2.8}$$

$$u(x, ks) = u((h + a_i)r - 0, ks - 0), \quad hr < x < (h + 1)r, \quad k = 1, 2, \ldots . \tag{2.9}$$

Due to (C-F-L) condition (2.6) these elementary waves do not interact within the layer. Thus the approximation solution is an exact solution except at the interfaces $t = ks$, $k = 1, 2, \ldots$. The following theorem is from Glimm [11]:

THEOREM 2.3. *Suppose that the initial data $u_0(x)$ is of small total variation T.V. Then the approximate solutions $u(x,t)$ are of small total variation $O(1)T.V.$ in x for all time t. Moreover, for almost all choices of the sequence $\{a_k\}_{k=1}^{\infty}$, the approximate solutions tend to an exact solution for a sequence of the mesh sizes r, s tending to zero with r/s fixed and r, s satisfying (C-F-L) condition. The exact solution $u(x,t)$ is of bounded variation in x for any time $t \geq 0$:*

$$\text{variation}_{-\infty<x<\infty}\ u(x,t) = O(1)T.V. \tag{2.10}$$

and is continuous in $L_1(x)$-norm:

$$\int_{-\infty}^{\infty} |u(x,t_1) - u(x,t_2)|dx = O(1)|t_1 - t_2|, \quad t_1,\ t_2 \geq 0. \tag{2.11}$$

The proof of the above Theorem is based on the proof of the non-increasingness of the Glimm's functional $F(u)(t)$ defined as follows:

$$\begin{aligned}
J(t) &\equiv \sum\{|\alpha| :\ \alpha \text{ any wave at time } t\},\\
D_d(t) &\equiv \sum\{|\alpha|\,|\beta| :\ \alpha \text{ and } \beta \text{ interacting waves of distinct}\\
&\qquad \text{characteristic families at time } t\},\\
D_s(t) &\equiv \sum_{i=1}^{n} D_s^i,\\
D_s^i(t) &\equiv \sum\{|\alpha||\beta|(-\min\{\Theta(\alpha,\beta),0\}) :\ \alpha \text{ and } \beta \text{ interacting}\\
&\qquad i\text{-waves at time } t\},\\
D(t) &\equiv D_d(t) + D_s(t),\\
F(u)(t) &\equiv J(t) + MD(t).
\end{aligned} \tag{2.12}$$

Here M is a sufficiently large constant and $\Theta(\alpha,\beta)$ is the interacting angle between α and β, cf. [11, 23].

It was shown in [16] that in fact the approximate solutions constructed by Glimm's scheme converges to a weak solution as long as the random sequence is equidistributed. In this wave tracing method, the waves are classified into the following three categories: surviving ones, canceled ones, and those produced by interactions. We summarize the result in [16] as follows:

THEOREM 2.4. *The waves in an approximate solution in a given a time zone $\Lambda = \{(x,t) : -\infty < x < \infty, K_1 s \leq t < K_2 s\}$ can be partitioned into subwaves of categories I, II or III with the following properties up to an error due to the random sequence:*
(i) *The subwaves in I are surviving. Given a subwave $\alpha(t)$, $K_1 s \leq t < K_2 s$ in I, write $\alpha \equiv \alpha(K_1 s)$ and denote by $|\alpha(t)|$ its strength and $\lambda(\alpha(t)$ its speed at time t, by$[\lambda(\alpha)]$ the variation of its speed and by $[\alpha]$ the variation of the jump of the states across it over the time interval $K_1 s \leq t < K_2 s$. Then*

$$\sum_{\alpha\in I}([\alpha] + |\alpha(K_1 s)|[\lambda(\alpha)]) = O(1)D(\Lambda). \tag{2.13}$$

(ii) *A subwave $\alpha(t)$ has positive initial strength $|\alpha(K_1 s)| > 0$, but is canceled in the zone Λ, $|\alpha(K_2 s)| = 0$. Moreover, the total strength and variation of the wave shape*

satisfy

$$\sum_{\alpha\in II}([\alpha]+|\alpha(K_1 s)|)\le C(\Lambda)+O(1)D(\Lambda). \tag{2.14}$$

(iii) *A subwave in III has zero initial strength* $|\alpha(K_1 s)|=0$, *and is created in the zone* Λ, $|\alpha(K_2 s)>0$. *Moreover, the total variation satisfies*

$$\sum_{\alpha\in III}([\alpha]+|\alpha(t)|)=O(1)D(\Lambda),\ K_1 s\le t<K_2 s, \tag{2.15}$$

where $C(\Lambda)$ *and* $D(\Lambda)$ *represent the cancellation and interaction potential change in the region* Λ *respectively.*

An application of the above theorem gives the following deterministic version of the Glimm scheme, cf. [16]:

THEOREM 2.5. *Suppose that the random sequence* a_k, $k=1,2,\dots$ *is equidistributed. Then the limiting function* $u(x,t)$ *of the Glimm scheme is a weak solution of the hyperbolic conservation laws.*

Theorem 2.4 was also applied to the study of the regularity and large time behaviour of the solutions and the convergence rate of the Glimm scheme, cf. [6, 18] and reference therein.

The application of Theorem 2.4 to L_1 stability of weak solutions is that waves can be viewed as linearly superimposed in each region $(p-1)Ns<t<pNs$ in the wave tracing method.

3. Nonlinear Functionals and Main Theorems

Given two solutions $u(x,t)$ and $v(x,t)$ of the system (1.1), we define their pointwise distance along the Hugoniot curves: solve the Riemann problem $(u(x,t), v(x,t))$ by discontinuity waves:

$$u_0=u(x,t),\ u_i\in H_i(u_{i-1}),\quad i=0,1,\dots,n,\ u_n=v(x,t). \tag{3.1}$$

Without loss of generality, we assume that the i-th component u^i of the vector u is a non-singular parameter along the i-th Hugoniot and rarefaction curves. We set

$$q_i(x,t)\equiv(u_i-u_{i-1})^i,\qquad \lambda_i(x)\equiv\lambda_i(u_{i-1}(x),u_i(x)), \tag{3.2}$$

This way of assigning the distance is convenient in that u^i is a conservative quantity and so it satisfies simple wave interaction estimates. Another advantage over choosing the Euclidean distance is that the strength of a shock (u_-,u_+) is the same as that of the rarefaction shock (u_+,u_-) in our measurement.

For an i-wave α^i in the solutions $u(x,t)$ or $v(x,t)$, we denote by $x(\alpha^i)=x(\alpha^i(t))$ its location at time t, and $q_j^{\pm}(\alpha^i)$ for $q_j(x(\alpha^i)\pm,t)$, $1\le j\le n$. For $j=i$ we also use the abbreviated notations $q^{\pm}(\alpha^i)=q_i^{\pm}(\alpha^i)$. The linear part $L[u,v]$ of the nonlinear functional $H[u,v]$ is equivalent to the $L_1(x)$ distance of the solutions:

$$\begin{aligned} L[u(\cdot,t),v(\cdot,t)]&\equiv\sum_{i=1}^{n}L_i[u(\cdot,t),v(\cdot,t)]\\ L_i[u(\cdot,t),v(\cdot,t)]&\equiv\int_{-\infty}^{\infty}|q_i(x,t)|dx. \end{aligned} \tag{3.3}$$

Without any ambiguity, we will use u and v to denote the approximate solutions in the Glimm scheme and also the corresponding weak solutions when the mesh sizes tend to zero. As in [5, 23], we will use the notations $J(u)$ and $J(v)$ to denote the waves in the solutions u and v at a given time, respectively. And $J \equiv J(u) \cup J(v)$. Moreover, α^i denotes a i-wave in J. The other two components of the nonlinear functional $H[u, v]$, the quadratic $Q_d(t)$ and the generalized entropies $E(t)$, are defined as follows:

$$Q_d(t) \equiv Q_d[u(\cdot,t), v(\cdot,t)] = \sum_{\alpha^i \in J} Q_d(\alpha^i)$$

$$Q_d(\alpha^i) = |\alpha^i| \Big(\sum_{j>i} \int_{-\infty}^{x(\alpha^i)} |q_j(x,t)|dx + \sum_{j<i} \int_{x(\alpha^i)}^{\infty} |q_j(x,t)|dx \Big) \tag{3.4}$$

$$E(t) \equiv E[u(\cdot,t), v(\cdot,t)] = \sum_{i=1}^{n} E^i(t),$$

$$\begin{aligned} E^i(t) = & \sum_{\alpha^i \in J(u)} |\alpha^i| \Big(\int_{x(\alpha^i)}^{\infty} |\min\{0, q_i(x,t)\}|dx + \int_{-\infty}^{x(\alpha^i)} \max\{0, q_i(x,t)\}dx \Big) \\ & + \sum_{\alpha^i \in J(v)} |\alpha^i| \Big(\int_{x(\alpha^i)}^{\infty} \max\{0, q_i(x,t)\}dx + \int_{-\infty}^{x(\alpha^i)} |\min\{0, q_i(x,t)\}|dx \Big). \end{aligned} \tag{3.5}$$

For any given time $T = MNs$ in the Glimm scheme through the wave tracing method, we define the main nonlinear functional $H(t)$ as follows:

$$H(t) \equiv H[u(\cdot,t), v(\cdot,t)] \equiv (1 + K_1 F(p-1)Ns))L(t) + K_2(Q_d(t) + E(t)),$$

for $t \in ((p-1)Ns, pNs)$, $p = 1, \ldots, M$. Notice here that the Glimm's functional $F = F(u) + F(v)$ is valued at the end time $t = (p-1)Ns$. The jump of the functionals $L(t)$, $Q_d(t)$ and $E(t)$ at each time step $t = pNs$, $p = 1, 2, \ldots, M$ due to wave interaction can be controlled by $O(1)[F(pNs) - F((p-1)Ns)]L(pNs)$.

Remark 3.1. If the distance between the two solutions is not measured by the Hugoniot curves but rarefaction wave curves as was done in Liu-Yang [22] for 2×2 system, then the functional $H[u(\cdot,t), v(\cdot,t)]$ has more components. One of which is called $L_h(t)$ which is used to control the error caused by the bifurcation of the shock wave curve from the rarefaction curve, and it is of the third order of the shock wave strength. The functional can be generalized to the general $n \times n$ system in the following form

$$H[u(\cdot,t), v(\cdot,t)] \equiv (1 + K_1 F)(L + L_h) + K_2(Q_d(t) + E(t)) + k_3 D(t),$$

where $D(t)$ is used to control the jumps of L_h due to the introduction of the 'domain of influence' for shock waves. The 'domain of influence' for shock wave can be defined when we consider the following two sets of n scalar functions:

$$\theta^i(x,t) \equiv \sum_{\alpha^i \in J(u), x(\alpha^i) < x} (q_i(x(\alpha^i)+,t) - q_i(x(\alpha^i)-,t)),$$

$$\eta^i(x,t) \equiv \sum_{\alpha^i \in J(v), x(\alpha^i) < x} (q_i(x(\alpha^i)+,t) - q_i(x(\alpha^i)-,t)),$$

where $q_i(x,t)$, $i = 1,2,\dots,n$, is defined as in (3.1) using rarefaction wave curves instead of Hugoniot curves.

To estimate $dL(t)/dt$ inside each region $(p-1)Ns < t < pNs$, the following two lemmas are needed.

LEMMA 3.1. *Let $\bar{u} \in \Omega$, $\xi, \xi' \in \mathbf{R}$, $k \in \{1,\dots,n\}$. Define the states and the wave speeds*

$$u = H_k(\xi)(\bar{u}), \quad u' = H_k(\xi')(u), \quad u'' = H_k(\xi+\xi')(\bar{u}),$$
$$\lambda = \lambda_k(\bar{u},u), \quad \lambda' = \lambda_k(u,u'), \quad \lambda'' = \lambda_k(\bar{u},u'').$$

Then we have

$$|(\xi+\xi')(\lambda''-\lambda') - \xi(\lambda-\lambda')| = O(1)\cdot|\xi\xi'|\,|\xi+\xi'|.$$

LEMMA 3.2. *If the values ξ, ξ_j, ξ_j', $j = 1,2,\dots,n$, satisfy*

$$H_n(\xi_n)\circ\cdots\circ H_1(\xi_1)(u) = \begin{cases} H_n(\xi_n')\circ\cdots\circ H_1(\xi_1')\circ H_i(\xi)(u), \text{ or} \\ H_i(\xi)\circ H_n(\xi_n')\circ\cdots\circ H_1(\xi_1')(u), \end{cases} \tag{3.6}$$

then

$$|\xi_i - \xi_i' - \xi| + \sum_{j\neq i}|\xi_j - \xi_j'| = O(1)|\xi|\bigg(|\xi_i'||\xi_i'+\xi| + \sum_{j\neq i}|\xi_j'|\bigg).$$

For the particular case where $\alpha^i = (u_-, u_+)$ is a shock in v with jump $[\alpha^i] \equiv (u_+ - u_-)^i$, the first part of Lemma 3.2 becomes

$$\begin{aligned} &\big|q^+(\alpha^i) - q^-(\alpha^i) - [\alpha^i]\big| + \sum_{j\neq i}|q_j^+(\alpha^i) - q_j^-(\alpha^i)| \\ &\quad = O(1)\cdot\Big(|q^-(\alpha^i)||q^-(\alpha^i)+[\alpha^i]| + \sum_{j\neq i}\big|q_j^-(\alpha^i)\big|\Big)|\alpha^i|, \\ &\quad = O(1)\cdot\Big(|q^+(\alpha^i)||q^+(\alpha^i)+[\alpha^i]| + \sum_{j\neq i}\big|q_j^+(\alpha^i)\big|\Big)|\alpha^i|. \end{aligned} \tag{3.7}$$

For definiteness, we set $[\alpha^i] < 0$ if α^i is a shock, and $[\alpha^i] > 0$ if α^i is a rarefaction wave. Recalling that $[\alpha^i] \in]0,\epsilon]$ when α^i is a rarefaction wave, using both parts of Lemma 3.2, we have the estimates

$$\begin{aligned} &\big|q^+(\alpha^i) - q^-(\alpha^i) - [\alpha^i]\big| + \sum_{j\neq i}|q_j^+(\alpha^i) - q_j^-(\alpha^i)| \\ &\quad = O(1)\cdot\Big(\epsilon + |q^-(\alpha^i)||q^-(k_\alpha^i)+[\alpha^i]| + \sum_{j\neq i}\big|q_j^-(\alpha^i)\big|\Big)|\alpha^i|, \\ &\quad = O(1)\cdot\Big(\epsilon + |q^+(\alpha^i)|\,|q^+(\alpha^i)+[\alpha^i]| + \sum_{j\neq i}\big|q_j^+(\alpha^i)\big|\Big)|\alpha^i|. \end{aligned} \tag{3.8}$$

The error $O(1)\epsilon$ due to rarefaction shocks and the one due to the random sequence tend to zero as the grid size tends to zero. Besides these errors, there are two main errors of the following order in estimating $dL(t)/dt$ when $t \in ((p-1)Ns, pNs)$:

$$E_1 = \sum_{\alpha^i\in J}|\alpha^i|\sum_{j\neq i}\big|q_j^{\pm}(\alpha^i)\big|; \quad \text{and} \quad E_2 = \sum_{\alpha^i\in J}|\alpha^i|\max\{q^+(\alpha^i)q^-(\alpha^i), 0\}.$$

By using the strict hyperbolicity of the system, it can be shown that the error term E_1 can be controlled by the good terms in $dQ_d(t)/dt$. And the error term E_2 can be controlled by the good terms from $dE(t)/dt$ by the genuine nonlinearity of the charateristic field. The reason that the cubic order error term E_2 can be controlled by the generalised entropy functional comes from the following theorem for convex scalar conservation laws.

THEOREM 3.1. *For a convex scalar conservation law, the generalized entropy functional is defined as follows:*

$$E(t) = \sum_{\alpha\in J_1} |\alpha|\Big(\int_{x(\alpha)}^{\infty} (u-v)_+(x,t)dx + \int_{-\infty}^{x(\alpha)} (u-v)_-(x,t)dx\Big)$$

$$+ \sum_{\alpha\in J_2} |\alpha|\Big(\int_{x(\alpha)}^{\infty} (v-u)_+(x,t)dx + \int_{-\infty}^{x(\alpha)} (v-u)_-(x,t)dx\Big), \tag{3.9}$$

for any two approximate solutions $u(x,t)$ and $v(x,t)$ in the Glimm's scheme through the wave tracing method with total variations bounded by T.V.. The generalised entropy functional satisfies

$$\frac{d}{dt}E(t) \le -C_1 \sum_{\alpha\in J} |\alpha| \max\{q_-(\alpha)q_+(\alpha), 0\} + O(1)T.V.\epsilon, \tag{3.10}$$

where

$$q_\pm(\alpha) = q_\pm(\alpha(t)) \equiv (u_1 - u_2)(x(\alpha(t)\pm, t)). \tag{3.11}$$

Here the summation is over all waves α at time t in both solutions.

REMARK 3.2. Since the derivative of the integral of a convex entropy with respect to time gives a negative of all shock waves strengths to the cubic power, the L_2 norm of a solution can be used when we consider the case when one of the solution is a constant. In fact, for $u(x,0) \in L_1(x)$, the nonlinear functional $H[u(x,t)]$ takes a form [21]:

$$H[u(\cdot,t)] \equiv (1 + K_1F)L(t) + K_2(Q_d(t) + ||u(\cdot,t)||^2_{L_2}).$$

REMARK 3.3. For the case when the Hugoniot curves coincide with the rarefaction wave curves, i.e. the Temple's class [28], the nonlinear functional $H[u(x,t), v(x,t)]$ takes a very simple form [19]:

$$H[u(\cdot,t), v(\cdot,t)] \equiv (1 + K_1F)L(t) + K_2Q_d(t).$$

We conclude the above discussion into the following main theorem.

THEOREM 3.2. *Suppose that the total variation of the initial data of the solutions is sufficiently small and bounded by T.V., and that $u_0(x) - v_0(x) \in L_1(R)$. Then, for the exact weak solutions $u(x,t)$ and $v(x,t)$ of (1.1) constructed by Glimm's scheme, there exists a constant G independent of time such that*

$$\|u(x,t) - v(x,t)\}_{L_1} \le G\|u(x,s) - v(x,s)\|_{L_1},$$

for any s, t, $0 \le s \le t < \infty$.

This theorem immediately implies the following theorem on uniqueness of the weak solution constructed by Glimm scheme.

THEOREM 3.3. *For any given initial data with total variation sufficiently small, the whole sequence of the approximate solutions contructed by the Glimm scheme converges to a unique weak solution of* (1.1) *as the mesh sizes tend to zero.*

References

[1] A. Bressan and R.M. Colombo, *The semigroup generated by* 2×2 *conservation laws*, Arch. Rational Mech. Anal. **133** (1995), 1–75.

[2] A. Bressan, *A locally contractive metric for systems of conservation laws*, Estratto dagli Annali Della Scuola Normale Superiore di Pisa, Scienze Fisiche e Matematiche-serie IV. vol. XXII. Fasc. 1 (1995).

[3] A. Bressan, G. Goatin and B. Piccoli, *Well posedness of the Cauchy problem for* $n \times n$ *systems of conservation laws*, Memoir Amer. math. Soc., to appear.

[4] A. Bressan and P. LeFloch, *Uniqueness of weak solutions to systems of conservation laws*, preprint S.I.S.S.A., Trieste 1996.

[5] A. Bressan, T.-P. Liu and T. Yang, L_1 *stability estimates for* $n \times n$ *conservation laws*, Arch. Rational Mech. Anal., to appear.

[6] A. Bressan and A. Marson, *Error bounds for a determimistic version of the Glimm scheme*, Arch. Rational Mech. Anal., to appear.

[7] C.M. Dafermos, *Polygonal approximations of solutions of the initial value problem for a conservation law*, J. Math. Anal. Appl. **38** (1972), 33–41.

[8] ———, *Entropy and the stability of classical solutions of hyperbolic systems of conservation laws*. In: Lecture Notes in Mathematics (T. Ruggeri ed.), Montecatini Terme, 1994, Springer.

[9] R. DiPerna, *Uniqueness of solutions to hyperbolic conservation laws*, Indiana Univ. Math. J. **28** (1979), 137–188.

[10] bysame, *Global existence of solutions to nonlinear hyperbolic systems of conservation laws*, J. Diff. Equa. **20** (1976), 187–212.

[11] J. Glimm, *Solutions in the large for nonlinear hyperbolic systems of equations*, Comm. Pure Appl. Math. **18** (1965), 697–715.

[12] J. Glimm and P. Lax, *Decay of solutions of systems of hyperbolic conservation laws*, Memoirs Amer. Math. Soc. **101**, 1970.

[13] P.D. Lax, *Hyperbolic systems of conservation laws II*, Comm. Pure Appl. Math. **10** (1957), 537–566.

[14] P.D. Lax, *Shock waves and entropy*. In: Contribution to Nonlinear Functional Analysis, (E. Zarantonello ed.), Academic Press, N.Y., 1971, pp.603–634.

[15] P. LeFloch and Z. P. Xin, *Uniqueness via the adjoint problems for systems of conservation laws*, Comm. Pure Appl. Math. XLVI (1993) 1499–1533 .

[16] T.-P. Liu, *The deterministic version of the Glimm scheme*, Comm. Math. Phys. **57** (1975), 135–148.

[17] ———, *Uniqueness of weak solutions of the Cuachy problem for general* 2×2 *conservation laws*, J. Diff. Equa. **20** (1976), 369–388.

[18] ———, *Admissible solutions of hyperbolic conservation laws*, Memoirs of the American Mathematical Society, Vol. 30, No. 240, 1981.

[19] T.-P. Liu and T. Yang, *Uniform* L_1 *boundedness of solutions of hyperbolic conservation laws*, Methods and Appl. Anal. **4** (1997), 339–355.

[20] ———, *A generalised enropy for scalar conservation laws*, preprint.

[21] ———, L_1 *stability of conservation laws with coinciding Hugoniot and characteristic curves*, Indiana Univ. Math. J.

[22] ———, L_1 *stability for* 2×2 *systems of hyperbolic conservation laws*, J. Amer. Math. Soc., to appear.

[23] ———, *Well-posedness theory for hyperbolic conservation laws*, to appear.

[24] O. Oleinik, *On the uniqueness of the generalized solution of the Cauchy problem for a nonlinear system of equations occuring in mechanics*, Usp. Mat. Nauk.(N.S.), 12(1957), 169–176. (in Russian)

[25] M. Schatzman, *Continuous Glimm functionals and uniqueness of solutios of the Riemann problem*, Indiana Univ. Math. J. **34** (1985).

[26] J. Smoller, *Shock Waves and Reaction-diffusion Equations*, Springer-Verlag, New York, 1982.

[27] B. Temple, *No L_1-contractive metrics for system of conservation laws*, Trans. Amer. Math. Soc. **288** (1985), 471–480.
[28] ———, *Systems of conservation laws with invariant submanifolds*, Trans. Amer. Math. Soc. **280** (1983), 781–795.

Department of Mathematics, Stanford University, Stanford, CA 94305-2060
E-mail address: liu@math.stanford.edu

Department of Mathematics, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong
E-mail address: matyang@math.cityu.edu.hk

Contemporary Mathematics
Volume **238**, 1999
B 0-8218-1196-7-03548-9

The Geometry of the Stream Lines of Steady States of the Navier–Stokes Equations

Tian Ma and Shouhong Wang

Abstract. It is proved in this article that for any external forcing in an open and dense subset of $C^\alpha(TM)$ $(0 < \alpha < 1)$, all steady state solutions of the two-dimensional Navier-Stokes equations are structurally stable.

1. Introduction

The motion of an incompressible fluid is governed by the Navier-Stokes (or Euler) equations, which form an infinite dimensional dynamical system. From the Lagrangian point of view, the velocity field v, which is a solution of the Navier-Stokes equations, determines the dynamics of the fluid particles in the physical space the fluid occupies. One of the main motivations of this article and accompanying articles is to study the geometrical/topological structure of two-dimensional fluid flows in the physical spaces.

The general philosophy we adopt in this project includes two aspects:

1. to develop a general (global) geometrical/topological theory of the velocity vector field $v(\cdot, t)$ at each time instant, treating the time t as a parameter, and then
2. allowing the time variable to change, to study the structural transitions of the velocity field v.

The study along the first direction was initialized in [**MW97, MW98**]. The main objective in this direction is to establish a geometrical/topological theory for divergence-free vector fields on general two-dimensional compact manifolds with or without boundary. The study in the second direction aims in particular the connections between the solutions of the Navier-Stokes (or Euler) equations and the dynamics of the velocity fields in the physical space. The main result in this

1991 *Mathematics Subject Classification.* Primary 35Q35; Secondary 76, 58F.

Key words and phrases. steady states, 2D Navier-Stokes equations, structural stability, Sard-Smale theorem.

The authors are grateful to an anonymous referee for his/her extremely careful reading of the earlier version of this article, and for his/her insightful comments. This work was supported in part by the Office of Naval Research under Grant NAVY-N00014-96-1-0425 and by the National Science Foundation under Grant NSF-DMS-9623071.

article provides an example of such connections by regarding the external forcing as a parameter.

One main result in [**MW97, MW98**] is a global structural stability theorem of divergence-free vector fields, providing necessary and sufficient conditions for structural stability of a divergence-free vector fields; see Theorem 2.3. The study of structural stability has been the main driving force behind much of the development of dynamical systems theory (see among others [**Sma67, PdM82, Pei62, Pug67, Rob70, Rob74, Shu78**]). We are interested in the structural stability of a divergence-free vector field with perturbations of divergence-free vector fields. We call this notion of structural stability the incompressibly structural stability or simply structural stability. Notice that the divergence-free condition changes completely the general features of structurally stable fields as compared to the situation when this condition is not present. The latter case was studied in a classical paper of M. Peixoto [**Pei62**]. The conditions for structural stability and genericity in Peixoto's theorem are: (i) the field can have only a finite number of singularities and closed orbits (critical elements) which must be hyperbolic; (ii) there are no saddle connections; (iii) the non wandering set consists of singular points and closed orbits.

The necessary and sufficient conditions for divergence-free vector fields we obtain in Theorem 2.3 are: (1) v is regular; (2) all interior saddle points of v are self-connected; and (3) each boundary saddle point is connected to boundary saddles on the same connected component of the boundary. The first condition here requires only regularity of the field and so it does not exclude centers which are not hyperbolic and excluded by (i) above. The second condition is of a completely different nature than the corresponding one in the Peixoto theorem. Namely, condition (ii) above excludes the possibility of saddle connections. In contrast, (2) amounts to saying that all interior saddles are self-connected! Namely, the interior saddles occur in graphs whose topological form is that of the number 8, being the singularities themselves hyperbolic. The condition (3) deals with singularities on the boundary, and we mention that similar condition appears in extensions of Peixoto's theorem to manifolds with boundary (see, *e.g.*, G. L. dos Reis [**dR78**] and M. J. Pacifico [**Pac84**]).

Moreover, a direct consequence of the Peixoto structural stability theorem and Theorem 2.3 here is that no divergence-free vector field is structurally stable under general C^r vector fields perturbations. Such a drastic change in the stable configurations is explained by the fact that **divergence-free fields preserve volume** and so attractors and sources can never occur for these fields. In particular, this makes it natural the restriction that saddles in the boundary must be connected with saddles in the boundary on the same connected component, in the third condition.

The main objective of this article is to study the structural stability of the solutions of the 2D Navier-Stokes equations. We prove in Theorem 3.1 that for any external forcing in an open and dense subset of C^α $(0 < \alpha < 1)$, all steady state solutions of the two-dimensional Navier-Stokes equations are structurally stable. Namely, the structurally stable steady states of a 2D incompressible fluid are generic.

The proof of the main theorem, Theorem 3.1, is accomplished by using $C^{2+\alpha}$ Schauder type of *a priori* estimates of the steady state solutions of the 2D Navier-Stokes equations with free boundary conditions, and the Sard-Smale theorem in

Banach spaces. We would like to mention that Foias and Temam were the first ones in [**FT77**] to use the Sard-Smale theorem to study solutions of the Navier-Stokes equations; they proved that the number of steady states of the Navier-Stokes equations is generically finite.

The paper is organized as follows. In Section 2, we recall the structural stability theorem of 2D divergence-free vector fields obtained in [**MW97**], announced in [**MW98**], along the main ideas of its proof. The main theorem, Theorem 3.1, and its proof are given in Section 3.

2. Structural Stability of Divergence-Free Vector Fields with Free Boundary Conditions

Let $M \subset \mathbb{R}^2$ be a closed and bounded domain with $C^{r+1}(r \geq 2)$ boundary. We remark that all results in this article hold true when M is a two-dimensional compact manifold with boundary, which is diffeomorphic to a sub-manifold of the unit sphere S^2.

Let TM for the tangent bundle of M, and $C^r(TM)$ be the space of all C^r vector fields on M. We set

$$\begin{aligned} D^r(TM) &= \{v \in C^r(TM) | \quad v_n|_{\partial M} = 0, \operatorname{div} v = 0\}, \\ B^r(TM) &= \{v \in D^r(TM) | \quad \frac{\partial v_\tau}{\partial n}|_{\partial M} = 0\}, \end{aligned} \tag{2.1}$$

where $v_n = v \cdot n, v_\tau = v \cdot \tau$, n and τ are the unit normal and tangent vectors on ∂M respectively. If $r = k + \alpha$ with $k \geq 0$ an integer and $0 < \alpha < 1$, then $v \in C^r(TM)$ means that $v \in C^k(TM)$ and all derivatives of v up to order k are α–Hölder continuous. By definition, vector fields in $B^r(TM)$ satisfy the following free boundary conditions:

$$v_n|_{\partial M} = 0, \qquad \frac{\partial v_\tau}{\partial n}|_{\partial M} = 0. \tag{2.2}$$

We have obtained in [**MW97**] necessary and sufficient conditions for structural stability of divergence–free vector fields in $D^r(TM)$. In order to facilitate the understanding of the structure of the solutions of the Navier-Stokes equations in the underlying physical space, we study in this section the structural stability of divergence–free vectors on M with free boundary conditions.

Definition 2.1. Two vector fields $u, v \in C^r(TM)$ are called topologically equivalent if there exists a homeomorphism $\phi : M \to M$ which takes the orbits of u to orbits of v, preserving orientation.

Definition 2.2. Let X be either $D^r(TM)$ or $B^r(TM)$. A vector field $v \in X$ is called structurally stable in X if there exists a neighborhood $\mathcal{O} \subset X$ of v such that for any $u \in \mathcal{O}$, u and v are topologically equivalent.

A point $p \in M$ is called a singular point of $v \in C^r(TM)$ if $v(p) = 0$; a singular point p of v is called nondegenerate if the Jacobian matrix $Dv(p)$ is invertible; v is called regular if all singular points of v are nondegenerate.

Let $v \in D^r(TM)$ be regular. It is easy to see the following basic facts (see [**MW97**]):

1. An interior non-degenerate singular point of v can either be a center or a saddle, and a nondegenerate boundary singularity must be a saddle;

2. Saddles of v must be connected to saddles. An interior saddle $p \in \overset{\circ}{M}$ is called *self-connected*, if p is connected only to itself, i.e. p occurs in a graph whose topological form is that of the number 8.

THEOREM 2.3. *Let $X = D^r(TM)$ or $X = B^r(TM)(r \geq 1)$, and $v \in X$. Then v is structurally stable in X if and only if*

(1) *v is regular;*
(2) *all interior saddles of v are self–connected; and*
(3) *each boundary saddle point is connected to boundary saddle points on the same connected component of the boundary.*

Moreover, the set of all structurally stable vector fields is open and dense in X.

IDEAS OF THE PROOF. When $X = D^r(TM)$, Theorem 2.3 is exactly Theorems 5.1 and 5.3 in [**MW97**]. When $X = B^r(TM)$, the proof of Theorem 2.3 is the same as that of Theorems 5.1 and 5.3 in [**MW97**]. The main ingredients of the proof include the following aspects; see [**MW97**] for details:

1. A global structural classification theorem: the topological set of orbits of a regular $v \in D^r(TM)(r \geq 1)$ consists of finite connected components of circle cells, circle bands and saddle connections. The largest neighborhood of a center of v containing closed orbits is called a *circle cell*; the largest neighborhood of a closed orbit, different from circle cells, is called a *circle band*.
2. The construction of a special class of turbular incompressible flows, and their applications to breaking saddle connections.
3. Extension of orbit preserving maps on the boundaries of circle cells and circle bands to the interiors of the circle cells and circle bands, preserving the orbits.

□

3. Genericity of Structurally Stable Steady States of the 2D Navier-Stokes Equations

3.1. The Main theorem. The main objective of this article is to study the generic properties of structurally stable solutions of the steady–state Navier–Stokes equations with free boundary conditions. The problem reads

$$-\mu\Delta u + (u \cdot \nabla)u + \nabla p = f, \quad \text{in } \overset{\circ}{M} \subset \mathbb{R}^2, \tag{3.1}$$

$$\text{div } u = 0, \qquad \text{in } \overset{\circ}{M}, \tag{3.2}$$

$$u_n = 0, \ \frac{\partial u_\tau}{\partial n}|_{\partial M} = 0, \qquad \text{on } \partial M. \tag{3.3}$$

Here $\overset{\circ}{M}$ stands for the interior of M. If $u \in B^{2+\alpha}(TM)$ satisfies (3.1), then u is a solution of the problem (3.1–3.3).

The main theorem in this article is

THEOREM 3.1. *For any $\mu > 0$, there is an open and dense set $\mathcal{F} \subset C^\alpha(TM)(0 < \alpha < 1)$ such that for each $f \in \mathcal{F}$, the solutions $u \in X = B^{2+\alpha}(TM)$ of (3.1–3.3) are structurally stable in X.*

3.2. Some A Priori Estimates. The proof of Theorem 3.1 is based on some *a priori* estimates of the steady state solutions of the Navier-Stokes equations, and an infinite dimensional version of the Sard theorem due to S. Smale [**Sma65**].

We start with some *a priori* estimates of the solutions of the following 2D Stokes equations:

$$\begin{cases} -\mu\Delta u + \nabla p = g, \quad \text{in } \overset{\circ}{M}, \\ \text{div } u = 0, \quad \text{in } \overset{\circ}{M}, \\ u_n = 0, \ \dfrac{\partial u_\tau}{\partial n}|_{\partial M} = 0, \qquad \text{on } \partial M. \end{cases} \tag{3.4}$$

It is easy to see that the pressure p can only be determined up to constants. Therefore, we let $C_0^{1+\alpha}(M) = C^{1+\alpha}(M)/\mathbb{R}$ be the space of all $C^{1+\alpha}$ functions module constants. Obviously,

$$C_0^{1+\alpha}(M) = \left\{ p \in C^{1+\alpha}(M) \mid \int_M p dM = 0 \right\}.$$

LEMMA 3.2. *Let $(u,p) \in B^{2+\alpha}(TM) \times C_0^{1+\alpha}(M)$ be a solution of (3.4), and $g \in C^\alpha(TM)$. Then*

$$\|u\|_{C^{2+\alpha}} + \|p\|_{C^{1+\alpha}} \le C\|g\|_{C^\alpha}, \tag{3.5}$$

where $c > 0$ is a constant.

PROOF. We proceed by applying the general regularity result for elliptic system of equations by Agmon–Doglis–Nirenberg [**ADN64**] to the above Stokes problem (3.4) with free boundary conditions.

In [**Tem84**], the ellipticity of (3.4) is verified. By Theorem 9.3 and Remark 2 on p. 74 of [**ADN64**], it remains to check the Complementary Boundary Conditions required there for the free boundary conditions (3.3).

Let $u = (u_1, u_2), u_3 = \frac{1}{\mu}p, v = (u_1, u_2, u_3)$. The principal part of (3.4) is

$$L(D)v = \begin{pmatrix} (\frac{\partial}{\partial x_1})^2 + (\frac{\partial}{\partial x_1})^2 & 0 & -\frac{\partial}{\partial x_1} \\ 0 & (\frac{\partial}{\partial x_1})^2 + (\frac{\partial}{\partial x_2})^2 & -\frac{\partial}{\partial x_2} \\ \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}.$$

The boundary operator is

$$B(D)v = \begin{pmatrix} n_1 & n_2 & 0 \\ \tau_1 n_1 \frac{\partial}{\partial x_1} + \tau_1 n_2 \frac{\partial}{\partial x_2} & \tau_2 n_1 \frac{\partial}{\partial x_1} + \tau_2 n_2 \frac{\partial}{\partial x_2} & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}.$$

For a vector $\xi = (\xi_1, \xi_2)$, the corresponding matrices of the above differential operators are given by

$$L(\xi) = \begin{pmatrix} \xi_1^2 + \xi_2^2 & 0 & \xi_1 \\ 0 & \xi_1^2 + \xi_2^2 & \xi_2 \\ -\xi_1 & -\xi_2 & 0 \end{pmatrix},$$

$$B(\xi) = \begin{pmatrix} n_1 & n_2 & 0 \\ \tau_1 n_1 \xi_1 + \tau_1 n_2 \xi_2 & \tau_2 n_1 \xi_1 + \tau_2 n_2 \xi_2 & 0 \end{pmatrix}.$$

For the normal and tangent vectors n, τ on the boundary and a parameter t, we have

$$L(\tau + tn) = \begin{pmatrix} (1+t^2) & 0 & \tau_1 + tn_1 \\ 0 & (1+t^2) & \tau_2 + tn_2 \\ -(\tau_1 + tn_1) & -(\tau_2 + tn_2) & 0 \end{pmatrix},$$

$$B(\tau + tn) = \begin{pmatrix} n_1 & n_2 & 0 \\ t\tau_1 & t\tau_2 & 0 \end{pmatrix}.$$

Hence we have

$$B(\tau + tn) \times L(\tau + tn) = \begin{pmatrix} n_1(1+t^2) & n_2(1+t^2) & t \\ t\tau_1(1+t^2) & t\tau_2(1+t^2) & t \end{pmatrix}.$$

It is easy to see that the algebraic equation

$$\det\ L(\tau + tn) = 0$$

has exactly two roots with positive imaginary part and these roots are all equal to $t^+ = i$. Therefore

$$M^+(\tau + tn) = (t - t^+)^2 = (t - i)^2.$$

Obviously the algebraic equation system

$$\begin{cases} n_1(1+t^2)c_1 + n_2(1+t^2)c_2 = 0, \\ t\tau_1(1+t^2)c_1 + t\tau_2(1+t^2)c_2 = 0 \pmod{M^+ = (t-i)^2}, \\ tc_1 + tc_2 = 0, \end{cases}$$

has only zero solution $c_1 = c_2 = 0$, and the Complementary Boundary conditions hold. The proof is complete.

□

The L_p estimates for the solutions of the Stokes equations with free boundary conditions (3.4) can also be obtained in the same fashion using Theorem 10.5 on p. 78 of [**ADN64**]. Notice that the same type of L_p estimates for the Stokes equations with Dirichlet boundary conditions are given in [**Tem84**].

LEMMA 3.3. *Let $(u, p) \in W^{2,p}(TM) \times W^{1,p}(M) (p \geq 2)$ be a solution of (3.4) with $\int_M p dM = 0$, and $g \in L^p(TM)$. Then*

$$\|u\|_{W^{2,p}} + \|p\|_{W^{1,p}} \leq c\|g\|_{L^p}, \tag{3.6}$$

where $c > 0$ is a constant.

Now we return to derive the $C^{2+\alpha}$–estimates for the steady state solutions of the Navier–Stokes equations with free boundary conditions (3.1)–(3.3).

LEMMA 3.4. *Let $(u, p) \in B^{2+\alpha}(TM) \times C_0^{1+\alpha}(M)$ be a solution of the stationary Navier-Stokes equations (3.1)–(3.3) with free boundary conditions, and $f \in C^\alpha(TM)$. Then*

$$\|u\|_{C^{2+\alpha}} + \|p\|_{C^{1+\alpha}} \leq C \left[\|f\|_{C^\alpha} + \|f\|_{L^p}^{4p}\right], \qquad p > \frac{2}{1-\alpha}. \tag{3.7}$$

PROOF. Let $g = f - (u \cdot \nabla)u$, then $g \in C^\alpha(TM)$. Lemma 3.2 yields that

$$\|u\|_{C^{2+\alpha}} + \|p\|_{C^{1+\alpha}} \leq C\|g\|_{C^\alpha} \leq C\left[\|f\|_{C^\alpha} + \|u\|^2_{C^{1+\alpha}}\right]. \tag{3.8}$$

By the Sobolev embedding theorem,

$$\|u\|_{C^{1+\alpha}} \leq C\|u\|_{W^{2,p}}, \qquad p > \frac{2}{1-\alpha}. \tag{3.9}$$

By Lemma 3.3

$$\begin{aligned} \|u\|_{W^{2,p}} &\leq C\|g\|_{L^p} \\ &\leq C\Big[\|f\|_{L^p} + \|u \cdot Du\|_{L^p}\Big] \\ &\leq C\Big[\|f\|_{L^p} + \|u\|_{L^{2p}} \cdot \|Du\|_{L^{2p}}\Big]. \end{aligned} \tag{3.10}$$

To estimate $\|Du\|_{L^{2p}}$, we recall a standard interpolation inequality for L^p norms (see (7.10) on p. 139 in [**GT77**]):

$$||u||_r \leq \varepsilon||u||_s + \varepsilon^{-\mu}||u||_t, \tag{3.11}$$

where $1 \leq t \leq r \leq s$ and

$$\mu = \left(\frac{1}{t} - \frac{1}{r}\right) / \left(\frac{1}{r} - \frac{1}{s}\right).$$

Therefore

$$\begin{aligned} \|Du\|_{L^{2p}} &\leq \varepsilon\|Du\|_{L^{4p}} + \varepsilon^{-\beta}\|Du\|_{L^2}, \\ &\leq C\varepsilon\|u\|_{W^{2,p}} + \varepsilon^{-\beta}\|Du\|_{L^2}, \end{aligned} \tag{3.12}$$

where $\varepsilon > 0$ is an arbitrary constant, and $\beta = 2(p-1)$. By the Poincaré inequalities and the $W^{1,2}$–estimates for the stationary solutions of the Navier–Stokes equations, we deduce that

$$\|u\|_{L^{2p}} \leq C\|Du\|_{L^2} \leq C\|f\|_{L^2}. \tag{3.13}$$

We infer then from (3.10–3.13) that

$$\|u\|_{W^{2,p}} \leq C\|f\|_{L^p} + C\varepsilon\|f\|_{L^2} \cdot \|u\|_{W^{2,p}} + C\varepsilon^{-\beta}\|f\|^2_{L^2}.$$

We take $C\varepsilon\|f\|_{L^2} = \frac{1}{2}$, then

$$\|u\|_{W^{2,p}} \leq C\|f\|_{L^p} + C\|f\|^{\beta+2}_{L^2} \leq C\|f\|^{2p}_{L^p}.$$

Therefore (3.7) follows. The proof is complete. □

3.3. The Sard-Smale Theorem on Banach Spaces. The proof of Theorem 3.1 relies on the Sard theorem on Banach spaces due to Smale [**Sma65**].

Let E_1 and E_2 be two Banach spaces. A map $G : E_1 \to E_2$ is called a completely continuous field if $G = L + H$, where $L : E_1 \to E_2$ is a linear isomorphism and $H : E_1 \to E_2$ is a compact operator. We note that a C^1 completely continuous field $G : E_1 \to E_2$ must be a Fredholm map of index zero.

Let $G : E_1 \to E_2$ be a C^1 completely continuous field. A point $u \in E_1$ is called a regular point of G if $G'(u) : E_1 \to E_2$ is an isomorphism, and is singular point if it is not a regular point. The image of a singular point under G is called a singular value of G, and the points in the complement of all singular values are called regular values of G. Notice that if $f \in E_2$ is not in the image $G(E_1)$, then f is automatically a regular value of G.

THEOREM 3.5. (Smale [**Sma65**]). *Let E_1 and E_2 be two Banach spaces and $G : E_1 \to E_2$ be a C^1 completely continuous field. Then the set of all regular values of G is dense in E_2. Moreover, if $f \in E_2$ is a regular value of G, then $G^{-1}(f)$ is discrete.*

3.4. Completely Continuous Fields Defined by the Navier–Stokes Equations. Let

$$G = L + H : B^{2+\alpha}(TM) \times C_0^{1+\alpha}(M) \to C^\alpha(TM) \tag{3.14}$$

be a map such that for $(u, p) \in B^{2+\alpha}(TM) \times C_0^{1+\alpha}(M)$,

$$\begin{cases} L(u,p) & = -\mu\Delta u + \nabla p, \\ H(u,p) & = (u \cdot \nabla)u. \end{cases} \tag{3.15}$$

We notice that $L : B^{2+\alpha}(TM) \times C_0^{1+\alpha}(M) \to C^\alpha(TM)$ is a bounded linear operator corresponding to the Stokes equation (3.4), and $H : B^{2+\alpha}(TM) \times C_0^{1+\alpha}(M) \to C^\alpha(TM)$ is a C^∞ nonlinear operator.

LEMMA 3.6. *The operator G is a completely continuous field. Moreover G is a surjective map, i.e. $G(B^{2+\alpha}(TM) \times C_0^{1+\alpha}(M)) = C^\alpha(TM)$.*

PROOF. It is well known that for any $g \in C^\infty(TM)$, the Stokes equation (3.6) has a unique solution $(u, p) \in C^\infty(TM) \times C^\infty(M)$ (for $p \in C^\infty(M)$ up to a constant). Therefore $L(B^{2+\alpha}(TM) \times C_0^{1+\alpha}(TM))$ is dense in $C^\alpha(TM)$. It follows from Lemma 3.2 that $L : B^{2+\alpha}(TM) \times C_0^{1+\alpha}(M) \to C^\alpha(TM)$ is an isomorphism. The compactness of $H : H^{2+\alpha}(TM) \times C_0^{1+\alpha}(M) \to C^\alpha(TM)$ is obvious.

Notice that for any $f \in C^\infty(TM)$ the Navier–Stokes equation (3.1–3.3) has a solution $(u, p) \in C^\infty(TM) \times C^\infty(M)$, and by Lemma 3.4, $G : B^{2+\alpha}(TM) \times C_0^{1+\alpha}(M) \to C^\alpha(TM)$ is surjective.

The proof is complete. □

THEOREM 3.7. *Let $G : B^{2+\alpha}(TM) \times C_0^{1+\alpha}(M) \to C^\alpha(TM)$ be defined by (3.14). Then the set of all regular values of G is open and dense in $C^\alpha(TM)$. Furthermore if $f \in C^\alpha(TM)$ is a regular value of G, then $G^{-1}(f)$ is nonempty and finite.*

PROOF. Let $\mathcal{R} \subset C^\alpha(TM)$ the set of regular values of G. By the Sard–Smale theorem and Lemma 3.6, $\mathcal{R}$ is dense in $C^\alpha(TM)$. Since G is surjective, we infer from (3.7) that for any $f \in \mathcal{R}, G^{-1}(f)$ is nonempty and finite.

Let E_1 and E_2 be two Banach spaces, and $L(E_1, E_2)$ be the space of all bounded linear operators from E_1 to E_2. It is known that the set of all isomorphisms from E_1 to E_2 is open in $L(E_1, E_2)$.

Let $f \in \mathcal{R}$, and $G^{-1}(f) = \{v_1, \cdots, v_n\}$. By the inverse function theorem, for each $v_i \in G^{-1}(f)(1 \le i \le n)$ there is a neighborhood $U_i \subset B^{2+\alpha}(TM) \times C_0^{1+\alpha}(M)$ of v_i and a neighborhood $\mathcal{F}_i \subset C^\alpha(TM)$ of f such that $G : U_i \to \mathcal{F}_i$ is a homeomorphism.

Therefore there is an open set $\mathcal{O} \subset \mathcal{F}_1 \cap \cdots \cap \mathcal{F}_n$ with $f \in \mathcal{O}$ such that $G^{-1}(\mathcal{O}) = V_1 + \cdots + V_n, V_i \cap V_j = \emptyset (i \ne j), v_i \in V_i (1 \le i \cdot j \le n)$, and for any $u \in \Sigma_{i=1}^n V_i, G'(u)$ is an isomorphism. Hence $\mathcal{O} \subset \mathcal{R}$ and $\mathcal{R} \subset C^\alpha(TM)$ is open.

The proof is complete. □

REMARK 3.8. The result of Theorem 3.7 in $W^{2,2}$ – space for the Dirichlet boundary conditions was obtained in [**FT77**].

3.5. Completion of the Proof of Theorem 3.1. Theorem 3.7 tells us that there is an open and dense set $\mathcal{R} \subset C^\alpha(TM)$ such that

$$G_0 = G|_{G^{-1}(\mathcal{R})} : G^{-1}(\mathcal{R}) \to \mathcal{R}$$

is an open map, i.e. G_0 maps open sets in $G^{-1}(\mathcal{R})$ to open sets in $\mathcal{R}$. Moreover, since $\mathcal{R}$ is open, $G^{-1}(\mathcal{R})$ is open in $B^{2+\alpha}(TM) \times C_0^{1+\alpha}(M)$.

By Theorem 2.3, the set of $X_0 \subset X = B^{2+\alpha}(TM)$ of all structurally stable vector fields in X is open and dense in X. Let $K = (X_0 \times C_0^{1+\alpha}(M)) \cap G^{-1}(\mathcal{R}) \subset X_0 \times C_0^{1+\alpha}(M)$, and $\mathcal{F} = G(K) \subset \mathcal{R}$. It is easy to see that $\mathcal{F}$ is open and dense in $\mathcal{R}$. Namely $\mathcal{F}$ is open and dense in $C^\alpha(TM)$.

The proof of Theorem 3.1 is complete.

Remark 3.9. We have in fact proved the following slightly stronger results: There exists an open and dense set $\mathcal{F}$ of $C^\alpha(TM)$ such that for any $f \in \mathcal{F}$,

1. the corresponding steady states are given by $v_i = (u_i, p_i) \in B^{2+\alpha}(TM) \times C_0^{1+\alpha}(M)$ $(i = 1, \cdots, I(f))$ for some integer $I(f)$,
2. there exist open neighborhoods $N_i \subset X_0 \times C_0^{1+\alpha}(M)$ of v_i, and an open neighborhood $N(f)$ of f in $C^\alpha(TM)$ such that for each $i = 1, \cdots, I(f)$,

$$G : N_i \to N(f)$$

are diffeomorphisms. Here $X_0 \subset B^{2+\alpha}(TM)$ is the set of all structurally stable vector fields in $B^{2+\alpha}(TM)$.

References

[ADN64] S. Agmon, A. Douglis, and L. Nirenberg. Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions. II. *Comm. Pure Appl. Math.*, 17:35–92, 1964.

[dR78] Genésio Lima dos Reis. Structural stability of equivariant vector fields. *An. Acad. Brasil. Ciênc.*, 50(3):273–276, 1978.

[FT77] C. Foiaş and R. Temam. Structure of the set of stationary solutions of the Navier-Stokes equations. *Comm. Pure Appl. Math.*, 30(2):149–164, 1977.

[GT77] D. Gilbarg and N.S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Springer-Verlag, New York, Heidelberg, Berlin, 1977.

[MW97] T. Ma and S. Wang. Structural classification and stability of divergence-free vector fields. *The Institute for Scientific Computing and Applied Mathematics Preprint*, Indiana University, 1997.

[MW98] T. Ma and S. Wang. Dynamics of incompressible vector fields. *Applied Mathematics Letters, to appear*, 1998.

[Pac84] M. J. Pacífico. Structural stability of vector fields on 3-manifolds with boundary. *J. Differential Equations*, 54(3):346–372, 1984.

[PdM82] J. Palis and W. de Melo. *Geometric theory of dynamical systems*. Springer-Verlag, New York, Heidelberg, Berlin, 1982.

[Pei62] M. Peixoto. Structural stability on two dimensional manifolds. *Topology*, 1:101–120, 1962.

[Pug67] Charles C. Pugh. The closing lemma. *Amer. J. Math.*, 89:956–1009, 1967.

[Rob70] C. Robinson. Generic properties of conservative systems, I, II. *Amer. J. Math.*, 92:562–603 and 897–906, 1970.

[Rob74] C. Robinson. Structure stability of vector fields. *Ann. of Math.*, 99:154–175, 1974.

[Shu78] Michael Shub. *Stabilité globale des systèmes dynamiques*, volume 56 of *Astérisque*. Société Mathématique de France, Paris, 1978. With an English preface and summary.

[Sma65] S. Smale. An infinite dimensional version of Sard's theorem. *Amer. J. Math.*, 87:861–866, 1965.

[Sma67] S. Smale. Differential dynamical systems. *Bull. AMS*, 73:747–817, 1967.

[Tem84] R. Temam. *Navier-Stokes Equations, Theory and Numerical Analysis, 3rd, rev. ed.* North Holland, Amsterdam, 1984.

Department of Mathematics & The Institute for Scientific Computing and Applied Mathematics, Indiana University, Bloomington, IN 47405 & Department of Mathematics, Sichuan University, Chengdu, P. R. China
E-mail address: tianma@cobray.math.indiana.edu

Department of Mathematics, Indiana University, Bloomington, IN 47405
E-mail address: showang@indiana.edu

Contemporary Mathematics
Volume **238**, 1999
B 0-8218-1196-7-03549-0

On Complex-Valued Solutions to a 2-D Eikonal Equation. Part One: Qualitative Properties

Rolando Magnanini & Giorgio Talenti

ABSTRACT. $w_x^2 + w_y^2 + n^2(x,y) = 0$ is a two-dimensional version of the *eikonal equation* appearing in the generalizations of geometrical optics that deal with diffraction. Here x and y denote rectangular coordinates in the Euclidean plane, and n is real-valued. A framework is proposed, which consists of Bäcklund transformations and second-order PDEs governing $\mathrm{Re}(w)$ and $\mathrm{Im}(w)$. Sample solutions are constructed in the case where n is constant. The critical points of $\mathrm{Re}(w)$ are the main motif. Theorems, focusing on the geometry of such critical points, are given.

1. Introduction

1.1. Formalities. Let x and y denote rectangular coordinates in Euclidean plane $\mathbb{R}^2$, and let n be a real-valued sufficiently well-behaved function of x and y. Assume $n \geq 0$ for definiteness. The following nonlinear first-order partial differential equation

$$\left(\frac{\partial w}{\partial x}\right)^2 + \left(\frac{\partial w}{\partial y}\right)^2 + n^2(x,y) = 0 \tag{1.1.1}$$

—all of whose solutions are *complex-valued*—is the theme of the present and a subsequent paper.

We offer motivations in Subsection 1.2 below, and devote the present subsection to sketching some features of equation (1.1.1) heuristically.

Let u and v be *real-valued* functions of x and y, and let

$$w = u + iv. \tag{1.1.2}$$

Then w is a solution to (1.1.1) if and only if u and v obey

$$\begin{aligned} u_x^2 + u_y^2 - v_x^2 - v_y^2 + n^2 &= 0, \\ u_x v_x + u_y v_y \qquad\qquad &= 0. \end{aligned} \tag{1.1.3}$$

1991 *Mathematics Subject Classification.* Primary 35J70, 35Q60; Secondary 49N60.

Key words and phrases. Partial differential equations, complex-valued solutions, Bäcklund transformations, level curves, critical points.

This work was supported by the Italian MURST.

(We let i denote $\sqrt{-1}$ throughout, and denote differentiations either by $\partial/\partial x$ and $\partial/\partial y$, or by subscripts.)

System (1.1.3), which is fully nonlinear, should be qualified *elliptic-parabolic* or *degenerate elliptic*. The real-valued solutions u and v to (1.1.3) such that the gradient of u is nowhere equal to zero are *elliptic*. A *degeneracy* occurs if a *critical point* of u, i.e. a point where $u_x = u_y = 0$, exists. (These critical points will prove central to subsequent developments.)

In effect,

$$\text{the characteristic determinant} = \begin{vmatrix} 2u_x & 2u_y & -2v_x & -2v_y \\ dx & dy & 0 & 0 \\ v_x & v_y & u_x & u_y \\ 0 & 0 & dx & dy \end{vmatrix},$$

the Jacobian determinant of

$$u_x^2 + u_y^2 - v_x^2 - v_y^2 + n^2, \quad u_x dx + u_y dy, \quad u_x v_x + u_y v_y, \quad v_x dx + v_y dy$$

with respect to u_x, u_y, v_x, v_y. Therefore

$$\text{characteristic determinant} = 2\left[(-u_x dy + u_y dx)^2 + (-v_x dy + v_y dx)^2\right].$$

Here dx and dy serve as auxiliary variables. The ensuing discriminant equals

$$-4\,(u_x v_y - u_y v_x)^2,$$

negative or zero. Recall that two distinct real characteristic roots stand for hyperbolicity, and two distinct characteristic roots having non-zero imaginary parts stand for ellipticity. Therefore all solutions to (1.1.3) are either elliptic or parabolic. Observe that $u_x v_y - u_y v_x = \partial(u,v)/\partial(x,y)$, the Jacobian determinant of u and v. In other words, the elliptic solutions to (1.1.3) are precisely those solution pairs u and v whose Jacobian determinant is nowhere equal to zero; a degeneracy occurs if such a Jacobian determinant has a zero. An algebraic identity gives

$$\left[\frac{\partial(u,v)}{\partial(x,y)}\right]^2 = (u_x^2 + u_y^2)(n^2 + u_x^2 + u_y^2)$$

if u and v satisfy (1.1.3). We infer that the Jacobian of a solution pair u and v vanishes exactly at the critical points of u. The above mentioned assertion follows.

System (1.1.3) can be decoupled, as can be seen in the following. The former equation in (1.1.3) simply relates the *length* of the gradients involved—it reads

$$(\text{length of the gradient of } u)^2 + n^2 = (\text{length of the gradient of } v)^2.$$

The latter equation in (1.1.3) informs us that the gradients of u and v are *orthogonal*—in other words, the level curves of u are curves of steepest descent of v, and the curves of steepest descent of u are level curves of v. (By definition, the curves of steepest descent of v are the orbits of the differential equation $dx : v_x(x,y) = dy : v_y(x,y)$, i.e. the trajectories of ∇v.) The same equation amounts to saying that either

$$\begin{bmatrix} v_x \\ v_y \end{bmatrix} : \sqrt{v_x^2 + v_y^2} = \pm \begin{bmatrix} -u_y \\ u_x \end{bmatrix} : \sqrt{u_x^2 + u_y^2}\,,$$

or $u_x = u_y = 0$.

Therefore system (1.1.3) can be recast as follows

$$\begin{bmatrix} v_x \\ v_y \end{bmatrix} = \pm\sqrt{1+\frac{n^2}{u_x^2+u_y^2}} \quad \begin{bmatrix} -u_y \\ u_x \end{bmatrix}, \tag{1.1.4}$$

provided that

$$u_x^2+u_y^2>0 \tag{1.1.5}$$

—i.e. elliptic solutions are dealt with. System (1.1.4), which reads also this way

$$dv = \pm\sqrt{1+\frac{n^2}{u_x^2+u_y^2}}\ (-u_y\ dx+u_x\ dy)\,,$$

is exact—hence determines v up to an additive constant on any simply connected domain where u is determined—if and only if u obeys both inequality (1.1.5) and

$$\frac{\partial}{\partial x}\left\{\sqrt{1+\frac{n^2}{u_x^2+u_y^2}}\ u_x\right\}+\frac{\partial}{\partial y}\left\{\sqrt{1+\frac{n^2}{u_x^2+u_y^2}}\ u_y\right\}=0, \tag{1.1.6}$$

a *nonlinear second-order partial differential equation in divergence form.*

On the other hand, (1.1.3) is equivalent to the pair made up by the following system

$$\begin{bmatrix} u_x \\ u_y \end{bmatrix} = \mp\sqrt{1-\frac{n^2}{v_x^2+v_y^2}} \quad \begin{bmatrix} -v_y \\ v_x \end{bmatrix} \tag{1.1.7}$$

and either the equation $v_x^2+v_y^2=n^2$ or the following inequality

$$v_x^2+v_y^2>n^2. \tag{1.1.8}$$

System (1.1.7), which reads also this way

$$du = \mp\sqrt{1-\frac{n^2}{v_x^2+v_y^2}}\ (-v_y\ dx+v_x\ dy)\,,$$

is exact if and only if either $v_x^2+v_y^2=n^2$ or v obeys both inequality (1.1.8) and

$$\frac{\partial}{\partial x}\left\{\sqrt{1-\frac{n^2}{v_x^2+v_y^2}}\ v_x\right\}+\frac{\partial}{\partial y}\left\{\sqrt{1-\frac{n^2}{v_x^2+v_y^2}}\ v_y\right\}=0, \tag{1.1.9}$$

another *nonlinear second-order partial differential equation in divergence form.*

Thus, system (1.1.3) and inequality (1.1.5) imply equation (1.1.6); moreover (1.1.3) implies (1.1.8) and (1.1.9). System (1.1.3) holds if v is given by (1.1.4) and u satisfies both inequality (1.1.5) and equation (1.1.6); alternatively, (1.1.3) holds if u is given by (1.1.7) and v satisfies both (1.1.8) and (1.1.9).

The map $u\mapsto v$ defined by (1.1.4) and the map $v\mapsto u$ defined by (1.1.7)—which are inverse of one another—pair u and v much in the same way as system (1.1.3) does. In other words, they pair the real part and the imaginary part of solutions to equation (1.1.1). Specifically, they convert any solution to equation (1.1.6) satisfying (1.1.5) into a solution to (1.1.9) satisfying (1.1.8), and vice versa. These maps can be viewed as *Bäcklund transformations* associated with the equations and

systems in hand. (Information on Bäcklund transformations can be found in [**AI**] and [**RS**].)

In conclusion, equation (1.1.1) can be approached by representing its solutions as in (1.1.2) and working out system (1.1.3). System (1.1.3) may be approached in either of the following ways: (i) solve first equation (1.1.6) subject to inequality (1.1.5), then perform the Bäcklund transformation defined by formulas (1.1.4); (ii) solve first equation (1.1.9) subject to inequality (1.1.8), then perform the Bäcklund transformation defined by formulas (1.1.7). (The former approach, which points to elliptic solutions, is preferred in the present paper.)

Note that in the borderline case, where n vanishes identically, systems (1.1.4) and (1.1.5) parallel Cauchy-Riemann equations, and (1.1.6) and (1.1.9) coincide with Laplace equation. Note also that the real-valued solutions to

$$\left(\frac{\partial v}{\partial x}\right)^2+\left(\frac{\partial v}{\partial y}\right)^2=n^2(x,y), \tag{1.1.10}$$

the eikonal equation of classical geometrical optics, all satisfy (1.1.9).

If suitable conditions are met, (1.1.6) and (1.1.9) can be recast in the form of *semilinear* second-order partial differential equations with *polynomial nonlinearities.* Equation (1.1.6) takes the following form

$$\begin{aligned}\left[\left(u_x^2+u_y^2\right)^2+n^2u_y^2\right]u_{xx}-2n^2u_xu_yu_{xy}+\\ \left[\left(u_x^2+u_y^2\right)^2+n^2u_x^2\right]u_{yy}+n\left(u_x^2+u_y^2\right)\left(n_xu_x+n_yu_y\right)=0\end{aligned} \tag{1.1.11}$$

if sufficiently smooth solutions are dealt with that obey (1.1.5). Equation (1.1.9) takes the following form

$$\begin{aligned}\left[\left(v_x^2+v_y^2\right)^2-n^2v_y^2\right]v_{xx}+2n^2v_xv_yv_{xy}+\\ \left[\left(v_x^2+v_y^2\right)^2-n^2v_x^2\right]v_{yy}-n\left(v_x^2+v_y^2\right)\left(n_xv_x+n_yv_y\right)=0\end{aligned} \tag{1.1.12}$$

if sufficiently smooth solutions are dealt with that obey (1.1.8).

It should be stressed that the equations just displayed are *not* equivalent to the original ones. Smooth solutions to (1.1.11) exist whose critical points form a set of measure zero and that *do not* satisfy (1.1.6) in the sense of distributions—they make the left-hand side of (1.1.6) a well-defined distribution which is supported by the set of the critical points, but is not zero. See Proposition 2.2.1.

Let the coefficients of u_{xx},u_{xy},u_{yy} appearing on the left-hand side of (1.1.11) be denoted by a_{11}, $2a_{12}$ and a_{22}. Then

$$\begin{vmatrix}a_{11}&a_{12}\\a_{12}&a_{22}\end{vmatrix}=\left(u_x^2+u_y^2\right)^3\cdot\left(u_x^2+u_y^2+n^2\right),$$

hence equation (1.1.11) should be qualified *elliptic-parabolic* or *degenerate-elliptic.* A real-valued solution u to (1.1.11) is *elliptic* if the gradient of u is nowhere equal to zero; a *degeneracy* occurs if a *critical point* of u exists.

If a_{11}, $2a_{12}$ and a_{22} denote the coefficients of v_{xx},v_{xy},v_{yy} appearing on the left-hand side of (1.1.12), then

$$\begin{vmatrix}a_{11}&a_{12}\\a_{12}&a_{22}\end{vmatrix}=\left(v_x^2+v_y^2\right)^3\cdot\left(v_x^2+v_y^2-n^2\right).$$

Hence the real-valued solutions v to (1.1.12) such that $v_x^2+v_y^2<n^2$ are *hyperbolic.* The real-valued twice-differentiable solutions to (1.1.10) are *parabolic* solutions to

(1.1.12). Solutions v to (1.1.12) such that $v_x^2 + v_y^2 > n^2$ are *elliptic.* If u and v satisfy system (1.1.3), then v is an *elliptic-parabolic* solution to (1.1.12).

Some sets of terms, appearing in equations (1.1.11) and (1.1.12), have a special geometric meaning. These equations read respectively

$$|\nabla u|\ \Delta u - n^2 \left[h - \nabla \log n \cdot \frac{\nabla u}{|\nabla u|} \right] = 0 \tag{1.1.13}$$

and

$$|\nabla v|\ \Delta v + n^2 \left[k - \nabla \log n \cdot \frac{\nabla v}{|\nabla v|} \right] = 0, \tag{1.1.14}$$

provided critical points are set aside. Here $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2$, the Laplace operator, and h and k are given either by

$$h = -\left(u_x^2 + u_y^2\right)^{-3/2} \left(u_y^2 u_{xx} - 2u_y u_x u_{xy} + u_x^2 u_{yy}\right) \tag{1.1.15}$$

and

$$k = -\left(v_x^2 + v_y^2\right)^{-3/2} \left(v_y^2 v_{xx} - 2v_y v_x v_{xy} + v_x^2 v_{yy}\right), \tag{1.1.16}$$

or more concisely by

$$h = -\operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) \qquad \text{and} \qquad k = -\operatorname{div}\left(\frac{\nabla v}{|\nabla v|}\right). \tag{1.1.17}$$

Frenet's formulas tell us that the principal normal to the orbits of the differential equation

$$u_x(x, y)\ dx + u_y(x, y)\ dy = 0$$

is precisely

$$(1/h)\frac{\nabla u}{|\nabla u|}.$$

(Recall that the principal normal points towards the center of curvature, and its length is the radius of curvature.) Thus the value of h at a point (x, y), where the gradient of u does not vanish, is a *signed curvature* at (x, y) of the *level curve* of u crossing (x, y). Similarly, the value of k at a point (x, y) is a *signed curvature* at (x, y) of the *level curve* of v crossing (x, y). (If u and v satisfy system (1.1.3), h is also a curvature of the curves of steepest descent of v and k is also a curvature of the curves of steepest descent of u.)

The last term on the left-hand side of (1.1.13) is related to the following Riemannian metric

$$n(x, y)\ \sqrt{(dx)^2 + (dy)^2}. \tag{1.1.18}$$

In the terminology of classical geometrical optics, the geodesics belonging to the metric (1.1.18) are nicknamed *rays.* The rays, which will play a central role in the sequel, are characterized by the following differential equation

$$n \begin{vmatrix} dx/ds & dy/ds \\ d^2x/ds^2 & d^2y/ds^2 \end{vmatrix} = \left[(dx/ds)^2 + (dy/ds)^2 \right] \begin{vmatrix} dx/ds & dy/ds \\ \partial n/\partial x & \partial n/\partial y \end{vmatrix} \tag{1.1.19}$$

—in other words, the principal normal to a ray obeys

$$\nabla \log n \cdot (\text{principal normal}) = 1. \tag{1.1.20}$$

Thus the value of

$$\nabla \log n \cdot \frac{\nabla u}{|\nabla u|}$$

at a point (x, y), where ∇u is different from 0, equals a *curvature* at (x, y) of the *ray* which is tangent at (x, y) to a level curve of u.

The last term appearing on the left-hand side of (1.1.14) can be treated in similar fashion.

1.2. Motivations. The electromagnetic field is governed by the following *Maxwell's equations*

$$\begin{aligned} &\frac{\partial}{\partial t}(\varepsilon \mathfrak{E}) = \operatorname{curl} \mathfrak{H}, \qquad \frac{\partial}{\partial t}(\mu \mathfrak{H}) = -\operatorname{curl} \mathfrak{E}, \\ &\operatorname{div}(\varepsilon \mathfrak{E}) = 0, \qquad\qquad \operatorname{div}(\mu \mathfrak{H}) = 0, \end{aligned} \tag{1.2.1}$$

in the case where standard physical conditions are met, the medium is isotropic and non-conducting, and no electric charges concur. (See [**Jo**], for instance.) Here $\mathfrak{E}, \mathfrak{H}, \varepsilon$, and μ denote the electric field, the magnetic field, the dielectric constant and the permeability, respectively; t stands for time, whereas the underlying space coordinates will be denoted by x, y, and z. Assume ε and μ are *constant in time* and *positive*, i.e. the medium is non-dissipative. Moreover, assume the electromagnetic field is *monochromatic*, i.e. both $\mathfrak{E}$ and $\mathfrak{H}$ depend upon an extra parameter ν —the *wave number*— in such a way that $\mathfrak{E} \cdot \exp(i\nu t)$ and $\mathfrak{H} \cdot \exp(i\nu t)$ do not depend on t. Then several theories apply that offer asymptotic expansions of the electromagnetic field as the wave number is large. (A relevant overview can be found in [**BM**]).

One of these theories is the classical geometrical optics, of course. Another, sometimes called *EWT (Evanescent Wave Tracking),* has been developed by L. Felsen and coworkers recently. Both geometrical optics and EWT result from an archetypal application of the WKBJ method (which provides asymptotic expansions of solutions to partial differential equations depending upon a large parameter, and is so called after Wentzel, Kramers, Brillouin, and Jeffreys), and rest upon the following Ansatz: a scalar field φ and a vector field $\mathbf{A}$, both independent of time t and wave number ν, obey

$$\mathfrak{E} = \exp\left[-i\nu t + i\nu\varphi(x, y, z)\right] \cdot \left[\mathbf{A}(x, y, z) + (\text{a remainder})\right], \tag{1.2.2}$$

where the remainder$= O(1/\nu)$ in some topology as $\nu \longrightarrow \infty$. It is usual to call φ the *eikonal.* The distinctive feature of EWT which makes it an extension of geometrical optics consists in allowing *the eikonal to take complex values.*

EWT points to detecting those properties of the elecromagnetic field that the leading term on the right-hand side of (1.2.2) potentially encodes under the hypotheses specified above. Though open to criticism, de facto EWT proves apt to account for phenomena of physical optics that are excluded from geometrical optics. For instance, EWT actually models *diffracted evanescent wave*— the fast decaying waves that appear beyond a caustic, into the region not reached by geometric optical rays. The imaginary part of the eikonal vanishes on the side of the caustic where geometrical optics prevails, and describes attenuation on the side where geometrical optics breaks down. Information on EWT can be found in [**CF1**], [**CF2**], [**EF**], [**ER**], [**Fe1**], [**HF**].

Partial differential equations governing φ and $\mathbf{A}$ are derived as follows. Define the *refractive index*, n, by

$$n = \sqrt{\varepsilon\mu}.$$

Eliminating $\mathfrak{H}$ from (1.2.1) gives the following equation

$$n^2\frac{\partial^2}{\partial t^2}\mathfrak{E} = \Delta\mathfrak{E} + 2\nabla\left(\frac{\nabla n}{n}\cdot\mathfrak{E}\right) - \left(\frac{\nabla\mu}{\mu}\cdot\nabla\right)\mathfrak{E} - (\mathfrak{E}\cdot\nabla)\left(\frac{\nabla\mu}{\mu}\right), \tag{1.2.3}$$

which governs the electric field. (Here Δ and ∇ denote the three-dimensional Laplace and gradient operators, respectively. A dot denotes either the product of numbers or the inner product of three-dimensional vectors;

$$\nabla\mu\cdot\nabla = \mu_x\frac{\partial}{\partial x} + \mu_y\frac{\partial}{\partial y} + \mu_z\frac{\partial}{\partial z},$$

the derivative along the curves of steepest descent of μ; $\mathfrak{E}\cdot\nabla$ stands for the derivative along the trajectories of $\mathfrak{E}$.) Plugging the right-hand side of (1.2.2) into (1.2.3) results into

$$\left(\frac{\partial\varphi}{\partial x}\right)^2 + \left(\frac{\partial\varphi}{\partial y}\right)^2 + \left(\frac{\partial\varphi}{\partial z}\right)^2 = n^2, \tag{1.2.4}$$

and

$$\left[2(\nabla\varphi\cdot\nabla) + \mu\ \operatorname{div}\left(\frac{\nabla\varphi}{\mu}\right)\right]\mathbf{A} + 2\left(\frac{\nabla n}{n}\cdot\mathbf{A}\right)\nabla\varphi = 0. \tag{1.2.5}$$

Equations (1.2.4) and (1.2.5) are called the *eikonal equation* and the *transport equation,* respectively. These equations are the proper key to EWT, provided *complex-valued solutions* are involved.

If a two-dimensional configuration is considered, the eikonal equation becomes

$$\left(\frac{\partial\varphi}{\partial x}\right)^2 + \left(\frac{\partial\varphi}{\partial y}\right)^2 = n^2(x,y). \tag{1.2.6}$$

Investigations on complex-valued solution to equation (1.2.6) can be found in [**ER**]. Related remarks appear in [**Kha**] and [**Hem**]. The present paper, where φ is replaced by $\pm i\varphi$ and (1.2.6) is recast in the form

$$\left(\frac{\partial\varphi}{\partial x}\right)^2 + \left(\frac{\partial\varphi}{\partial y}\right)^2 + n^2(x,y) = 0, \tag{1.2.7}$$

is an attempt to go further in the same direction.

The authors are indebted to Professor G.A.Viano from the University of Genova, and Professors I.Montrosset and R.Zich from the Technical University of Torino for helpful conversations about the subject of this subsection.

1.3. Summary of Results. The present paper is devoted to examples and qualitative properties of solutions. Existence theorems will appear in a subsequent article.

Solutions to equation (1.1.1), to either equations (1.1.6) or (1.1.11), and to either equations (1.1.9) or (1.1.12) can be displayed in closed form in the case where n is constant. An ad hoc tool was pointed out by L. Felsen and coworkers, another is the classical Legendre transformation. Section 2 discusses this. Solutions having special traits are collected there—e.g. solutions to (1.1.11) whose critical points form a *continuum.*

Theorem 3.1.1 claims that if n is strictly positive, w is a smooth solution to equation (1.1.1) and u is the *real part* of w then *the critical points of u are not isolated. They spread precisely along the rays.*

Theorem 3.1.1 expresses a distinctive property of equation (1.1.1). In the language of EWT, it informs us that any *complex ray* passing through a point, where the first-order derivatives of the eikonal take real values, necessarily coincides with a *geometric optical ray*—provided a two-dimensional configuration is in hand and singularities are shaken off. To put it roughly, complex rays develop precisely where geometrical optics breaks down. A closely related statement appears in [**ER**, §3.2] without proof.

Theorem 3.1.2 belongs to the same vein. It basically shows that equation (1.1.11), unlike more conventional second-order partial differential equations, prevents its solutions from having *isolated critical points.* Suppose u is smooth and real-valued, and satisfies either equation (1.1.6) or equation (1.1.11) where no critical point occurs. If ∇u vanishes at some point and the Hessian matrix of u is different from zero there, then ∇u *vanishes everywhere on a ray* passing through that point.

2. Sample Solutions

2.1. Background. Throughout this section we assume

$$n(x,y) \equiv 1, \tag{2.1.1}$$

and display model solutions to the equations in hand. Under assumption (2.1.1) these equations read as follows:

$$w_x^2 + w_y^2 + 1 = 0, \tag{2.1.2}$$

$$\frac{\partial}{\partial x}\left\{\sqrt{1+\frac{1}{u_x^2+u_y^2}}\;u_x\right\} + \frac{\partial}{\partial y}\left\{\sqrt{1+\frac{1}{u_x^2+u_y^2}}\;u_y\right\} = 0, \tag{2.1.3}$$

$$\left[\left(u_x^2+u_y^2\right)^2+u_y^2\right]u_{xx} - 2u_xu_yu_{xy} + \left[\left(u_x^2+u_y^2\right)^2+u_x^2\right]u_{yy} = 0, \tag{2.1.4}$$

$$\frac{\partial}{\partial x}\left\{\sqrt{1-\frac{1}{v_x^2+v_y^2}}\;v_x\right\} + \frac{\partial}{\partial y}\left\{\sqrt{1-\frac{1}{v_x^2+v_y^2}}\;v_y\right\} = 0, \tag{2.1.5}$$

$$\left[\left(v_x^2+v_y^2\right)^2-v_y^2\right]v_{xx} + 2v_xv_yv_{xy} + \left[\left(v_x^2+v_y^2\right)^2-v_x^2\right]v_{yy} = 0, \tag{2.1.6}$$

$$\begin{bmatrix} v_x \\ v_y \end{bmatrix} = \pm\sqrt{1+\frac{1}{u_x^2+u_y^2}}\;\begin{bmatrix} -u_y \\ u_x \end{bmatrix}. \tag{2.1.7}$$

2.2. Complex Source-Point Solutions. A recipe for constructing special solutions was exploited by L. Felsen and coworkers in analysis of electromagnetic waves and reviewed in [**Fe**]. It rests upon the observation that if a partial differential equation has a favorable structure and appropriate solutions depend analytically upon extra parameters then analytic continuation with respect to such parameters leads to new solutions to the same equation.

The Euclidean distance from a given point depends analytically on the coordinates of such a point and obeys a partial differential equation which is germane

to the present discussion. Thus the recipe quoted above suggests that if a and b are constant the function defined by

$$w(x,y) = i \cdot \sqrt{(x-a)^2 + (y-b)^2} \tag{2.2.1}$$

satisfies equation (2.1.2) irrespective of whether a and b are real or complex. In the terminology of [**Fe**] a *complex source-point solution* to (2.1.2) is at hand.

Denote the real and the imaginary part of w by u and v, respectively. As seen in Subsection 1.1.1, u satisfies both equations (2.1.3) and (2.1.4) in the region where no critical point occurs; v obeys $v_x^2 + v_y^2 \geq 1$ and equation (2.1.5), and satisfies equation (2.1.6) in the region where $v_x^2 + v_y^2 > 1$; u and v are related by Bäcklund transformation (2.1.7) in the region where u has no critical point.

Let $a = 0$ and $b = i$, for instance. Then (2.2.1) reads

$$w(x,y) = \sqrt{1 - x^2 - y^2 + 2iy}. \tag{2.2.2}$$

The following equations and the following properties ensue.

$$\sqrt{2} \cdot u(x,y) = \sqrt{\sqrt{(1-x^2-y^2)^2 + 4y^2} + 1 - x^2 - y^2}\ , \tag{2.2.3}$$

$$\sqrt{2} \cdot v(x,y) = \ \mathrm{sgn}(y) \cdot \sqrt{\sqrt{(x^2+y^2-1)^2 + 4y^2} + x^2 + y^2 - 1}\ . \tag{2.2.4}$$

(i) $0 \leq u \leq 1$. The level sets of u can be described as follows. The set where $u = 0$ is

$$\{(x,y) : |x| \geq 1, y = 0\}.$$

The set where u equals a constant C, $0 < C < 1$, is the *hyperbola* defined by

$$\frac{x^2}{1-C^2} - \frac{y^2}{C^2} = 1.$$

The set where $u = 1$ is the y–axis.

(ii) The level sets of v can be described as follows. The set where $v = 0$ is

$$\{(x,y) : |x| \leq 1, y = 0\}.$$

The set where v equals a constant C, $C \neq 0$, is the *arc of ellipse* defined by

$$\frac{x^2}{1+C^2} + \frac{y^2}{C^2} = 1 \quad \text{and} \quad \mathrm{sgn}(y) = \mathrm{sgn}(C).$$

(iii) u is smooth in

$$\{(x,y) : \text{either } \ |x| < 1 \text{ or } y \neq 0\},$$

the region where $u > 0$. u is everywhere continuous, but fails to be differentiable at points where $|x| \geq 1$ and $y = 0$. The singularities of ∇u are detailed by

$$\sqrt{x^2 - 1} \cdot \mathrm{sgn}(y) \cdot \nabla u(x,y) \longrightarrow \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{2.2.5}$$

as $|x| > 1$ and $y \to 0$, and

$$\sqrt{(|x|-1)^2 + y^2} \cdot |\nabla u(x,y)| \longrightarrow 1/\sqrt{2} \tag{2.2.6}$$

as (x,y) approaches either $(1,0)$ or $(-1,0)$.

(iv) v is smooth in $\{(x,y) : \text{either } |x|<1 \text{ or } y\neq 0\}$, and is discontinuous elsewhere.

(v) Let C be any strictly positive constant. The set where

$$|\nabla u(x,y)| = C$$

is made up by two circles with two points removed—the circles are defined by

$$x^2+y^2+\left(\sqrt{1+C^{-2}}+1/\sqrt{1+C^{-2}}\right)\cdot|x|+1=0,$$

and the points are $(\pm\sqrt{1+C^{-2}},0)$.

(vi) The set of the *critical points* of u is the y–axis. Moreover

$$\frac{\nabla u(x,y)}{|\nabla u(x,y)|} = \operatorname{sgn}(x)\cdot\left\{-\begin{bmatrix}1\\0\end{bmatrix}+O(x)\right\} \tag{2.2.7}$$

as x approaches 0 and y is constant.

Note that *u has no isolated critical point*, and *the set of the critical points of u is a perfect straight line*—not an accident, as theorems from Section 3 below will clarify. The same straight line is *the locus of points where u attains its greatest value* — the customary strong maximum principle, as stated in [**CH**, Chapter 4, Section 6] for instance, does not apply here.

PROPOSITION 2.2.1. *Function u given by (2.2.3) is a smooth solution to (2.1.4) in*

$$\{(x,y) : \textit{either } |x|<1 \textit{ or } y\neq 0\};$$

u fails to obey equation (2.1.3) in the sense of distributions in any open set $\mathcal{O}$ such that

$$\mathcal{O}\cap\{(x,y) : \textit{either } x=0 \textit{ or } |x|\geq 1 \textit{ and } y=0\}\neq\emptyset.$$

PROOF. Observe that u satisfies both (2.1.3) and (2.1.4) precisely in

$$\{(x,y) : x\neq 0 \text{ and either } |x|<1 \text{ or } y\neq 0\}.$$

Since u is smooth across the axis where $x=0$, the former conclusion follows.

Relevant behavior tracts of $|\nabla u|$ and $\nabla u\cdot|\nabla u|^{-1}$ at points where either $x=0$ or $|x|\geq 1$ and $y=0$ are fixed by (2.2.5), (2.2.6) and (2.2.7). Thus an integration by parts yields

$$\int\sqrt{1+|\nabla u|^2}\,\frac{\nabla u}{|\nabla u|}\cdot\nabla\varphi dxdy = \tag{2.2.8}$$
$$2\int_{-\infty}^{\infty}\varphi(0,y)dy-2\left(\int_{-\infty}^{-1}+\int_{1}^{\infty}\right)\varphi(x,0)\frac{|x|}{\sqrt{x^2-1}}dx$$

if φ is any infinitely differentiable compactly supported real-valued function. In other words, in the sense of distributions

$$\operatorname{div}\left(\sqrt{1+|\nabla u|^2}\,\frac{\nabla u}{|\nabla u|}\right)$$

equals a *non-zero measure* supported by

$$\{(x,y) : \text{ either } x=0 \text{ or } |x|\geq 1 \text{ and } y\neq 0\}.$$

The second conclusion follows. □

2.3. Radial Solutions. If R is any real-valued solution to the following ordinary differential equation

$$r \cdot \frac{dR}{dr} \cdot \frac{d^2R}{dr^2} + \left(\frac{dR}{dr}\right)^2 + 1 = 0, \tag{2.3.1}$$

the function defined by

$$u(x, y) = R\Big(\sqrt{x^2 + y^2}\Big) \tag{2.3.2}$$

—invariant under the group of rotations about the origin—satisfies equation (2.1.3) in the region where $(dR/dr)\Big(\sqrt{x^2 + y^2}\Big) \neq 0$, and satisfies equation (2.1.4) everywhere. Equation (2.3.1), which reads also as follows

$$\frac{d}{dr}\left\{r\sqrt{1 + \left(\frac{dR}{dr}\right)^2}\right\} = 0,$$

tells us that any tangent straight line to the graph of R meets the R-axis at a point whose distance from the point of contact is constant. Hence any orbit of (2.3.1) is a *tractrix* asymptotic to the R-axis. The graph given by (2.3.2) is a surface generated by revolving a tractrix about its asymptote—a dilated and translated *pseudosphere.*

Recall that the pseudosphere is the simplest among the surfaces of revolution in Euclidean 3-dimensional space whose Gauss curvature is -1. Its Riemannian structure is identical with that of the Poicaré half-plane. See [**Lau**, Chapter 2, Section 6.4] and [**DoC**, Section 5-10].

An integration gives

$$u(x, y) = A\left\{\sqrt{1 - \frac{r^2}{A^2}} + \log \frac{r}{|A| + \sqrt{A^2 - r^2}}\right\} + B, \tag{2.3.3}$$

where

$$r = \sqrt{x^2 + y^2},$$

A is a constant different from 0—either the radius of the disk where u is defined, or its negative—and B is any constant.

Equation (2.3.3) implies

$$|\nabla u(x, y)| = \sqrt{A^2/r^2 - 1}.$$

Thus *the set of the critical points of u is the boundary circle where $x^2 + y^2 = A^2$ —a continuum.*

If R obeys the following ordinary differential equation

$$r \cdot \frac{dR}{dr} \cdot \frac{d^2R}{dr^2} + \left(\frac{dR}{dr}\right)^2 - 1 = 0, \tag{2.3.4}$$

slightly different from (2.3.1), the function defined by

$$v(x, y) = R\Big(\sqrt{x^2 + y^2}\Big) \tag{2.3.5}$$

satisfies equation (2.1.6). Equation (2.3.4) reads

$$\frac{d}{dr}\left\{r^2\left[\left(\frac{dR}{dr}\right)^2 - 1\right]\right\} = 0,$$

hence an integration informs us that either

$$v(x,y) = A\left\{\sqrt{1+\frac{r^2}{A^2}} + \log\frac{r}{|A|+\sqrt{A^2+r^2}}\right\} + B \tag{2.3.6}$$

or

$$v(x,y) = A\left\{\sqrt{\frac{r^2}{A^2}-1} - \arctan\sqrt{\frac{r^2}{A^2}-1}\right\} + B, \tag{2.3.7}$$

where

$$r = \sqrt{x^2+y^2}$$

and A and B are constant.

Equations (2.3.6) and (2.3.7) give

$$|\nabla v(x,y)| = \sqrt{1+A^2/r^2}$$

and

$$|\nabla v(x,y)| = \sqrt{1-A^2/r^2},$$

respectively. Thus equations (2.3.6) and (2.3.7) define *elliptic and hyperbolic radial solutions* to equation (2.1.6), respectively.

Observe in passing that equation (2.3.6) and

$$\begin{bmatrix} u_x \\ u_y \end{bmatrix} = \mp\sqrt{1-\frac{1}{v_x^2+v_y^2}} \quad \begin{bmatrix} -v_y \\ v_x \end{bmatrix},$$

the inverse of Bäcklund transformation (2.1.7), give

$$u(x,y) = C\arctan(y/x) + D; \tag{2.3.8}$$

Bäcklund transformation (2.1.7) and equation (2.3.3) give

$$v(x,y) = C\arctan(y/x) + D \tag{2.3.9}$$

—here C and D are constants. Formulas (2.3.8) and (2.3.9) respectively define the solutions to equations (2.1.4) and (2.1.6) that are invariant under the group of homothetic transformations. The graphs of these solutions are *right helicoids*—surfaces that are also harmonic, ruled and minimal.

2.4. Legendre Transformation. The Legendre transformation and its connection with partial differential equations are presented in detail in [**CH**, Chapter 1, Section 6]. Wide applications to fluid dynamics, calculus of variations and convex analysis are known—see [**Be**], [**KS**] and [**Ro**], for instance. Succinct directions appear in the next paragraphs.

Let u be a twice continuously differentiable real-valued function defined in some open subset of Euclidean plane $\mathbb{R}^2$. Suppose ∇u is a bijection, and that

$$u_{xx}u_{yy} - u_{xy}^2 \neq 0$$

everywhere—in a word, let ∇u be a diffeomorphism.

The *hodograph* of u is the range of ∇u. U is the *Legendre transform* of u if:

(i) the domain of U is the hodograph of u;

(ii) for any (p, q) from the hodograph of u, the negative of $U(p, q)$ is the height above the origin of the tangent plane to the graph of u whose normal parallels $(p, q, -1)$. In formulas:

$$p = u_x(x, y), \quad q = u_y(x, y), \quad U(p, q) = xp + yq - u(x, y). \tag{2.4.1}$$

If u is defined in the whole of $\mathbb{R}^2$ and grows fast enough at infinity, then

$$U(p, q) = \max\left\{xp + yq - u(x, y) : (x, y) \in \mathbb{R}^2\right\}$$

—in the terminology of convex analysis, U coincides with the *Fenchel conjugate* of u.

The differentials of U, p and q are related by $dU = xdp + ydq$, as one may immediately infer from (2.4.1). In other words, (2.4.1) implies

$$x = U_p\ (p, q), \quad y = U_q\ (p, q).$$

The following set

$$\begin{aligned} &p = u_x(x, y), \quad q = u_y(x, y), \\ &x = U_p\ (p, q), \quad y = U_q\ (p, q), \\ &u(x, y) + U(p, q) = xp + yq, \end{aligned} \tag{2.4.2}$$

results. An alternative arrangement of (2.4.2), which will be useful later, reads

$$\begin{bmatrix} p & x \\ q & y \\ U(p,q) & u(x,y) \end{bmatrix} = \tag{2.4.3}$$

$$\left\{ \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ x & y & p & q \end{bmatrix} \begin{bmatrix} \partial/\partial x & 0 \\ \partial/\partial y & 0 \\ 0 & \partial/\partial p \\ 0 & \partial/\partial q \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} \right\} \begin{bmatrix} u(x,y) & 0 \\ 0 & U(p,q) \end{bmatrix}.$$

Equations (2.4.2) provide a *parametrization* of the graph of u—the apposite parameters are p and q, which coincide with the first-order partial derivatives of u. Note the following: a unit normal vector field, relevant to this parametrization, has these components

$$p \cdot (p^2 + q^2 + 1)^{-1/2}, \quad q \cdot (p^2 + q^2 + 1)^{-1/2}, \quad -(p^2 + q^2 + 1)^{-1/2};$$

the first and the second fundamental forms are given by

$$I = \begin{bmatrix} U_{pp} & U_{pq} \\ U_{pq} & U_{qq} \end{bmatrix} \begin{bmatrix} 1 + p^2 & pq \\ pq & 1 + q^2 \end{bmatrix} \begin{bmatrix} U_{pp} & U_{pq} \\ U_{pq} & U_{qq} \end{bmatrix},$$

$$II = (p^2 + q^2 + 1)^{-1/2} \begin{bmatrix} U_{pp} & U_{pq} \\ U_{pq} & U_{qq} \end{bmatrix};$$

the mean and the Gauss curvature obey

$$\text{mean curvature} = \frac{(1 + p^2)U_{pp} + 2pqU_{pq} + (1 + q^2)U_{qq}}{(p^2 + q^2 + 1)^{3/2}\left(U_{pp}U_{qq} - U_{pq}^2\right)},$$

$$(\text{Gauss curvature})^{-1} = (p^2 + q^2 + 1)^2\left(U_{pp}U_{qq} - U_{pq}^2\right).$$

Equations (2.4.2) show that the Legendre transform of U equals u—the Legendre transformation is *involutory.* A bijection from $\mathbb{R}^2$ into itself is a gradient if and only if the inverse mapping is a gradient: equations (2.4.2) inform us

that the gradient of u and the gradient of U are inverse mappings of one another. The following equation

$$\begin{bmatrix} u_{xx}(x,y) & u_{xy}(x,y) \\ u_{xy}(x,y) & u_{yy}(x,y) \end{bmatrix} \begin{bmatrix} U_{pp}(p,q) & U_{pq}(p,q) \\ U_{pq}(p,q) & U_{qq}(p,q) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{2.4.4}$$

results, provided x, y, p and q are related by

$$\begin{bmatrix} p \\ q \end{bmatrix} = \nabla u(x,y) \quad \text{or} \quad \begin{bmatrix} x \\ y \end{bmatrix} = \nabla U(p,q).$$

Equation (2.4.4) amounts to the following set

$$\begin{aligned} &\big[u_{xx}(x,y)u_{yy}(x,y) - u_{xy}^2(x,y)\big] \cdot \big[U_{pp}(p,q)U_{qq}(p,q) - U_{pq}^2(p,q)\big] = 1, \\ &u_{xx}(x,y) : U_{qq}(p,q) = -u_{xy}(x,y) : U_{pq}(p,q) = u_{yy}(x,y) : U_{pp}(p,q), \end{aligned} \tag{2.4.5}$$

which provides expressions of the second-order derivatives of u in terms of the second-order derivatives of U, and viceversa.

PROPOSITION 2.4.1. *Let u and U be a pair of Legendre transforms. Then u satisfies equation (2.1.4) if and only if U satisfies*

$$\big[\big(p^2+q^2\big)^2 + p^2\big]U_{pp} + 2pqU_{pq} + \big[\big(p^2+q^2\big)^2 + q^2\big]U_{qq} = 0. \tag{2.4.6}$$

PROOF. Combine (2.1.4) and formulas (2.4.1) to (2.4.5). □

Observe that the change of variables specified by

$$0 < \lambda < \infty, \quad 0 \le \mu < 2\pi, \quad p + iq = \sinh\lambda \cdot e^{i\mu}, \tag{2.4.7}$$

converts (2.4.6) into

$$\left(\frac{\partial^2}{\partial\lambda^2} + \frac{\partial^2}{\partial\mu^2}\right) U\big(\sinh\lambda \cdot e^{i\mu}\big) = 0, \tag{2.4.8}$$

Laplace's equation. (For notational convenience, the point whose rectangular coordinates are p and q is treated here as identical to the complex number $p + iq$.) Observe also that (2.4.3) gives

$$\begin{bmatrix} \sinh\ \lambda\cdot\cos\ \mu & x \\ \sinh\ \lambda\cdot\sin\ \mu & y \\ U\big(\sinh\ \lambda\cdot e^{i\mu}\big) & u(x,y) \end{bmatrix} = \tag{2.4.9}$$

$$\left\{ \begin{bmatrix} 1 & 0 & \frac{\cos\mu}{\cosh\lambda} & -\frac{\sin\mu}{\sinh\lambda} \\ 0 & 1 & \frac{\sin\mu}{\cosh\lambda} & \frac{\cos\mu}{\sinh\lambda} \\ x & y & \tanh\lambda & 0 \end{bmatrix} \begin{bmatrix} \partial/\partial x & 0 \\ \partial/\partial y & 0 \\ 0 & \partial/\partial\lambda \\ 0 & \partial/\partial\mu \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} \right\} \times$$

$$\begin{bmatrix} u(x,y) & 0 \\ 0 & U\big(\sinh\ \lambda\cdot e^{i\mu}\big) \end{bmatrix}.$$

Proposition 2.4.1 together with equation (2.4.8) and (2.4.9) enable one to exhibit particular solutions to equation (2.1.4) in a closed form.

EXAMPLE 1. The function defined by $0 < \lambda < \infty,\ 0 \le \mu < 2\pi$ and

$$4\,U\big(\sinh\lambda\cdot e^{i\mu}\big) = \sinh(2\lambda)\cdot\sin(2\mu)$$

satisfies (2.4.8). U obeys the following equation too

$$U(p,q) = pq\cdot\sqrt{1+1/(p^2+q^2)}$$

and equation (2.4.6). Moreover,

$$9/4 \ge -(\text{Hessian determinant of } U) \ge (p^2+q^2)/(1+p^2+q^2).$$

The Legendre transform of this U is a solution u to equation (2.1.4), whose properties are itemized below. The first property is supplied by formula (2.4.9), the others follow successively.

(i) Function u and its first-order derivatives are represented by

$$x = \frac{\sin\mu}{2\cosh\lambda}\cdot[\cosh(2\lambda)-\cos(2\mu)],\ y = \frac{\cos\mu}{2\cosh\lambda}\cdot[\cosh(2\lambda)+\cos(2\mu)],$$
$$u(x,y) = \frac{1}{4}\cdot\tanh\lambda\cdot[\cosh(2\lambda)-1]\cdot\sin(2\mu), \tag{2.4.10}$$
$$u_x(x,y) = \sinh\lambda\cdot\cos\mu,\quad u_y(x,y) = \sinh\lambda\cdot\sin\mu.$$

(ii) The pair made up of the first and the second equation in (2.4.10) defines a smooth diffeomeorphism. The relevant domain is specified by

$$0 < \lambda < \infty,\quad 0 \le \mu < 2\pi;$$

the relevant range is the exterior of the standard *astroid* that is represented by

$$(x^2+y^2-1)^3 + 27x^2y^2 = 0.$$

(iii) u is smooth in the exterior of the astroid. The gradient of u is continuous up to the astroid, but the second-order derivatives of u blow up there.

(iv) Let C be any constant. The set where μ equals C, that is where

$$u_x(x,y) : \cos C = u_y(x,y) : \sin C,$$

is a part of the *hyperbola* where

$$\sin(2C)\cdot(x^2+y^2) - 2xy = \frac{1}{2}\sin(2C)\cdot[1+\cos(4C)]$$

and whose envelope coincides with the above astroid—the part in question consists of a subarc leaving the astroid non-tangentially.

(v) Let C be any strictly positive constant. The set where λ equals C, that is where

$$|\nabla u(x,y)| = \sinh C,$$

is a *hypotrochoid*—the hypotrochoid is the roulette of a point that is attached to a circle rolling inside another circle, see [**Law**, Section 6.2]. Here

radius of the fixed circle $= (4\cosh C - 1/\cosh C)/3$,
radius of the rolling circle=(radius of the fixed circle)/4
distance of the tracing point from the center of the rolling circle
$= 1/(4\cosh C)$.

The hypotrochoid in hand has neither self-intersections nor cusps, is close to the above mentioned astroid if C is small, looks alike a large circle if C is large, is contained in the annulus where

$$\cosh C - 1/(2\cosh C) \le \sqrt{x^2+y^2} \le \cosh C,$$

and has the following algebraic equation

$$\big[1+\cosh(2C)-2(x^2+y^2)\big]\cdot\big[1+\cosh(4C)-2(x^2+y^2)\big]^2 = 16\,\frac{[1+2\cosh(2C)]^3}{1+\cosh(2C)}x^2y^2.$$

(vi) The critical points of u form the boundary astroid—thus *the set of the critical points of u is a continuum.*

PROPOSITION 2.4.2. *Let v and V be a pair of Legendre transforms. Then v satisfies equation (2.1.6) if and only if V satisfies*

$$\big[(p^2+q^2)^2-p^2\big]V_{pp} - 2pqV_{pq} + \big[(p^2+q^2)^2-q^2\big]V_{qq} = 0. \tag{2.4.11}$$

PROOF. Combine (2.1.6) and the analogs of formulas (2.4.1) to (2.4.5). □

Equation (2.4.11) is *elliptic* in the region where $p^2+q^2>1$, and is *hyperbolic* in the disk where $p^2+q^2<1$. Its *characteristic curves* are the circles specified by

$$p^2+q^2-p\cdot\cos C - q\cdot\sin C = 0$$

and $C=$ constant, and the circle specified by

$$p^2+q^2=1.$$

Incidentally, the change of variables defined by

$$-\infty<\lambda<\infty,\quad 0\le\mu<2\pi,\quad p+iq=\cosh\Big(2\sqrt{\lambda}\Big)\cdot e^{i\mu}$$

converts equation (2.4.11) into

$$\left(\lambda\frac{\partial^2}{\partial\lambda^2} + \frac{1}{2}\frac{\partial}{\partial\lambda} + \frac{\partial^2}{\partial\mu^2}\right)V\Big(\cosh\Big(2\sqrt{\lambda}\Big)\cdot e^{i\mu}\Big) = 0, \tag{2.4.12}$$

a Tricomi-type equation.

PROPOSITION 2.4.3. *Suppose u and U are a pair of Legendre transforms, and that u is free from critical points. Let v and V be another pair of Legendre transforms. Then u and v obey Bäcklund transformation (2.1.7) if and only if U and V obey the following system*

$$\begin{bmatrix} -\dfrac{\partial}{\partial\mu} \\ \dfrac{\partial}{\partial\lambda} \end{bmatrix} U\big(\sinh\lambda\cdot e^{i\mu}\ \big) \pm \begin{bmatrix} \dfrac{\partial}{\partial\lambda} \\ \dfrac{\partial}{\partial\mu} \end{bmatrix} V\Big(\cosh\lambda\cdot e^{i(\mu\pm\pi/2)}\Big) = 0 \tag{2.4.13}$$

for every λ and μ such that $\sinh\lambda\cdot e^{i\mu}$ belongs to the hodograph of u—in other words,

$$(\lambda,\mu)\mapsto U\big(\sinh\lambda\cdot e^{i\mu}\ \big)$$

and

$$(\lambda,\mu)\mapsto \pm V\Big(\cosh\lambda\cdot e^{i(\mu\pm\pi/2)}\Big)$$

are conjugate harmonic.

PROOF. The Bäcklund transformation involved informs us that

$$\nabla v = \mathbf{A} \circ \nabla u, \tag{2.4.14}$$

the composite mapping of $\mathbf{A}$ and ∇u. Here $\mathbf{A}$ is the mapping defined by

$$\mathbf{A}(p,q) = \pm\sqrt{1+(p^2+q^2)^{-1}} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix}, \tag{2.4.15}$$

a diffeomorphism from $\mathbb{R}^2 \setminus \{\text{origin}\}$ onto $\mathbb{R}^2 \setminus \{\text{unit disk}\}$.

The very definition of Legendre transform implies that u and v obey (2.4.14) if and only if U and V obey

$$\nabla U = (\nabla V) \circ \mathbf{A}. \tag{2.4.16}$$

Equation (2.4.16) can be recast as follows

$$(\nabla \mathbf{A}(p,q))^T \begin{bmatrix} \dfrac{\partial}{\partial p} \\ \dfrac{\partial}{\partial q} \end{bmatrix} U(p,q) = \begin{bmatrix} \dfrac{\partial}{\partial p} \\ \dfrac{\partial}{\partial q} \end{bmatrix} (V \circ \mathbf{A})(p,q), \tag{2.4.17}$$

since

$$\begin{bmatrix} \dfrac{\partial}{\partial p} \\ \dfrac{\partial}{\partial q} \end{bmatrix} (V \circ \mathbf{A})(p,q) = (\nabla \mathbf{A}(p,q))^T (\nabla V)(\mathbf{A}(p,q))$$

by the chain rule. Observe that

$$\nabla \mathbf{A}(p,q) = \frac{\pm 1}{\sqrt{(p^2+q^2)^3(1+p^2+q^2)}} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} (p^2+q^2)^2+q^2 & -pq \\ -pq & (p^2+q^2)^2+p^2 \end{bmatrix} \tag{2.4.18}$$

according to (2.4.15), and recall that the superscript T stands for transpose.

System (2.4.13) is nothing but a more readable version of equation (2.4.17), and can be obtained via a change of variables.

Let $\mathbf{B}$ be the diffeomorphism from $\mathbb{R}^2 \setminus \{\text{origin}\}$ onto itself defined by

$$\mathbf{B}(\lambda,\mu) = \sinh\lambda \begin{bmatrix} \cos\mu \\ \sin\mu \end{bmatrix} \tag{2.4.19}$$

and let

$$\mathbf{C} = \mathbf{A} \circ \mathbf{B}. \tag{2.4.20}$$

Equations (2.4.18) and (2.4.19) give

$$\nabla \mathbf{A}(\mathbf{B}(\lambda,\mu)) = \pm \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \cos\mu & -\sin\mu \\ \sin\mu & \cos\mu \end{bmatrix} \begin{bmatrix} \tanh\lambda & 0 \\ 0 & \coth\lambda \end{bmatrix} \begin{bmatrix} \cos\mu & \sin\mu \\ -\sin\mu & \cos\mu \end{bmatrix},$$

and consequently

$$(\nabla \mathbf{B}(\lambda,\mu))^{-1} \cdot \nabla \mathbf{A}(\mathbf{B}(\lambda,\mu)) \cdot \nabla \mathbf{B}(\lambda,\mu) = \pm \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{2.4.21}$$

Equations (2.4.20) and (2.4.21) enable one to rewrite (2.4.17) thus

$$\pm \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}^T \begin{bmatrix} \dfrac{\partial}{\partial \lambda} \\ \dfrac{\partial}{\partial \mu} \end{bmatrix} (U \circ \mathbf{B})(\lambda, \mu) = \begin{bmatrix} \dfrac{\partial}{\partial \lambda} \\ \dfrac{\partial}{\partial \mu} \end{bmatrix} (V \circ \mathbf{C})(\lambda, \mu). \tag{2.4.22}$$

Since (2.4.15) and (2.4.20) give

$$\mathbf{C}(\lambda, \mu) = \cosh \lambda \begin{bmatrix} \cos(\mu \pm \pi/2) \\ \sin(\mu \pm \pi/2) \end{bmatrix},$$

the conclusion follows from (2.4.22). □

Proposition 2.4.3 enables one to exhibit pairs of Bäcklund transforms in closed form. The following example can be worked out this way.

EXAMPLE 2. Let u be as in Example 1 above and let v be given by (2.1.7). Then a parametric representation of v and its first-order derivatives looks as follows

$$\begin{gathered} 0 < \lambda < \infty, \quad 0 \le \mu < 2\pi, \\ x = \frac{\sin \mu}{2 \cosh \lambda} \cdot [\cosh(2\lambda) - \cos(2\mu)], \quad y = \frac{\cos \mu}{2 \cosh \lambda} \cdot [\cosh(2\lambda) + \cos(2\mu)], \\ v(x, y) = \pm \frac{1}{4} [\cosh(2\lambda) + 2] \cdot \cos(2\mu) + (\text{Constant}), \\ v_x(x, y) = \mp \cosh \lambda \cdot \sin \mu, \quad v_y(x, y) = \pm \cosh \lambda \cdot \cos \mu. \end{gathered} \tag{2.4.23}$$

2.5. Developable Graphs. The case where the Hessian determinant vanishes identically can be characterized by saying that the graph in hand is a *developable surface* — see [**Lau**, Chapter 2, Sections 3.6, 3.7 and 6.3]. This case was ruled out by the methods of Subsection 2.4, and is covered by the next proposition. There equation (2.1.6) is dealt with, and hyperbolic solutions come out.

PROPOSITION 2.5.1. *Suppose the graph of v is developable, but is not a plane. Then v satisfies equation (2.1.6) if and only if the hodograph of v is included in a characteristic curve of equation (2.4.11)—in other words, either a constant C exists such that*

$$v_x^2 + v_y^2 - v_x \cdot \cos C - v_y \cdot \sin C = 0$$

or

$$v_x^2 + v_y^2 = 1.$$

PROOF. In view of the assumptions made, a real-valued smooth function f of two real variables p and q exists such that

$$f(v_x, v_y) = 0 \tag{2.5.1}$$

and ∇f vanishes nowhere; moreover,

$$v_{xx}^2 + 2v_{xy}^2 + v_{yy}^2 \neq 0.$$

Equation (2.5.1) implies

$$v_{xx} : \left[\frac{\partial f}{\partial q}(v_x, v_y) \right]^2 = -v_{xy} : \left[\frac{\partial f}{\partial q}(v_x, v_y) \cdot \frac{\partial f}{\partial p}(v_x, v_y) \right] = v_{yy} : \left[\frac{\partial f}{\partial p}(v_x, v_y) \right]^2.$$

Hence equations (2.1.6) and (2.5.1) consist with one another if and only if f obeys

$$\big[\big(p^2+q^2\big)^{\,2}-p^2\big]\left(\frac{\partial f}{\partial q}\right)^2-2pq\frac{\partial f}{\partial q}\frac{\partial f}{\partial p}+\big[\big(p^2+q^2\big)^{\,2}-q^2\big]\left(\frac{\partial f}{\partial p}\right)^2=0. \tag{2.5.2}$$

Equation (2.5.2) means that any level curve of f is a characteristic curve of equation (2.4.11). □

3. Critical Points

3.1. The main results of the present section, which appear in the following two theorems and in ensuing generalizations, bring *critical points* into relation with *rays.* Recall that rays can be defined as either geodesics belonging to the metric displayed in (1.1.18) or orbits of the differential equations (1.1.19) and (1.1.20). Recall also that a critical point of a smooth real-valued function u is said to be *degenerate* if the Hessian determinant of u vanishes at that point. (Any non-degenerate critical point is isolated, all non-isolated critical points are degenerate.)

THEOREM 3.1.1. *Assume n is strictly positive, w is a smooth solution to (1.1.1), and u is the real part of w. If ∇u vanishes at some point, then ∇u vanishes everywhere on a ray passing through that point.*

THEOREM 3.1.2. *Suppose n is strictly positive; suppose u is smooth and real-valued, and satisfies either (1.1.6) or (1.1.11) in every open set where $\nabla u \neq 0$. Assertions:*

(i) *Any critical point of u is degenerate.*
(ii) *If $u_x = u_y = 0$ and $u_{xx}^2 + 2u_{xy}^2 + u_{yy}^2 > 0$ at some point, then $u_x = u_y = 0$ everywhere on a smooth curve passing through that point.*
(iii) *If $u_x = u_y = 0$ and $u_{xx}^2 + 2u_{xy}^2 + u_{yy}^2 > 0$ at every point of a smooth curve, then this curve is a ray.*

A convenient working context is introduced in the next Subsection; theorems are demonstrated in Subsections 3.3 that include Theorems 3.1.1 and 3.1.2 as special cases.

3.2. Let f be a *Young function*—i.e. a map from $[0,\infty[$ to $[0,\infty[$ that is convex and vanishes at 0. Let n be a *strictly positive* function of x and y. Suppose f and n are *smooth.* Let u and v stand for smooth real-valued functions of x and y.

The following 2×2 system of partial differential equations

$$|\nabla v| = nf'\left(\frac{|\nabla u|}{n}\right),\quad \nabla u\cdot\nabla v = 0 \tag{3.2.1}$$

is degenerate elliptic—degeneracies occur at the critical points of u, and at those points where either $f'\left(\frac{|\nabla u|}{n}\right)$ or $f''\left(\frac{|\nabla u|}{n}\right)$ vanishes.

The following partial differential equation

$$\operatorname{div}\left\{nf'\left(\frac{|\nabla u|}{n}\right)\frac{\nabla u}{|\nabla u|}\right\}=0 \tag{3.2.2}$$

is either elliptic or degenerate elliptic—degeneracies prevail if f' fails to vanish at 0 and critical points of u occur.

Comments follow.

Equation (3.2.2) can be recast as follows

$$\left[f''(\rho)(\cos\mu)^2+\frac{f'(\rho)}{\rho}(\sin\mu)^2\right]u_{xx} \tag{3.2.3}$$
$$+2\left[f''(\rho)-\frac{f'(\rho)}{\rho}\right]\cos\mu\sin\mu\ u_{xy}+\left[f''(\rho)(\sin\mu)^2+\frac{f'(\rho)}{\rho}(\cos\mu)^2\right]u_{yy}$$
$$=\left[f''(\rho)-\frac{f'(\rho)}{\rho}\right]\nabla u\cdot\nabla\log n$$

if sufficiently smooth solutions are dealt with and ρ and μ are defined thus

$$n\rho=|\nabla u|,\quad u_x:\cos\mu=u_y:\sin\mu.$$

The left-hand side of (3.2.3) is the trace of the following matrix

$$\begin{bmatrix}\cos\mu & -\sin\mu\\ \sin\mu & \cos\mu\end{bmatrix}\begin{bmatrix}f''(\rho) & 0\\ 0 & f'(\rho)/\rho\end{bmatrix}\begin{bmatrix}\cos\mu & \sin\mu\\ -\sin\mu & \cos\mu\end{bmatrix}\begin{bmatrix}u_{xx} & u_{xy}\\ u_{xy} & u_{yy}\end{bmatrix},$$

which makes the elliptic character of equation (3.2.3) apparent.

The following equation

$$f''(\rho)\ \Delta u+[\rho f''(\rho)-f'(\rho)]\left[nh-\nabla n\cdot\frac{\nabla u}{|\nabla u|}\right]=0, \tag{3.2.4}$$

which proves decisive in subsequent developments, results from reassembling terms in (3.2.3). Here h is the signed curvature of the level curves of u defined by formulas (1.1.15) and (1.1.17).

Statements (i) and (ii) below correlate system (3.2.1) with equation (3.2.2). They show the following: if both u and v obey (3.2.1) and either $f'(0)=0$ or $\nabla u\neq 0$, then u obeys (3.2.2); if either $f'(0)=0$ or $\nabla u\neq 0$, and in addition u obeys (3.2.2) and v is a suitable Bäcklund transform of u, then u and v obey (3.2.1).

(i) The mapping $u\mapsto v$ defined by

$$\begin{bmatrix}v_x\\ v_y\end{bmatrix}=\pm\frac{n}{|\nabla u|}f'\left(\frac{|\nabla u|}{n}\right)\begin{bmatrix}-u_y\\ u_x\end{bmatrix} \tag{3.2.5}$$

correlates the first component of any appropriate solution pair to system (3.2.1) with the second component of the same pair. If u and v satisfy (3.2.1) and either $f'(0)=0$ or $\nabla u\neq 0$, then u and v satisfy (3.2.5); conversely, if either $f'(0)=0$ or $\nabla u\neq 0$ and u and v satisfy (3.2.5), then u and v satisfy (3.2.1).

(ii) Equation (3.2.2) is a necessary condition for system (3.2.5) to be exact.

System (3.2.1) basically amounts to Cauchy-Riemann equations if $f(\rho)=\rho^2/2$ for every nonnegative ρ. System (3.2.1), equation (3.2.2) and equation (3.2.3) coincide with (1.1.3), (1.1.6) and (1.1.11), respectively, if f is specified by

$$f(\rho)=\tfrac{1}{2}\left[\rho\sqrt{\rho^2+1}+\log\left(\rho+\sqrt{\rho^2+1}\right)\right] \tag{3.2.6}$$

for every nonnegative ρ.

3.3. In the present subsection we deal with system (3.2.1), equation (3.2.2) and equation (3.2.3).

THEOREM 3.3.1. *Suppose $f'(0) > 0$, $f''(0) = 0$. Let u and v be smooth real-valued solutions to (3.2.1). If ∇u vanishes at some point, then ∇u vanishes everywhere on a ray passing through that point.*

PROOF. The former equation in (3.2.1) and the hypotheses made on f and n inform us that

$$\nabla v \neq 0 \tag{3.3.1}$$

everywhere. We start by exploiting (3.3.1) and

$$u_x v_x + u_y v_y = 0, \tag{3.3.2}$$

the latter equation in (3.2.1).

Let

$$\boldsymbol{l} = \frac{\nabla v}{|\nabla v|}$$

and

$$k = -\operatorname{div} \boldsymbol{l}$$

—$\boldsymbol{l}$ is a unit vector field tangent to the curves of steepest descent of v, k is a signed curvature of the level curves of v. Condition (3.3.1) ensures that $\boldsymbol{l}$ and k are smooth, as are the level curves and the curves of steepest descent of v. Note the following alternative formula

$$k = -|\nabla v|^{-3} \left(\begin{bmatrix} v_{yy} & -v_{xy} \\ -v_{xy} & v_{xx} \end{bmatrix} \nabla v, \nabla v \right).$$

We have

$$\frac{\partial}{\partial \boldsymbol{l}} |\nabla u|^2 = \frac{2}{|\nabla v|} \left(\begin{bmatrix} u_{xx} & u_{xy} \\ u_{xy} & u_{yy} \end{bmatrix} \nabla v, \nabla u \right),$$

because

$$\nabla |\nabla u|^2 = 2 \begin{bmatrix} u_{xx} & u_{xy} \\ u_{xy} & u_{yy} \end{bmatrix} \nabla u.$$

Equation (3.3.2) implies

$$\begin{bmatrix} u_{xx} & u_{xy} \\ u_{xy} & u_{yy} \end{bmatrix} \nabla v + \begin{bmatrix} v_{xx} & v_{xy} \\ v_{xy} & v_{yy} \end{bmatrix} \nabla u = 0$$

as well as

$$\pm \nabla u = |\nabla u| \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \frac{\nabla v}{|\nabla v|}.$$

We deduce successively that

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{l}} |\nabla u|^2 = -\frac{2}{|\nabla v|} \left(\begin{bmatrix} v_{xx} & v_{xy} \\ v_{xy} & v_{yy} \end{bmatrix} \nabla u, \nabla u \right) = \\ - \frac{2}{|\nabla v|^3} \left(\begin{bmatrix} v_{yy} & -v_{xy} \\ -v_{xy} & v_{xx} \end{bmatrix} \nabla v, \nabla v \right) |\nabla u|^2. \end{aligned}$$

Above,

$$\frac{\partial}{\partial \boldsymbol{l}} = \frac{1}{|\nabla v|} \left(v_x \frac{\partial}{\partial x} + v_y \frac{\partial}{\partial y} \right),$$

the directional derivative along the trajectories of $\boldsymbol{l}$, and we have shown that

$$\frac{\partial}{\partial \boldsymbol{l}} |\nabla u|^2 - 2k |\nabla u|^2 = 0. \tag{3.3.3}$$

Equation (3.3.3) forces $|\nabla u|$ to vanish everywhere on any curve of steepest descent of v which crosses a critical point of u. We claim that any such curve is a ray.

Let us exploit more closely the following hypothesis

$$f''(0) = 0$$

and the former equation in (3.2.1)

$$|\nabla v| = nf'\left(\frac{|\nabla u|}{n}\right). \tag{3.3.4}$$

Equation (3.3.4) gives

$$\nabla|\nabla v| = [f'(\rho) - \rho f''(\rho)]\nabla n + f''(\rho)\begin{bmatrix} u_{xx} & u_{xy} \\ u_{xy} & u_{yy} \end{bmatrix}\frac{\nabla u}{|\nabla u|},$$

since ρ was defined by $n\rho = |\nabla u|$ and

$$\nabla|\nabla u| = \begin{bmatrix} u_{xx} & u_{xy} \\ u_{xy} & u_{yy} \end{bmatrix}\frac{\nabla u}{|\nabla u|}.$$

It follows that

$$|\nabla v| = f'(0)n \quad \text{and} \quad \nabla|\nabla v| = f'(0)\nabla n \quad \text{at any critical point of } u. \tag{3.3.5}$$

A curve of steepest descent of v is an orbit of the following system

$$\frac{d}{ds}\begin{bmatrix} x \\ y \end{bmatrix} = \nabla v(x, y),$$

which implies that

$$(dx/ds)^2 + (dy/ds)^2 = |\nabla v|^2$$

and that

$$\frac{d^2}{ds^2}\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{2}\nabla|\nabla v|^2.$$

Suppose ∇u vanishes at every point of a curve of steepest descent of v. This curve obeys

$$(dx/ds)^2 + (dy/ds)^2 = [f'(0)n]^2,$$

and

$$\begin{vmatrix} dx/ds & dy/ds \\ d^2x/ds^2 & d^2y/ds^2 \end{vmatrix} = [f'(0)]^2 n \begin{vmatrix} dx/ds & dy/ds \\ \partial n/\partial x & \partial n/\partial y \end{vmatrix}$$

because of (3.3.5). Consequently, the same curve satisfies (1.1.19)—that is the equation defining the rays.

The proof is complete. □

THEOREM 3.3.2. *Suppose $f'(0) > 0$, $f''(0) = 0$. Suppose u is smooth and real-valued, and satisfies either equation (3.2.2) or equation (3.2.3) in every open set where $\nabla u \neq 0$. Assertions:*

(i) *Any critical point of u is degenerate.*

(ii) *If $u_x = u_y = 0$ and $u_{xx}^2 + 2u_{xy}^2 + u_{yy}^2 > 0$ at some point, then $u_x = u_y = 0$ everywhere on a smooth curve passing through that point.*

(iii) *If $u_x = u_y = 0$ and $u_{xx}^2 + 2u_{xy}^2 + u_{yy}^2 > 0$ at every point of a smooth curve, then this curve is a ray.*

PROOF OF (I). Assume $u_x(0,0) = u_y(0,0) = 0$; let

$$u_{11} = u_{xx}(0,0),\ u_{12} = u_{xy}(0,0),\ u_{22} = u_{yy}(0,0),$$

and assume

$$u_{11}u_{22} - u_{12}^2 \neq 0 \tag{3.3.6}$$

by contradiction.

These hypotheses imply that the origin is a saddle or an extremum point depending on whether the determinant $u_{11}u_{22} - u_{12}^2$ is negative or positive. The following proposition can be found in [**Gou**, §42]. If $u_{11}u_{22} - u_{12}^2 < 0$, the set of points (x, y) such that $u(x, y) = u(0, 0)$ and $x^2 + y^2$ is small enough consists of two smooth branches crossing the origin at different slopes; if $u_{11}u_{22} - u_{12}^2 > 0$, the set in question consists of the origin only.

The behavior of h near the origin can easily be derived: $h(x, y)$ oscillates between $-\infty$ and $+\infty$ or blows up as (x, y) approaches $(0, 0)$ depending on whether $u_{11}u_{22} - u_{12}^2$ is negative or positive. Recall that h stands for a signed curvature of the level curves of u.

More precise information can be derived as follows. Formula (1.1.15) gives

$$h = -\left(u_{xx}u_{yy} - u_{xy}^2\right) \cdot |\nabla u|^{-3} \ \cdot \left(\begin{bmatrix} u_{xx} & u_{xy} \\ u_{xy} & u_{yy} \end{bmatrix}^{-1} \nabla u, \nabla u \right)$$

at every point where the gradient of u does not vanish and the Hessian matrix of u is non-singular. Taylor's formula gives

$$\nabla u(x, y) = \begin{bmatrix} u_{11} & u_{12} \\ u_{12} & u_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + O(x^2 + y^2)$$

as (x, y) approaches $(0, 0)$. Let λ, κ_1 and κ_2 obey

$$\begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ u_{12} & u_{22} \end{bmatrix} \begin{bmatrix} \cos\lambda & -\sin\lambda \\ \sin\lambda & \cos\lambda \end{bmatrix} = \begin{bmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{bmatrix},$$

and let ρ and μ be the polar coordinates defined by

$$0 < \rho < \infty, \quad 0 \leq \mu < 2\pi, \quad x = \rho \cdot \cos\mu, \quad y = \rho \cdot \sin\mu.$$

We deduce

$$h(x, y) = -\frac{\kappa_1\kappa_2}{\rho} \ \frac{\kappa_1(\cos(\lambda - \mu))^2 + \kappa_2(\sin(\lambda - \mu))^2}{\left[(\kappa_1 \cos(\lambda - \mu))^2 + (\kappa_2 \sin(\lambda - \mu))^2\right]^{3/2}} + O(1) \tag{3.3.7}$$

as ρ approaches 0 and μ remains constant.

Formula (3.3.7) portrays the behavior of h near the origin. It shows that $h(x, y)$ does *not* stay bounded as (x, y) approaches $(0, 0)$ if (3.3.6) is in force.

On the other hand, equation (3.2.4) holds in a punctured neighborhood of $(0, 0)$. This equation and the hypotheses made on f and n tell us that $h(x, y)$ *must* remain bounded as (x, y) approaches $(0, 0)$.

This is a contradiction, which concludes the proof of (i). □

PROOF OF (II). Let

$$u_x(0,0) = u_y(0,0) = 0$$

and

$$u_{xx}^2(0,0) + 2u_{xy}^2(0,0) + u_{yy}^2(0,0) > 0.$$

By assertion (i), the Hessian matrix of u is singular at $(0,0)$. Hence $\Delta u(0,0) \neq 0$, so that u_{xx} and u_{yy} do not vanish at $(0,0)$ simultaneously. Let

$$u_{yy}(0,0) \neq 0,$$

for instance.

By the implicit function theorem, if δ is positive and small enough then

$$\left\{(x,y) \in \mathbb{R}^2 : u_y(x,y) = 0 \text{ and } x^2 + y^2 < \delta^2\right\},$$

a relevant portion of the set of points where u_y vanishes, is a smooth curve.

We claim that a positive number δ exists such that

$$\left\{(x,y) \in \mathbb{R}^2 : u_y(x,y) = 0 \text{ and } x^2 + y^2 < \delta^2\right\} \subseteq \left\{(x,y) \in \mathbb{R}^2 : u_x(x,y) = 0\right\}.$$

In effect, both equations (3.2.2) and (3.2.3) give

$$|u_x| f''(\rho) u_{xx} + n f'(\rho) u_{yy} + [f'(\rho) - \rho f''(\rho)] \cdot u_x \cdot n_x = 0$$

and

$$n \cdot \rho = |u_x|$$

at every point where $u_x \neq 0$ and $u_y = 0$. These equations and the hypotheses made on f and n tell us the following. If a sequence of points (x_m, y_m) had the properties

$$x_m^2 + y_m^2 \to 0, \quad u_x(x_m, y_m) \neq 0, \quad u_y(x_m, y_m) = 0,$$

the following equation would ensue

$$u_{yy}(0,0) = 0,$$

a contradiction.

The claim follows. Assertion (ii) follows too. □

PROOF OF (III). Let γ denote the curve in question. The Hessian matrix of u is singular at every point of γ, but is not zero. Hence

$$\Delta u \neq 0$$

and a scalar field λ exists such that

$$\begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix} \begin{bmatrix} u_{xx} & u_{xy} \\ u_{xy} & u_{yy} \end{bmatrix} \begin{bmatrix} \cos\lambda & -\sin\lambda \\ \sin\lambda & \cos\lambda \end{bmatrix} = \begin{bmatrix} \Delta u & 0 \\ 0 & 0 \end{bmatrix}$$

at every point of γ. Curve γ obeys

$$-dx : \sin\lambda = dy : \cos\lambda,$$

since $u_{xx}dx + u_{xy}dy = u_{xy}dx + u_{yy}dy = 0$ along γ; further differentiations show that γ obeys also

$$-\frac{1}{\Delta u}\left\{u_{xxx}(\sin\lambda)^2 - 2u_{xxy}\sin\lambda\ \cos\lambda + u_{xyy}(\cos\lambda)^2\right\} \times (\text{princ. norm.}) =$$
$$\cos\lambda \begin{bmatrix} \cos\lambda \\ \sin\lambda \end{bmatrix}$$

and

$$-\frac{1}{\Delta u}\ \left\{u_{xxy}(\sin\lambda)^2 - 2u_{xyy}\sin\lambda\ \cos\lambda + u_{yyy}(\cos\lambda)^2\right\} \times (\text{princ. norm.}) = \sin\lambda\ \begin{bmatrix}\cos\lambda\\ \sin\lambda\end{bmatrix}.$$

Therefore the following equations

$$-\frac{1}{\Delta u}\Big\{\left[u_{xxx}(\sin\lambda)^2 - 2\ u_{xxy}\sin\lambda\cos\lambda + u_{xyy}(\cos\lambda)^2\right]\cos\lambda + \left[u_{xxy}(\sin\lambda)^2 - 2u_{xyy}\sin\lambda\cos\lambda + u_{yyy}(\cos\lambda)^2\right]\sin\lambda\Big\} \times (\text{princ. norm.}) = \begin{bmatrix}\cos\lambda\\ \sin\lambda\end{bmatrix} \tag{3.3.8}$$

and

$$u_{xxx}\ (\sin\lambda)^3 - 3u_{xxy}\ (\sin\lambda)^2\cos\lambda + 3u_{xyy}\sin\lambda(\cos\lambda)^2 - u_{yyy}(\cos\lambda)^3 = 0 \tag{3.3.9}$$

hold everywhere on γ.

Consider any point of γ. Without loss of generality we may assume that it coincides with $(0,0)$ and that λ equals $\pi/2$ there. Let

$$A = \Delta u(0,0),\quad B = u_{xxy}(0,0),\quad C = u_{xyy}(0,0),\quad D = u_{yyy}(0,0).$$

Equation (3.3.8) gives

$$-(B/A) \times [(\text{principal normal at } (0,0)] = \begin{bmatrix}0\\ 1\end{bmatrix} \tag{3.3.10}$$

—that is

$$\text{unit normal to } \gamma \text{ at } (0,0) = \begin{bmatrix}0\\ 1\end{bmatrix}$$

and

$$\text{signed curvature of } \gamma \text{ at } (0,0) = -B/A.$$

Equation (3.3.9) implies $u_{xxx}(0,0) = 0$. Hence Taylor's formula gives

$$\begin{aligned}
u(x,y) &= u(0,0) + \frac{A}{2}y^2 + \frac{1}{6}\big(3Bx^2y + 3Cxy^2 + Dy^3\big) + O(\rho^4),\\
u_x(x,y) &= Bxy + \frac{C}{2}y^2 + O(\rho^3),\\
u_y(x,y) &= Ay + \frac{1}{2}\big(Bx^2 + 2Cxy + Dy^2\big) + O(\rho^3),\\
u_{xx}(x,y) &= By + O(\rho^2),\\
u_{xy}(x,y) &= Bx + Cy + O(\rho^2),\\
u_{yy}(x,y) &= A + Cx + Dy + O(\rho^2)
\end{aligned}$$

as ρ approaches 0. We deduce

$$\frac{\nabla u(x,y)}{|\nabla u(x,y)|} = \operatorname{sgn}(A\sin\mu)\ \begin{bmatrix}0\\ 1\end{bmatrix} + O(\rho) \tag{3.3.11}$$

as ρ approaches 0, μ is constant, $\mu \neq 0$ and $\mu \neq \pi$. We deduce also

$$h(x,y) = -(B/A)\operatorname{sgn}(A\sin\mu) + O(\rho) \tag{3.3.12}$$

as ρ approaches 0, μ is constant, $\mu \neq 0$ and $\mu \neq \pi$. We used the polar coordinates defined in the proof above, and used formula (1.1.15).

Formulas (3.3.10), (3.3.11) and (3.3.12) show that

$$(1/h)\frac{\nabla u}{|\nabla u|} \text{ approaches the principal normal to } \gamma \tag{3.3.13}$$

as (x, y) approaches γ non tangentially.

Equation (3.2.4) informs us that

$$h(x,y) - \frac{\nabla n(x,y)}{n(x,y)} \cdot \frac{\nabla u(x,y)}{|\nabla u(x,y)|} \to 0 \tag{3.3.14}$$

as (x, y) approaches any critical point.

We infer from (3.3.13) and (3.3.14) that γ obeys equation (1.1.20). In other words, γ is a ray. □

THEOREM 3.3.3. *Suppose*

$$n(x,y) \equiv 1, \tag{3.3.15}$$

and let u and U be a pair of Legendre transforms. Then u satisfies equation (3.2.3) if and only if U satisfies

$$\left[f'(\rho) \cdot \frac{\partial^2}{\partial \rho^2} + f''(\rho) \cdot \left(\frac{\partial}{\partial \rho} + \frac{1}{\rho} \cdot \frac{\partial^2}{\partial \mu^2} \right) \right] U\big(\rho \cdot e^{i\mu}\big) = 0 \tag{3.3.16}$$

for every ρ and μ such that $0 < \rho < \infty$, $0 \leq \mu < 2\pi$ and $\rho \cdot e^{i\mu}$ belongs to the hodograph of u.

PROOF. Combine equation (3.2.3) and formulas (2.4.1) to (2.4.5)—the basic formulas about the Legendre transformation. Use polar coordinates in the hodograph of u and represent the Hessian matrix of U in these coordinates. □

References

[AI] R.L. Anderson & N.H. Ibragimov, *Lie-Bäcklund Transformations in Applications*, Studies in Applied Mathematics, SIAM, 1979.

[Be] L. Bers, *Mathematical aspects of subsonic and transonic gas dynamics*, Chapman & Hall, 1958.

[BM] D. Bouche & F. Molinet, *Méthodes asymptotiques en électromagnétisme*, Springer-Verlag, 1994.

[CF1] S. Choudhary & L.B. Felsen, *Asymptotic theory for inhomogeneous waves*, IEEE Trans. Antennas Propaga. **AP-21** (1973), 827–842.

[CF2] S. Choudhary & L.B. Felsen, *Analysis of Gaussian beam propagation and diffraction by inhomogeneous wave tracking*, Proc. IEEE **62** (1974), 1530–1541.

[CH] R. Courant & D. Hilbert, *Methods of Mathematical Physics, vol. 2*, Interscience, 1962.

[DoC] M.P. DoCarmo, *Differential Geometry of Curves and Surfaces*, Prentice Hall, 1976.

[EF] P.D. Einzinger & L.B. Felsen, *Evanescent waves and complex rays*, IEEE Trans. Ant. Prop. **AP-30, 4** (1982), 594–605.

[ER] P. Einzinger & S. Raz, *On the asymptotic theory of inhomogeneous wave tracking*, Radio Science **15** (1980), 763–771.

[Fe1] L.B. Felsen, *Evanescent waves*, J. Opt. Soc. Amer. **66** (1976), 751–760.

[Fe2] L.B. Felsen, *Complex-source-point solutions of the field equations and their relation to the propagation and scattering of Gaussian beams*, Symposia Mathematica **18** (1976), 39–56.

[Gou] É. Goursat, *Cours d'Analyse Mathématique*, Gauthier-Villars, 1956.

[Hem] J.E. Hemmati, *Entire solutions of first-order nonlinear partial differential equations*, Proc. Amer. Math. Soc. **125** (1997), 1483–1485.
[HF] E. Heyman & L.B. Felsen, *Evanescent waves and complex rays for modal propagation in curved open waveguides*, SIAM J. Appl. Math. **43** (1983), 855–884.
[Jo] D.S. Jones, *The Theory of Electromagnetism*, Pergamon Press, 1964.
[Kha] D. Khavinson, *A note on entire solutions of the eiconal equation*, Amer. Math. Monthly **102** (1995), 159–161.
[KS] D. Kinderlehrer & G. Stampacchia, *An introduction to variational inequalities and their applications*, Academic Press, 1980.
[Lau] D. Laugwitz, *Differential and Riemannian Geometry*, Academic Press, 1965.
[Law] J.D. Lawrence, *A catalog of special plane curves*, Dover Publications, 1972.
[RS] C. Rogers & W.F. Shadwick, *Bäcklund Transformations and Their Applications*, Mathematics in Science and Engineering, vol. 161, Academic Press, 1982.

Dipartimento di Matematica "U. Dini", Università di Firenze, viale Morgagni 67A, 50134 Firenze, Italy
E-mail address: magnan@udini.math.unifi.it

Dipartimento di Matematica "U. Dini", Università di Firenze, viale Morgagni 67A, 50134 Firenze, Italy
E-mail address: talenti@udini.math.unifi.it

Contemporary Mathematics
Volume **238**, 1999
B 0-8218-1196-7-03550-7

On Diffusion-Induced Grain-Boundary Motion

Uwe F. Mayer and Gieri Simonett

Abstract. We consider a sharp interface model which describes diffusion-induced grain-boundary motion in a poly-crystalline material. This model leads to a fully nonlinear coupled system of partial differential equations. We show existence and uniqueness of smooth solutions.

1. Introduction

In this paper we consider a model which describes diffusion-induced grain-boundary motion of a surface which separates different grains in a poly-crystalline material. Let Γ_0 be a compact closed hypersurface in $\mathbb{R}^n$ which is the boundary of an open domain, and let $u_0 : \Gamma_0 \to \mathbb{R}$ be a given function. Then we are looking for a family $\Gamma := \{\Gamma(t)\,;\, t \geq 0\}$ of hypersurfaces and a family of functions $\{u(\cdot,t) : \Gamma(t) \to \mathbb{R}\,;\, t \geq 0\}$ such that the following system of equations holds:

$$\begin{aligned} V &= -H_\Gamma - f(u), & \Gamma(0) &= \Gamma_0, \\ \dot{u} &= \Delta_\Gamma u - VH_\Gamma u + Vu + g(u), & u(0) &= u_0. \end{aligned} \tag{1.1}$$

Here $V(t)$ denotes the normal velocity of Γ at time t, while $H_{\Gamma(t)}$ and $\Delta_{\Gamma(t)}$ stand for the mean curvature and the Laplace-Beltrami of $\Gamma(t)$, respectively. The symbol $\dot{u}$ denotes the derivative of u along flow lines which are orthogonal to $\Gamma(t)$, see the definition in (2.6). We assume that

$$f, g \in C^\infty(\mathbb{R}, \mathbb{R}) \quad \text{and} \quad f(0) = 0, \quad g(0) = U.$$

In two dimensions, the interface $\Gamma(t)$ represents the boundary of a grain of a thin poly-crystalline material with vapor on top (in the third dimension). The vapor in the third dimension contains a certain solute which is absorbed by the interface and which diffuses along the interface. Furthermore, as the interface moves, some of the solute will be deposited in the bulk through which the interface has passed. The chemical composition of the newly created crystal behind the advancing grain will be different from that in front, because atoms of the solute have been deposited there. For this physical background we consider only convex curves, and we choose the signs so that a family of shrinking curves has negative normal velocity. A high concentration u of the solute in the interface increases the velocity, because the

1991 *Mathematics Subject Classification.* Primary 35R35; Secondary 35K45, 35K55.

The research of the second author has been partially supported by the Vanderbilt University Research Council and by NSF Grant DMS-9801337.

interface tries to reduce that concentration by depositing the solute in the regions it passes through. In addition, the stretching or shrinking of Γ during its motion induces a change in the concentration of the solute.

This situation results in the following terms: $V = -H_\Gamma$ is the usual motion by mean curvature that models motion driven purely by surface tension, and the term $f(u)$ results from the deposition effect. Here, $f(u) = u^2$ is reasonable [**9**]. As for the second equation, $\Delta_\Gamma u$ describes diffusion on a manifold, $-VH_\Gamma u$ indicates the concentration change due to the change of the length of the interface, Vu describes the reduction of the solute due to deposition, and $g(u)$ results from absorption of the solute from the vapor. Physically, $g(u) = U - u$ is meaningful, where U is the concentration of the solute in the vapor [**9**].

DIGM is known to be an important component of many complicated diffusion processes in which there are moving grain boundaries; see [**3**] and the references cited therein. In this type of phenomenon, the free energy of the system can be reduced by the incorporation of some of the solute into one or both of the grains separated by the grain boundary. In the DIGM mechanism, this transfer is accomplished by the disintegration of one grain and the simultaneous building up of the adjacent grain, the solute being added during the build-up process. This results in the migration of the grain boundary [**9**]. The possibility of reducing the free energy this way does not automatically imply that migration actually takes place; mechanisms for this to happen have been proposed, including the recent one in [**3**].

In [**3**], a thermodynamically consistent phase-field model for DIGM is suggested. This model has two phase fields, one being the concentration of the solute, and the other one being an order parameter which distinguishes the two crystal grains by the values $+1$ and -1, and which takes intermediate values in the grain boundary.

In this paper we consider a sharp interface model for DIGM. The same model has been studied in [**14**], where existence and uniqueness of classical Hölder solutions is proved. Here we improve this result considerably. Namely, we prove that solutions are in fact smooth in space and time.

Let $\Gamma := \{\Gamma(t)\,;\, t \in [0,T)\}$ be a family of closed compact embedded hypersurfaces in $\mathbb{R}^n$ and let

$$\mathcal{M}_0 := \bigcup_{t\in[0,T)} \Gamma(t)\times\{t\}\,, \qquad \mathcal{M} := \bigcup_{t\in(0,T)} \Gamma(t)\times\{t\}\,. \tag{1.2}$$

Finally, let u be a function on $\mathcal{M}_0$. Then we call (Γ, u) a smooth C^∞-solution of (1.1) on $[0,T)$ if the following properties hold:

- $\mathcal{M}$ is an n-dimensional manifold of class C^∞ in $\mathbb{R}^{n+1}$ and $u|_{\mathcal{M}} \in C^\infty(\mathcal{M})$,
- $\mathcal{M}_0$ is a C^1-manifold with boundary $\mathcal{M}_0 \cap (\mathbb{R}^n \times \{0\})$ and $u \in C^1(\mathcal{M}_0)$,
- $\mathcal{M}_0 \cap (\mathbb{R}^n \times \{0\}) = \Gamma_0 \times \{0\} \equiv \Gamma_0$ and $u|_{\Gamma_0} = u_0$,
- the pair (Γ, u) satisfies system (1.1).

We are now ready to state our main theorem on existence and uniqueness of smooth solutions for (1.1)

THEOREM 1.1. *Let $\beta \in (0,1)$ be given and suppose that $\Gamma_0 \in C^{2+\beta}$ and that $u_0 \in C^{2+\beta}(\Gamma_0)$. Then system* (1.1) *has a smooth solution* (Γ, u) *on* $[0,T)$ *for some* $T > 0$. *The solution is unique in the class* (4.1).

A detailed analysis shows that (1.1) is a fully nonlinear coupled system, where the fully nonlinear character comes in through the term $VH_\Gamma u$. It is shown in [**14**] that (1.1) admits classical solutions which are smooth in time and $C^{2+\alpha}$ in space for given initial data in $C^{2+\beta}$, where $0 < \alpha < \beta < 1$.

In order to investigate system (1.1) we represent the moving hypersurface $\Gamma(t)$ as a graph over a fixed reference manifold Σ and then transform (1.1) to an evolution equation over Σ. This leads to a fully nonlinear system which is parabolic (in the sense that the linearization generates an analytic semigroup on an appropriate function space), as is shown in [**14**]. Since the fully nonlinear term occurs on the cross diagonal we will be able to combine maximal regularity results and bootstrapping arguments to show that solutions immediately regularize for positive times.

System (1.1) reduces to the well-known mean curvature flow

$$V = -H_\Gamma, \qquad \Gamma(0) = \Gamma_0\,, \tag{1.3}$$

if $u_0 = 0$ and $U = 0$, since $u \equiv 0$ then solves the second equation of (1.1). It is well-known that solutions of the mean curvature flow (1.3) remain strictly convex if Γ_0 is strictly convex, and that $\Gamma(t)$ shrinks to a point in finite time [**10, 12**]. Moreover, embedded curves in the plane always become convex before they shrink to a point [**11**]. We do not know if similar properties hold true for system (1.1).

2. Motion of the Interface

In this section we briefly introduce the mathematical setting in order to reformulate (1.1) as an evolution equation over a fixed reference manifold. Here we follow [**14**], see also [**5, 6, 7, 8**] for a similar situation.

Let Σ be a smooth compact closed hypersurface in $\mathbb{R}^n$, and assume that Γ_0 is close in a C^1 sense to this fixed reference manifold Σ. Let ν be the unit normal field on Σ. We choose $a > 0$ such that

$$X : \Sigma \times (-a, a) \to \mathbb{R}^n\,, \qquad X(s, r) := s + r\nu(s)$$

is a smooth diffeomorphism onto its image $\mathcal{R} := \mathrm{im}(X)$, that is,

$$X \in \mathit{Diff}^\infty(\Sigma \times (-a, a), \mathcal{R}).$$

This can be done by taking $a > 0$ sufficiently small so that Σ has a tubular neighborhood of radius a. It is convenient to decompose the inverse of X into $X^{-1} = (S, \Lambda)$, where

$$S \in C^\infty(\mathcal{R}, \Sigma) \qquad \text{and} \qquad \Lambda \in C^\infty(\mathcal{R}, (-a, a)).$$

$S(x)$ is the nearest point on Σ to $x \in \mathcal{R}$, and $\Lambda(x)$ is the signed distance from x to Σ, that is, to $S(x)$. Moreover, $\mathcal{R}$ consists of those points in $\mathbb{R}^n$ with distance less than a to Σ.

Let $T > 0$ be a fixed number. In the sequel we assume that $\Gamma := \{\Gamma(t),\, t \in [0, T)\}$ is a family of graphs in normal direction over Σ. To be precise, we ask that there is a function $\rho : \Sigma \times [0, T) \to (-a, a)$ such that

$$\Gamma(t) = \mathrm{im}\big([s \mapsto X(s, \rho(s, t))]\big), \quad t \in [0, T).$$

$\Gamma(t)$ can then also be described as the zero-level set of the function

$$\Phi_\rho \,:\, \mathcal{R} \times [0, T) \to \mathbb{R}\,, \qquad \Phi_\rho(x, t) := \Lambda(x) - \rho(S(x), t); \tag{2.1}$$

one has $\Gamma(t) = \Phi_\rho(\cdot,t)^{-1}(0)$ for any fixed $t \in [0,T)$. Hence, the unit normal field $N(x,t)$ on $\Gamma(t)$ at x can be expressed as

$$N(x,t) = \frac{\nabla_x \Phi_\rho(x,t)}{|\nabla_x \Phi_\rho(x,t)|}\,, \tag{2.2}$$

and the normal velocity V of Γ at time t and at the point $x = X(s,\rho(s,t))$ is given by

$$V(x,t) = \frac{\partial_t \rho(s,t)}{|\nabla_x \Phi_\rho(x,t)|}\,. \tag{2.3}$$

We can now explain the precise meaning of the derivative $\dot u(x,t)$ for $x \in \Gamma(t)$. Given $x \in \Gamma(t)$, let $\{z(\tau,x) \in \mathbb{R}^n\,;\, \tau \in (-\varepsilon,\varepsilon)\}$ be a flow line through x such that

$$z(\tau,x) \in \Gamma(t+\tau), \quad \dot z(\tau) = (VN)(z(\tau),t+\tau), \quad \tau \in (-\varepsilon,\varepsilon), \quad z(0) = x. \tag{2.4}$$

The existence of a unique trajectory $\{z(\tau,x) \in \mathbb{R}^n\,;\, \tau \in (-\varepsilon,\varepsilon)\}$ with the above properties is established in the next result.

LEMMA 2.1. *Suppose $\rho \in C^2(\Sigma \times (0,T))$ and let $\Gamma(t) := \Phi_\rho(\cdot,t)^{-1}(0)$ for t in $(0,T)$. Then for every $x \in \Gamma(t)$ there exist an $\varepsilon > 0$ and a unique solution $z(\cdot,x) \in C^1((-\varepsilon,\varepsilon),\mathbb{R}^n)$ of (2.4).*

PROOF. This result is proved in [**14**, Lemma 2.1]. For the reader's convenience we include a short proof. Observe that (2.4) is equivalent to the ordinary differential equation

$$(\dot z, \dot t) = ((VN)(z,t),1)\,, \quad (z(0),t(0)) = (x,t) \tag{2.5}$$

on the manifold $\mathcal{M} = \bigcup_{t\in(0,T)} \Gamma(t) \times \{t\}$. We show that

$$((VN)(x,t),1) \in T_{(x,t)}(\mathcal{M}) \quad \text{for any} \quad (x,t) \in \mathcal{M}\,.$$

For this let $\Psi_\rho := \Phi_\rho|_{\mathcal{R}\times(0,T)}$ and observe that $\mathcal{M} = \Psi_\rho^{-1}(0)$, so that the vector

$$(\nabla_x \Phi_\rho(x,t), -\partial_t \rho(S(x),t))$$

is orthogonal to $\mathcal{M}$ at $(x,t) \in \mathcal{M}$. Using the definition of Φ_ρ it can easily be seen that $\partial_\nu \Phi_\rho = 1$, and hence the vector displayed above is nonzero. By (2.2) and (2.3) we have

$$\big(((VN)(x,t),1)|(\nabla_x \Phi_\rho(x,t), -\partial_t\rho(S(x),t))\big) = 0\,, \quad (x,t) \in \mathcal{M}\,,$$

showing that $((VN)(x,t),1)$ is tangential to $\mathcal{M}$ at (x,t). We can now conclude that there is an $\varepsilon > 0$ such that (2.5) has a unique solution

$$[\tau \mapsto (z(\tau,x), t+\tau)] \in C^1((-\varepsilon,\varepsilon),\mathcal{M})\,.$$

It follows that $[\tau \mapsto z(\tau,x)] \in C^1((-\varepsilon,\varepsilon),\mathbb{R}^n)$ is the unique solution of (2.4). □

Let $(x,t) \in \mathcal{M}$ be given. Then we define

$$\dot u(x,t) := \frac{d}{d\tau} u(z(\tau,x), t+\tau)\Big|_{\tau=0}\,. \tag{2.6}$$

We now introduce the pull-back function v of u,

$$v : \Sigma \times [0,T) \to \mathbb{R}, \quad v(s,t) := u(X(s,\rho(s,t)),t). \tag{2.7}$$

Since $u(x,t) = v(S(x),t)$, it follows from (2.4) and (2.6) that

$$\dot u(x,t) = \frac{d}{d\tau} v(S(z(\tau,x)), t+\tau)\Big|_{\tau=0} = (\nabla_x v(S(x),t)|N(x,t))V(x,t) + \frac{dv}{dt}(S(x),t).$$

Note that this formula also makes sense if $t = 0$ and $x \in \Gamma(0)$, whereas we required $t > 0$ in (2.6). We take this last formula as new definition for $\dot{u}$, that is, we set

$$\dot{u}(x,t) := (\nabla_x v(S(x),t)|N(x,t))V(x,t) + \frac{dv}{dt}(S(x),t), \quad (x,t) \in \mathcal{M}_0\,. \tag{2.8}$$

Finally, we set

$$\begin{aligned} L(\rho)(s,t) &:= |\nabla_x \Phi_\rho(x,t)|\big|_{x=X(s,\rho(s,t))}, \\ I(\rho,v)(s,t) &:= (\nabla_x v(S(x),t)|N(x,t))\big|_{x=X(s,\rho(s,t))}, \end{aligned} \tag{2.9}$$

for $(s,t) \in \Sigma \times [0,T)$ and we obtain

$$\dot{u}(x,t)\Big|_{x=X(s,\rho(s,t))} = \frac{dv}{dt}(s,t) + I(\rho,v)(s,t)V(x,t)\Big|_{x=X(s,\rho(s,t))}. \tag{2.10}$$

3. The Transformed Equations

Given an open set $U \subset \mathbb{R}^n$, let $h^s(U)$ denote the little Hölder spaces of order $s > 0$, that is, the closure of $BUC^\infty(U)$ in $BUC^s(U)$, the latter space being the Banach space of all bounded and uniformly Hölder continuous functions of order s. If Σ is a (sufficiently) smooth submanifold of $\mathbb{R}^n$ then the spaces $h^s(\Sigma)$ are defined by means of a smooth atlas for Σ. It is known that $BUC^t(\Sigma)$ is continuously embedded in $h^s(\Sigma)$ whenever $t > s$. Moreover, the little Hölder spaces have the interpolation property

$$(h^s(\Sigma), h^t(\Sigma))_\theta = h^{(1-\theta)s+\theta t}(\Sigma), \quad \theta \in (0,1), \tag{3.1}$$

whenever $s, t, (1-\theta)s + \theta t \in \mathbb{R}^+ \setminus \mathbb{N}$, and where $(\cdot,\cdot)_\theta$ denotes the continuous interpolation method of DaPrato and Grisvard [**4**], see also [**1, 2, 13**].

In the following we fix $t \in (0,T)$ and drop it in our notation. Given $\alpha \in (0,1)$ and $k \in \mathbb{N}$ we set

$$\begin{aligned} U(k,\alpha) &:= \{\rho \in h^{k+\alpha}(\Sigma)\,;\, \|\rho\|_{C(\Sigma)} < a\} \\ \mathbb{U}(k,\alpha) &:= U(k,\alpha) \times h^{k+\alpha}(\Sigma). \end{aligned} \tag{3.2}$$

Clearly, the sets $U(k,\alpha)$ and $\mathbb{U}(k,\alpha)$ are open in $h^{k+\alpha}(\Sigma)$ and in $(h^{k+\alpha}(\Sigma))^2$, respectively. Given $\rho \in U(k,\alpha)$, we introduce the mapping

$$\theta_\rho : \Sigma \to \mathbb{R}^n, \qquad \theta_\rho(s) := X(s,\rho(s)) \text{ for } s \in \Sigma\,, \qquad \rho \in U.$$

It follows that θ_ρ is a well-defined $(k+\alpha)$-diffeomorphism from Σ onto $\Gamma_\rho := \mathrm{im}(\theta_\rho)$. Let

$$\theta_\rho^* u := u \circ \theta_\rho \quad \text{for} \quad u \in C(\Gamma_\rho), \quad \theta_*^\rho v := v \circ \theta_\rho^{-1} \quad \text{for} \quad v \in C(\Sigma),$$

be the pull-back and the push-forward operator, respectively. Given $\rho \in U(k,\alpha)$, $k \geq 2$, we denote by Δ_{Γ_ρ} and H_{Γ_ρ} the Laplace–Beltrami operator and the mean curvature, respectively, of Γ_ρ. Finally we set

$$\Delta_\rho := \theta_\rho^* \Delta_{\Gamma_\rho} \theta_*^\rho, \qquad H(\rho) := \theta_\rho^* H_{\Gamma_\rho}.$$

We will now consider the smoothness properties the substitution operators induced by the local functions f and g, and of the operators L and I introduced in (2.9). Moreover we investigate the structure of the transformed operators Δ_ρ and $H(\rho)$.

For this, we introduce the set $\mathcal{H}(E_1, E_0)$ of generators of analytic semigroups. To be more precise, we assume that E_0 and E_1 are Banach spaces such that E_1

is densely injected in E_0, and we use the symbol $\mathcal{H}(E_1, E_0)$ to denote the set of all linear operators $A \in \mathcal{L}(E_1, E_0)$ such that $-A$ is the generator of a strongly continuous analytic semigroup on E_0. It is known that $\mathcal{H}(E_1, E_0)$ is an open subset of $\mathcal{L}(E_1, E_0)$, which will be given the relative topology of $\mathcal{L}(E_1, E_0)$.

LEMMA 3.1. *Assume that $\alpha \in (0,1)$ and $k \in \mathbb{N}$.*

(a) $[v \mapsto (f(v), g(v))] \in C^\infty(h^{k+\alpha}(\Sigma), h^{k+\alpha}(\Sigma) \times h^{k+\alpha}(\Sigma))$.

(b) $[\rho \mapsto L(\rho)] \in C^\infty(U(k+1,\alpha), h^{k+\alpha}(\Sigma))$.

(c) $[(\rho, v) \mapsto I(\rho, v)] \in C^\infty(\mathbb{U}(k+1,\alpha), h^{k+\alpha}(\Sigma))$.

(d) *There exists a function*

$$C \in C^\infty(U(k+2,\alpha), \mathcal{H}(h^{k+2+\alpha}(\Sigma), h^{k+\alpha}(\Sigma)))$$

such that $\Delta_\rho v = -C(\rho)v$ for $(\rho, v) \in \mathbb{U}(k+2,\alpha)$.

(e) *There exist functions*

$$P \in C^\infty(U(k+1,\alpha), \mathcal{H}(h^{k+2+\alpha}(\Sigma), h^{k+\alpha}(\Sigma))),$$
$$K \in C^\infty(U(k+1,\alpha), h^{k+\alpha}(\Sigma))$$

such that $H(\rho) = P(\rho)\rho + K(\rho)$ for $\rho \in U(k+2,\alpha)$. Furthermore,

(f) $[\rho \mapsto L(\rho)P(\rho)] \in C^\infty(U(k+1,\alpha), \mathcal{H}(h^{k+2+\alpha}(\Sigma), h^{k+\alpha}(\Sigma)))$.

PROOF. This follows by similar arguments as in the proofs of [**14**, Lemmas 3.1–3.3], and of [**5**, Section 2]. □

We are now ready to investigate the transformed system of equations

$$\begin{aligned} \frac{d\rho}{dt} &= -L(\rho)P(\rho)\rho - L(\rho)K(\rho) - L(\rho)f(v), && \rho(0) = \rho_0, \\ \frac{dv}{dt} &= \Delta_\rho v + \Big(I(\rho, v) + H(\rho)v - v\Big)\Big(H(\rho) + f(v)\Big) + g(v), && v(0) = v_0. \end{aligned} \tag{3.3}$$

In the following, we call (ρ, v) a smooth solution of (3.3) on $[0, T)$ if

$$(\rho, v) \in C^1(\Sigma \times [0,T), \mathbb{R}^2) \cap C^\infty(\Sigma \times (0,T), \mathbb{R}^2), \tag{3.4}$$

and if (ρ, v) satisfies system (3.3).

LEMMA 3.2. *(1.1) and (3.3) are equivalent: Smooth solutions of (1.1) give rise to smooth solutions of (3.3), and vice-versa.*

PROOF. This can be proved similarly as in [**14**, Lemma 4.1]. □

4. Existence and Uniqueness of Smooth Solutions

THEOREM 4.1. *Let $\mathbb{V} := \mathbb{U}(2,\alpha)$. Given any $w_0 := (\rho_0, v_0) \in \mathbb{V}$ there exists a number $T = T(w_0) > 0$ such that system (3.3) has a unique maximal smooth solution*

$$(\rho(\cdot, w_0), v(\cdot, w_0)) \in C([0,T), \mathbb{V}) \cap C^1([0,T), h^\alpha(\Sigma) \times h^\alpha(\Sigma)) \cap C^\infty(\Sigma \times (0,T), \mathbb{R}^2).$$

The map $[w_0 \mapsto (\rho(\cdot, w_0), v(\cdot, w_0))]$ defines a smooth semiflow on $\mathbb{V}$.

PROOF. It follows from [**14**, Theorem 4.3] that (3.3) has a unique maximal solution

$$(\rho(\cdot,w_0),v(\cdot,w_0))\in C([0,T),\mathbb{V})\cap C^1([0,T),h^\alpha(\Sigma)). \tag{4.1}$$

Moreover, [**14**, Equation (4.7)] shows that the solution has the additional smoothness property

$$(\rho(\cdot,w_0),v(\cdot,w_0))\in C^\infty((0,T),h^{2+\alpha}(\Sigma)\times h^{2+\alpha}(\Sigma)). \tag{4.2}$$

Let $T_0\in(0,T)$ be fixed and choose $\tau\in[0,T_0)$. We consider the linear parabolic equation

$$\frac{d\rho}{dt}+A(t)\rho=F(t),\quad \tau<t\le T_0,\quad \rho(\tau)=\rho(\tau,w_0), \tag{4.3}$$

on $h^{1+\alpha}(\Sigma)$, with

$$\begin{aligned}A(t)&:=L(\bar\rho(t))P(\bar\rho(t)),\\ F(t)&:=-L(\bar\rho(t))K(\bar\rho(t))-L(\bar\rho(t))f(\bar v(t)),\end{aligned}$$

for $t\in[\tau,T_0]$, where $\bar\rho(t):=\rho(t,w_0)$ and $\bar v(t):=v(t,w_0)$. If follows from (4.1) that $\rho(\tau,w_0)\in h^{2+\alpha}(\Sigma)$. Moreover, (4.1) and Lemma 3.1 with $k=1$ yield

$$(A,F)\in C([\tau,T_0],\mathcal{H}(h^{3+\alpha}(\Sigma),h^{1+\alpha}(\Sigma))\times h^{1+\alpha}(\Sigma)). \tag{4.4}$$

Let $X_0:=h^{1+\gamma}(\Sigma)$ and $X_1:=h^{3+\gamma}(\Sigma)$ for some fixed $\gamma\in(0,\alpha)$. It follows with the same arguments as in Lemma 3.1 that $A(t)\in\mathcal{H}(X_1,X_0)$ for $t\in[\tau,T_0]$. Next, note that the interpolation result (3.1) implies that

$$X_\theta:=(X_0,X_1)_\theta\doteq h^{1+\alpha}(\Sigma)\quad\text{if}\quad\theta=(\alpha-\gamma)/2\,,$$

where $\doteq$ indicates that the spaces are equal, except for equivalent norms. Let $A_\theta(t)$ denote the maximal X_θ-realization of $A(t)$, where $A(t)$ is considered as an operator in $\mathcal{L}(X_1,X_0)$, and let $X_{1+\theta}(A(t))$ denote its domain, equipped with the graph norm. Using

$$A(t)\in\mathcal{H}(h^{3+\alpha}(\Sigma),h^{1+\alpha}(\Sigma))\text{ and }A_\theta(t)\in\mathcal{H}(X_{1+\theta}(A(t)),X_\theta),$$

we readily infer that

$$X_{1+\theta}(A(t))\doteq X_{1+\theta}(A(\tau))\doteq h^{3+\alpha}(\Sigma)\quad\text{for}\quad t\in[\tau,T_0].$$

It follows from the maximal regularity result [**1**, Remark III.3.4.2.(c)], from (3.1) and [**1**, Theorem III.2.3.3] with $E_0:=h^{1+\alpha}(\Sigma)$, $E_1:=h^{3+\alpha}(\Sigma)$ and $(\rho,\mu)=(0,1/2)$, and from [**1**, Proposition III.2.1.1] that equation (4.3) admits a unique solution

$$\rho\in C([\tau,T_0],h^{2+\alpha}(\Sigma))\cap C((\tau,T_0],h^{3+\alpha}(\Sigma))\cap C^1((\tau,T_0],h^{1+\alpha}(\Sigma)). \tag{4.5}$$

It is a consequence of (4.5) and of (4.3) that ρ satisfies

$$\rho\in C([\tau,T_0],h^{2+\alpha}(\Sigma))\cap C^1([\tau,T_0],h^\alpha(\Sigma)),$$

so that ρ has at least the same regularity as $\rho(\cdot,w_0)$. Moreover, ρ solves the same equation on $h^\alpha(\Sigma)$ as $\rho(\cdot,w_0)$ for $t\in[\tau,T_0]$, and we conclude that $\rho=\rho(\cdot,w_0)|_{[\tau,T_0]}$. Since τ and T_0 can be chosen arbitrarily we obtain

$$\rho(\cdot,w_0)\in C((0,T),h^{3+\alpha}(\Sigma))\cap C^1((0,T),h^{1+\alpha}(\Sigma)). \tag{4.6}$$

Now we use (4.6) to show that $v(\cdot, w_0)$ also enjoys better regularity properties in the space variable than stated in (4.1). Let $\tau \in (0, T_0)$ be fixed and consider the linear parabolic equation

$$\frac{dv}{dt} + B(t)v = G(t), \quad \tau < t \le T_0, \quad v(\tau) = v(\tau, w_0), \tag{4.7}$$

on $h^{1+\alpha}(\Sigma)$, with

$$B(t) := C(\bar\rho(t)),$$
$$G(t) := \Big(I(\bar\rho(t), \bar v(t)) + H(\bar\rho(t))\bar v(t) - \bar v(t)\Big)\Big(H(\bar\rho(t)) + f(\bar v(t))\Big) + g(\bar v(t)),$$

for $t \in [\tau, T_0]$, where $\bar\rho(t) := \rho(t, w_0)$ and $\bar v(t) := v(t, w_0)$. It is a consequence of (4.1), (4.6), and of Lemma 3.1 with $k = 1$, that

$$(B, G) \in C([\tau, T_0], \mathcal{H}(h^{3+\alpha}(\Sigma), h^{1+\alpha}(\Sigma)) \times h^{1+\alpha}(\Sigma)), \tag{4.8}$$

and that $v(\tau, w_0) \in h^{2+\alpha}(\Sigma)$. As above we infer that (4.7) has a unique solution

$$v \in C([\tau, T_0], h^{2+\alpha}(\Sigma)) \cap C((\tau, T_0], h^{3+\alpha}(\Sigma)) \cap C^1((\tau, T_0], h^{1+\alpha}(\Sigma)). \tag{4.9}$$

This allows us to conclude, once again, that

$$v(\cdot, w_0) \in C((0, T), h^{3+\alpha}(\Sigma)) \cap C^1((0, T), h^{1+\alpha}(\Sigma)). \tag{4.10}$$

In a next step we use (4.6) and (4.10) to deduce that $\rho(\cdot, \rho_0)$ has more regularity than noted in (4.6). It should be observed that this time we need to choose $\tau \in (0, T_0)$, whereas $\tau = 0$ was admissible in (4.3)–(4.5). To be more precise, we consider

$$\frac{d\rho}{dt} + A(t)\rho = F(t), \quad \tau < t \le T_0, \quad \rho(\tau) = \rho(\tau, w_0),$$

as an evolution equation on $h^{2+\alpha}(\Sigma)$. It follows from (4.6), (4.10) and Lemma 3.1 with $k = 2$ that

$$(A, F) \in C([\tau, T_0], \mathcal{H}(h^{4+\alpha}(\Sigma), h^{2+\alpha}(\Sigma)) \times h^{2+\alpha}(\Sigma)),$$

and that $\rho(\tau, w_0) \in h^{3+\alpha}(\Sigma)$. We conclude by similar arguments as above—involving maximal regularity—that the solution of (4.3) satisfies

$$\rho \in C([\tau, T_0], h^{3+\alpha}(\Sigma)) \cap C((\tau, T_0], h^{4+\alpha}(\Sigma)) \cap C^1((\tau, T), h^{2+\alpha}(\Sigma)),$$

and that $\rho = \rho(\cdot, w_0)|_{[\tau, T_0]}$. Since τ and T_0 are arbitrary we get

$$\rho(\cdot, w_0) \in C((0, T), h^{4+\alpha}(\Sigma)) \cap C^1((0, T), h^{2+\alpha}(\Sigma)). \tag{4.11}$$

We can repeat the arguments and we arrive, after m steps, to the conclusion

$$(\rho(\cdot, w_0), v(\cdot, w_0)) \in C((0, T), (h^{m+2+\alpha}(\Sigma))^2) \cap C^1((0, T), (h^{m+\alpha}(\Sigma))^2). \tag{4.12}$$

Let $j \in \mathbb{N}$ be a number such that $2j \le m$. Then one also obtains

$$(\rho(\cdot, w_0), v(\cdot, w_0)) \in C^j((0, T), (h^{m-2j+\alpha}(\Sigma))^2). \tag{4.13}$$

In order to show (4.13), let us assume that we have already verified that

$$(\rho(\cdot, w_0), v(\cdot, w_0)) \in C^{j-1}((0, T), (h^{m-2(j-1)+\alpha}(\Sigma))^2)$$

for some j in $\{2, \cdots, m\}$. Then it follows from Lemma 3.1 with $k = m - 2j$ that

$$(A, F),\ (B, G) \in C^{j-1}((0, T), \mathcal{L}(h^{m+2-2j+\alpha}(\Sigma), h^{m-2j+\alpha}(\Sigma)) \times h^{m-2j+\alpha}(\Sigma)).$$

Now we go back to the evolution equation (4.3) and (4.7), respectively, and conclude that

$$\left(\frac{d}{dt}\rho(\cdot,w_0),\frac{d}{dt}v(\cdot,w_0)\right)\in C^{j-1}((0,T),(h^{m-2j+\alpha}(\Sigma))^2),$$

and hence $(\rho(\cdot,w_0),v(\cdot,w_0))\in C^j((0,T),(h^{m-2j+\alpha}(\Sigma))^2)$. Since $m\in\mathbb{N}$ can be chosen arbitrarily we have proved that

$$(\rho(\cdot,w_0),v(\cdot,w_0))\in C^\infty((0,T),C^\infty(\Sigma)\times C^\infty(\Sigma)). \tag{4.14}$$

This completes the proof of Theorem 4.1 □

REMARKS 4.2. **(a)** The strategy for the bootstrapping arguments in the proof of Theorem 4.1 relies on the following observation: The equation for ρ in (3.3), while coupled, is quasilinear in ρ and involves no derivatives of v. Therefore, if we insert $v(\cdot,w_0)$ into the first equation, we can take advantage of the regularizing effect to establish more regularity for $\rho(\cdot,w_0)$. As the equation for v is also quasilinear once ρ is frozen, we can now use that $\rho(\cdot,w_0)$ has more regularity to improve the regularity of $v(\cdot,w_0)$. These steps can then be repeated.

(b) The bootstrapping arguments used in the proof of Theorem 4.1 could also be based on [**1**, Theorem II.1.2.1]. Indeed, it follows from equations (3.1), (4.1), and from [**1**, Proposition II.1.1.2] that

$$(\rho(\cdot,w_0),v(\cdot,w_0))\in C^{1-\theta}([0,T),h^{2+\gamma}(\Sigma)) \text{ if } \gamma\in(0,\alpha) \text{ and } \theta=1-(\alpha-\gamma)/2\,.$$

A slightly modified version of Lemma 3.1 then yields

$$(A,F)\in C^{1-\theta}([0,T),\mathcal{H}(h^{3+\gamma}(\Sigma),h^{1+\gamma}(\Sigma))\times h^{1+\gamma}(\Sigma)).$$

Theorem II.1.2.1 in [**1**] shows that the solution of the linear parabolic equation (4.3) has better regularity properties than stated in (4.1). One can then go on and reiterate the arguments.

(c) It is important to note that system (1.1) or (3.3), respectively, is fully nonlinear. This indicates that one needs maximal regularity results—which compensate for the loss of derivatives—in order to get a solution via a fixed point argument. This has been achieved in [**14**].

(d) Theorem 1.1 allows to construct solutions even if they are not represented as graphs over the initially fixed reference manifold Σ. Indeed, we can take $\Gamma(t_1)$ for some $t_1\in[0,T)$ as new reference manifold and then get solutions on a time interval $[t_1,t_2]$. Thus we are not restricted to hypersurfaces which are graphs over a fixed manifold.

PROOF OF THEOREM 1.1. Let Γ_0 be a given compact, closed $C^{2+\beta}$-manifold in $\mathbb{R}^n$. As in Section 2 we find a smooth reference manifold Σ and a function $\rho_0\in C^{2+\beta}(\Sigma)$ such that

$$\Gamma_0=\mathrm{im}([s\mapsto X(s,\rho_0(s))])\,.$$

Since $C^{2+\beta}(\Sigma)\subset h^{2+\alpha}(\Sigma)$ for $\alpha\in(0,\beta)$ we also have that $\rho_0\in U(2,\alpha)$. Given $u_0\in C^{2+\beta}(\Gamma_0)$, let $v_0:\Sigma\to\mathbb{R}$ be defined by $v_0(s):=u_0(X(s,\rho_0(s)))$ for $s\in\Sigma$. We can conclude that $v_0\in h^{2+\alpha}(\Sigma)$. Theorem 4.1 yields the existence of a unique solution

$$(\rho(\cdot,w_0),v(\cdot,w_0))\in C([0,T),\mathbb{V})\cap C^1([0,T),h^\alpha(\Sigma)\times h^\alpha(\Sigma))\cap C^\infty(\Sigma\times(0,T))$$

for system (3.3), where we have set $w_0=(\rho_0,v_0)$. Clearly, this solution also satisfies the regularity assumptions in (3.4). Lemma 3.2 then shows that (1.1) has a

classical solution on $[0, T)$. The solution is unique in the class (4.1), as follows from Theorem 4.1, and the proof of Theorem 1.1 is now completed. □

Acknowledgment. We are grateful to Paul Fife for sharing with us his insight into DIGM.

References

1. H. Amann, *Linear and Quasilinear Parabolic Problems. Vol. I: Abstract Linear Theory*, Birkhäuser Verlag, Basel, 1995.
2. S. B. Angenent, *Nonlinear analytic semiflows*, Proc. Roy. Soc. Edinburgh Sect. A **115** (1990), no. 1-2, 91–107.
3. J. W. Cahn, P. C. Fife, and O. Penrose, *A phase-field model for diffusion-induced grain-boundary motion*, Acta mater. **45** (1997), no. 10, 4397–4413.
4. G. Da Prato and P. Grisvard, *Equations d'évolution abstraites non linéaires de type parabolique*, Ann. Mat. Pura Appl. (4) **120** (1979), 329–396.
5. J. Escher, U. F. Mayer, and G. Simonett, *The surface diffusion flow for immersed hypersurfaces*, SIAM J. Math. Anal. **29** (1998), no. 6, 1419–1433.
6. J. Escher and G. Simonett, *Classical solutions for Hele-Shaw models with surface tension*, Adv. Differential Equations **2** (1997), no. 4, 619–642.
7. ———, *A center manifold analysis for the Mullins-Sekerka model*, J. Differential Equations **143** (1998), no. 2, 267–292.
8. ———, *The volume preserving mean curvature flow near spheres*, Proc. Amer. Math. Soc. **126** (1998), no. 9, 2789–2796.
9. P. C. Fife, *Private communication*, (1998).
10. M. Gage and R. S. Hamilton, *The heat equation shrinking convex plane curves*, J. Differential Geom. **23** (1986), no. 1, 69–96.
11. M. A. Grayson, *The heat equation shrinks embedded plane curves to round points*, J. Differential Geom. **26** (1987), no. 2, 285–314.
12. G. Huisken, *Flow by mean curvature of convex surfaces into spheres*, J. Differential Geom. **20** (1984), no. 1, 237–266.
13. A. Lunardi, *Analytic Semigroups and Optimal Regularity in Parabolic Problems*, Birkhäuser Verlag, Basel, 1995.
14. U. F. Mayer and G. Simonett, *Classical solutions for diffusion-induced grain-boundary motion*, submitted (1998).

Department of Mathematics, Vanderbilt University, Nashville, TN 37240, USA
E-mail address: mayer@math.vanderbilt.edu

Department of Mathematics, Vanderbilt University, Nashville, TN 37240, USA
E-mail address: simonett@math.vanderbilt.edu

Contemporary Mathematics
Volume **238**, 1999
B 0-8218-1196-7-03551-9

Local Estimates for Solutions to Singular and Degenerate Quasilinear Parabolic Equations

Mike O'Leary

1. Introduction and Results

We shall obtain $L_{q,loc}(\Omega_T)$ and $L_{\infty,loc}(\Omega_T)$ estimates for a class of equations modeled after

$$u_t - \operatorname{div}(|\nabla u|^{p-2}\nabla u) = f(x,t) + \operatorname{div}\mathbf{g}. \tag{1}$$

If $p > 2$ the equation is degenerate, while if $p < 2$ the problem is singular. In particular, we shall study solutions of equations of the form

$$u_t - \operatorname{div} a(x,t,u,\nabla u) = b(x,t,u,\nabla u) \tag{2}$$

on domains $\Omega_T = \Omega \times (0,T)$ where $\Omega \subset \mathbf{R}^N$ and the equation satisfies the following structure conditions for each $(x,t,u,\mathbf{v}) \in \Omega \times (0,T) \times \mathbf{R} \times \mathbf{R}^N$

(H1) $1 < p \le \delta < p\left(\frac{N+2}{N}\right) \equiv m$, $c_i \ge 0$ for $0 \le i \le 5$, $c_0 > 0$, and $\phi_j \ge 0$ for $0 \le j \le 2$,

(H2) $a(x,t,u,\mathbf{v})\cdot\mathbf{v} \ge c_o|\mathbf{v}|^p - c_3|u|^\delta - \phi_o(x,t)$,

(H3) $|a(x,t,u,\mathbf{v})| \le c_1|\mathbf{v}|^{p-1} + c_4|u|^{\delta(1-\frac{1}{p})} + \phi_1(x,t)$,

(H4) $|b(x,t,u,\mathbf{v})| \le c_2|\mathbf{v}|^{p(1-\frac{1}{\delta})} + c_5|u|^{\delta-1} + \phi_2(x,t)$,

(H5) $\phi_1 \in L_{\frac{p}{p-1},loc}(\Omega_T)$,

(H6) $\phi_o \in L_{\mu,loc}(\Omega_T)$ with $\mu > 1$, and $\phi_1, \phi_2 \in L_{s,loc}(\Omega_T)$ with $s > \frac{m}{m-1}$,

while on the solution u we assume

(H7) For every $0 \le t_1 < t_2 \le T$ and for every $\Omega' \Subset \Omega$

$$\operatorname*{ess\,sup}_{t_1<t<t_2} \int_{\Omega'} |u(x,t)|^2\,dx + \int_{t_1}^{t_2}\int_{\Omega'} |\nabla u|^p\,dx\,dt < \infty,$$

(H8) $u \in L_{r,loc}(\Omega_T)$ for some $r > \frac{N}{p}(2-p)$.

By a weak solution of (2) we mean a function u that satisfies H8 and for which

$$\iint_{\Omega_T} \{-u\psi_t + a(x,t,u,\nabla u)\cdot\nabla\psi\}\,dx\,dt = \iint_{\Omega_T} b(x,t,u,\nabla u)\psi\,dx\,dt \tag{3}$$

for all $\psi \in C_0^\infty(\Omega_T)$.

Our main result is the following.

1991 *Mathematics Subject Classification.* Primary: 35K65, 35B45.

THEOREM 1. *Let u be a weak solution of (2), and suppose that H1-H8 are satisfied.*

If $\min\{s,\mu\} > (N+p)/p$, then $u \in L_{\infty,loc}(\Omega_T)$;

if $\min\{s,\mu\} = (N+p)/p$, then $u \in L_{q,loc}(\Omega_T)$ for all $q < \infty$;

if $\min\{s,\mu\} < (N+p)/p$, then $u \in L_{q,loc}(\Omega_T)$ for all $q < q^*$, where

$$q^* = \min\left\{ \frac{m-(1+\frac{p}{N})}{1-(1-\frac{1}{s})(1+\frac{p}{N})}, \frac{m}{1-(1-\frac{1}{\mu})(1+\frac{p}{N})} \right\}. \tag{4}$$

Moreover, the resulting bounds are independent of $\|\phi_1\|_{L_{\frac{p}{p-1},loc}(\Omega_T)}$.

Regularity properties of solutions of these types of equations have been extensively studied; an excellent reference is the book of DiBenedetto [**5**]. More specifically, Hölder continuity of solutions was proven in the degenerate case by DiBenedetto and Friedman [**6, 7**], while in the singular case by Y.Z. Chen and DiBenedetto in [**3, 4**]. Local boundedness of solutions under appropriate structure conditions was proven by Porzio [**14**] and these results have been extended to equations with more general structure in [**1, 8, 9, 11, 12, 15, 17, 18**].

The results contained in this paper have the following new features. First, to the best of this author's knowledge, this is the only result which yields information about the degree of local integrability of solutions which are not necessarily bounded. Secondly, this result extends the class of equations for which the local boundedness of solutions is guaranteed. Indeed, for the case $p > \frac{2N}{N+2}$, in [**5**, Chp. 5, Thm 3.1] boundedness of solutions was proven only if

$$\phi_1^{\frac{p}{p-1}}, \phi_2^{\frac{\delta}{\delta-1}} \in L_{s,loc}(\Omega_T) \qquad \text{for } s > \frac{N+p}{p}. \tag{5}$$

In the case $p \le \frac{2N}{N+2}$, local boundedness was proven in [**5**, Chp. 5, Thm. 5.1] only if the problem had homogeneous structure, meaning (H2), (H3) and (H4) are replaced by the requirements $a(x,t,u,\mathbf{v})\cdot\mathbf{v} \ge c_o|\mathbf{v}|^p$, $|a(x,t,u,\mathbf{v})| \le c_1|\mathbf{v}|^{p-1}$ and $b(x,t,u,\mathbf{v}) = 0$; moreover further global information was required, to the effect that the solution could be approximated weakly in $L_{r,loc}(\Omega_T)$ by bounded solutions. Only under these additional conditions, now no longer necessary, was boundedness proven.

We remark that the results of this note are almost optimal in the sense that they almost agree with the results of the linear case ($p = 2$). In particular, in [**10**, Chp. 3, Secs. 8,9] it is shown that solutions of linear problems of the form

$$u_t - \{a^{ij}(x,t)u_{x_j} + a^i(x,t)u\}_{x_i} + b^i(x,t)u_{x_i} + a(x,t)u = \phi(x,t) + \phi^i_{x_i} \tag{6}$$

when $\phi \in L_{s,loc}(\Omega_T)$ and $\phi^i \in L_{\mu,loc}(\Omega_T)$ are in $L_\infty(\Omega_T)$ when $\min\{s,\mu\} > (N+p)/p$, while they are in $L_{q,loc}(\Omega_T)$ for all $q < \infty$ if $\min\{s,\mu\} = (N+p)/p$, and are in $L_{q^*,loc}(\Omega_T)$ otherwise, where q^* is the number in Theorem 1 with $p = 2$.

A few comments on our hypotheses are now in order. The assumption H5 is made only to ensure that terms of the form $a(x,t,u,\nabla u)\cdot\nabla u$ are integrable. This information is needed only qualitatively and the resulting bounds are independent of $\|\phi_1\|_{\frac{p}{p-1}}$. The restriction on s in H6 is exactly that which is needed to ensure that $q^* > m$; recall that H7 and the Sobolev embedding theorem will imply that $u \in L_{m,loc}(\Omega_T)$. Finally, it is noted in [**5**] that the requirement H8 is necessary to prove boundedness of the solutions.

2. Proof of the $L_{q,loc}(\Omega_T)$ Estimates for $q < \infty$

The first step in our proof is the following local energy estimate.

PROPOSITION 2. *Suppose that u is a solution of (2) and that H1-H8 are satisfied. Then for any $Q_R(x_o,t_o) \equiv B_R(x_o)\times(t_o-R^p,t_o) \Subset \Omega_T$, for any $0<\sigma<1$, and for any $k>0$ we have*

$$\begin{aligned}\left[\iint_{Q_{\sigma R}}(u\mp k)_\pm^m\,dx\,dt\right]^{\frac{1}{1+p/N}} &\le \frac{\gamma}{(1-\sigma)^pR^p}\iint_{Q_R}(u\mp k)_\pm^2\,dx\,dt\\ &+\frac{\gamma}{(1-\sigma)^pR^p}\iint_{Q_R}(u\mp k)_\pm^p\,dx\,dt+\gamma\iint_{Q_R}|u|^\delta\chi[(u\mp k)_\pm>0]\,dx\,dt\\ &+\gamma\left[\frac{\|\phi_1\|_{L_s(Q_R)}}{(1-\sigma)R}+\|\phi_2\|_{L_s(Q_R)}\right]\left[\iint_{Q_R}(u\mp k)_\pm^{\frac{s}{s-1}}dx\,dt\right]^{1-\frac{1}{s}}\\ &+\gamma\|\phi_o\|_{L_\mu(Q_R)}(\mathrm{meas}[(u\mp k)_\pm>0])^{1-\frac{1}{\mu}}\end{aligned}\tag{7}$$

where γ depends only on c_i, N, p, δ, s and μ, but is independent of k.

This is a standard result proven by using a smooth cutoff approximation of $(u\mp k)_\pm$ as a testing function; for details see [**13**] or [**5**, Chp. 5, Prop. 6.1].

Our plan is to start with the assumption that $u\in L_{\beta,loc}(\Omega_T)$ for some $\beta\ge m$. We shall then estimate (7) in terms of $\|u\|_{L_\beta(Q_R)}$ and powers of k. This will give us an estimate of the form $|u|_{L^{\mathrm{weak}}_{\alpha(\beta)}(Q_{\sigma R})}\le C$ for some function $\alpha(\beta)$, which will give us our $L_{q,loc}(\Omega_T)$ estimates for $q<\infty$.

Indeed, recall that a measurable function u is an element of $L_q^{\mathrm{weak}}(\mathcal{U})$ if and only if

$$|u|^q_{L_q^{\mathrm{weak}}}\equiv\sup_{k>0}k^q\,\mathrm{meas}[|u|>k]<\infty.\tag{8}$$

Moreover, $L_q(\mathcal{U})\subset L_q^{\mathrm{weak}}(\mathcal{U})\subset L_{q'}(\mathcal{U})$ for all $q'<q$ provided $\mathcal{U}$ is bounded. More details about the spaces $L_q^{\mathrm{weak}}(\mathcal{U})$ can be found in [**2**, Chp. 1] or [**16**, IX.4].

As a consequence, our knowledge that $u\in L^{\mathrm{weak}}_{\alpha(\beta),loc}(\Omega_T)$ lets us conclude that $u\in L_{q,loc}(\Omega_T)$ for all $q<\alpha(\beta)$. Repeating this process then tells us that $u\in L_{q,loc}(\Omega_T)$ for all $q<(\alpha\circ\alpha\circ\cdots\circ\alpha)(\beta)$. Carefully calculating $\alpha(\beta)$ and iterating shall then give us our $L_{q,loc}(\Omega_T)$ estimates.

Indeed, estimate the left side of (7) with the $(u-k)_+$ choice as

$$\iint_{Q_{\sigma R}}(u-k)_+^m\,dx\,dt\ge k^m\operatorname*{meas}_{Q_{\sigma R}}[u>2k].\tag{9}$$

On the other hand, if $\theta<\beta$ then

$$\begin{aligned}\iint_{Q_R}(u-k)_+^\theta\,dx\,dt&\le\left[\iint_{Q_R}(u-k)_+^\beta\,dx\,dt\right]^{\frac{\theta}{\beta}}(\mathrm{meas}[u>k])^{1-\frac{\theta}{\beta}}\\ &\le\|u\|^\theta_{L_\beta(Q_R)}\left[\frac{1}{k^\beta}|u|^\beta_{L^{\mathrm{weak}}_\beta(Q_R)}\right]^{1-\frac{\theta}{\beta}}\le\|u\|^\beta_{L_\beta(Q_R)}\left(\frac{1}{k}\right)^{\beta-\theta}.\end{aligned}\tag{10}$$

Using this in (7) then gives us an estimate of the form

$$\left(k^m \operatorname*{meas}_{Q_{\sigma R}}[u>2k]\right)^{\frac{1}{1+p/N}} \le \gamma\|u\|^{\beta}_{L_\beta(Q_R)}\left[\left(\frac{1}{k}\right)^{\beta-2}+\left(\frac{1}{k}\right)^{\beta-p}+\left(\frac{1}{k}\right)^{\beta-\delta}\right] + \gamma\|u\|^{\beta(1-\frac{1}{s})}_{L_\beta(Q_R)}\left(\frac{1}{k}\right)^{\beta(1-\frac{1}{s})-1} + \gamma\|u\|^{\beta(1-\frac{1}{\mu})}_{L_\beta(Q_R)}\left(\frac{1}{k}\right)^{\beta(1-\frac{1}{\mu})} \tag{11}$$

for a γ that also depends on σ, R, $\|\phi_o\|_{L_\mu(Q_R)}$ and $\|\phi_1,\phi_2\|_{L_s(Q_R)}$. If we repeat this process for $(u+k)_-$, we obtain the estimate

$$\operatorname*{meas}_{Q_{\sigma R}}[u>k] \le \gamma\Bigg[\left(\frac{1}{k}\right)^{(\beta-2)(1+\frac{p}{N})+m}+\left(\frac{1}{k}\right)^{(\beta-\delta)(1+\frac{p}{N})+m} + \left(\frac{1}{k}\right)^{[\beta(1-\frac{1}{s})-1](1+\frac{p}{N})+m}+\left(\frac{1}{k}\right)^{\beta(1-\frac{1}{\mu})(1+\frac{p}{N})+m}\Bigg] \tag{12}$$

for all $k\ge 1$, where γ now also depends on $\|u\|_{L_\beta(Q_R)}$. As a consequence

$$|u|_{L^{\text{weak}}_{\alpha(\beta)}(Q_{\sigma R})} \le C \tag{13}$$

where

$$\alpha_1(\beta)=(\beta-2)\left(1+\frac{p}{N}\right)+m, \quad \alpha_3(\beta)=\left[\beta\left(1-\frac{1}{s}\right)-1\right]\left(1+\frac{p}{N}\right)+m,$$
$$\alpha_2(\beta)=(\beta-\delta)\left(1+\frac{p}{N}\right)+m, \quad \alpha_4(\beta)=\beta\left(1-\frac{1}{\mu}\right)\left(1+\frac{p}{N}\right)+m,$$

and

$$\alpha(\beta)=\min\{\alpha_1(\beta),\alpha_2(\beta),\alpha_3(\beta),\alpha_4(\beta)\}. \tag{14}$$

For the iteration, we start by setting

$$\beta_o=\max\{2,m,r\} \tag{15}$$

because the Sobolev embedding theorem and our hypotheses guarantee that $u\in L_{\beta_o,loc}(\Omega_T)$. We shall analyze the sequence of iterations $(\alpha\circ\alpha\circ\cdots\circ\alpha)(\beta_o)$ by cases.

Case 1: α_1. Because we can rewrite $\alpha_1(\beta)$ as

$$\alpha_1(\beta)=\left(1+\frac{p}{N}\right)\beta+(p-2), \tag{16}$$

we see that $\alpha_1(\beta)>\beta$ if and only if

$$\beta>\frac{N}{p}(2-p). \tag{17}$$

As a consequence if $\beta_o>\frac{N}{p}(2-p)$, then the sequence $\beta_o,\alpha_1(\beta_o),\alpha_1(\alpha_1(\beta_o)),\cdots$ will tend to infinity. Indeed, the above shows that the sequence is monotone increasing, so if it tended to a finite limit, that limit would be a fixed point of α_1 greater than β_0. Since there are no such fixed points, we can conclude that the sequence tends to infinity.

That the requirement $\beta_o>\frac{N}{p}(2-p)$ is satisfied is an immediate consequence of H7 and the fact that $\beta\ge r$.

Case 2: α_2. Now $\alpha_2(\beta) > \beta$ if and only if

$$\beta > \delta - \frac{N}{p}(m-\delta). \tag{18}$$

Then because $\beta_o \geq m > \delta > \delta - \frac{N}{p}(m-\delta)$, we conclude that the sequence $\beta_o, \alpha_2(\beta_o), \alpha_2(\alpha_2(\beta_o)), \cdots$ tends to infinity for the same reasons as case 1.

Case 3: α_3. Here the situation is somewhat different. Since $m - (1 + \frac{p}{N}) > 0$, we see that the sequence $\beta_o, \alpha_3(\beta_o), \alpha_3(\alpha_3(\beta_o)), \cdots$ tends to infinity provided $(1 - \frac{1}{s})(1 + \frac{p}{N}) \geq 1$ or equivalently if $s \geq \frac{N+p}{p}$. If not, we see that $\alpha_3(\beta) > \beta$ if and only if

$$\beta < \frac{m - (1 + \frac{p}{N})}{1 - (1 - \frac{1}{s})(1 + \frac{p}{N})} \tag{19}$$

so that $\beta_o, \alpha_3(\beta_o), \alpha_3(\alpha_3(\beta_o)), \cdots$ tends to

$$q_s^* \equiv \frac{m - (1 + \frac{p}{N})}{1 - (1 - \frac{1}{s})(1 + \frac{p}{N})}. \tag{20}$$

Case 4: α_4. This is handled in much the same fashion as case 3. Indeed if $\mu \geq \frac{N+p}{p}$ then $\beta_o, \alpha_4(\beta_o), \alpha_4(\alpha_4(\beta_o)), \cdots$ tends to infinity; otherwise it tends to

$$q_\mu^* \equiv \frac{m}{1 - (1 - \frac{1}{\mu})(1 + \frac{p}{N})}. \tag{21}$$

Since $q^* = \min\{q_s^*, q_\mu^*\}$, we have our $L_{q,loc}(\Omega_T)$ estimates for $q < \infty$.

3. The $L_{\infty,loc}(\Omega_T)$ Estimates

The boundedness of the solutions shall now be proven using the usual DeGiorgi methods, coupled with an interpolation in the case when $m \leq 2$.

Indeed, let $Q_\rho(x_o, t_o) \Subset \Omega_T$, fix $0 < \sigma < 1$, and let $k > 0$ be chosen later. Then set

$$\rho_n = \sigma\rho + \frac{1-\sigma}{2^n}\rho, \qquad k_n = k\left(1 - \frac{1}{2^{n+1}}\right), \tag{22}$$

and let $Q^n = Q_{\rho_n}(x_o, t_o)$. We now apply (7) where we replace k by k_{n+1}, and Q_R by Q^n, and $Q_{\sigma R}$ by Q^{n+1}. This then gives us

$$\begin{aligned}\left[\iint_{Q^{n+1}} (u-k_{n+1})_+^m \, dx\, dt\right]^{\frac{1}{1+p/N}} &\leq \frac{\gamma 2^{np}}{(1-\sigma)^p\rho^p} \iint_{Q^n} u - k_{n+1})_+^2 \, dx\, dt \\ &+ \frac{\gamma 2^{np}}{(1-\sigma)^p\rho^p} \iint_{Q^n} (u-k_{n+1})_+^p \, dx\, dt + \gamma \iint_{Q^n} |u|^\delta \chi[u > k_{n+1}] \, dx\, dt \\ &+ \frac{\gamma 2^n}{(1-\sigma\rho)} \left[\iint_{Q^n} (u-k_{n+1})_+^{\frac{s}{s-1}} dx\, dt\right]^{1-\frac{1}{s}} + \gamma(\text{meas}\, A_{n+1})^{1-\frac{1}{\mu}}\end{aligned} \tag{23}$$

where

$$A_{n+1} = \{(x,t) \in Q^n : u(x,t) > k_{n+1}\}. \tag{24}$$

Note that

$$\text{meas}\, A_{n+1} \leq \left(\frac{2^{n+2}}{k}\right)^\theta \iint_{Q^n} (u-k_n)_+^\theta \, dx\, dt \tag{25}$$

for each $\theta \geq 1$; this follows from the fact that

$$\iint_{Q^n} (u-k_n)_+^\theta \, dx\, dt \geq (k_{n+1}-k_n)^\theta \operatorname{meas} A_{n+1}. \tag{26}$$

What happens next depends on the parameter m.

Case 1: $m > 2$. In this instance we shall obtain an iterative inequality for

$$Y_n \equiv \fint\!\!\!\fint_{Q^n} (u-k_n)_+^m \, dx\, dt = \frac{1}{\operatorname{meas} Q^n} \iint_{Q^n} (u-k_n)_+^m \, dx\, dt \tag{27}$$

with the aid of (23). Indeed,

$$\begin{aligned} &\frac{\gamma 2^{np}}{(1-\sigma)^p \rho^p} \iint_{Q^n} (u-k_{n+1})_+^2 \, dx\, dt \\ &\qquad \leq \frac{\gamma 2^{np}}{(1-\sigma)^p \rho^p} \left[\iint_{Q^n} (u-k_{n+1})_+^m \, dx\, dt \right]^{\frac{2}{m}} (\operatorname{meas} A_{n+1})^{1-\frac{2}{m}} \\ &\qquad \leq \frac{\gamma 2^{np}}{(1-\sigma)^p \rho^p} \left(\frac{2^{n+2}}{k} \right)^{m-2} \iint_{Q^n} (u-k_n)_+^m \, dx\, dt \\ &\qquad \leq \frac{\gamma}{(1-\sigma)^p k^{m-2}} \rho^N 2^{(p+m-2)n} Y_n \end{aligned} \tag{28}$$

while similarly

$$\frac{\gamma 2^{np}}{(1-\sigma)^p \rho^p} \iint_{Q^n} (u-k_{n+1})^p \, dx\, dt \leq \frac{\gamma}{(1-\sigma)^p k^{m-p}} \rho^N 2^{mn} Y_n. \tag{29}$$

Further, because

$$\frac{u(x,t)}{u(x,t)-k_n} \leq \frac{k_{n+1}}{k_{n+1}-k_n} \tag{30}$$

for all $(x,t) \in [u > k_{n+1}]$, we have the estimate

$$\iint_{Q^n} |u|^\delta \chi[u > k_{n+1}] \, dx\, dt \leq \gamma \frac{1}{k^{m-\delta}} \rho^{N+p} 2^{mn} Y_n. \tag{31}$$

Finally, because $\frac{s}{s-1} \leq m$

$$\begin{aligned} &\frac{\gamma 2^n}{(1-\sigma)\rho} \left[\iint_{Q^n} (u-k_{n+1})_+^{\frac{s}{s-1}} \right]^{1-\frac{1}{s}} \\ &\qquad \leq \frac{\gamma}{(1-\sigma) k^{m\left(1-\frac{1}{s}\right)-1}} \rho^{(N+p)\left(1-\frac{1}{s}\right)-1} 2^{\left[m\left(1-\frac{1}{s}\right)-1\right]n} Y_n^{1-\frac{1}{s}}, \end{aligned} \tag{32}$$

and

$$\gamma \left(\operatorname{meas} A_{n+1}\right)^{1-\frac{1}{\mu}} \leq \frac{\gamma}{k^{m\left(1-\frac{1}{\mu}\right)}} \rho^{(N+p)\left(1-\frac{1}{\mu}\right)} 2^{m\left(1-\frac{1}{\mu}\right)n} Y_n^{1-\frac{1}{\mu}}. \tag{33}$$

Combining these results gives us the estimate

$$
\begin{aligned}
\rho^{N+p} Y_{n+1} \le & \frac{\gamma \rho^{N+p} 2^{(p+m-2)\left(1+\frac{p}{N}\right)n}}{(1-\sigma)^{p+N} k^{(m-2)\left(1+\frac{p}{N}\right)}} Y_n^{1+\frac{p}{N}} \\
& + \frac{\gamma \rho^{N+p} 2^{(p+m)\left(1+\frac{p}{N}\right)n}}{(1-\sigma)^{p+N} k^{(m-p)\left(1+\frac{p}{N}\right)}} Y_n^{1+\frac{p}{N}} \\
& + \frac{\gamma \rho^{(N+p)\left(1+\frac{p}{N}\right)} 2^{m\left(1+\frac{p}{N}\right)n}}{k^{(m-\delta)\left(1+\frac{p}{N}\right)}} Y_n^{1+\frac{p}{N}} \\
& + \frac{\gamma \rho^{\left[(N+p)\left(1-\frac{1}{s}\right)-1\right]\left(1+\frac{p}{N}\right)} 2^{\left[m\left(1-\frac{1}{s}\right)-1\right]\left(1+\frac{p}{N}\right)n}}{(1-\sigma)^{1+\frac{p}{N}} k^{\left[m\left(1-\frac{1}{s}\right)-1\right]\left(1+\frac{p}{N}\right)}} Y_n^{\left(1-\frac{1}{s}\right)\left(1+\frac{p}{N}\right)} \\
& + \frac{\gamma \rho^{(N+p)\left(1-\frac{1}{\mu}\right)\left(1+\frac{p}{N}\right)} 2^{m\left(1-\frac{1}{\mu}\right)\left(1+\frac{p}{N}\right)n}}{k^{m\left(1-\frac{1}{\mu}\right)\left(1+\frac{p}{N}\right)}} Y_n^{\left(1-\frac{1}{\mu}\right)\left(1+\frac{p}{N}\right)}.
\end{aligned} \tag{34}
$$

Then, because $m(1-\frac{1}{s})-1>0$, we find that there are constants A and B independent of n and k so that

$$
Y_{n+1} \le AB^n Y_n^{1+\frac{p}{N}} + AB^n Y_n^{\left(1-\frac{1}{s}\right)\left(1+\frac{p}{N}\right)} + AB^n Y_n^{\left(1-\frac{1}{\mu}\right)\left(1+\frac{p}{N}\right)}. \tag{35}
$$

Our assumptions on s and μ imply that $(1-\frac{1}{s})(1+\frac{p}{N})>1$ and $(1-\frac{1}{\mu})(1+\frac{p}{N})>1$, so that standard results on fast geometric convergence imply that $Y_n \to 0$ if Y_o is sufficiently small. By choosing k sufficiently large, we can make Y_o sufficiently small and guarantee that

$$
Y_\infty = \iint\!\!\!\!\!\!\int_{Q_{\sigma R}} (u-k)_+^m \, dx \, dt = 0, \tag{36}
$$

making u bounded above.

Similar considerations for $(u+k)_-$ show that u is bounded below.

Case 2: $m \le 2$. Let $\lambda > \max\{2, m\}$ be chosen later and set

$$
Y_n = \iint\!\!\!\!\!\!\int_{Q^n} (u-k_n)_+^\lambda \, dx \, dt; \tag{37}
$$

this is well defined thanks to our $L_{q,loc}(\Omega_T)$ estimates. Now for $\Lambda > \lambda > m$, the convexity inequality implies

$$
\begin{aligned}
Y_{n+1} &= \iint\!\!\!\!\!\!\int_{Q^{n+1}} (u-k_{n+1})_+^\lambda \, dx \, dt \\
&\le \frac{1}{\operatorname{meas} Q^{n+1}} \left(\iint_{Q^{n+1}} (u-k_{n+1})_+^\Lambda \, dx \, dt \right)^{\frac{\lambda}{\Lambda}\theta} \\
&\qquad \times \left(\iint_{Q^{n+1}} (u-k_{n+1})_+^m \, dx \, dt \right)^{\frac{\lambda}{m}(1-\theta)}
\end{aligned} \tag{38}
$$

where

$$
\theta = \frac{\frac{1}{m}-\frac{1}{\lambda}}{\frac{1}{m}-\frac{1}{\Lambda}} = \frac{\Lambda}{\lambda}\frac{\lambda-m}{\Lambda-m}. \tag{39}
$$

As a consequence,

$$\iint_{Q^{n+1}} (u-k_{n+1})_+^m \, dx\, dt \geq [\operatorname{meas} Q^{n+1}]^{\frac{\Lambda-m}{\Lambda-\lambda}} \frac{1}{\|u\|_{L_\Lambda(Q_\rho)}^{\frac{\lambda-m}{\Lambda-\lambda}\Lambda}} Y_{n+1}^{\frac{\Lambda-m}{\Lambda-\lambda}} \tag{40}$$

which estimates the left side of (23). The right side is estimated in the same fashion as case 1, so that

$$\frac{\gamma 2^{np}}{(1-\sigma)^p \rho^p} \iint_{Q^n} (u-k_{n+1})_+^2 \, dx\, dt \leq \frac{\gamma \rho^N 2^{(p+\lambda-2)n}}{(1-\sigma)^p k^{\lambda-2}} Y_n, \tag{41}$$

$$\frac{\gamma 2^{np}}{(1-\sigma)^p \rho^p} \iint_{Q^n} (u-k_{n+1})_+^p \, dx\, dt \leq \frac{\gamma \rho^N 2^{\lambda n}}{(1-\sigma)^p k^{\lambda-p}} Y_n, \tag{42}$$

$$\gamma \iint_{Q^n} |u|^\delta \chi[u>k_{n+1}] \, dx\, dt \leq \frac{\gamma \rho^{N+p} 2^{\lambda n}}{k^{\lambda-\delta}} Y_n, \tag{43}$$

further

$$\frac{\gamma 2^n}{(1-\sigma)\rho} \left[\iint_{Q^n} (u-k_{n+1})_+^{\frac{s}{s-1}} \, dx\, dt \right]^{1-\frac{1}{s}} \leq \frac{\gamma \rho^{(N+p)\left(1-\frac{1}{s}\right)} 2^{\left[\lambda\left(1-\frac{1}{s}\right)-1\right]n}}{(1-\sigma) k^{\lambda\left(1-\frac{1}{s}\right)-1}} Y_n^{1-\frac{1}{s}}, \tag{44}$$

and

$$\gamma \left(\operatorname{meas} A_{n+1}\right)^{1-\frac{1}{\mu}} \leq \frac{\gamma \rho^{(N+p)\left(1-\frac{1}{\mu}\right)} 2^{\lambda\left(1-\frac{1}{\mu}\right)n}}{k^{\lambda\left(1-\frac{1}{\mu}\right)}} Y_n^{1-\frac{1}{\mu}}. \tag{45}$$

If we make these substitutions, we shall find constants A and B independent of n and k so that

$$Y_{n+1} \leq AB^n Y_n^{\left(1+\frac{p}{N}\right)\left(\frac{\Lambda-\lambda}{\Lambda-m}\right)} + AB^n Y_n^{\left(1+\frac{p}{N}\right)\left(1-\frac{1}{s}\right)\left(\frac{\Lambda-\lambda}{\Lambda-m}\right)} + AB^n Y_n^{\left(1+\frac{p}{N}\right)\left(1-\frac{1}{\mu}\right)\left(\frac{\Lambda-\lambda}{\Lambda-m}\right)}. \tag{46}$$

Then because

$$\lim_{\Lambda\to\infty} \frac{\Lambda-\lambda}{\Lambda-m} = 1 \tag{47}$$

we can choose Λ and λ sufficiently large that both $(1+\frac{p}{N})(1-\frac{1}{s})(\frac{\Lambda-\lambda}{\Lambda-m}) > 1$ and $(1+\frac{p}{N})(1-\frac{1}{\mu})(\frac{\Lambda-\lambda}{\Lambda-m}) > 1$. The usual results on fast geometric convergence let us proceed as we did in case 1, giving us our result.

References

[1] D. Andreucci, *L^∞_{loc}-estimates for local solutions of degenerate parabolic equations*, SIAM J. Math. Anal. **22** (1991), no. 1, 138–145.

[2] J. Bergh and J. Löfström, *Interpolation spaces*, Springer–Verlag, New York, 1976.

[3] Y. Z. Chen and E. DiBenedetto, *On the local behavior of solutions of singular parabolic equations*, Arch. Rat. Mech. Anal. **103** (1988), no. 4, 319–345.

[4] ———, *Boundary estimates for solutions of nonlinear degenerate parabolic systems*, J. Reine Angew. Math. **395** (1989), 102–131.

[5] E. DiBenedetto, *Degenerate parabolic equations*, Springer-Verlag, New York, 1993.

[6] E. DiBenedetto and A. Friedman, *Regularity of solutions of non–linear degenerate parabolic systems*, J. Reine Angew. Math. **349** (1984), 83–128.

[7] ———, *Hölder estimates for non–linear degenerate parabolic systems*, J. Reine Angew. Math. **357** (1985), 1–22.
[8] A. V. Ivanov, *Maximum modulus estimates for generalized solutions to doubly nonlinear parabolic equations*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **221** (1995), 83–113.
[9] ———, *On the maximum principle for doubly nonlinear parabolic equations*, J. Math. Sci. **80** (1996), no. 6, 2236–2254.
[10] O. A. Ladyzhenskaya, V. A. Solonnikov, and N. N. Ural'ceva, *Linear and quasilinear equations of parabolic type*, Trans. AMS, vol. 23, American Mathematical Society, Providence, RI, 1968.
[11] X. T. Liang, *Estimation of the maximum modulus of solutions to doubly nonlinear parabolic systems*, Acta Sci. Natur. Univ. Sunyatseni **33** (1994), no. 4, 111–113.
[12] G. M. Lieberman, *Maximum estimates for solutions of degenerate parabolic equations in divergence form*, J. Diff. Eq. **113** (1994), 543–571.
[13] M. O'Leary, *Integrability and boundedness of solutions to singular and degenerate quasilinear parabolic equations*, Differential Integral Equations (to appear).
[14] M. M. Porzio, L^{∞}_{loc}*-estimates for degenerate and singular parabolic equations*, Nonlin. Anal. **17** (1991), no. 11, 1093–1107.
[15] ———, L^{∞}_{loc} *estimates for a class of doubly nonlinear parabolic equations with sources*, Rend. Mat. Appl. (7) **16** (1996), 433–456.
[16] M. Reed and B. Simon, *Fourier analysis, self-adjointness*, Methods of Modern Mathematical Physics, vol. 2, Academic Press, New York, 1975.
[17] L. Xiting and W. Zaide, *The a priori estimate of the maximum modulus to solutions of doubly nonlinear parabolic equations*, Comment. Math. Univ. Carolinae **38** (1997), no. 1, 109–119.
[18] ———, *Local boundedness for solutions of doubly nonlinear parabolic equations*, Rev. Mat. Univ. Complut. Madrid **10** (1997), no. 1, 179–190.

Department of Mathematics, Towson University, Towson, MD 21252

Contemporary Mathematics
Volume **238**, 1999
B 0-8218-1196-7-03552-0

The Geometry of Wulff Crystal Shapes and Its Relations with Riemann Problems

Danping Peng, Stanley Osher, Barry Merriman, and Hong-Kai Zhao

Abstract. In this paper we begin to explore the mathematical connection between equilibrium shapes of crystalline materials (Wulff shapes) and shock wave structures in compressible gas dynamics (Riemann problems). These are radically different physical phenomena, but the similar nature of their discontinuous solutions suggests a connection.

We show there is a precise sense in which any two dimensional crystalline form can be described in terms of rarefactions and contact discontinuities for an associated scalar hyperbolic conservation law. As a byproduct of this connection, we obtain a new analytical formula for crystal shapes in two dimension. We explore a possible extension to high dimensions.

We also formulate the problem in the level set framework and present a simple algorithm using the level set method to plot the approximate equilibrium crystal shape corresponding to a given surface energy function in two and three dimensions.

Our main motivation for establishing this connection is to encourage a transfer of theoretical and numerical techniques between the rich but disparate disciplines of crystal growth and gas dynamics. The work reported here represents a first step towards this goal.

1. Introduction

In this paper we develop a mathematical connection between two quite different physical phenomena: the shapes of crystalline materials, and dynamics of shock waves in a gas.

Both of these phenomena have long research histories: The problem of determining the equilibrium shape of a perfect crystal was posed and first solved by Wulff in 1901 [**28**]. In nature this ideal "Wulff shape" (see figure 1) is observed in crystals that are small enough to relax to their lowest energy state without becoming stuck in local minima.

The problem of determining the dynamics of a gas initialized with an arbitrary initial jump in state was posed and partially solved by Riemann in 1860 [**21**]. Solutions to this "Riemann problem" can be observed experimentally in shock tubes,

1991 *Mathematics Subject Classification.* Primary: 35L65, 49L25. Secondary: 52A39, 65M06.

Research supported by DARPA/NSF VIP grant, NSF grant DMS9615854 and NSF grant DMS9706827.

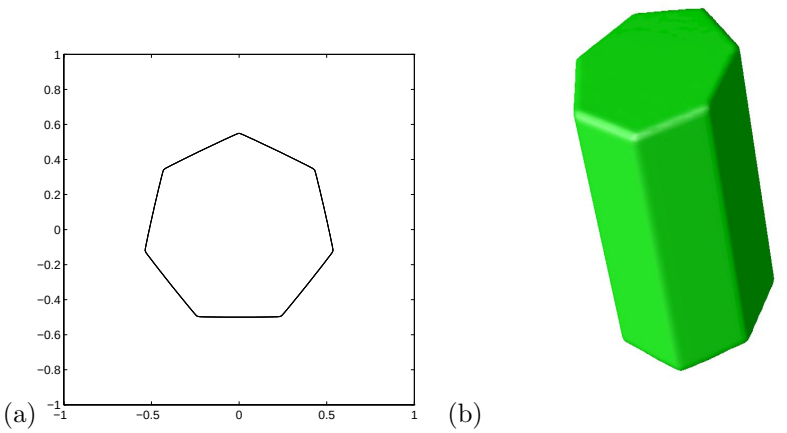

FIGURE 1. *(a) A 2D Wulff crystal. (b) A 3D Wulff crystal.*

where a membrane separating gases in different uniform states is rapidly removed. The Riemann problem has since been generalized to mean the solution of any system of hyperbolic conservation laws subject to initial data prepared with a single jump in state separating two regions with different constant states.

The intuitive link between Wulff crystals and Riemann problems is the similar nature of the discontinuous solutions. Crystalline shapes are characterized by perfectly flat faces—facets—separated by sharp edges, whereas shocked gases are characterized by regions of constant pressure separated by steep jumps (see figure 2). It is tempting to imagine that the sharp edges in a crystal shape can be though of as shock waves in some sense.

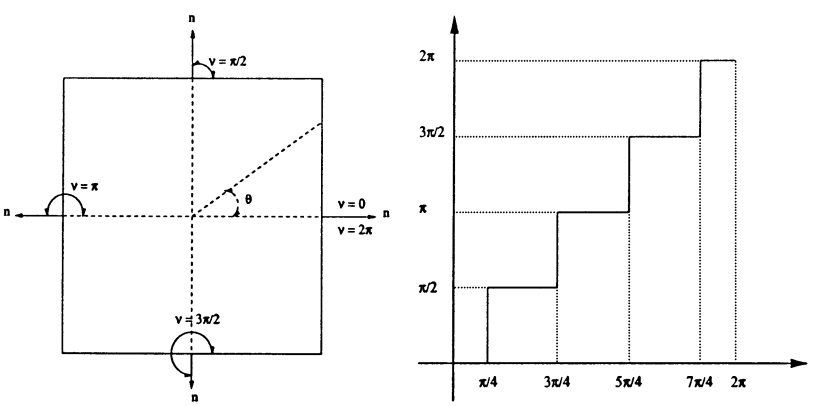

FIGURE 2. *The left is a 2D square Wulff crystal. The right is the plot of the angle between outward normal and the horizontal axis vs the polar angle.*

To explore this shock wave-crystal edge analogy more precisely, we represent the crystal surface in terms of its unit normal vector. The normal has regions of

constant direction separated by jumps in direction, which suggests it may satisfy the same sort of equations—hyperbolic conservation laws—that govern shocks in nonlinear gas dynamics. The analogy continues to hold if we consider the most general behavior of Wulff crystals and Riemann problems: Wulff shapes are constructed entirely from facets, rounded faces and sharp edges, for which the normal direction has regions of constancy, smooth variation, and isolated jumps. Correspondingly, the solution to a Riemann problem for any hyperbolic conservation law is constructed entirely from constant states, rarefactions (smooth variation), and shocks or contacts (isolated jumps).

We will show that the precise connection is this: the normal vector to the Wulff shape of a crystal in *two dimensions* is the time self-similar solution of an associated Riemann problem for a hyperbolic conservation law. In this representation, it does indeed turn out that crystal facets are the constant states in the Riemann problem and the curved faces are the rarefactions, but the sharp corners are contact discontinuities, not shocks.

The most immediate consequences of this representation is a new analytical formula for the Wulff shape, derived using formulas and methods from the theory of Riemann problems.

For clarity, we will summarize the analytical results here; these are explained in detail as they are derived in the main text. Let the Wulff shape be described in a polar coordinate system with the origin at its center. Choose the horizontal axis such that it intersects with the Wulff shape at a point where the unit normal to the Wulff shape coincides with the horizontal axis. (For example, the horizontal axis can be chosen as the line emanating from the origin and passing through the minima of the surface tension.) The boundary curve of the shape can be parameterized by giving the angle ν between the normal and the horizontal axis as a function of the polar angle θ. This curve $\nu(\theta)$ gives the time self-similar viscosity solution $\nu(\xi, t) = \nu(\xi/t)$ to the hyperbolic conservation law

$$\nu_t + F(\nu)_\xi = 0 \tag{1.1}$$

with flux function

$$F(\nu) = \frac{1}{2}\nu^2 + \int_0^\nu \tan^{-1}\left(\frac{\hat{\gamma}'(u)}{\hat{\gamma}(u)}\right) du \tag{1.2}$$

and initial data

$$\nu(\xi < 0, t = 0) = 0, \tag{1.3}$$

$$\nu(\xi > 0, t = 0) = 2\pi, \tag{1.4}$$

where $\gamma(\nu)$ is the crystalline surface tension as a function of the surface normal direction, $\hat{\gamma}(\nu)$ is the Frank convexification of $\gamma(\nu)$ (described in section 3.4) and $\tan^{-1}$ has range $[-\pi/2, \pi/2]$. The new formula for the Wulff shape that results from this connection is

$$\nu(\theta) = -\frac{d}{d\theta} \min_{0 \le \nu \le 2\pi} [F(\nu) - \theta\nu], \tag{1.5}$$

where F is the flux function from the conservation law.

The primary goal of this paper is to expose the connection between faceted crystal shapes and shock waves and related phenomena from gas dynamics. Because readers familiar with the theory of Wulff shapes come from a material science background, they are unlikely to know the theory of Riemann problems from the

field of gas dynamics, and vice versa. To fill in these likely gaps, we will present the elementary background for both problems prior to deriving our new results. Most of our proof will be somewhat formal, referring the mathematically inclined readers to relevant publications for rigorous treatment. In addition to making the present paper more readable, we hope this inclusive approach will foster future interaction between these two disparate research communities.

The paper is organized as follows: we start with the essential background on the Wulff problem and the Riemann problem, emphasizing their similarities. Then we show how to represent the Wulff shape as the solution to a Riemann problem, via two seemingly quite different approaches: the first approach starts from the Euler-Lagrange equation of the surface energy and connects it with the Riemann problem of a scalar conservation law under a suitable choice of variables. The other approach uses the self-similar growth property of the Wulff shape and shows that it solves a Riemann problem for the same conservation law. In the process, we develop the new formula (1.5). We present two simple illustrative examples and comment on further possible extensions of these ideas. We then formulate the Wulff problem in the level set setting and use it to derive some theoretical results about Wulff shape. This method is also a convenient and versatile tool for plotting the Wulff shape of a given surface tension function in both two and three dimensions. We present numerous examples demonstrating this and verify some recently obtained theoretical results in [**19**] concerning the Wulff shape in the numerical section. The appendix contains proofs of some results in the main text that require certain degrees of technicality.

2. The Wulff Problem and the Legendre Transformation

This section will briefly review and develop some general results about the Wulff crystal shape that are valid in any dimension. The next section will concentrate on the Wulff crystal shape in 2D.

2.1. The Formulation of Wulff Problem. The Wulff problem is to determine the equilibrium shape of a perfect crystal of one material in contact with a single surrounding medium. The equilibrium shape is determined by minimizing the total system energy, which is composed of contributions from the bulk and surface of the crystal. If we fix the bulk energy, the problem becomes that of finding a shape of given volume with minimal surface energy.

If the surface energy density—that is, the "surface tension"—is constant, the solution is the shape of minimal surface area, which is a circle in 2D and a sphere in 3D. However, in many solid materials the surface tension depends on how the surface is directed relative to the bulk crystalline lattice, due to the detailed structure of the bonding between atoms. Assuming some standard orientation of the bulk lattice, the surface tension, γ, will be a definite function of the normal vector to the surface, $\hat{n}$, say $\gamma = \gamma(\hat{n})$. In that case, if the material is bounded by a surface Γ, the total surface energy is

$$E = \int_{\Gamma} \gamma(\hat{n}) dA, \tag{2.1}$$

which must be minimized subject to the constraint of constant volume enclosed by Γ.

This problem makes sense both in two and three dimensions, and essentially 2D crystals do arise experimentally in the growth of thin films [**7**]. The formula we derive in this section applies equally well in both dimensions. In the next section, we will concentrate our attention on the 2D problem, where we can make the precise connection to a Riemann problem. In this case, $\gamma(\hat{n})$ is the energy per unit length on the boundary, and we we seek to determine the bounding curve, Γ, of minimal surface energy that encloses a given area.

2.2. Wulff's Geometric Construction of the Solution. Wulff presented the solution to this minimization problem as an ingenious geometric construction, based on the geometry of the surface tension. Let $\gamma : S^{d-1} \to R^+$ be the surface tension which is a continuous function, where $d = 2$ or 3. Wulff's construction is as follows (refer to figure 3):

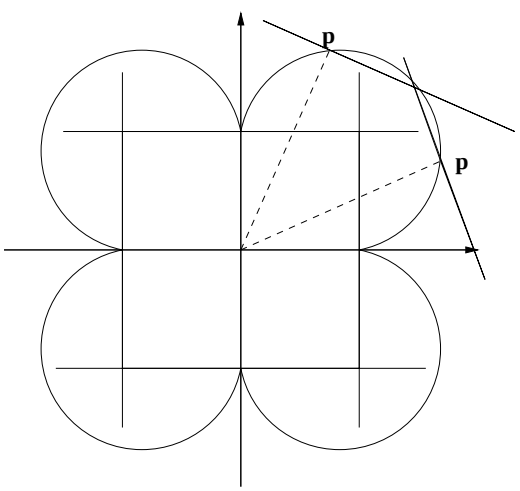

FIGURE 3. *Wulff's geometric construction.*

- **Step 1.** Construct a "polar plot" of $\gamma(\nu)$. In 2D, this is simply the curve defined in r-ν polar coordinates by $r = \gamma(\nu)$, $0 \leq \nu \leq 2\pi$. In 3D, this is a surface around the origin in sphere coordinates r-ν.
- **Step 2.** For each point P on the polar plot, construct the hyperplane through P and normal to the radial vector emanating from the origin to P. (Note this is typically *not* the tangent plane to the polar plot at P.)
- **Step 3.** Construct the inner (convex) envelope of this family of hyperplanes. This is the minimizing crystal *shape*, and rescaling it to have the proper volume yields the solution to the constrained problem.

We will call the geometric shape obtained through the above procedure Wulff crystal shape or simply Wulff shape. It is easy to see that the region enclosed by Wulff shape is

$$W = \{\mathbf{x} \in R^d : \mathbf{x} \cdot \theta \leq \gamma(\theta), \text{ for all } \theta \in S^{d-1}\}, \tag{2.2}$$

which is convex.

It is possible to write an analytic expression for the envelope of a family of smoothly parameterized lines or planes, and doing so yields formula (2.3) in any

dimension. In particular, we get a simple formula (3.1) in 2D. The stipulation from step 3 to use the "inner" envelope means that the multivalued swallowtails occurring in the envelope equations must be clipped off to obtain the true shape.

It is easy to see that the construction places facets in directions of local minima of surface tension, which is a sensible energy-reducing strategy. Indeed the entire process is simply to position a planar face at every possible orientation, with distance from the origin proportional to its energy, and then simply take the innermost set of facets as the crystal shape. However, it is difficult to prove rigorously why Wulff's construction gives the minimal energy shape. J. E. Taylor [**27**] and others [**2, 9**] have given general proofs that this construction does yield a minimizer of the energy, and this shape is unique up to translations. See also the recent paper [**19**] of Osher and Merriman.

2.3. The Wulff Shape and the Legendre Transformation. Wulff's geometric construction described above can be mathematically formalized by the use of the Legendre transformation, which we define below.

DEFINITION 1. *Let $\zeta : S^{d-1} \to R^+$ be a continuous function.*

1. *The first Legendre transformation of ζ is:*

$$\zeta_*(\nu) = \inf_{\substack{\theta\cdot\nu>0 \\ |\theta|=1}} \left[\frac{\zeta(\theta)}{(\theta\cdot\nu)}\right]. \tag{2.3}$$

2. *The second Legendre transformation of ζ is:*

$$\zeta^*(\nu) = \sup_{\substack{\theta\cdot\nu>0 \\ |\theta|=1}} \left[\zeta(\theta)(\theta\cdot\nu)\right]. \tag{2.4}$$

The geometric interpretation of Legendre transformation should be clear from figure 4 and the remarks below.

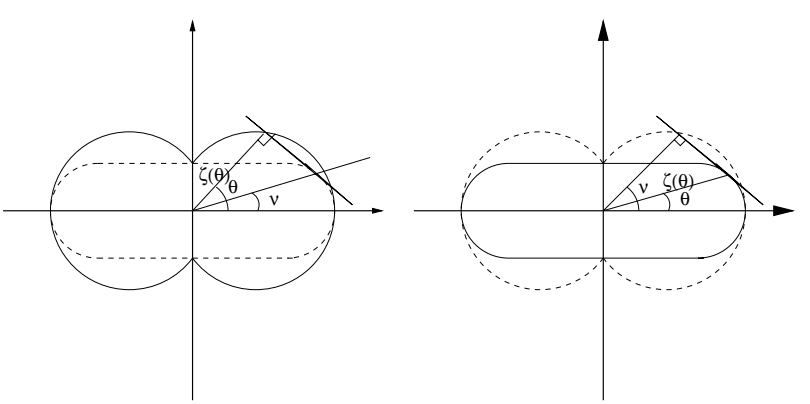

FIGURE 4. *Left: the first Legendre transformation. The solid line is the plot of ζ, and the dashed line is the plot of ζ_*, the corresponding Wulff shape. Right: second Legendre transformation. The solid line is the plot of ζ, and the dashed line is the plot of ζ^*, the support function.*

Remark 1. The first Legendre transformation $\zeta_*(\nu)$ gives the Wulff crystal shape. This is easy to see from equation (2.2) in polar coordinates:

$$\begin{aligned} W &= \{(r,\nu) : r(\nu\cdot\theta) \le \zeta(\theta), \text{ for all } \theta \in S^{d-1}\} \\ &= \{(r,\nu) : r \le \inf_{\substack{\theta\cdot\nu>0 \\ |\theta|=1}} \left[\frac{\zeta(\theta)}{(\theta\cdot\nu)}\right]\} \\ &= \{(r,\nu) : r \le \zeta_*(\nu)\}. \end{aligned}$$

Remark 2. The second Legendre transformation $\zeta^*(\nu)$ gives the support function of the region enclosed by the polar plot of ζ. Recall that the support function p_Ω of a bounded region Ω which contains the origin is defined by

$$p_\Omega(\nu) = \max\{\mathbf{x}\cdot\nu : \mathbf{x}\in\Omega\}, \text{ for } \nu \in S^{d-1}. \tag{2.5}$$

We will see shortly that the first and second Legendre transformation are dual to each other in certain sense. The following relations are obvious by definition:

LEMMA 1.

$$\zeta_*(\nu) \le \zeta(\nu) \le \zeta^*(\nu) \tag{2.6}$$

$$\tfrac{1}{\zeta^*} = \left(\tfrac{1}{\zeta}\right)_*, \ \tfrac{1}{\zeta_*} = \left(\tfrac{1}{\zeta}\right)^*. \tag{2.7}$$

Since ζ is defined on a curved manifold S^{d-1}, sometimes it is convenient to study the extension of ζ to the whole R^d. We extend $\zeta : S^{d-1} \to R^+$ to R^d by defining

$$\bar\zeta(x) = |x|\zeta(\frac{x}{|x|}), \text{ for } x \in R^d, \text{ with } \bar\zeta(0) = 0. \tag{2.8}$$

Such an extension $\bar\zeta$ is homogeneous of degree 1. If $\bar\zeta$ is differentiable, we have the following important relation due to Euler:

$$\sum_{j=1}^{n} x_j \frac{\partial\bar\zeta}{\partial x_j}(x) = \bar\zeta(x). \tag{2.9}$$

Note that each of the first partial derivatives of $\bar\zeta$ is homogeneous of degree 0, and the second partial derivatives are homogeneous of degree -1. We will abuse the notation and write $\bar\zeta$ as ζ when no ambiguity arise.

DEFINITION 2. *ζ is convex if the polar plot of $\frac{1}{\zeta}$ is convex. ζ is polar convex if the polar plot of ζ is convex.*

The following lemma gives a necessary and sufficient condition for ζ to be convex in terms of its extension.

LEMMA 2. *ζ is convex if and only if its homogeneous extension of degree 1 $\bar\zeta : R^d \to R^+$ is a convex function on R^d.*

See Appendix I for the proof.

As we have seen, the Wulff shape $W = \{\ x : |x| \le \zeta_*(\frac{x}{|x|})\ \}$ is always convex. By definition, ζ_* is polar convex. From lemma 1, ζ^* is convex. We put these facts in

LEMMA 3. *ζ_* is always polar convex and ζ^* is always convex.*

Now let us introduce

$$\hat{\zeta}(\nu) = (\zeta_*)^*(\nu) \ , \ \check{\zeta}(\nu) = (\zeta^*)_*(\nu). \tag{2.10}$$

From the definitions and the lemma above, we know that $\hat{\zeta}$ is always convex and $\check{\zeta}$ polar-convex.

DEFINITION 3. *We call $\hat{\zeta}(\nu)$ the Frank convexification of ζ, and $\check{\zeta}(\nu)$ the polar convexification of ζ.*

We proceed to prove the following important relations:

LEMMA 4.

$$\hat{\zeta}(\nu) \le \zeta(\nu) \le \check{\zeta}(\nu), \tag{2.11}$$

$$\left(\frac{1}{\zeta}\right)^{\wedge} = \frac{1}{\check{\zeta}}, \quad \left(\frac{1}{\zeta}\right)^{\vee} = \frac{1}{\hat{\zeta}}. \tag{2.12}$$

Proof: From the definition,

$$\begin{aligned} \hat{\zeta}(\nu) &= \sup_{\theta\cdot\nu>0} [\zeta_*(\theta)(\theta\cdot\nu)] = \sup_{\theta\cdot\nu}[(\inf_{\eta\cdot\theta>0} \frac{\zeta(\eta)}{(\eta\cdot\theta)}(\theta\cdot\nu)] \\ &\le \sup_{\theta\cdot\nu>0} [\frac{\zeta(\nu)}{(\nu\cdot\theta)}(\theta\cdot\nu)] = \zeta(\nu). \end{aligned}$$

The other inequality can be proved similarly.

The key ingredient in proving the two equalities is the repeated use of the duality relations (2.7):

$$\begin{aligned} \left(\frac{1}{\zeta}\right)^{\wedge} &= \left(\left(\frac{1}{\zeta}\right)_*\right)^* = \left(\frac{1}{\zeta^*}\right)^* = 1/\frac{1}{\left(\frac{1}{\zeta^*}\right)^*} \\ &= 1/\left(\frac{1}{\frac{1}{\zeta^*}}\right)_* = \frac{1}{(\zeta^*)_*} = \frac{1}{\check{\zeta}}. \end{aligned}$$

The other equality follows from the above and the duality relations (2.7). □

Both $\hat{\zeta}$ and $\check{\zeta}$ have simple geometric interpretations. From the definition, the steps used to obtain $\check{\zeta}$ can be described in words as following: draw the polar-plot Γ of ζ. Let Ω be the region enclosed by Γ. Through each point on Γ, draw the hyperplane(s) tangent to Γ which lie outside Ω. Note that such a plane may not exist, such as at points that curved inward, and may not be unique, such as at the points that bulge outward. This corresponds to the steps used in constructing the support function. Then find the inner envelope of all such tangent planes. This corresponds to the construction of the Wulff shape of the support function. The inner envelope is the smallest convex set that contains Ω, i.e. the convexification of Ω. We thus get a simple procedure to obtain $\check{\zeta}$: plot the graph of ζ in polar coordinates, convexify the plot and obtain a convex graph. This is the polar plot of $\check{\zeta}$. Using the duality relation (2.12), we can obtain $\hat{\zeta}$ by first drawing the polar plot of $\frac{1}{\zeta}$, then convexifying the region enclosed to get $\left(\frac{1}{\zeta}\right)^{\vee} = \frac{1}{\hat{\zeta}}$, then inverting it to get $\hat{\zeta}$. See the figures in section 5.5. These arguments show that

LEMMA 5. *If ζ is convex, then $\hat{\zeta} = \zeta$. If ζ is polar-convex, then $\check{\zeta} = \zeta$.*

We now show the following important

THEOREM 1. 1. *The Wulff shape of ζ and $\hat{\zeta}$ are the same. That is,*

$$\left(\hat{\zeta}\right)_* = \zeta_*. \tag{2.13}$$

2. *The support function of ζ and $\check{\zeta}$ are the same. That is,*

$$\left(\check{\zeta}\right)^* = \zeta^*. \tag{2.14}$$

Proof: We already know that $\hat{\zeta} \leq \zeta$. Hence $(\hat{\zeta})_* \leq \zeta_*$. On the other hand,

$$\hat{\zeta}(\theta) = \max_{\substack{\theta\cdot\nu>0 \\ |\nu|=1}} [\zeta_*(\nu)(\theta\cdot\nu)] \geq \zeta_*(\nu)(\nu\cdot\theta), \text{for all } \nu \in S^{d-1}.$$

Using this inequality and the definitions, we have:

$$(\hat{\zeta})_*(\nu) = \inf_{\substack{\theta\cdot\nu>0 \\ |\theta|=1}} \left[\frac{\hat{\zeta}(\theta)}{(\theta\cdot\nu)}\right] \geq \inf_{\substack{\theta\cdot\nu>0 \\ |\theta|=1}} \left[\frac{\zeta_*(\nu)(\nu\cdot\theta)}{(\theta\cdot\nu)}\right] = \zeta_*(\nu).$$

□

From the above discussion, we see that given a convex body $K \subset R^d$, the surface tension whose Wulff shape is K is not uniquely defined. However, if we require that the surface tension is convex, then it is uniquely determined by K. If the boundary of K is given by $r : S^{d-1} \to R^+$, then the convex surface tension function whose Wulff shape is K is given by the second Legendre transformation $\gamma = r^*$.

Now let us further assume that $\zeta : S^{d-1} \to R^+$ is a C^1 function. We extend ζ to R^d to be a homogeneous function of degree 1. Then the first Legendre transformation can be rewritten as

$$\zeta_*(\theta) = \inf_{\mathbf{x}\cdot\theta>0} \left[\frac{\zeta(\mathbf{x})}{\mathbf{x}\cdot\theta}\right] = \inf_{\mathbf{x}\cdot\theta>0} \zeta\left(\frac{\mathbf{x}}{\mathbf{x}\cdot\theta}\right). \tag{2.15}$$

Suppose the infimum is reached at certain $\mathbf{x} = \mathbf{x}(\theta) \in R^d$, and let $\mathbf{p} = \frac{\mathbf{x}}{\mathbf{x}\cdot\theta}$. We have

$$\begin{aligned}
0 &= \frac{\partial}{\partial x_i}\zeta\left(\frac{\mathbf{x}}{\mathbf{x}\cdot\theta}\right) = \sum_j \frac{\partial\zeta}{\partial p_j}\frac{\partial p_j}{\partial x_i} \\
&= \sum_j \frac{\partial\zeta}{\partial p_j}\left(\frac{\delta_{ij}}{\mathbf{x}\cdot\theta} - \frac{\theta_i x_j}{(\mathbf{x}\cdot\theta)^2}\right) \\
&= \frac{\partial\zeta}{\partial p_i}\frac{1}{\mathbf{x}\cdot\theta} - \left(\sum_j \frac{\partial\zeta}{\partial p_j}p_j\right)\frac{\theta_i}{\mathbf{x}\cdot\theta} \\
&= \frac{1}{\mathbf{x}\cdot\theta}\left[\frac{\partial\zeta}{\partial p_i} - \zeta(\mathbf{p})\theta_i\right],
\end{aligned}$$

using relation (2.9). Thus we have

$$\zeta(\mathbf{p})\theta_i = \frac{\partial\zeta}{\partial p_i}(\mathbf{p}).$$

Since $\mathbf{p} = \frac{\mathbf{x}}{\mathbf{x}\cdot\theta} = \frac{\hat{n}}{\hat{n}\cdot\theta}$, where $\hat{n} = \hat{n}(\theta)$ is the unit normal to the Wulff shape at $\mathbf{R} = \zeta_*(\theta)\theta$, and $\zeta_*(\theta) = \frac{\zeta(\hat{n})}{\hat{n}\cdot\theta}$, we get

$$\zeta_*(\theta)\theta = D\zeta(\hat{n}), \tag{2.16}$$

$$\zeta_*(\theta) = |D\zeta(\hat{n})|, \tag{2.17}$$

$$\theta = \frac{D\zeta(\hat{n})}{|D\zeta(\hat{n})|}. \tag{2.18}$$

For a given θ, the $\hat{n}$ determined by (2.18) may not be unique unless ζ is strictly convex.

Equation (2.16) gives us a convenient way to get the Wulff shape for a given surface tension function γ. We simply draw the surface (in 3D) or curve (in 2D) parameterized by $\hat{n}$. The surface or curve will generally self-intersect, except when γ is convex. We may clip off the intersecting part, and obtain the Wulff crystal shape. See next section for examples in 2D.

There are equally simple relations for the second Legendre transformation. From the duality relations (2.7), we have

$$\begin{aligned} \frac{1}{\zeta^*(\hat{n})} &= |D\frac{1}{\zeta}(\theta)|, \\ \hat{n} &= \frac{D\zeta^{-1}}{|D\zeta^{-1}|}(\theta). \end{aligned}$$

Note that here it is $1/\zeta$ that is extended as a homogeneous function of degree 1, and hence ζ itself is extended as a homogeneous function of degree -1. From the above relations, we get

$$\zeta^*(\hat{n})\hat{n} = -\frac{\zeta^2 D\zeta}{|D\zeta|^2}(\theta), \tag{2.19}$$

$$\zeta^*(\hat{n}) = \frac{\zeta^2}{|D\zeta|}(\theta), \tag{2.20}$$

$$\hat{n} = -\frac{D\zeta}{|D\zeta|}(\theta). \tag{2.21}$$

Let us look at two examples.

Example 1. In 2D, the surface tension function γ is usually given in term of the angle ν of normal $\hat{n}$ to a fixed horizontal axis, i.e. $\gamma = \gamma(\nu)$. We extend γ as a homogeneous function of degree 1 in the following way

$$\gamma(x, y) = \sqrt{x^2 + y^2}\gamma(\tan^{-1}\frac{y}{x}), \tag{2.22}$$

and easily get

$$\gamma_*(\theta)\hat{n}(\theta) = D\gamma(\nu) = \gamma(\nu)\hat{n}(\nu) + \gamma'(\nu)\hat{\tau}(\nu), \tag{2.23}$$

where

$$\begin{aligned} \hat{n}(\nu) &= (\cos\nu, \sin\nu), \\ \hat{\tau}(\nu) &= (-\sin\nu, \cos\nu). \end{aligned}$$

When γ is convex, that is when $\gamma+\gamma'' \geq 0$, equation (2.23) is a parameterization of the Wulff shape in term of ν and the first Legendre transformation of γ is given by

$$\gamma_*(\theta) = \sqrt{\gamma^2(\nu) + \gamma'^2(\nu)} \tag{2.24}$$

where ν is determined by

$$\theta = \nu + \tan^{-1}\left(\frac{\gamma'(\nu)}{\gamma(\nu)}\right). \tag{2.25}$$

See section 3 for details.

To find the second Legendre transformation, we extend γ to the whole space as a homogeneous function of degree -1 by defining

$$\gamma(x,y) = \frac{1}{\sqrt{x^2+y^2}}\gamma(\tan^{-1}\frac{y}{x}). \tag{2.26}$$

From the general relations (2.19)–(2.21), we obtain

$$\gamma^*(\nu)\hat{n}(\nu) = \frac{\gamma^2(\theta)}{\gamma^2(\theta)+\gamma'^2(\theta)}[\gamma(\theta)\hat{n}(\theta) - \gamma'(\theta)\hat{\tau}(\theta)], \tag{2.27}$$

$$\gamma^*(\nu) = \frac{\gamma^2(\theta)}{\sqrt{\gamma^2(\theta)+\gamma'^2(\theta)}}, \tag{2.28}$$

where θ is determined by

$$\nu = \theta - \tan^{-1}\left(\frac{\gamma'(\theta)}{\gamma(\theta)}\right). \tag{2.29}$$

Example 2. In 3D, suppose the surface tension function is given in terms of spherical coordinates, $\gamma = \gamma(\nu,\psi)$, where $0 \le \nu < 2\pi$ and $-\frac{\pi}{2} < \psi < \frac{\pi}{2}$. We extend γ to the whole space by defining

$$\gamma(x,y,z) = \sqrt{x^2+y^2+z^2}\,\gamma(\tan^{-1}\frac{y}{x}, \tan^{-1}\frac{z}{\sqrt{x^2+y^2}}), \tag{2.30}$$

and direct calculation gives:

$$D\gamma = \gamma(\nu,\psi)\hat{r} + \frac{1}{\cos\psi}\frac{\partial\gamma}{\partial\nu}\hat{\nu} + \frac{\partial\gamma}{\partial\psi}\hat{\psi}, \tag{2.31}$$

where

$$\begin{aligned} \hat{r} &= (\cos\psi\cos\nu, \cos\psi\sin\nu, \sin\psi), \\ \hat{\nu} &= (-\sin\nu, \cos\nu, 0), \\ \hat{\psi} &= (-\sin\psi\cos\nu, -\sin\psi\sin\nu, \cos\psi). \end{aligned}$$

Equation (2.31) is a parameterization of Wulff shape in terms of ν and ψ when γ is convex. The first Legendre transformation is given by

$$\gamma_*(\theta,\phi) = \sqrt{\gamma^2(\nu,\psi) + \left|\frac{1}{\cos\psi}\frac{\partial\gamma}{\partial\nu}\right|^2 + \left|\frac{\partial\gamma}{\partial\psi}\right|^2}, \tag{2.32}$$

where ν and ψ are implicitly defined by equation (2.18).

2.4. Growing a Wulff Crystal. Now we consider a surprising and interesting property of objects moving outward with normal velocity equal to the surface tension function γ. This is discussed in more detail in section 7.3 below.

In [**19**], Osher and Merriman proved a generalization of a conjecture made by Chernov [**4, 5**]. Namely, starting from any (not necessarily convex or even connected) region, if we grow it with normal velocity equal to (not necessarily convex) $\gamma(\nu) : S^{d-1} \to R^+$, where $\nu \in S^{d-1}$ is the unit normal direction, the region asymptotes to a single Wulff shape corresponding to the surface tension γ. This is not totally surprising, because if we start with a convex region whose

support function is $p(\nu)$ and move it outward with normal velocity $\gamma(\nu)$, with $\gamma(\nu)$ convex, the evolution of $p(\nu, t)$, the support function of the growing region at time t, satisfies:

$$\begin{cases} \frac{\partial p}{\partial t}(\nu, t) = \gamma(\nu), \\ p(\nu, 0) = p(\nu). \end{cases} \tag{2.33}$$

Thus the evolving region has support function

$$p(\nu, t) = p(\nu) + t\gamma(\nu),$$

so the growing region asymptotes to the Wulff shape associated with $\gamma(\nu)$. This argument is only valid for convex initial shape and convex γ. In particular, it shows that a growing Wulff shape under this motion just expands itself similarly, since $p(\nu) = \gamma(\nu)$ for a convex γ. This is also true for a nonconvex γ. See [**19**] and section 7.3 below for more details.

2.5. Typical Forms for Surface Tension. From Wulff's construction, we see that the crystalline form depends on the geometry of the polar plot of the surface tension. While Wulff's construction is valid for an arbitrary surface tension function, the polar plots of physically relevant surface tensions have several characteristic features. These are worth noting in order to appreciate the crystalline forms in nature and also in order to formulate representative examples.

A physical surface tension should have reflection symmetry, $\gamma(\hat{n}) = \gamma(-\hat{n})$. Further, it is known that modeling a crystalline material as a regular lattice of atoms with given bonding energies between neighbors necessarily leads to a continuum limit in which the polar plot of γ consists of portions of spheres (circles in 2D) passing through the origin [**10**]. In particular, a 2D plot consists of outward bulging circular arcs that meet at inward pointing cusps. A simple example coming from a square lattice (X-Y) model of a 2D crystal is $\gamma(\nu) = |\sin(\nu)| + |\cos(\nu)|$. The polar plot consists of four semicircular arcs arranged in a clover-leaf fashion. The cubic lattice (X-Y-Z) model of a 3D crystal is $\gamma(\hat{n}) = |\hat{n}_x| + |\hat{n}_y| + |\hat{n}_z|$. Its plot in spherical coordinates consists of eight spherical pieces in a similar fashion. Its Wulff shape is a cube. See figure 5.

Physical surface tensions depend on the material temperature as well, and increasing temperature tends to smooth out the cusps in the surface tension plot.

From Wulff's construction, we can see that each cusp in the γ plot will result in a facet on the crystal shape. Increasing the material temperature smoothes out the cusps, which in turn rounds out the original facets of the Wulff shape.

3. The Wulff Crystal Shape in 2D

In 2D, a unit vector can be represented by its angle to a horizontal axis. Many of the general results developed in the last section have interesting concrete expressions. Although we can obtain many of the results in this section by a simple change of variables and then apply the general results, as we did in the example above, we will see that 2D Wulff problem has its own fascinating properties which may be missed by this "general-to-special" approach. Instead, we will use the general results as a guideline and develop the 2D theory from the ground up. A comprehensive discussion of the 2D Wulff problem and related matters can be found in the book of Gurtin [**11**].

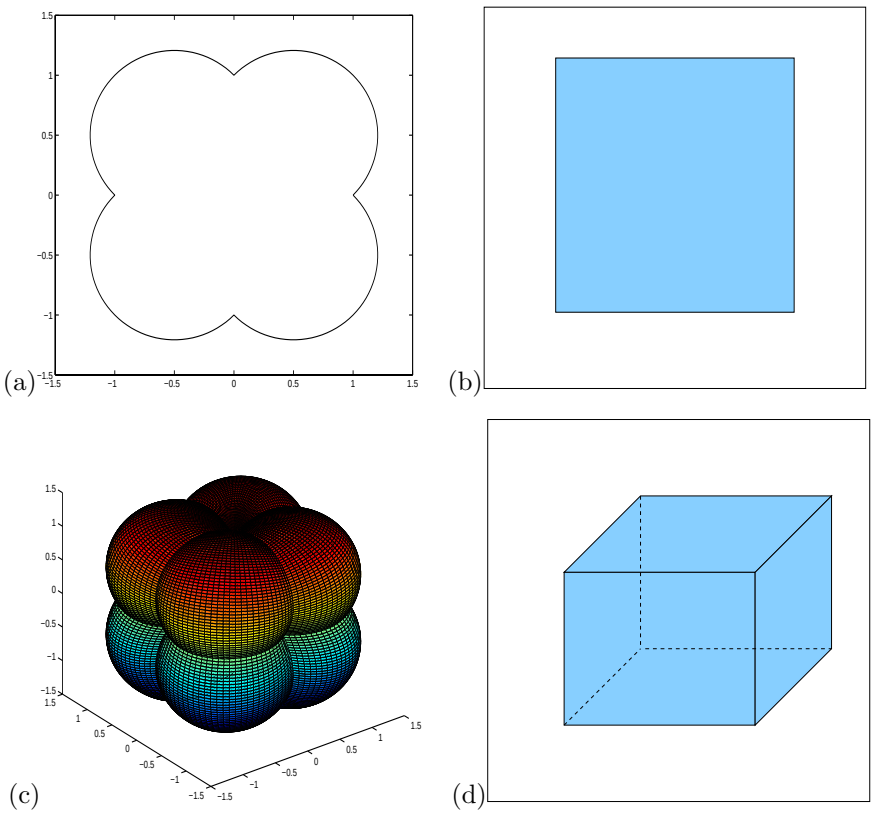

FIGURE 5. *(a) Plot of* $\gamma(\nu) = |\cos\nu| + |\sin\nu|$. *(b) Wulff shape of* γ *on the left. (c) Plot of* $\gamma(\hat{n}) = |\hat{n}_x| + |\hat{n}_y| + |\hat{n}_z|$. *(d) Wulff shape of* γ *on the left.*

Choosing the angle between the outward unit normal to the horizontal axis as parameter, the 2D version of the Legendre transformations of a function $\zeta : S^1 \to R^+$ are

$$\zeta_*(\theta) = \inf_{\theta-\frac{\pi}{2}<\nu<\theta+\frac{\pi}{2}} \left[\frac{\zeta(\nu)}{\cos(\theta-\nu)}\right], \tag{3.1}$$

$$\zeta^*(\theta) = \sup_{\theta-\frac{\pi}{2}<\nu<\theta+\frac{\pi}{2}} \left[\zeta(\nu)\cos(\theta-\nu)\right]. \tag{3.2}$$

3.1. The Legendre Transformation in 2D. We briefly review some basic facts about plane curves. For convenience, we will change the notation of the Legendre transformation in the section. Given $r : S^1 \to R^+$, a continuous function. Let $\mathbf{r} : S^1 \to R^2$ be the polar plot of r, i.e. $\mathbf{r}(\theta) = r(\theta)(\cos\theta, \sin\theta)$, and denote the resulting curve as Γ. The second Legendre transformation of r is the support function of Γ and will be denoted as p. On the other hand, given a positive continuous function $p : S^1 \to R^+$, its first Legendre transformation gives the Wulff shape and will be denoted as r.

We will use s to denote the arclength parameter of Γ, and θ the angle between $\mathbf{r}$ and the horizontal x-axis. Let $\hat{\tau} = \frac{d\mathbf{r}}{ds}$ be the tangent vector and $\hat{n}$ be the

outwards unit normal. Let $\nu \in [0, 2\pi)$ be the angle between $\hat{n}$ and x-axis.Then $\hat{n} = (\cos\nu, \sin\nu), \hat{\tau} = (-\sin\nu, \cos\nu)$.

The curvature of the curve Γ has a simple expression in term of ν:

$$\kappa = \frac{d\nu}{ds}. \tag{3.3}$$

Recall that the support function of the curve Γ is defined as

$$p(\nu) = \max_{\theta}\{\mathbf{r}(\theta) \cdot \hat{n}(\nu)\}. \tag{3.4}$$

Suppose the maximum is obtained at $\theta = \theta(\nu)$. Differentiate with respect to ν, we get:

$$p'(\nu) = \mathbf{r}(\theta) \cdot \hat{\tau}(\nu). \tag{3.5}$$

Note that $\mathbf{r} = (\mathbf{r} \cdot \hat{n})\hat{n} + (\mathbf{r} \cdot \hat{\tau})\hat{\tau}$. Combining the definition of p and the above equation, we get

$$\mathbf{r}(\theta) = p(\nu)\hat{n}(\nu) + p'(\nu)\hat{\tau}(\nu), \tag{3.6}$$

which gives us a simple way to express the curve if we know its support function.

Differentiating equation (3.5) with respect to ν gives

$$p''(\nu) = \frac{1}{\kappa} - p(\nu), \tag{3.7}$$

or equivalently:

$$\kappa = \frac{1}{p(\nu) + p''(\nu)}. \tag{3.8}$$

This gives us a convenient way to express the curvature of a curve given its support function.

Recall from the last section that for a positive function on S^1 to be a support function, it must be convex in the sense that the polar plot of its reciprocal be convex. The curvature of the polar plot of $1/p$ is easily shown to be

$$\kappa = \frac{p^3(p + p'')}{(p^2 + p'^2)^{3/2}}. \tag{3.9}$$

Thus p is convex if and only if $p + p'' \geq 0$.

From the above results, we can find explicit formulae for the first and second Legendre transformations in 2D. Given $p : S^1 \to R^+$, its first Legendre transformation $r(\theta) = p_*(\theta)$ is simply:

$$r(\theta) = \sqrt{p^2(\nu) + p'^2(\nu)}. \tag{3.10}$$

To determine ν for a given θ, let

$$\alpha = \tan^{-1}\left(\frac{p'(\nu)}{p(\nu)}\right). \tag{3.11}$$

We have

$$\begin{aligned} \mathbf{r}(\theta) &= r(\theta)\left[\frac{p(\nu)}{r(\theta)}\hat{n}(\nu) + \frac{p'(\nu)}{r(\theta)}\hat{\tau}(\nu)\right] \\ &= r(\theta)(\cos(\alpha + \nu), \sin(\alpha + \nu)). \end{aligned}$$

So

$$\theta = \nu + \tan^{-1}\left(\frac{p'(\nu)}{p(\nu)}\right), \tag{3.12}$$

which implicitly defines ν for a given θ. To ensure the inverse exists, we need

$$\frac{\partial\theta}{\partial\nu} = \frac{p(p+p'')}{p^2+p'^2} \geq 0,$$

or equivalently

$$p(\nu) + p''(\nu) \geq 0. \tag{3.13}$$

That is to say, p has to be convex.

On the other hand, suppose we are given $r : S^1 \to R^+$. To find its second Legendre transformation, i.e. its support function p, we note that

$$\mathbf{r}'(\theta) = r'(\theta)\hat{n}(\theta) + r(\theta)\hat{\tau}(\theta).$$

Define

$$\beta = \tan^{-1}\left(\frac{r'(\theta)}{r(\theta)}\right). \tag{3.14}$$

We have the following expression for the tangent vector at $\mathbf{r}(\theta)$

$$\begin{aligned}\hat{\tau}(\nu) &= \frac{\mathbf{r}'(\theta)}{|\mathbf{r}'(\theta)|} = \sin\beta\hat{n}(\theta) + \cos\beta\hat{\tau}(\theta)\\ &= (-\sin(\theta-\beta), \cos(\theta-\beta)).\end{aligned}$$

Thus

$$\nu = \theta - \tan^{-1}\left(\frac{r'(\theta)}{r(\theta)}\right), \tag{3.15}$$

which determines θ for a given ν. To ensure that the inverse exists, we require

$$\frac{\partial\nu}{\partial\theta} = \frac{r^2+2|r'|^2-r''r}{r^2+|r'|^2} = \frac{R(R+R'')}{R^2+|R'|^2} \geq 0,$$

or equivalently

$$R(\theta) + R''(\theta) \geq 0,$$

where $R = \frac{1}{r}$. This is equivalent to that r is polar convex.

The support function is found to be

$$\begin{aligned}p(\nu) &= \mathbf{r}(\theta)\cdot\hat{n}(\nu) = r(\theta)\hat{n}(\theta)\cdot\hat{n}(\nu)\\ &= r(\theta)\cos(\theta-\nu) = r(\theta)\cos(\beta)\\ &= \frac{r^2(\theta)}{\sqrt{r^2(\theta)+r'^2(\theta)}},\end{aligned}$$

and

$$\begin{aligned}\mathbf{p}(\nu) &= p(\nu)(\cos(\theta-\beta), \sin(\theta-\beta))\\ &= \frac{r^2(\theta)}{r^2(\theta)+r'^2(\theta)}\left[r(\theta)\hat{n}(\theta) - r'(\theta)\hat{\tau}(\theta)\right],\end{aligned}$$

where θ is defined by equation (3.15).

We summarize the results in this section in the following

THEOREM 2. 1. *Given $p : S^1 \to R^+$, a continuous piecewise differentiable convex function, its first Legendre transformation is*

$$r(\theta) = \sqrt{p^2(\nu)+p'^2(\nu)}$$

and

$$\mathbf{r}(\theta) = p(\nu)\hat{n}(\nu) + p'(\nu)\hat{\tau}(\nu),$$

where ν is determined by

$$\theta = \nu + \tan^{-1}\left(\frac{p'(\nu)}{p(\nu)}\right)$$

for a given θ.

2. *Given $r : S^1 \to R^+$, a continuous piecewise differentiable polar convex function, its second Legendre transformation is*

$$p(\nu) = \frac{r^2(\theta)}{\sqrt{r^2(\theta) + r'^2(\theta)}}$$

and

$$\mathbf{p}(\nu) = \frac{r^2(\theta)}{r^2(\theta) + r'^2(\theta)}\left[r(\theta)\hat{n}(\theta) - r'(\theta)\hat{\tau}(\theta)\right],$$

where θ is determined by

$$\nu = \theta - \tan^{-1}\left(\frac{r'(\theta)}{r(\theta)}\right)$$

for a given ν.

3.2. The Euler-Lagrange Equation. We can apply standard variational calculus to obtain the Euler-Lagrange equations for the minimizing boundary curve in the 2D problem. For our purposes, these equations are best expressed when the curve is parameterized in terms of the angle, ν, between its unit normal vector and some fixed axis, as a function of arc length s along the curve, as was done in the previous section. The curve is completely specified by $\nu(s)$.

In this parameterization, the Euler-Lagrange equation become

$$(\gamma(\nu) + \gamma''(\nu))\nu_s = \lambda, \tag{3.16}$$

where λ is the constant Lagrange multiplier associated with the volume constraint. It is worth noting that $\nu_s = \kappa$, the curvature of the curve; in particular, this shows that when the surface tension is constant, the solution is of constant curvature, i.e. a circle.

For the derivation of the Euler-Lagrange equation of the surface energy, which contains (3.16) as a special case, see Appendix III.

3.3. Multivalued Solutions. Equation (3.6) give us a simple ways to obtain the Wulff shape given the surface tension function γ. We can represent the Wulff shape of γ using ν as parameter:

$$\mathbf{x}(\nu) = \gamma(\nu)\hat{n}(\nu) + \gamma'(\nu)\hat{\tau}(\nu). \tag{3.17}$$

This is true only when γ is convex, i.e. $\gamma + \gamma'' \geq 0$. In this case, the curve Γ defined by $\mathbf{x}(\nu)$ is convex. When the convexity condition fails, the curve will self-intersect and thus have swallowtails. See figure 6. This can be easily seen by noting that

$$\mathbf{x}'(\nu) = (\gamma(\nu) + \gamma''(\nu))\hat{\tau}(\nu). \tag{3.18}$$

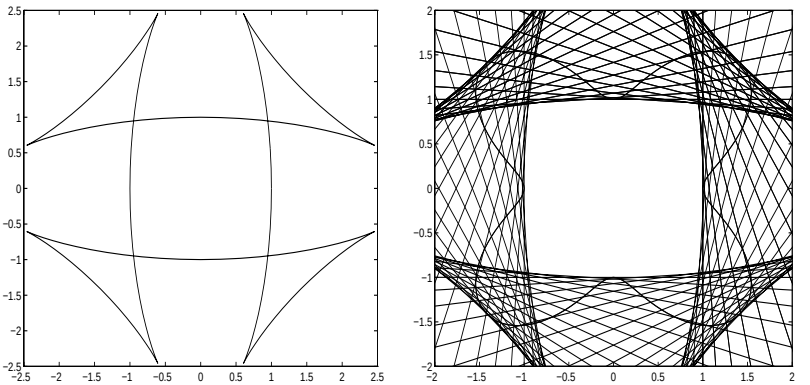

FIGURE 6. *Left: Plot of formula (3.17) when* $\gamma(\nu) = 1 + \sin^2(2\nu)$. *Swallowtails appear since this* γ *is not convex. Right: Wulff shape from Wulff's geometric construction. Notice that by clipping off the swallowtails in the graph on the left, we get the true Wulff shape.*

It is clear that the curve kinks and reverses direction whenever $\gamma(\nu) + \gamma''(\nu)$ changes sign, as it does at each corner of a swallowtail. Suppose the curve $\mathbf{x}$ self-intersects at $\nu = \nu_L$ and $\nu = \nu_R$, then the following condition must be satisfied

$$\gamma(\nu_L)\hat{n}(\nu_L) + \gamma'(\nu_L)\hat{\tau}(\nu_L) = \gamma(\nu_R)\hat{n}(\nu_R) + \gamma'(\nu_R)\hat{\tau}(\nu_R). \tag{3.19}$$

In this case, we can obtain the Wulff shape simply by clipping off the swallowtail.

There is yet another way to obtain the Wulff shape from γ. The Euler-Lagrange equation can be written as a simple first order ordinary differential equation (taking λ as 1, which is amounts to a rescaling of the length of the curve Γ)

$$\frac{d\nu}{ds} = \frac{1}{(\gamma(\nu) + \gamma''(\nu))}, \tag{3.20}$$

which completely specifies the curve up to a scaling.

If $\gamma + \gamma''$ does not change sign, there are two possibilities. If $\gamma + \gamma''$ stays positive, the right hand side is always finite and can be integrated to compute a convex shape (since the curvature $\kappa = \nu_s$ is positive). This is the unique solution to Wulff's problem. If $\gamma + \gamma'' \geq 0$, and becomes 0 at some points, the resulting curve is still convex, and yet has kinks at points where $\gamma + \gamma''$ vanishes.

However, if $\gamma + \gamma''$ does change sign, the curve one obtains from this integration will not be convex, can have kinks (points where curvature $\kappa = \nu_s$ is infinite), and will typically cross over itself, so that it does not even define a possible material shape. The result can be considered as a multiple valued solution to the problem, since in polar coordinates with origin at the crystal center it corresponds to having a multivalued radius as a function of polar angle. This multivalued solution can be regularized to obtain the desired solution by "clipping off" the non-physical parts of the shape created by self-crossings.

3.4. Frank's Convexification of Surface Tension. Many different surface tension functions can lead to the same Wulff shape. This is clear from Wulff's

geometric construction, which effectively ignores the behavior of the parts of the γ polar plot farthest from the origin, i.e. the high energy parts of the surface tension function. Thus in general we have the freedom of using a surface tension that is *equivalent* to the original γ, in the sense that it has the same Wulff shape.

The breakdown of equation (3.17) and (3.20) ultimately stems from a change in sign of $\gamma + \gamma''$. We would thus like to use our freedom to define an equivalent surface tension, $\hat{\gamma}$, for which

$$\hat{\gamma} + \hat{\gamma}'' \geq 0. \tag{3.21}$$

It turns out there is a classical procedure known as Frank convexification which yields such an equivalent surface tension.

The Frank convexification of γ, denoted $\hat{\gamma}$, involves two Legendre transformations and appears complicated. See formula (2.10) in section 2.3. But there exists a simple geometric procedure to obtain $\hat{\gamma}$ from γ by taking the polar plot of $1/\gamma(\nu)$, forming its outer convex hull, and defining this to be the polar plot of $1/\hat{\gamma}(\nu)$. The results in section 2.3 shows that the relationship between the surface tension plot and the Wulff shape becomes a standard geometric duality when viewed under the inversion mapping. See also the article of Frank [**10**].

Now let us take a closer look at the above procedure. The normal direction to the curve $r(\nu) = \frac{1}{\gamma(\nu)}$ is

$$\theta = \nu + \tan^{-1}\left(\frac{\gamma'(\nu)}{\gamma(\nu)}\right). \tag{3.22}$$

Thus

$$\theta_\nu = \frac{\gamma(\gamma + \gamma'')}{\gamma^2 + \gamma'^2}. \tag{3.23}$$

The curve fails to be convex only when $\gamma + \gamma'' < 0$.

There are basically two situations where this can happen. One situation is that there is a region on the plot of $r = \gamma(\nu)$ that "bumps" out. For the plot of $r = \frac{1}{\gamma(\nu)}$, it corresponds to a region that curves inward. Thus the convexifying curve $r = \frac{1}{\hat{\gamma}(\nu)}$ is a straight line, tangent to the curve $r = \frac{1}{\gamma(\nu)}$ at two points $(\nu_L, \frac{1}{\gamma(\nu_L)})$ and $(\nu_R, \frac{1}{\gamma(\nu_R)})$. The other situation is that the plot of $r = \gamma(\nu)$ has a inward cusp at $\nu = \nu_M$. Then the curve $r = \frac{1}{\gamma(\nu)}$ has a kink at $\nu = \nu_M$, and this curve can be convexified by two lines which meet at the tip $(\nu_M, \frac{1}{\gamma(\nu_M)})$, and are tangent to the curve at two points $(\nu_L, \frac{1}{\gamma(\nu_L)})$ and $(\nu_R, \frac{1}{\gamma(\nu_R)})$. See figure 7.

In the first case, the polar plot of $\hat{\gamma}$ contains a circular arc whose extension passes through the origin and

$$\hat{\gamma}(\nu) = \frac{\gamma(\nu_R)\sin(\nu - \nu_L) + \gamma(\nu_L)\sin(\nu_R - \nu)}{\sin(\nu_R - \nu_L)}, \text{ for } \nu_L \leq \nu \leq \nu_R. \tag{3.24}$$

One can easily derive the following jump conditions

$$\gamma(\nu_L)\cos\nu_L - \gamma'(\nu_L)\sin\nu_L = \gamma(\nu_R)\cos\nu_R - \gamma'(\nu_R)\sin\nu_R, \tag{3.25}$$

$$\gamma(\nu_L)\sin\nu_L + \gamma'(\nu_L)\cos\nu_L = \gamma(\nu_R)\sin\nu_R + \gamma'(\nu_R)\cos\nu_R \tag{3.26}$$

from the second order contact of the line with the original curve. Note that these equations are exactly the self-intersection condition (3.19).

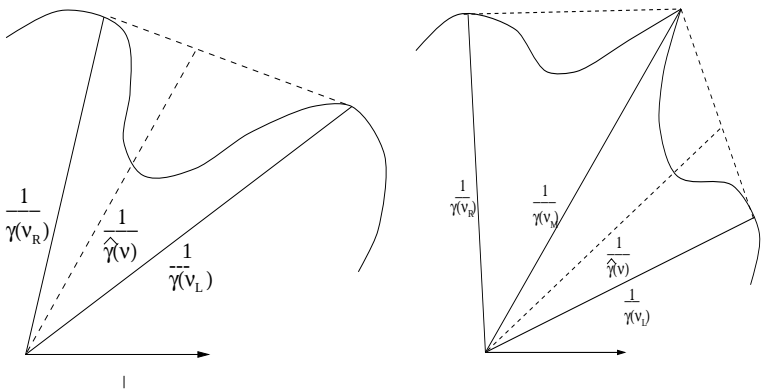

FIGURE 7. *Left: the first case; Right: the second case.*

In the second case, the polar plot of $\hat{\gamma}$ has two circular arcs and meet at $\nu = \nu_M$ and form a cusp there, and we have

$$\hat{\gamma}(\nu) = \frac{\gamma(\nu_L)\sin(\nu_M - \nu) + \gamma(\nu_M)\sin(\nu - \nu_L)}{\sin(\nu_M - \nu_L)}, \text{ for } \nu_L \le \nu \le \nu_M, \tag{3.27}$$

$$\hat{\gamma}(\nu) = \frac{\gamma(\nu_M)\sin(\nu_R - \nu) + \gamma(\nu_R)\sin(\nu - \nu_M)}{\sin(\nu_R - \nu_M)}, \text{ for } \nu_M \le \nu \le \nu_R. \tag{3.28}$$

At ν_M the convexified surface tension $\hat{\gamma}$ is continuous and $\hat{\gamma}(\nu_M) = \gamma(\nu_M)$, but γ and $\hat{\gamma}$ each have a jump in derivative there. The following inequalities are satisfied:

$$\gamma'(\nu_M^-) \le \hat{\gamma}'(\nu_M^-) < 0 < \hat{\gamma}'(\nu_M^+) \le \gamma'(\nu_M^+), \tag{3.29}$$

where

$$\hat{\gamma}'(\nu_M^-) = \frac{\gamma(\nu_M)\cos(\nu_L - \nu_M) - \gamma(\nu_L)}{\sin(\nu_M - \nu_L)},$$

$$\hat{\gamma}'(\nu_M^+) = \frac{\gamma(\nu_R) - \gamma(\nu_M)\cos(\nu_L - \nu_M)}{\sin(\nu_R - \nu_M)}.$$

In both cases, we have the following inequality

$$\gamma(\nu) \ge \hat{\gamma}(\nu), \text{ for } \nu_L \le \nu \le \nu_R. \tag{3.30}$$

We note that when this convexified surface tension is used within the general formula (3.17) for the multivalued solution to Wulff's problem, the resulting curve

$$\mathbf{x}(\nu) = \hat{\gamma}(\nu)\hat{n}(\nu) + \hat{\gamma}'(\nu)\hat{\tau}(\nu) \tag{3.31}$$

has no self-intersections and thus is the correct Wulff shape. In the first case discussed above, $\hat{\gamma}(\nu) + \hat{\gamma}''(\nu) = 0$ on $[\nu_L, \nu_R]$, thus the Wulff shape defined by (3.31) has a sharp corner where the normal jumps from ν_L to ν_R. In the second case, $\hat{\gamma}(\nu) + \hat{\gamma}''(\nu) = 0$ on $[\nu_L, \nu_M)$ and $(\nu_M, \nu_R]$ and $\hat{\gamma}(\nu_M) + \hat{\gamma}(\nu_M) = \infty$. The normal to the Wulff shape jumps from ν_L to ν_M, forming a corner there, then stays equal to ν_M and forms a facet, and then jumps again from ν_M to ν_R, forming another corner. Thus the second case corresponds to two corners connected with a facet.

The Frank convexified surface tension provides the basis for our general Riemann problem representation of the Wulff shape.

3.5. Two Formulae for the Normal Direction. Recall that the Wulff shape is given by the first Legendre transformation

$$\gamma_*(\theta) = \inf_{\theta-\frac{\pi}{2}<\nu<\theta+\frac{\pi}{2}} \left[\frac{\gamma(\nu)}{\cos(\theta-\nu)}\right].$$

So the problem of finding the Wulff shape for a given γ is reduced to find the $\nu(\theta)$ for a given θ where the infimum is reached. This section concerns finding explicit formulae for $\nu(\theta)$ which we will see shortly is closely connected with the Riemann problem for a scalar conservation law.

We point out some technical difficulties here. First of all, for a given θ, there may exist more than one ν that minimizes $\frac{\gamma(\nu)}{\cos(\theta-\nu)}$. This occurs partly because γ may not be convex. We can get rid of this difficulty by replacing γ by $\hat{\gamma}$, since γ and $\hat{\gamma}$ have the same Wulff shape. But even if this convexity condition is satisfied, there is still no uniqueness when $\gamma(\nu)+\gamma''(\nu)=0$. Such a situation arise at the corner of a Wulff shape. We deal with this ambiguity by requiring $\nu(\theta)$ to be increasing and continuous from the right in θ. Secondly, it is a subtle matter how to choose the range for the normal angle ν and the polar angle θ of the convex Wulff shape so that under the above assumptions on γ and $\nu(\theta)$, the one to one correspondence between ν and θ is naturally defined. This can be achieved by choosing the horizontal axis so that it intersects with the Wulff shape at a (global) minima of the surface tension. At this point, both the normal angle and the polar angle are 0 or 2π. Since the Wulff shape is convex, $\nu(\theta)$ must be a nondecreasing function in θ. Thus the ν–θ correspondence can be chosen as a function from $[0,2\pi]$ to itself. This will be our choice of horizontal axis in our theoretical analysis below.

Our first expression of $\nu(\theta)$ is

LEMMA 6. *For each $\theta \in [0,2\pi)$, there is a unique $\nu=\nu(\theta)$ that is increasing, continuous from the right and is implicitly defined by*

$$\theta = \nu + \tan^{-1}\left(\frac{\hat{\gamma}'(\nu)}{\hat{\gamma}(\nu)}\right). \tag{3.32}$$

We postpone the proof and introduce

DEFINITION 4.

$$F(\nu) = \frac{1}{2}\nu^2 + \int_0^\nu \tan^{-1}\left(\frac{\hat{\gamma}'(u)}{\hat{\gamma}(u)}\right) du. \tag{3.33}$$

We notice that

$$F'(\nu) = \nu + \tan^{-1}\left(\frac{\hat{\gamma}'(\nu)}{\hat{\gamma}(\nu)}\right), \tag{3.34}$$

$$F''(\nu) = \frac{\hat{\gamma}(\nu)(\hat{\gamma}(\nu)+\hat{\gamma}''(\nu))}{\hat{\gamma}^2(\nu)+\hat{\gamma}'^2(\nu)}. \tag{3.35}$$

Since $\theta = F'(\nu) \geq 0$ and $\hat{\gamma}+\hat{\gamma}'' \geq 0$, F is always nondecreasing and convex.

Our second expression for $\nu(\theta)$ is

THEOREM 3.

$$\nu(\theta) = -\frac{d}{d\theta}\min_{0\leq\nu\leq 2\pi}\left[F(\nu)-\theta\nu\right]. \tag{3.36}$$

We outline the proof of the above two results here. Refer to figure 8 to get the idea. Lemma 6 follows from the fact that $F'(\nu)$ is an increasing function in ν and formula (3.12).

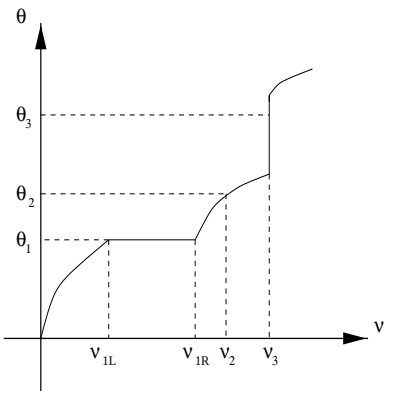

FIGURE 8. *Plot of* $\theta = F'(\nu)$ *vs* ν.

To prove Theorem 3, suppose the infimum of $F(\nu) - \theta\nu$ is reached at $\tilde{\nu}$. The first order condition

$$\frac{d}{d\nu}[F(\nu) - \theta\nu] \mid_{\nu=\tilde{\nu}} = F'(\tilde{\nu}) - \theta = 0$$

gives

$$\theta = \tilde{\nu} + \tan^{-1}\left(\frac{\hat{\gamma}'(\tilde{\nu})}{\hat{\gamma}(\tilde{\nu})}\right).$$

So $\tilde{\nu} = \nu(\theta)$ except at points where $F'(\nu) = \theta$ over some interval $\nu \in [\nu_L, \nu_R]$. (As a function of θ, ν jumps from ν_L to ν_R.) Such θ are isolated. Ignoring this situation, we have

$$\begin{aligned} \text{rhs} &= -\frac{d}{d\theta}\left[F(\nu(\theta)) - \theta\nu(\theta)\right] \\ &= -F'(\nu(\theta))\nu'(\theta) + \nu(\theta) + \theta\nu'(\theta) \\ &= \nu(\theta). \end{aligned}$$

So our result is valid except at countable isolated points. The values of $\nu(\theta)$ at the jumps are uniquely determined by the requirement that ν is increasing and continuous from the right.

4. The Riemann Problem

4.1. The Riemann Problem Formulation. The original Riemann problem was to determine the 1D dynamics of a gas when the initial data consists of constant states to the left and right separated by a single discontinuous jump in value.

The equations of motion for a gas are generally formulated as integral conservation laws for mass, momentum and energy. In one spatial dimension (1D), these state that the rate of change of the amount of conserved quantity contained in any

interval $[x_1, x_2]$ is due to the difference between the flux out at x_2 and the flux in at x_1:

$$\frac{d}{dt}\int_{x_1}^{x_2} u(x,t)dx = f(u(x_2,t)) - f(u(x_1,t)), \tag{4.1}$$

where $u(x,t)$ is the density of the conserved quantity, and $f(u)$ is the corresponding flux function.

When the solution u is smooth, by letting $x_1 - x_2$ become infinitesimal these integral conservation laws can be reduced to differential equations. The result is a system of hyperbolic conservation laws of the form

$$u_t + f(u)_x = 0 \tag{4.2}$$

for the conservative convective transport of mass, momentum and energy.

Riemann's problem was to find the solution of equation (4.2) for arbitrary piecewise constant initial data

$$u = u_L, x < 0, \tag{4.3}$$

$$u = u_R, x > 0, \tag{4.4}$$

where u_L and u_R are the constant states to the left and right of the origin.

The problem as posed is physically idealized, since the conservation law (4.2) does not include any viscous or diffusive transport effects. In a real gas the viscous effects are usually small, but they do play a role when the states have steep spatial gradients as in the Riemann problem. Indeed, it turns out that idealized Riemann problem allows multiple solutions. The unique physically relevant one is the "viscosity solution", i.e. the limiting solution as viscosity goes to zero $u = \lim_{\epsilon\downarrow 0} u^\epsilon$ from the viscous version of the conservation law:

$$u_t^\epsilon + f(u^\epsilon)_x = \epsilon u_{xx}^\epsilon. \tag{4.5}$$

In contrast, this regularized equation has unique well behaved solutions for any $\epsilon > 0$.

Both the Riemann problem and the viscosity solution make sense for general systems of hyperbolic conservation laws, and the names commonly refer to this more general context. The Riemann problem solutions provide insight into the fundamental propagation of discontinuities in the system. For our purposes, we will only need to consider a single conservation equation, so that u is a scalar state, with scalar flux $f(u)$.

A comprehensive discussion of the the Riemann problem for gas dynamics and related matters can be found in the text of Courant and Friedrichs [**6**].

4.2. Multiple-Valued Solutions. The solutions to the Riemann problem have a simple form in which a disturbance emanates from initial discontinuity at $x = 0$. These solutions can be found by assuming the time self-similar form $u(x,t) = u(x/t)$, which implies the graph of $u(x,t)$ has the same shape at all times, differing only by a spatial rescaling. Substituting this form into the conservation law 4.2 results in the equation

$$(-\theta + f'(u))u_\theta = 0, \tag{4.6}$$

where $\theta = x/t$ is the similarity variable. The formal solution consists of regions on the left and right where u is constant with values u_L and u_R, joined by a region in which $u(\theta) = (f')^{-1}(\theta)$. If $f''(u)$ changes sign between u_L and u_R, then the

inverse of f' is multivalued and this u can be considered a multivalued solution of the Riemann problem.

Such a multivalued solution is not physically meaningful, so some additional principle is required to extract a single valued solution by "clipping off" extra values. However, from a plot of the multivalued solution it is not immediately obvious where to clip.

The proper single-valued, self-similar solution to the Riemann problem is given analytically by

$$u(\theta) = -\frac{d}{d\theta} \min_{u_L \le u \le u_R} (f(u) - \theta u), \text{ if } u_L \le u_R. \tag{4.7}$$

$$u(\theta) = -\frac{d}{d\theta} \max_{u_L \ge u \ge u_R} (f(u) - \theta u), \text{ if } u_L \ge u_R. \tag{4.8}$$

This formula was first derived by Osher in [**15**] and [**16**]; It can be understood as an analytical interpretation of the geometric construction given in the next section.

Note that in (4.7) f can be replaced by the convexified $\hat{f}$. In the case when $u_L < u_R$, $\hat{f}$ is defined by

$$f_*(\theta) = \min_{u_L \le u \le u_R} [f(u) - \theta u], \tag{4.9}$$

$$\hat{f}(u) = \max_{-\infty < \theta < \infty} [f_*(\theta) + \theta u]. \tag{4.10}$$

The case $u_L > u_R$ can be defined similarly.

It follows that $\hat{f}(u) \equiv f(u)$ if $f''(u) \ge 0$ on the interval $u_L \le u \le u_R$. $\hat{f}$ is always convex and has a nondecreasing derivative.

The solution to the Riemann problem at $\theta = \frac{x}{t}$ is defined as follows.

- **Case 1.** There exists a unique $u(\theta)$ such that $\hat{f}'(u(\theta)) = \theta$ and $\hat{f}''(u) > 0$ in a neighborhood of $u(\theta)$. In this case $u(\theta) = (\hat{f}')^{-1}(\theta)$. This point lies in a rarefaction fan.
- **Case 2.** $\hat{f}'(u)$ is constant over $a_L \le u \le a_R$ ($\hat{f}(u)$ is linear over $a_L \le u \le a_R$) and $\theta = \hat{f}'(a_L) = \hat{f}'(a_R)$. In this case, the resulting solution has a jump at θ: $u(\theta^-) = u_L$ and $u(\theta^+) = u_R$. This increasing jump in u corresponds to a contact discontinuity.
- **Case 3.** $\hat{f}'$ has an increasing jump at $u = u_0$ ($\hat{f}$ has a kink at $u = u_0$). Then

  $$u(\theta) = u_0, \text{ for } \hat{f}'(u_0^-) < \theta < \hat{f}'(u_0^+).$$

  This corresponds to a constant state.

We shall show that these three situations corresponds to three scenarios in Wulff crystal shape. The rarefaction wave corresponds to regions where the angle of the normal increases smoothly with the polar angle. The contact discontinuities correspond to the corner on a Wulff crystal where the angle of the normal jumps. The constant states correspond to the facets, where the normal to the Wulff shape points to a constant direction as the polar angle increases.

4.3. Geometric Construction of the Solution. The general conservation law (4.2) can formally be written as the convection equation $u_t + v(u)u_x = 0$, where the convective velocity is $v(u) = f'(u)$. The solutions to this can be visualized by letting each value u on the graph move horizontally with constant speed $v(u)$.

Based on this we obtain a simple geometric construction of the (possibly multiple-valued) solution to the Riemann problem. Each value from the initial step function u simply moves at its constant speed; in particular, each u value from the "step" itself, where u ranges between u_L to u_R at the single point $x = 0$, will propagate at its constant speed $v(u)$. Thus the resulting graph of u at any $t > 0$ will, when turned on its side, simply reproduce the graph of $v(u)$, u between u_L and u_R. This implies that $u(x, t)$ will be a multivalued solution of the Riemann problem if the graph of $v(u)$ is not monotone, i.e. if $v' = f''$ changes sign, as mentioned in the previous section.

To extract the physically correct single valued solution—the viscosity solution—we apply a geometric generalization of the conservation of the area under the graph of u implied by the original conservation law (4.2): At each overhang in the multiple-valued graph, introduce a jump that clips off the same amount overhanging area as it fills in on the underhang. Refer to figure 9.

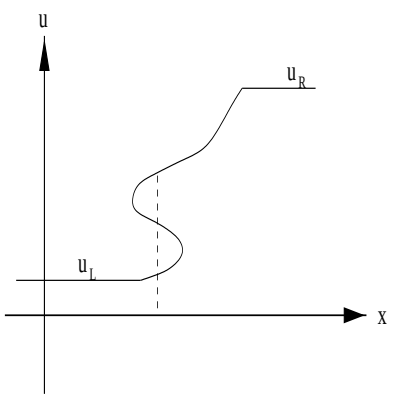

FIGURE 9. *The clipping procedure to the multivalued solution.*

The application of this clipping procedure to the multivalued solution at any time $t > 0$ will yield the proper single valued, time-self similar solution. Due to the self-similarity in time, the same shape results independent of t. Note the profile consists of constant regions to the far left and right, smooth regions where no clipping was necessary—"rarefactions"—and jumps where a clip was performed. These jumps in turn are classified as a "contact" if the velocity $v(u)$ is the same on each side of the jump, or a "shock" if the velocity causes u values one one side to overtake those on the other. Thus values appear to flow into a shock from both sides as time goes by.

While this clipping procedure is reasonable from the perspective of conserving u, it is not so easy to understand why it yields the true viscosity solution, i.e. the solution selected by the action of a small viscous dissipation.

5. The 2D Wulff Crystal as the Solution of a Riemann Problem

From the summaries of the Wulff Crystal and the Riemann problems, we can see a number of points of similarity in addition to the discontinuous nature of the solutions. Both problems admit self-similar solutions. Both are generally formulated in terms of integral equations. Both lead to governing differential equations

that formally have multiple-valued "solutions". In both cases the multiple-valued solutions occur due to a lack of convexity, in the sense that a second derivative changes sign (in the Wulff problem its the sign of $\gamma + \gamma''$, in the Riemann problem, it is f''). And in both cases, there is a geometric construction that effectively truncates these multi valued solutions to yield the unique physical solution.

With this background in place, we are prepared to discuss the precise connection and differences between the two problems. There are several approaches that we can connect the Wulff shape with a Riemann problem of a scalar conservation law.

5.1. From the Euler-Lagrange Equation to a Scalar Conservation Law.

5.1.1. *The Basic Connection.* The first precise formal connection comes from rewriting the Euler-Lagrange equation (3.16) from Wulff's problem as the equation for the time self-similar solution of a Riemann problem 4.6. To do this, we define a function "flux function" $F(\nu)$ by the relation (assuming $\lambda = 1$)

$$F'' = \gamma + \gamma''. \tag{5.1}$$

Then the Euler-Lagrange equation (3.16) can be written as

$$(F'(\nu))_s = 1, \tag{5.2}$$

and integrating this yields

$$(F'(\nu(s))) - s = C, \tag{5.3}$$

where C is the constant of integration. By appropriate choice of the origin for the arclength parameter s, we can have $C = 0$. In this normalization, multiplying through by ν_s yields

$$(F'(\nu) - s)\nu_s = 0, \tag{5.4}$$

which is identical in form to the time self-similar equation (4.6). This in turn is the equation for the Riemann problem for the conservation law

$$\nu_t + (F(\nu))_x = 0. \tag{5.5}$$

Thus at least formally the normal angle $\nu(s)$ is the time self-similar solution to a Riemann problem for this conservation law. This would explain the crystal facets as constant states, the smooth faces as rarefactions, and the jumps in normal angle at crystal edges as shocks or contacts. However, this formal connection is not generally valid, because the differential equations used in the derivation only govern smooth solutions, i.e. crystal shapes with no edges and Riemann problems with no jumps. Whether a crystal shape with edges is the solution to this Riemann problem must be investigated separately. It turns out that the conditions at jumps are different in the two problems, as described in the next subsection. Thus to completely realize the Wulff crystal as a Riemann problem solution requires a more subtle connection.

5.1.2. *Differences Between Wulff and Riemann Jump Conditions.* If the solution to a Riemann problem contains a propagating discontinuous jump, the differential equation for the conservation law (4.2) is not applicable at that point. However, the more general integral conservation law (4.1) still holds, and applied to a small interval containing the jump it yields the Rankine-Hugoniot jump condition

$$V(u^+ - u^-) = f(u^+) - f(u^-) \tag{5.6}$$

where V is the constant propagation speed of the discontinuity and u^+ and u^- are the right and left side values of u at the jump. This condition constrains the allowed jumps in a Riemann problem.

Similarly, if a Wulff crystal has a sharp edge with a jump in normal angle ν, the Euler-Lagrange equation (3.16) does not apply at that point. In this case, we can still identify a condition that governs the allowed jumps in angle. In the formal solution curve (3.17), the sharp corners on a crystal occur at points of self-intersection of the curve. These points separate the primary crystal shape from the artificial "swallowtail" shaped appendages that must be removed. Thus the jump condition at the edge is simply the condition for self intersection of this curve at two distinct normal angles ν_L and ν_R (corresponding to the normal direction on either side of the edge): $x(\nu_L) = x(\nu_R)$, or by the formula (3.17)

$$\gamma(\nu_L)\hat{n}(\nu_L) + \gamma'(\nu_L)\hat{\tau}(\nu_L) = \gamma(\nu_R)\hat{n}(\nu_R) + \gamma'(\nu_R)\hat{\tau}(\nu_R). \tag{5.7}$$

Taking the two components of this vector condition yields two scalar jump conditions.

If we compare the jump conditions for the specific Riemann problem (5.5) we formally associated with the Wulff crystal (i.e. $f = F$ from (5.1)), and the jump conditions (5.7) that hold for the true Wulff shape, it turns out they are different. The former has one constraint while the latter has two, and in addition, they allow different jumps. As we will see, the additional constraint comes from the fact that the allowed jumps is a contact discontinuity and must satisfy

$$f'(u_L) = f'(u_R) = \frac{f(u_R) - f(u_L)}{u_R - u_L}. \tag{5.8}$$

One can easily check that the flux F defined above does not possess this property at the corner.

This difference in jump conditions means that the discontinuous physical solutions to the Riemann problem (5.5) for $\nu(s)$ do not yield the correct normal angle function for Wulff shape. Thus for crystals with corners, a more careful construction is required to realize them as the solution to a Riemann problem.

The origin of this difference for discontinuous solutions can be traced back to the viscosity regularization used to define the unique solution of Riemann problem for conservation law (4.2). Evidently, this is *not* the proper regularization technique for use on the Euler-Lagrange differential equations for the Wulff problem. In retrospect this is not so surprising, since a proper regularizing correction for these equations should be derived by adding a physically reasonable energy penalty term to the original crystal energy (2.1), and using the variational calculus to derive the corresponding additional term in the Euler-Lagrange equations. The proper form of such a regularizing energy correction is considered in Gurtin's book [**11**].

5.1.3. *Reparameterization of Euler-Lagrange Equation.* In order to represent an arbitrary Wulff crystal as a solution to a related Riemann problem, we must take advantage of two additional degrees of freedom in the basic derivation. This added freedom will allow us to determine a flux for the Riemann problem such that the time self similar solution matching both the smooth parts and the jumps in the crystal shape.

The first freedom is that surface tension function γ can be replaced by an equivalent (i.e. resulting in the same Wulff shape) yet convex function $\hat{\gamma}$. This

choice of $\hat{\gamma}$ will free us from considering self-intersection. But facets and jumps are still allowed.

The second degree of freedom is the choice of parameterization of the Wulff shape curve. So far we have used arc length, s, but if we used any other parameterization, $\alpha(s)$. the change of variables in the Euler-Lagrange equation (3.16) would have the general form (using the equivalent $\hat{\gamma}$ instead of γ)

$$J(\hat{\gamma}(\nu) + \hat{\gamma}''(\nu))\nu_\alpha = 0, \tag{5.9}$$

where $J = \alpha_s$. If $J = J(\nu)$, we can follow the derivation of the basic Riemann problem from section 5.1.1 and conclude that the flux function $F(\nu)$ given by

$$F'' = J(\hat{\gamma} + \hat{\gamma}'') \tag{5.10}$$

defines a Riemann problem whose time self similar solutions match the smooth behavior of the Wulff shape $\nu(\alpha)$. The remaining freedom in choice of J can be used to match the proper jump conditions at the crystal shape corners.

In principle, this condition gives a set of equations defining the change of parameterization, J, and thus the flux F. In practice it would be tedious to blindly attempt to solve these equations. Fortunately, the proper reparameterization is one of the obvious possibilities, namely a change to standard polar coordinates for the Wulff shape curve. If we parameterize the normal direction at a point on the Wulff curve by the polar angle for that point, $\nu = \nu(\theta)$, we can show by using chain rule that

$$\theta_s = \frac{\hat{\gamma}}{(\hat{\gamma}')^2 + \hat{\gamma}^2}. \tag{5.11}$$

The corresponding flux function from (5.10) is then defined by

$$F'' = \frac{\hat{\gamma}(\hat{\gamma} + \hat{\gamma}'')}{\hat{\gamma}^2 + \hat{\gamma}'^2}. \tag{5.12}$$

This can be integrated to obtain an explicit formula for the flux function,

$$F(\nu) = \frac{1}{2}\nu^2 + \int_0^\nu \tan^{-1}\left(\frac{\hat{\gamma}'(u)}{\hat{\gamma}(u)}\right) du. \tag{5.13}$$

The time self similar viscosity solution to the appropriate Riemann problem for the corresponding hyperbolic conservation law $\nu_t + F(\nu)_\xi = 0$ is exactly the Wulff shape parameterized as $\nu(\theta)$.

We now verify by directly checking that the contact jump conditions (5.8) agree with the Wulff jump conditions (5.7) in this case. Suppose ν jumps from ν_L to ν_R. Then from the discussion in section 4.2, $\hat{\gamma}(\nu) + \hat{\gamma}''(\nu) \equiv 0$ for $\nu_L \le \nu \le \nu_R$ and F is linear over this interval. Note that

$$\theta = F'(\nu) = \nu + \tan^{-1}\left(\frac{\hat{\gamma}'(\nu)}{\hat{\gamma}(\nu)}\right).$$

So the first equality in contact jump condition (5.8) means

$$\theta_L = \theta_R.$$

The conclusion follows from (3.24) in section 3.4.

In retrospect, the correct form of flux is also the most natural one if we write the corresponding conservation law as

$$\frac{\partial \nu}{\partial t} + F'(\nu)\frac{\partial \nu}{\partial \xi} = 0 \tag{5.14}$$

whose characteristic equations are

$$\begin{cases} \frac{d\xi}{dt} = \theta(\nu), \\ \frac{d\nu}{dt} = 0, \end{cases} \tag{5.15}$$

which simply say that ν is constant along the ray emanating from the origin with polar angle $\theta = \xi/t$. This is obviously true for the self-similar growth of Wulff crystal shape.

5.2. Self-Similar Growth of Wulff Shape and Riemann Problem. In section 2.3, we have seen that Wulff shape growing with the normal velocity equal to its surface tension is a simple self-similar dilation. We now try to find the evolution equation that governs the normal angle. We assume γ is convex in this section. Otherwise just replace γ with its Frank convexification $\hat{\gamma}$.

To start with, choose the x-axis so that it intersects with Wulff shape at a minima of the surface energy. From Wulff's construction, the normal at this point and the x-axis coincide. Since the growth is self similar, this point will remain on the x-axis. As before, let ν be the normal angle to the positive x-axis and s the arclength parameter. In Appendix II, we derive the evolution equation of ν to be

$$\frac{\partial \nu}{\partial t} + \int_0^{\nu} [\gamma(u) + \gamma''(u)]\, du \frac{\partial \nu}{\partial s} = 0. \tag{5.16}$$

If we let $F''(\nu) = \gamma(\nu) + \gamma''(\nu)$, then the above equation becomes

$$\frac{\partial \nu}{\partial t} + \frac{F(\nu)}{\partial s} = 0. \tag{5.17}$$

This is the same as (5.5) which we derived above.

Note that using the arclength s as a parameter is not a good choice, because for a self similar growth, the point on the interface that moves on a straight line away from the origin corresponds to different values of s at different time. This issue actually predicts a problem with this connection. As we have seen before, the above conservation law does not give the right solution.

The correct equation can be obtained by a change of variables in the equation (5.17) which governs the evolution of the angle of the normal. We introduce the following new set of variables:

$$\begin{cases} \tau = t, \\ \xi = t\theta(t, s) \end{cases} \tag{5.18}$$

where $\theta(t, s)$ is defined implicitly by:

$$\frac{s}{t} = \int_0^{\nu(\theta)} [\gamma(\nu) + \gamma''(\nu)]\, d\nu \tag{5.19}$$

and $\nu(\theta)$ in turn is defined by

$$\theta = \nu + \tan^{-1}\left(\frac{\gamma'(\nu)}{\gamma(\nu)}\right).$$

The equation under this new set of variables is

$$\frac{\partial \nu}{\partial \tau} + \frac{\partial F(\nu)}{\partial \xi} = 0, \tag{5.20}$$

where $F(\nu) = \frac{1}{2}\nu^2 + \int_0^{\nu} \tan^{-1}\left(\frac{\gamma'(u)}{\gamma(u)}\right) du$. This coincides with (5.13) above. See Appendix II for the derivation.

5.3. Main Theorem and Its Consequences. We have at least formally demonstrated through two quite different approaches that the Wulff shape is connected with the Riemann problem for a scalar conservation law. This is the major result of this paper. We summarize it in the following theorem and explore some of its consequences.

THEOREM 4. *Let $\gamma : S^1 \to R^+$ be continuous and let its Frank convexification $\hat{\gamma}$ be piecewise differentiable. Let W be the Wulff shape corresponding to surface tension γ, as defined by Wulff's construction, and $\nu(\theta)$ be the angle of the outward normal to W as a function of polar angle θ, in the polar coordinate system with origin at the centroid of W, and the horizontal axis passes through a global minima of the surface tension. Then for all θ where $\nu(\theta)$ is well-defined and differentiable*

$$\nu(\theta) = -\frac{d}{d\theta} \min_{0 \le \nu \le 2\pi} [F(\nu) - \theta\nu], \tag{5.21}$$

where F is the function on $[0, 2\pi]$ defined by

$$F(\nu) = \frac{\nu^2}{2} + \int_0^\nu \tan^{-1}\left(\frac{\hat{\gamma}'(\alpha)}{\hat{\gamma}(\alpha)}\right) d\alpha. \tag{5.22}$$

Furthermore, $\nu(\xi, t) \equiv \nu(\frac{\xi}{t})$ is the time self-similar viscosity solution to the Riemann problem

$$\nu_t + (F(\nu))_\xi = 0, \tag{5.23}$$

$$\nu(\xi < 0, t = 0) = 0, \tag{5.24}$$

$$\nu(\xi > 0, t = 0) = 2\pi. \tag{5.25}$$

The proof follows from Theorem 3 and Osher's formula (4.7) for the Riemann problem for a scalar conservation law.

This theorem serves as a bridge that connect the world of gas dynamics, which has a long history and has been extensively studied (See the book by Courant and Friedrich [**6**]) with the fascinating world of crystal shapes which is characterized by facets, edges and corners. We can characterize these shapes in term of the flux F, which is a convex function. The facet corresponds to a kink in the graph of F in R^2, which in turn corresponds to constant states in the world of gas dynamics; The corner corresponds to a piece of a straight line in the graph of F, which in the world of gas dynamics corresponds to contact jumps; We observe rounded edges when a crystal melts, and the sharp corners become smooth out. These regions correspond to the smooth region in the graph of F, and, in the conservation law analogue, they correspond to rarefaction waves.

We can also characterize these phenomena with the polar plot of $\hat{\gamma}$. Here the facets correspond to cusps, and the corners correspond to circular arcs in the polar plot of $\hat{\gamma}$.

Conceptually, we have fully clarified the initial intuitive similarity between these disparate problems: at least in 2D, it is completely accurate to say that the corners on a crystal are contact discontinuities, the smooth faces are rarefactions, the facets are constant states, for a generally discontinuous solution of a hyperbolic conservation law.

5.4. A Convex Example. We consider the surface tension

$$\gamma(\nu) = |\cos(\nu)| + |\sin(\nu)|. \tag{5.26}$$

This is an important example, since this surface tension arises in the continuum limit of the simple X-Y lattice model of a crystal. However, it is also quite simple to analyze. We will also remark on how it relates to the general case where appropriate.

Note that the key measure of convexity, $\gamma + \gamma''$, vanishes almost everywhere, but does not change sign. In fact, as a distribution it is

$$\gamma(\nu) + \gamma''(\nu) = \sum_{k=0}^{3} \delta(\nu - k\frac{\pi}{2}) \tag{5.27}$$

$$\geq 0. \tag{5.28}$$

Because this quantity appears in the Euler-Lagrange equation (3.16) and in the generalized solution (3.17), these are both degenerate. The solution curve $x(\nu) = \gamma(\nu)\hat{n}(\nu) + \gamma'(\nu)\hat{\tau}(\nu)$ is readily computed, and its image consist of just four isolated points, shown in figure 10 (c),

$$x(\nu) = \begin{cases} (+1,+1), & 0 < \nu < \frac{\pi}{2}. \\ (-1,+1), & 0 < \nu < \pi. \\ (-1,-1), & 0 < \nu < \frac{3\pi}{2}. \\ (+1,-1), & 0 < \nu < 2\pi. \end{cases} \tag{5.29}$$

Connecting these dots into a continuous curve yields a square, which is the Wulff shape. Note that in this example, no multivalued swallowtails occur because there is no sign change in $\gamma + \gamma''$. Also, the image curve consists only of isolated vertices, since, by equation (3.18), the tangent vector x' is proportional to $\gamma + \gamma''$ and thus vanishes almost everywhere.

Wulff's geometric construction also leads to the same shape. The surface tension polar plot is the "four leaf clover" shown in figure 10 (a), consisting of four symmetrically positioned arcs of circles that, if extended, would pass through the origin.

Wulff's geometric construction places one facet at each cusp in the polar plot, and together these form a square as shown in figure 10 (d). The virtual facets placed at all other points along the polar plot lie entirely outside this square, and so the inner envelope defining the Wulff shape is the square itself. The simplicity of the the construction is due to the fact that the polar plot is composed of circular arcs; these are always "dual" to a single vertex in a polygonal Wulff shape (refer to [**10**] for the general properties of this duality).

Next we consider the details of the Riemann problem construction. The first step is to compute the flux function (5.13). Recall the that the flux function is based on the Frank convexified surface tension, $\hat{\gamma}$. However, the surface tension function in this example is already "convex", in the appropriate sense, i.e. $\gamma + \gamma'' \geq 0$. Thus $\hat{\gamma} = \gamma$, and this is a major source of simplification over the general surface tension case. Note that in general the Frank convexified surface tension will replace any nonconvex portion of the polar plot (i.e. segment where $\gamma + \gamma'' < 0$ with the arc of a circle passing through the origin, since that is the curve of neutral convexity (i.e. with $\gamma + \gamma'' = 0$. Because of this, the surface tension used in this example is representative of what generally occurs after convexification.

To compute the flux function, it greatly simplifies the trigonometry to note that

$$\gamma(\nu) = \sqrt{2}\cos(\nu - \phi(\nu)) \tag{5.30}$$

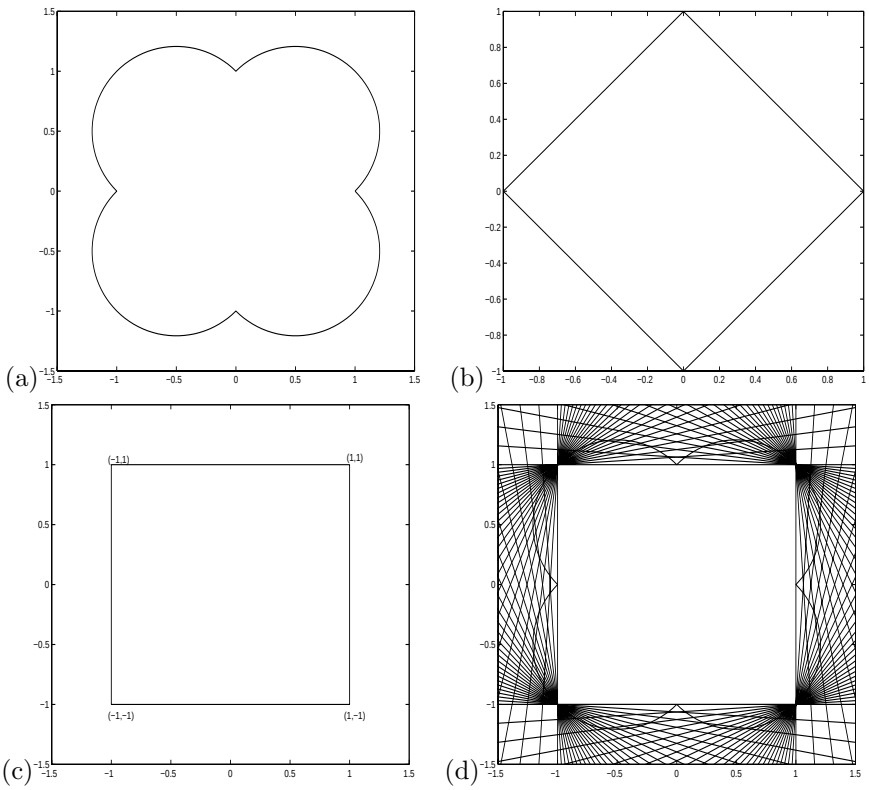

FIGURE 10. *(a) Plot of surface tension γ. (b) Plot of $\frac{1}{\gamma}$. (c) Plot of $\gamma(\nu)\hat{n}(\nu)+\gamma'(\nu)\hat{\tau}(\nu)$. (d) Crystal shape from Wulff's geometric construction.*

where the phase shift $\phi(\nu)$ is

$$\phi(\nu)=\frac{\pi}{4}(2k-1),(k-1)\frac{\pi}{2}<\nu<k\frac{\pi}{2}, \tag{5.31}$$

or, more succinctly,

$$\phi(\nu)=\frac{\pi}{4}(2[\frac{\nu}{\pi/2}]-1), \tag{5.32}$$

where $[x]$ denotes the least integer $\geq x$.

From this we get that

$$\frac{\gamma'}{\gamma}=\tan(-\nu+\phi(\nu)), \tag{5.33}$$

or,

$$\tan^{-1}(\frac{\gamma'}{\gamma})=-\nu+\phi(\nu). \tag{5.34}$$

Applying this in the flux formula 5.13, we get

$$F(\nu) = \frac{\nu^2}{2} + \int_0^\nu -y + \phi(y)dy \tag{5.35}$$

$$= \int_0^\nu \phi(y)dy \tag{5.36}$$

$$= \frac{\pi}{4}([u]^2 + (1 - 2[u])([u] - u)), \tag{5.37}$$

where $u = \frac{\nu}{\pi/2}$. The graph of F is shown in figure 11. F can easily described by noting that it is a piecewise linear function that linearly interpolates the values $F(k\frac{\pi}{2}) = \frac{\pi}{4}k^2$, for integers $k = 0, 1, 2, \ldots$. These values in turn lie on the parabola $f(\nu) = \nu^2/\pi$. Note that the linear segment of the graph beginning at $\nu = k\frac{\pi}{2}$ has slope $(2k+1)\frac{\pi}{4}$. Considering the relation to the general surface tension case, note that the piecewise linear segments of the graph of the flux correspond to portions of the polar plot where $\hat{\gamma} + \hat{\gamma}'' = 0$ (or, geometrically, the surface tension polar plot is a circular arc), and that these will be present wherever the surface tension required convexification. Thus they will be a typical feature of the general case.

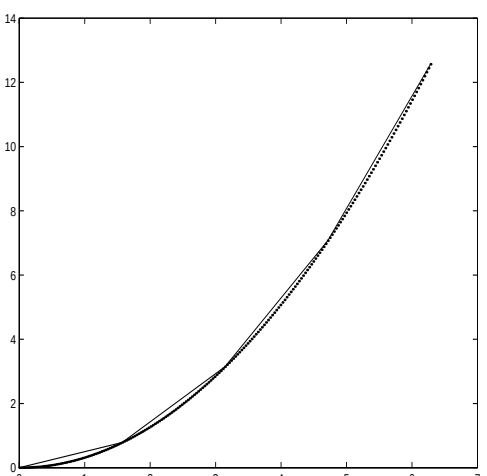

FIGURE 11. *The flux function (the solid line). The dashed line is the graph of* ν^2/π.

With the flux F in hand, we can now work out the analytic form of the Riemann problem solution from formula (5.21). The first step is to find, for a given θ in $[0, 2\pi]$, the minimal value of $F(\nu) - \theta\nu$. Note this function is also a piecewise linear (in ν) function inscribed in a parabola, and so its minimum will be at the vertex where its slope switches from negative to positive as indicated in figure 11. This in turn will occur where $F(\nu)$ changes from having slope less that θ to slope greater than θ. Call this point $\nu_{\min}(\theta)$. It can be described precisely as follows: if θ lies between $\theta_1 = (2k-1)\frac{\pi}{4}$ and $\theta_2 = (2k+1)\frac{\pi}{4}$, then the transition in the slope of F will occur at $\nu_{\min}(\theta) = (k-1)\frac{\pi}{2}$, at which point the slope changes from θ_1 to θ_2. Note that

for a given θ, k is simply the *nearest integer* to $\frac{\theta}{\pi/2}$. Thus we can write

$$\nu_{\min}(\theta) = (N(\frac{\theta}{\pi/2}) - 1)\frac{\pi}{2}, \tag{5.38}$$

where $N(x)$ is the nearest integer to x. In particular, the minimizing argument is a piecewise constant function of θ.

Continuing to unravel the Riemann problem solution formula (5.21), we see that for θ in an interval for which the minimizing argument $\nu_{\min}(\theta)$ remains *constant* with value ν_{min}, we have

$$\min_{0\le\nu\le 2\pi}(F(\nu) - \theta\nu) = F(\nu_{min}) - \theta\nu_{min} \tag{5.39}$$

and thus the solution to the Riemann problem for that range of θ is

$$\nu(\theta) = -\frac{d}{d\theta}\min_{0\le\nu\le 2\pi}(F(\nu) - \theta\nu) \tag{5.40}$$

$$= -\frac{d}{d\theta}(F(\nu_{min}) - \theta\nu_{min}) \tag{5.41}$$

$$= \nu_{min}. \tag{5.42}$$

Applying this over the respective θ intervals corresponding to $\nu_{min} = 0, \pi/2, \pi, 3\pi/2$, we obtain the complete Riemann problem solution as

$$\nu(\theta) = \begin{cases} 0, & 0 \le \theta < \pi/4. \\ \pi/2, & \pi/4 < \theta < 3\pi/4. \\ \pi, & 3\pi/4 < \theta < 5\pi/4. \\ 3\pi/2, & 5\pi/4 < \theta < 7\pi/4. \\ 2\pi, & 7\pi/4 < \theta \le 2\pi. \end{cases} \tag{5.43}$$

This is precisely the angle of the normal vector (to the x-axis) as a function of polar angle θ for a square shape centered at the origin. Thus the solution to the Riemann problem describes the square Wulff shape.

Finally, we can also recover the Wulff shape via the geometric solution to the Riemann problem. For this, we first graph the initial data for $\nu(\xi, t)$, which has left and right states 0 and 2π with the jump at $\xi = 0$. Then we graph $v(\nu) = F'(\nu)$ along the ν axis. In this case, $v(\nu)$ is the piecewise constant function with values $(2k+1)\pi/4$ over the ν intervals $(k\pi/2, (k+1)\pi/2)$, for $k = 0, 1, 2, 3$. Because of the convexity of the flux $F(\nu)$, there are no overhangs in the resulting plot, i.e. it defines a single valued function of $\nu(\xi)$ over the ξ axis. This function is the self-similar solution, $\nu(x, t) = \nu(x/t)$. We see as before that $\nu(\theta)$ is the same function found via the analytic solution to the Riemann problem, and thus it again describes the square Wulff shape. Regarding the general case, note that the flux will always be convex, since $F'' = \frac{\hat{\gamma}(\hat{\gamma}+\hat{\gamma}'')}{\hat{\gamma}^2+\hat{\gamma}'^2} \ge 0$. Thus in this geometric construction, the graph of $v(\nu)$ will always result in a single valued $\nu(\xi)$, and there will be no need for the equal area procedure of clipping off multivalued overhangs as described in the general geometric algorithm for solving the Riemann problem.

5.5. A Nonconvex Example. Now let us consider the following surface tension

$$\gamma(\nu) = 1 + |\sin(2\nu)|. \tag{5.44}$$

The Wulff shape of this γ is also a square. See figure 12 (d). This surface tension is nonconvex, since

$$\gamma(\nu)+\gamma''(\nu)=\begin{cases} 1-3\sin(2\nu), & \text{for } \nu\in[0,\frac{\pi}{2}]\cup[\pi,\frac{3\pi}{2}]. \\ 1+3\sin(2\nu), & \text{for } \nu\in[\frac{\pi}{2},\pi]\cup[\frac{3\pi}{2},2\pi]. \end{cases} \tag{5.45}$$

changes sign as ν goes from 0 to 2π. It turn out that its Frank convexification $\hat{\gamma}(\nu)=|\cos\nu|+|\sin(\nu)|$, which is exactly the surface tension that we discussed in the section above.

Replace γ by $\hat{\gamma}$, we are back to the example in the last section. Refer to figure 12.

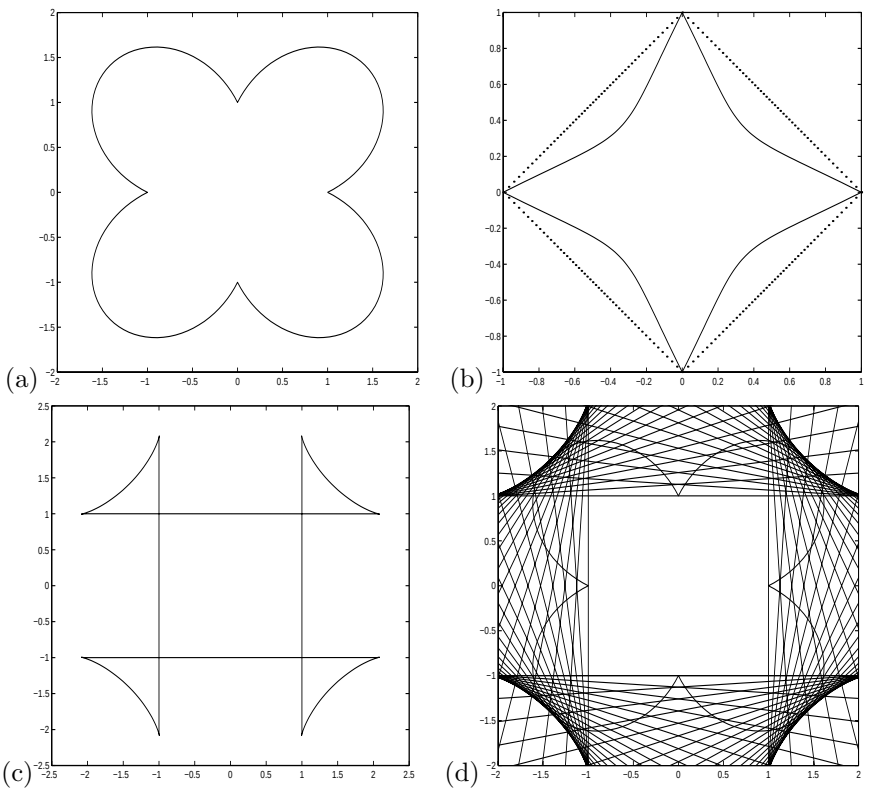

FIGURE 12. *(a) Plot of surface tension γ. (b) The solid line is the plot of $\frac{1}{\gamma}$, and the dashed line is the plot of $\frac{1}{\hat{\gamma}}$. (c) Plot of $\gamma(\nu)\hat{n}(\nu)+\gamma'(\nu)\hat{\tau}(\nu)$. The self-intersection of the plot indicates that this γ is nonconvex. (d) The Wulff crystal shape from Wulff's geometric construction.*

6. Some Comments on The Wulff Problem in Higher Dimensions

We have seen in section 2.3 that the growth of Wulff crystal shape with its (convexified) surface energy is simply a self-similar dilation. Suppose we grow a crystal from a infinitesimal initial Wulff shape, and at time $t=1$ the Wulff shape

is given by $W(\theta) = \gamma_*(\theta)$, then the unit outwards normal at a certain later time satisfies

$$\hat{n}(t, t\mathbf{W}(\theta)) = \hat{n}(1, \mathbf{W}(\theta)).$$

Denote $t\mathbf{W}(\theta)$ as ξ, and differentiate with respect to t, we get

$$\hat{n}_t + \mathbf{W}(\theta) \cdot \nabla_\xi \hat{n} = 0, \tag{6.1}$$

where $\nabla_\xi \hat{n}$ is the gradient of $\hat{n}$. Recall that $\mathbf{W}(\theta) = D\gamma(\hat{n})$, we get the following

$$\frac{\partial \hat{n}}{\partial t} + \sum_{k=1}^{n} \frac{\partial \gamma}{\partial n_k}(\hat{n}) \frac{\partial \hat{n}}{\partial \xi_k} = 0. \tag{6.2}$$

This is a system of hyperbolic equations. The question of whether this system can be transformed into a system of conservation laws through a choice of suitable variables is still open.

At present, the relation between 3D Wulff shapes and nonlinear wave dynamics is unclear. However, the original intuitive connection between crystals and shock waves remains compelling in 3D, and the possibility of some such relation calls for further investigation.

7. The Level Set Formulation for the Wulff Problem

The level set method of Osher and Sethian [**17**] has been very successful as a computational tool in capturing the moving interfaces, especially when the interface undergoes topological changes. It is also useful for the theoretical analysis of the variational problem associated with Wulff crystals. We now briefly review this method and apply it to the Wulff problem.

7.1. The Level Set Representation of Surface Energy. Suppose Ω is a open region in R^d which may be multi-connected. Let $\Gamma = \partial\Omega$ be its boundary. We define an auxiliary function ϕ so that

$$\begin{cases} \phi(x) < 0, \text{ if } x \in \Omega. \\ \phi(x) = 0, \text{ if } x \in \Gamma. \\ \phi(x) > 0, \text{ otherwise.} \end{cases} \tag{7.1}$$

For example, we can choose ϕ to be the signed distance function to the interface Γ. Indeed, for computational accuracy, this is the most desirable case. We call ϕ the level set function of Γ.

Many geometric quantities have simple expressions in terms of level set function. For example, the outward unit normal direction $\hat{n} = \frac{\nabla\phi}{|\nabla\phi|}$, the mean curvature $\kappa = \nabla \cdot \frac{\nabla\phi}{|\nabla\phi|}$, and the area element (or arclength element) $dA = \delta(\phi)|\nabla\phi|dx$. The surface energy over Γ can be expressed as

$$E(\phi) = \int \gamma(\frac{\nabla\phi}{|\nabla\phi|})\delta(\phi)|\nabla\phi|dx \tag{7.2}$$

where δ is the 1 dimensional δ function.

7.2. The Euler-Lagrange Equation for the Wulff Problem. Once we write the surface energy in terms of the level set function, the Wulff problem becomes to find the particular level set function that minimizes the surface energy subject to the constraint that its zero contour enclosed a fixed volume. We extend γ to the whole space as a homogeneous function of degree 1(which we still denote as γ) and introduce a Lagrange multiplier λ. The Lagrangian is:

$$\mathcal{L}(\phi, \lambda) = \int \gamma\left(\frac{\nabla\phi}{|\nabla\phi|}\right)\delta(\phi)|\nabla\phi|dx - \lambda\int H(-\phi)dx, \tag{7.3}$$

where $H(\phi)$ is the Heaviside function which is 0 for $\phi < 0$ and 1 otherwise.

In Appendix III, we show that the Euler-Lagrange equation for (7.3) is

$$\sum_{j=1}^{n}\frac{\partial}{\partial x_j}\left[\frac{\partial\gamma}{\partial p_j}(\frac{\nabla\phi}{|\nabla\phi|})\right] = \lambda, \tag{7.4}$$

or in a more compact form

$$\nabla\cdot\left[D\gamma(\frac{\nabla\phi}{|\nabla\phi|})\right] = \lambda, \tag{7.5}$$

where the constant λ is chosen so that the volume is as given.

Note the denominator in the above expression is simply the perimeter (in 2D) or area (in 3D) of Γ. In 2D, equation (7.4) becomes the familiar formula (3.16).

The gradient flow of the Wulff energy is

$$\phi_t = |\nabla\phi|\left[\nabla D\gamma\left(\frac{\nabla\phi}{|\nabla\phi|}\right) - \lambda\right], \tag{7.6}$$

where λ is given by

$$\lambda = \frac{\int\nabla\cdot\left[D\gamma(\frac{\nabla\phi}{|\nabla\phi|})\right]\delta(\phi)|\nabla\phi|dx}{\int\delta(\phi)|\nabla\phi|dx}, \tag{7.7}$$

so that the area is fixed and the energy is decreasing under the gradient flow.

Equation (7.6) is fully nonlinear weakly parabolic type equation when γ is convex in the sense defined in section 2.3, and is of mixed type when γ is not. How to regularize the variational problem by adding an appropriate penalty term is an interesting question. We shall discuss this issue in future work. See Gurtin's book [**11**] for some discussions of this matter.

7.3. The Hamilton-Jacobi Equation for a Growing Wulff Crystal. Now let the interface move with normal velocity equal to V, which might depend on some local and global properties of the interface Γ. Denote the boundary at a later time t as $\Gamma(t)$, and the associated level set function as $\phi(t, x)$. Let $x(t)$ be a particle trajectory on the interface. By definition, $\phi(t, x(t)) = 0$. By differentiating with respect to t, and noting that $V = \dot{x}(t)\cdot\frac{\nabla\phi(x)}{|\nabla\phi(x)|}$, we get

$$\phi_t + V|\nabla\phi| = 0. \tag{7.8}$$

This is a Hamilton-Jacobi equation if V depends only on x, t, and $\nabla\phi$. The location of the interface is find by solving this equation and then finding its zero level set $\{x : \phi(t, x) = 0\}$. Thus a vast wealth of recent extensive theoretical and numerical research on Hamilton-Jacobi equations can be applied to the moving interface problem.

It was shown in [**19**] that Wulff shape growing with normal velocity equal to surface tension is a self-similar dilation. For any other shape (which may be multiply connected), one can place two concentric Wulff shapes, such that one is contained by this shape, and the other contains this shape, and then let them grow with surface tension. Since the arbitrary shape will always be confined between the two Wulff shapes by the comparison principle for the viscosity solutions to Hamilton-Jacobi equations, one immediately concludes that the asymptotic shape growing from any initial configuration is a Wulff crystal shape. For details of the proof with error bounds, see the recent paper [**19**] by Osher and Merriman. This approach give us a very convenient way to find the Wulff shape numerically for a given surface energy, especially in 3D. The next section contains many examples demonstrating this.

By embedding the interface problem into a one dimensional higher space, it appears that a substantial increase in computation cost is incurred. This is not true, because we are only interested in the behavior of the zero level set. A localized method can be used to lower the computational expense. This is discussed in [**1, 25**] and a more recent paper [**20**]. The method in [**20**] is the one that we used in our numerical examples below.

8. Numerical Examples

We present in this section some numerical results obtained by solving equation (7.8) with $V = \frac{\nabla\phi}{|\nabla\phi|}$, that is,

$$\phi_t + \gamma(\frac{\nabla\phi}{|\nabla\phi|})|\nabla\phi| = 0, \quad x \in R^d, t > 0 \tag{8.1}$$

with a fast localized level set method coupled with a PDE based re-initialization step developed in [**20**] using the ENO [**18**] or WENO [**14**] schemes for Hamilton-Jacobian equations.

First, let us briefly review the numerical schemes that we shall use below for a general Hamilton-Jacobi equation:

$$\phi_t + H(\nabla\phi) = 0, \qquad x \in R^d, t > 0. \tag{8.2}$$

To simplify notation, we will only write down the formulae for the 2D case. The extension to higher dimensions is straightforward.

The semi-discrete version of (8.2) in 2D is:

$$\frac{\partial\phi_{ij}}{\partial t} = -\hat{H}(\phi^+_{x,ij}, \phi^-_{x,ij}, \phi^+_{y,ij}, \phi^-_{y,ij}), \tag{8.3}$$

where $\phi^{\pm}_{x,ij}$ and $\phi^{\pm}_{y,ij}$ are one-sided approximations to the partial derivatives ϕ_x and ϕ_y at (x_i, y_j), respectively. $\hat{H}$ is a numerical Hamiltonian that is monotone and consistent with H. See [**8**] or [**18**] for more details.

In our computations, $\phi^{\pm}_{x,ij}$ and $\phi^{\pm}_{y,ij}$ are calculated with the 3rd order ENO scheme of Osher and Shu [**18**] or the 5th order WENO scheme of Jiang and Peng [**14**] for Hamilton-Jacobi equations, and $\hat{H}$ is chosen as the following Lax-Friedrichs (LF) flux:

$$\hat{H}^{LF}(u^+, u^-, v^+, v^-) = H(\frac{u^+ + u^-}{2}, \frac{v^+ + v^-}{2}) - \frac{\alpha}{2}(u^+ - u^-) - \frac{\beta}{2}(v^+ - v^-) \tag{8.4}$$

where α and β are artificial viscosities defined by:

$$\alpha = \max_{\substack{u\in[A,B]\\ v\in[C,D]}} |H_1(u,v)|, \qquad \beta = \max_{\substack{u\in[A,B]\\ v\in[C,D]}} |H_2(u,v)|. \tag{8.5}$$

Here $H_1 = \partial H/\partial u$, $H_2 = \partial H/\partial v$, $[A,B]$ and $[C,D]$ are the range of $u^{\pm}$ and $v^{\pm}$, respectively.

Solutions to (8.1) often will become either too flat or too steep near the interface $\{\phi = 0\}$ even if the initial data is a perfect signed distance function. In order to avoid numerical difficulties and retain accuracy, an additional operation, which is called re-initialization, is needed to reset ϕ to be a distance function again. This becomes essential for the localized level set method of [**20**]. In [**26**], a PDE based re-initialization method was proposed. By solving the following equation:

$$\begin{cases} \phi_t + sign(\phi_0)(|\nabla\phi| - 1) = 0 \quad \text{in } R^d \times R_+, \\ \phi(x,0) = \phi_0(x) \end{cases} \tag{8.6}$$

to steady state, the original level set function ϕ_0 becomes a distance function to the front defined by $\{\phi_0 = 0\}$. For (8.6), we use the Godunov numerical Hamiltonian:

$$H^{God}(u^+,u^-,v^+,v^-) = \tag{8.7}$$
$$\begin{cases} s\sqrt{[\max((u^+)^-,(u^-)^+)]^2 + [\max((v^+)^-,(v^-)^+)]^2}, & \text{if } \phi^0_{ij} \geq 0. \\ s\sqrt{[\max((u^+)^+,(u^-)^-)]^2 + [\max((v^+)^+,(v^-)^-)]^2}, & \text{otherwise,} \end{cases}$$

where $\phi^0_{ij} = \phi_0(x_i,y_j)$, $(a)^+ = \max(a,0)$,$(a)^- = \max(-a,0)$, and $s = \phi_0/\sqrt{\phi_0^2 + \Delta x}$ is an approximation to $sign(\phi_0)$.

For the time discretization, we use the 3rd order TVD Runge-Kutta scheme developed in [**23**]. Consider the following ODE:

$$\frac{d\phi}{dt} = L(\phi), \qquad \phi(0) = \phi_0. \tag{8.8}$$

The 3^{rd} order TVD Runge-Kutta method at the n^{th} step is:

$$\begin{aligned} \phi^{(1)} &= \phi^n + \Delta t L(\phi^n), \\ \phi^{(\frac{1}{2})} &= \phi^n + \frac{\Delta t}{4}\{L(\phi^n) + L(\phi^{(1)})\}, \\ \phi^{n+1} &= \phi^n + \frac{\Delta t}{6}\{L(\phi^n) + 4L\phi^{(\frac{1}{2})} + L(\phi^{(1)})\}. \end{aligned} \tag{8.9}$$

In all the examples below except for the first one, the computation is performed in the region $[-1,1]^2$ in 2D, and $[-1,1]^3$ in 3D. The time step Δt is chosen as $.1\Delta x$, and for the re-initialization step it is $.5\Delta x$. Since the computation is only done near the front in both the approximation to (8.1) and (8.6), we observe a considerable speed up of approximately 7 times over the global method. In example 2, 3, 4 and 5, we start from a circle or sphere purely for simplicity in preparing the initial data. It is interesting to see initial objects merge and asymptote to the Wulff shape. This is displayed in figure 28 in example 6, where we start from a multiply connected initial shape.

Example 1. To test our main result Theorem 4 in section 5.3, we solve the scalar conservation law (5.23)—(5.25) directly with the 3rd order ENO scheme for conservation laws developed in [**24**] by Shu and Osher , for the case $\gamma(\nu) = |\cos\nu| + |\sin\nu|$. We have found the flux function $F(\nu)$ for this problem in section 5.4. See figure 13.

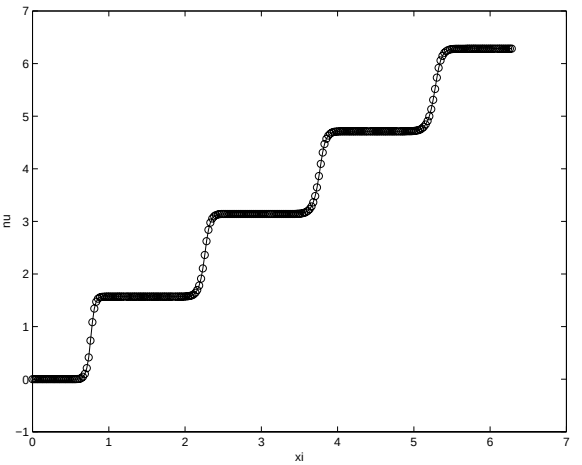

FIGURE 13. *Solution of the conservation law computed with the 3rd order ENO scheme. The computation is done on* $[0, 2\pi]$ *with* 256 *grid points to time* $t = 1.5$.

Example 2. (2D) We start from a circle and evolve it with normal velocity equal to $\gamma(\nu)$, where ν is the angle between the outward normal direction $\hat{n} = \frac{\nabla\phi}{|\nabla\phi|}$ and the x-axis, $-\pi \le \nu \le \pi$. We use a 200×200 grid. The pictures in figure 14–22 on the left are the crystalline shapes obtained from Wulff's construction, the corresponding pictures on the right are the shapes obtained from evolution. We print out the evolving shapes every 50 time steps.

Example 3. (3D) We start from a sphere and evolve it with normal velocity equal to $\gamma(\nu, \varphi)$, where ν and φ are the spherical coordinates, $-\pi \le \nu \le \pi$, $-\frac{\pi}{2} \le \varphi \le \frac{\pi}{2}$. We use a grid of $100\times100\times100$. We choose $\gamma(\nu, \varphi) = \gamma(\nu)h_i(\varphi)$ for $i = 1, 2$ and 3. In figure 23, $h_1(\varphi) = 1+2|\sin(\varphi)|$, and the corresponding Wulff shapes are prisms with different bases that depend on $\gamma(\nu)$. In figure 24, $h_2(\varphi) = 1 + 2\sqrt{|\sin\frac{3}{2}(\varphi + \frac{\pi}{2})|}$, and the corresponding Wulff shapes are pyramids with different bases depending on $\gamma(\nu)$. $h_3(\varphi) = 1 + 2\sqrt{|\sin(|\varphi| - \frac{\pi}{6})|}$, and the corresponding Wulff shapes are bi-pyramids with various bases depending on $\gamma(\nu)$.

Example 4. We define

$$\gamma(\hat{n}) = 1 + 2\sqrt{|\max_{1\le i\le 20} \hat{n} \cdot \mathbf{v}_i - 1|}$$

where the $\mathbf{v}_i$'s are the twenty vertices of a regular polygon of 12 faces that inscribes a unit sphere. We can expect that with the given surface intensity γ, the Wulff shape obtained from Wulff's geometric construction is a regular polygon of 12 faces, a soccer ball like object. We demonstrate this conjecture by starting with a sphere, growing it with the above defined γ. See figure 25 for the numerical result. We use a $100 \times 100 \times 100$ grid in our computation.

Example 5. In this example, we study the behavior of the ratio $E/V^{1-1/d}$ in the evolution process. Here $E = \int_\gamma \gamma(\hat{n})dA$ is surface energy, V is the volume enclosed by the surface. In a recent paper [**19**] of Osher and Merriman, it was shown that,

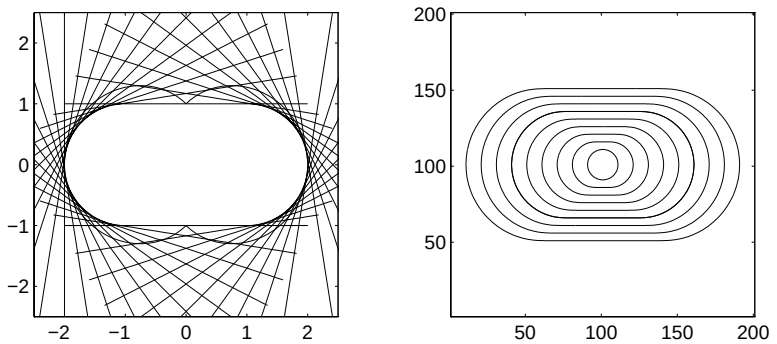

FIGURE 14. $\gamma(\nu) = 1 + |\sin(\nu + \frac{\pi}{2})|$.

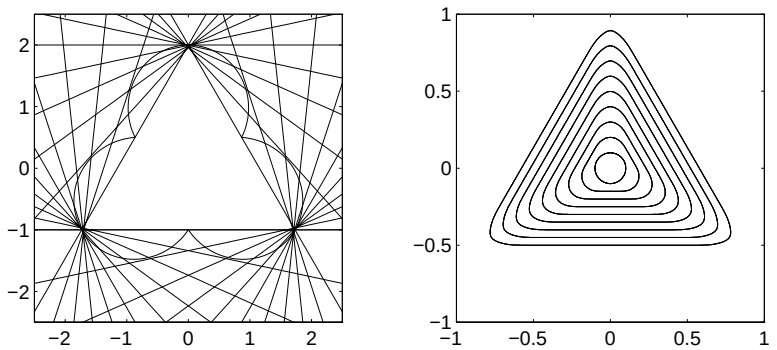

FIGURE 15. $\gamma(\nu) = 1 + |\sin(\frac{3}{2}(\nu + \frac{\pi}{2}))|$.

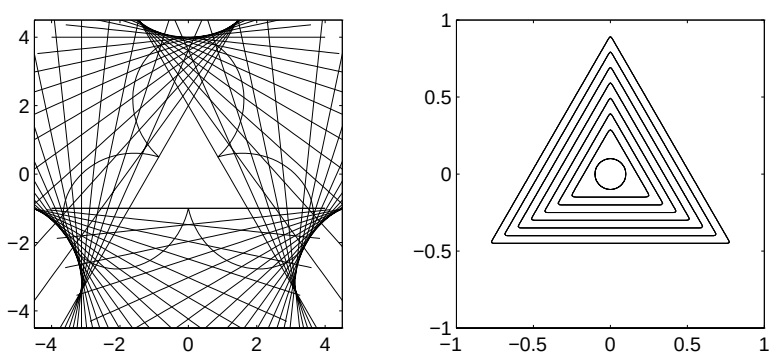

FIGURE 16. $\gamma(\nu) = 1 + 3|\sin(\frac{3}{2}(\nu + \frac{\pi}{2}))|$.

starting from a convex initial shape, this ratio decreases to its minimum as a shape grows outward normal to itself with velocity $\gamma(\nu)$, and the decreasing is strict unless the shape is the Wulff shape. This was proven for a general, not necessarily convex,

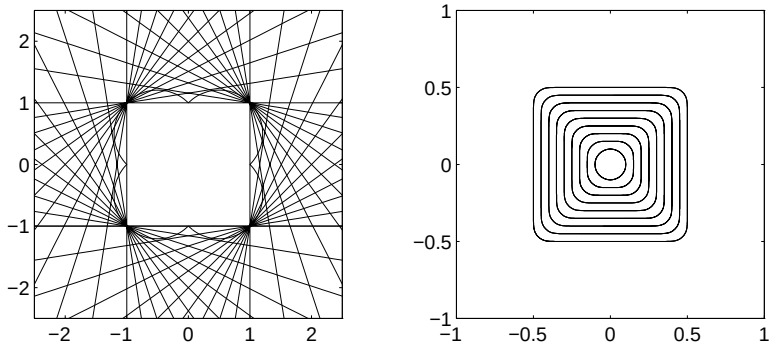

FIGURE 17. $\gamma(\nu) = |\cos(\nu)| + |\sin(\nu)|$.

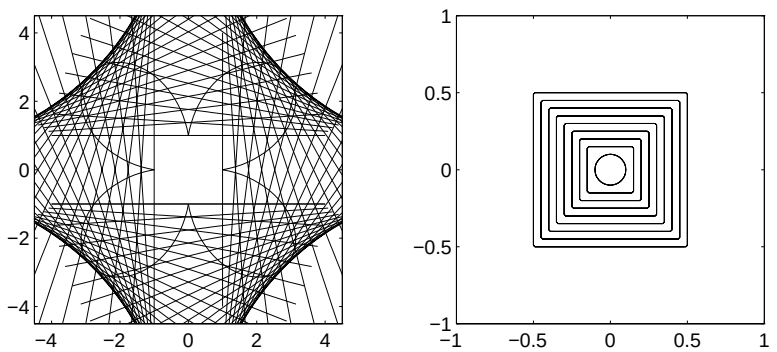

FIGURE 18. $\gamma(\nu) = 1 + 3|\sin(2\nu)|$.

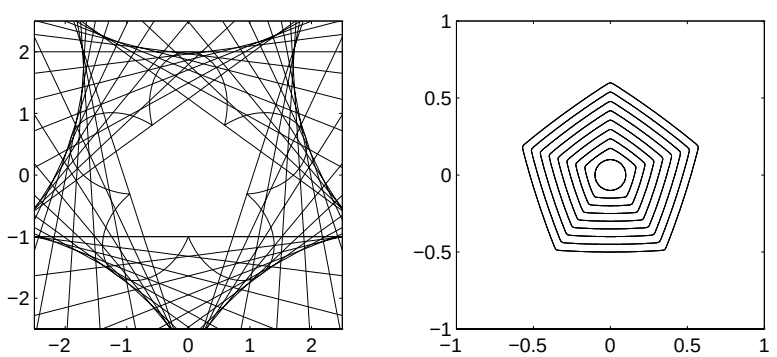

FIGURE 19. $\gamma(\nu) = 1 + |\sin(\frac{5}{2}(\nu + \frac{\pi}{2}))|$.

γ. In the level set formulation,

$$\frac{E}{V^{1-1/d}} = \frac{\int \gamma(\frac{\nabla\phi}{|\nabla\phi|})\delta(\phi)|\nabla\phi|dx}{\int H(-\phi)dx} \tag{8.10}$$

where $\delta(\phi)$ is the 1D δ function, $H(\phi)$ is 1D Heaviside function, d = 2 or 3 is the dimension.

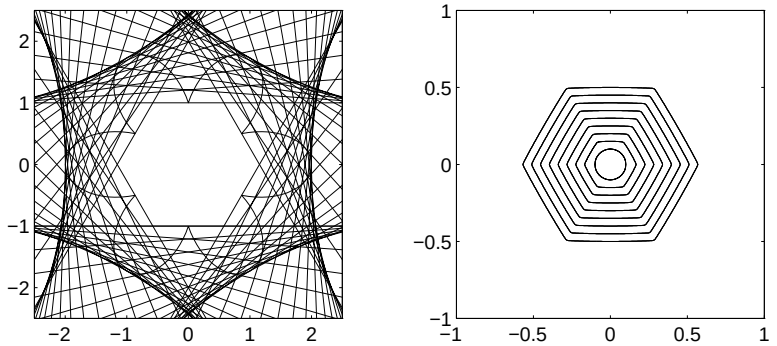

FIGURE 20. $\gamma(\nu) = 1 + |\sin(3(\nu + \frac{\pi}{2}))|$.

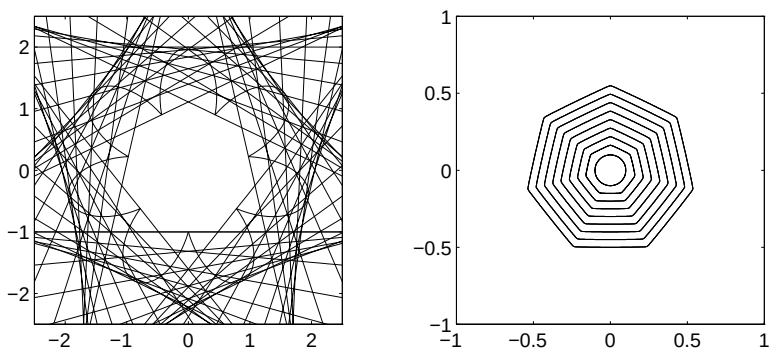

FIGURE 21. $\gamma(\nu) = 1 + |\sin(\frac{7}{2}(\nu + \frac{\pi}{2}))|$.

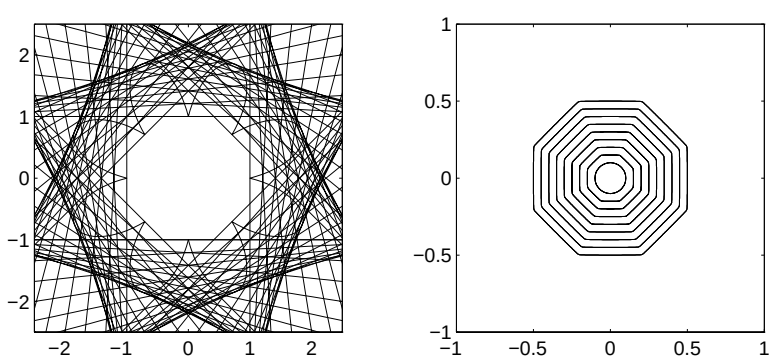

FIGURE 22. $\gamma(\nu) = 1 + |\sin(4\nu)|$.

In our computation, $\delta(\phi)$ is approximated by:

$$\delta(\phi) = \begin{cases} 0 & \text{if } |\phi| \geq \epsilon, \\ -\frac{1}{6\epsilon}(1 + \cos(\frac{\pi x}{\epsilon})) & \text{if} |\phi| \geq \frac{\epsilon}{2}, \\ -\frac{1}{6\epsilon}(1 + \cos(\frac{\pi x}{\epsilon})) + \frac{4}{3\epsilon}(1 + \cos(\frac{2\pi x}{\epsilon})) & \text{otherwise.} \end{cases} \tag{8.11}$$

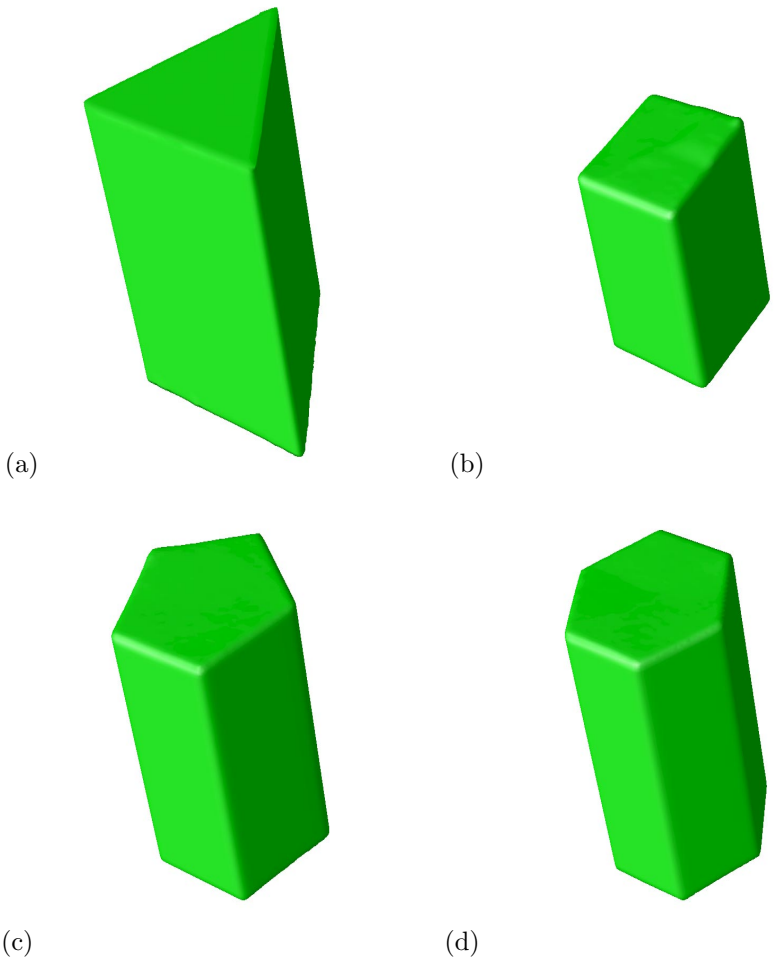

FIGURE 23. Wulff shape of prism. (a) $\gamma(\nu,\varphi) = (1+3|\sin(\frac{3}{2}(\nu+\frac{\pi}{2}))|)h_1(\varphi)$. (b) $\gamma(\nu,\varphi) = (1+|\sin(2(\nu+\frac{\pi}{2}))|)h_1(\varphi)$. (c) $\gamma(\nu,\varphi) = (1+|\sin(\frac{5}{2}(\nu+\frac{\pi}{2}))|)h_1(\varphi)$. (d) $\gamma(\nu,\varphi) = (1+|\sin(3(\nu+\frac{\pi}{2}))|)h_1(\varphi)$.

The Heaviside function $H(\phi)$ is approximated by:

$$H(\phi) = \begin{cases} 0 & \text{if } \phi \le -\epsilon, \\ -\frac{1}{6}(1+\frac{x}{\epsilon}+\frac{1}{\pi}\sin(\frac{\pi x}{\epsilon})) & \text{if } \phi \le -\frac{\epsilon}{2}, \\ -\frac{1}{6}(1+\frac{x}{\epsilon}+\frac{1}{\pi}\sin(\frac{\pi x}{\epsilon}))+\frac{1}{3}(2+\frac{4x}{\epsilon}+\frac{1}{\pi}\sin(\frac{2\pi x}{\epsilon})) & \text{if } \phi \le \frac{\epsilon}{2}, \\ -\frac{1}{6}(1+\frac{x}{\epsilon}+\frac{1}{\pi}\sin(\frac{\pi x}{\epsilon}))+\frac{4}{3} & \text{if } x \le \epsilon, \\ 1 & \text{otherwise.} \end{cases} \tag{8.12}$$

where $\epsilon = 3\Delta x$.

We start with a circle in 2D and a sphere in 3D. As was show in [**19**], the ratio decreases, and is a convex function of time. If we start from a nonconvex shape, our computations seem to show that this ratio also decreases.

Example 6. In this example, we start from a nonconvex, multiply connected shape and show how it grows, merges and finally asymptotes to a Wulff shape.

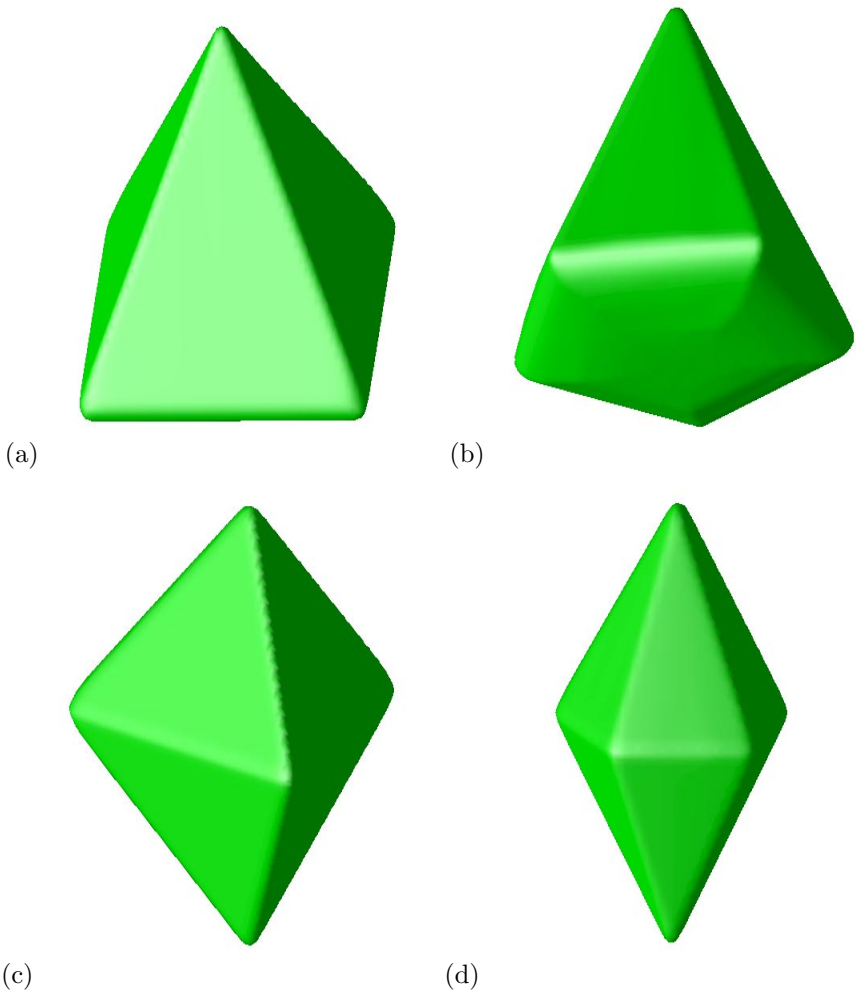

FIGURE 24. *Wulff shape of pyramid and bi-pyramid. (a)*$\gamma(\nu,\varphi) = (1+|\sin(2(\nu+\frac{\pi}{2}))|)h_2(\varphi)$. *(b)*$\gamma(\nu,\varphi) = (1+|\sin(\frac{5}{2}(\nu+\frac{\pi}{2}))|)h_2(\varphi)$. *(c)*$\gamma(\nu,\varphi) = (1+|\sin(\frac{3}{2}(\nu+\frac{\pi}{2}))|)h_3(\varphi)$. *(d)*$\gamma(\nu,\varphi) = (1+|\sin(\frac{5}{2}(\nu+\frac{\pi}{2}))|)h_3(\varphi)$.

This demonstrates the versatility and simplicity of our method. See figure 28 and 29.

9. Appendix

I. Proof of Lemma 2. In this appendix, we prove the following result stated in section 2.3.

Lemma 2 *γ is convex if and only if its homogeneous extension of degree 1 $\bar{\gamma}$: $R^d \to R^+$ is a convex function on R^d.*

FIGURE 25. *A regular 12 polygon grown from a sphere with surface energy.*

Proof: Suppose the homogeneous extension of degree 1 $\bar{\gamma} : R^d \to R^+$ is a convex function on R^d. Hence the region $K = \{ x : \bar{\gamma}(x) \leq 1 \}$ is convex. But $K = \{ x : |x|\gamma(\frac{x}{|x|}) \leq 1 \} = \{ x : |x| \leq \frac{1}{\gamma(\frac{x}{|x|})} \}$ which is the region enclosed by the polar plot of $\frac{1}{\gamma}$. By definition, γ is convex.

On the other hand, suppose $\gamma : S^{d-1} \to R^+$ is convex, i.e. $K = \{ x : |x| \leq \gamma^{-1}(\frac{x}{|x|} \}$ is convex. Note $K = \{ x : |x|\gamma(\frac{x}{|x|}) \leq 1 \} = \{ x : \bar{\gamma}(x) \leq 1 \}$ since $\bar{\gamma}$ is a degree 1 homogeneous function. We further conclude that

$$K_c = \{ x : \bar{\gamma}(x) \leq c \}$$

is convex for any $c > 0$. We claim that this implies $\bar{\gamma}$ is convex over R^d.

Refer to figure 30. Pick any two points P and Q from R^d, and a arbitrary $t \in (0, 1)$. Without loss of generality, let us assume $\bar{\gamma}(P) < \bar{\gamma}(Q)$. Let Γ_0, Γ_t and Γ_1 be the level contour of $\bar{\gamma}$ taking values $\bar{\gamma}(P), (1-t)\bar{\gamma}(P) + t\bar{\gamma}(Q)$ and $\bar{\gamma}(Q)$, respectively. Let the origin of R^d be denoted as O, and the half line OP emanating from O intersect Γ_t at T and Γ_1 at R, and the line segment $\overline{OQ}$ intersect Γ_0 at S and Γ_t at U. Denote $\alpha = \frac{\bar{\gamma}(Q)}{\bar{\gamma}(P)}$. Then since $\bar{\gamma}$ is homogeneous function of degree 1, $R = \alpha P$ and $T = (1-t)P + tR$. Note that since $\frac{|P|}{|R|} = \frac{|S|}{|Q|}$, we have $PS \parallel RQ$. Similarly, $TU \parallel RQ$. Suppose $\overline{PQ}$ intersects $\overline{TU}$ at W, then $\frac{|PW|}{|PQ|} = \frac{|PT|}{|PR|} = t$. Hence $W = (1-t)P + tQ$. Since the region K_t enclosed by Γ_t is convex, we have $W \in K_t$ and therefore $\bar{\gamma}(W) \leq \bar{\gamma}(T)$, which is $\bar{\gamma}((1-t)P+tQ) \leq (1-t)\bar{\gamma}(P)+t\bar{\gamma}(Q)$.

II. The Evolution Equation for the Normal Angle in 2D. In this section, we derive the evolution equation which governs the motion of the normal angle of a growing shape in 2D. It is stated without proof in section 5.2 for the special case when the curve is a Wulff shape. A good reference on this topic is [**12**]. Let $\mathbf{r} : S^1 \to R^2$ be a smooth simple closed curve in 2D that is parameterized by α.

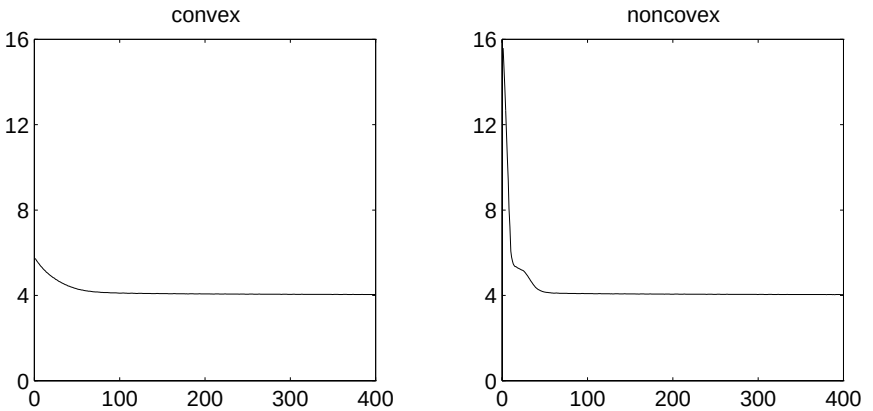

FIGURE 26. 2D.$\gamma(\nu) = |\cos(\nu)| + |\sin(\nu)|$.

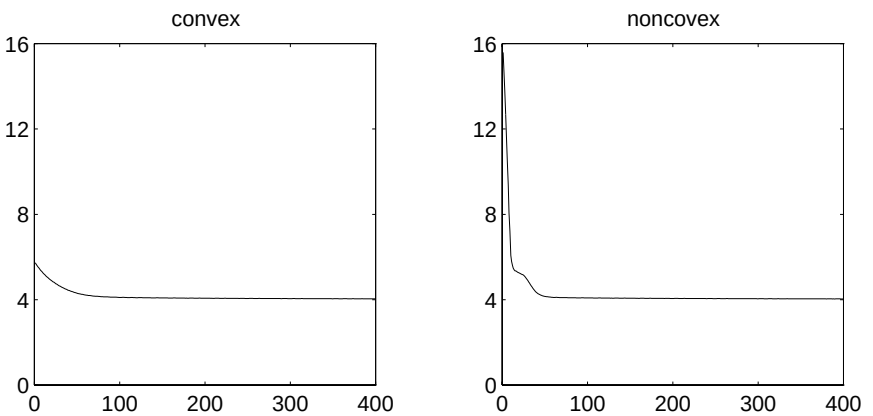

FIGURE 27. 3D.$\gamma(\hat{n}) = |n_x| + |n_y| + |n_z|$.

Let the curve move with normal velocity V, which may depend on some local or global properties of the curve. If we denote the curve at some later time t as $\mathbf{r}(t, \alpha)$, then

$$\frac{\partial \mathbf{r}}{\partial t} = V\hat{n}, \tag{9.1}$$

where the partial derivative $\frac{\partial}{\partial t}$ is taken for α fixed. Similarly, the partial derivative $\frac{\partial}{\partial \alpha}$ is taken for t fixed. We want to make this point clear since some confusion may arise in the following analysis.

Let $w(t, \alpha) = |\frac{\partial \mathbf{r}}{\partial \alpha}(t, \alpha)|$ and s be the arclength parameter, which is only defined up to a constant. However $\frac{\partial}{\partial s}$ is well defined in the following sense

$$\frac{\partial}{\partial s} = \frac{1}{w(t, \alpha)} \frac{\partial}{\partial \alpha}. \tag{9.2}$$

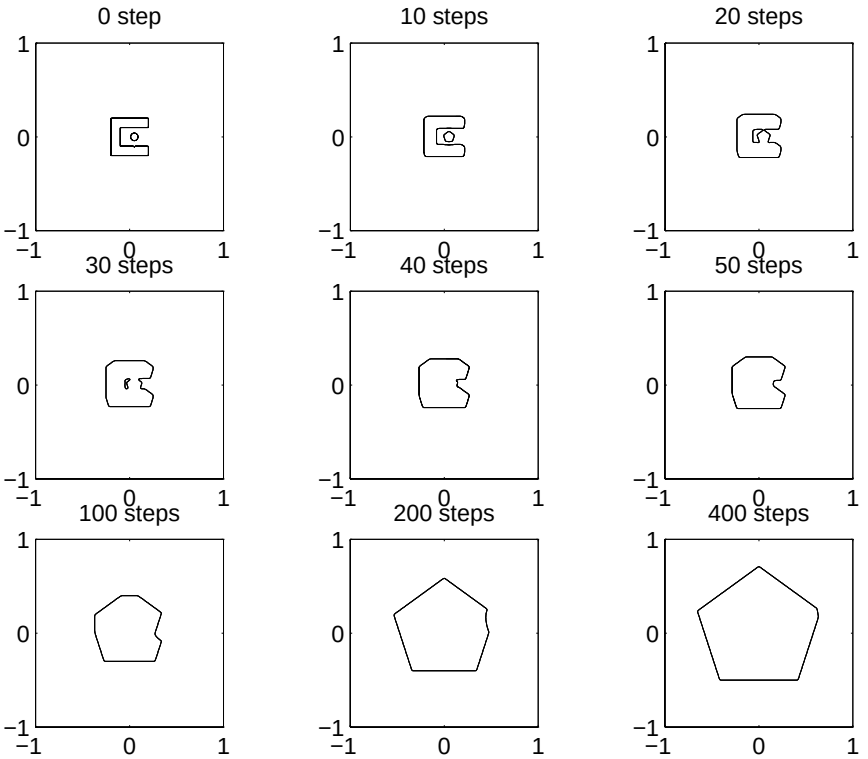

FIGURE 28. *The growing and merging of the initial nonconvex and multiply connected shape into the Wulff shape.* $\gamma(\nu) = 1 + |\sin(\frac{5}{2}(\nu + \frac{\pi}{2})|$.

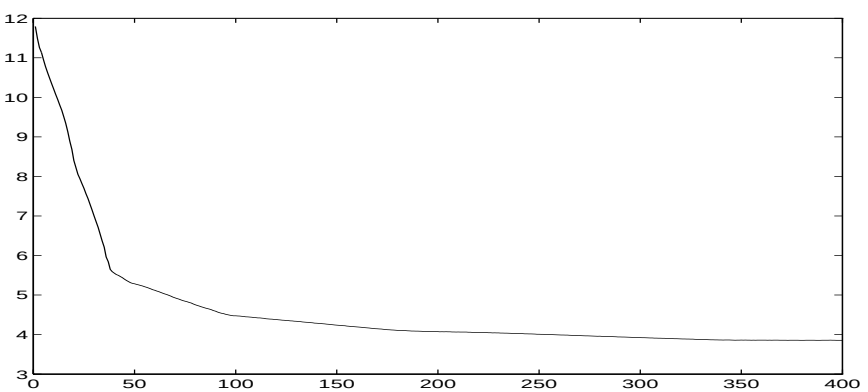

FIGURE 29. *The change of energy and area ratio in the above process.*

Note that t and s may not be independent variables, and thus

$$\frac{\partial}{\partial t}\frac{\partial}{\partial s} \neq \frac{\partial}{\partial s}\frac{\partial}{\partial t}. \tag{9.3}$$

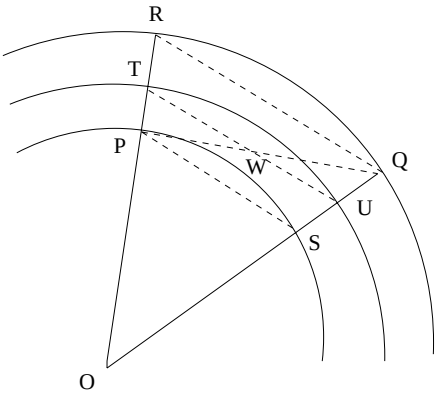

FIGURE 30. Contour of a homogeneous function of degree 1.

Let $\hat{\tau} = \frac{\partial \mathbf{r}}{\partial s}$ be the unit tangent vector, and $\hat{n}$ the unit outwards normal. Denote by θ the angle between $\mathbf{r}$ and the positive x-axis, and ν the angle between $\hat{n}$ and the positive x-axis. The Frechet formulae give us:

$$\left\{ \begin{array}{rcl} \frac{\partial \hat{\tau}}{\partial \alpha} & = & -w\kappa\hat{n}, \\ \frac{\partial \hat{n}}{\partial \alpha} & = & w\kappa\hat{\tau}. \end{array} \right. \tag{9.4}$$

Using these relations, it is easy to show that:

$$\frac{\partial w}{\partial t} = \kappa V w \tag{9.5}$$

and

$$\frac{\partial}{\partial t}\frac{\partial}{\partial s} = \frac{\partial}{\partial s}\frac{\partial}{\partial t} - \kappa V \frac{\partial}{\partial s}. \tag{9.6}$$

For proofs of the above results and more details, please refer to [**12**].

Applying the above results to $\hat{n}$ and $\hat{\tau}$, we get

LEMMA 7.

$$\frac{\partial \hat{n}}{\partial t} = -\frac{\partial V}{\partial s}\hat{\tau}, \quad \frac{\partial \hat{\tau}}{\partial t} = \frac{\partial V}{\partial s}\hat{n}. \tag{9.7}$$

Proof:

$$\begin{array}{rcl} \frac{\partial \hat{\tau}}{\partial t} & = & \frac{\partial}{\partial t}\frac{\partial \gamma}{\partial s} = \frac{\partial}{\partial s}\frac{\partial \gamma}{\partial t} - \kappa V \frac{\partial \gamma}{\partial s} \\ & = & \frac{\partial}{\partial s}(V\hat{n}) - \kappa V \hat{\tau} = \frac{\partial V}{\partial s}\hat{n} \\ 0 & = & \frac{\partial}{\partial t} < \hat{n}, \hat{\tau} > = < \frac{\partial \hat{n}}{\partial t}, \hat{\tau} > + < \hat{n}, \frac{\partial \hat{\tau}}{\partial t} > \\ & = & < \frac{\partial \hat{n}}{\partial t}, \hat{\tau} > + \frac{\partial V}{\partial s}. \end{array}$$

Hence the first equality. □

From the above Lemma, we immediately get:

$$\frac{\partial \nu}{\partial t} = -\frac{\partial V}{\partial s}. \tag{9.8}$$

Now we introduce the time/arclength coordinate system:

$$\begin{cases} \tau = t \\ s = s(t,\alpha) = \int_0^\alpha w(t,\alpha)d\alpha \end{cases} \tag{9.9}$$

In the this system, we have

$$\begin{aligned} \frac{\partial \nu}{\partial t} &= \frac{\partial \nu}{\partial \tau} + \frac{\partial \nu}{\partial s}\frac{\partial s}{\partial t}, \\ \frac{\partial s}{\partial t} &= \int_0^\alpha \frac{\partial w}{\partial t} d\alpha = \int_0^\alpha \kappa V w d\alpha \\ &= \int_0^s V\kappa ds = \int_{\nu_0}^\nu V d\nu, \end{aligned}$$

where ν_0 is the normal angle of the reference point with $\alpha = 0$. Note that

$$\frac{\partial V}{\partial s} = \frac{\partial V}{\partial \nu}\frac{\partial \nu}{\partial s}.$$

We thus obtain the evolution equation for ν in this system:

$$\frac{\partial \nu}{\partial \tau} + \left[\int_{\nu_0}^\nu V(\nu)d\nu + \frac{\partial V}{\partial \nu}\right]\frac{\partial \nu}{\partial s} = 0. \tag{9.10}$$

In the self-similar growth of Wulff crystals, the velocity $V = \gamma(\nu)$ and we can choose the reference point so that $\nu_0 \equiv 0$ and $\gamma'(0) = 0$. We replace τ by t and get

$$\frac{\partial \nu}{\partial t} + \int_0^\nu \left[\gamma(u) + \gamma''(u)\right] du \frac{\partial \nu}{\partial s} = 0. \tag{9.11}$$

This conservation law gives incorrect jump conditions at corners. The correct equation can be obtained by a change of variables in the equation (9.11) that governing the evolution of normal angle. We introduce the following new set of variables:

$$\begin{cases} \tau = t, \\ \xi = t\theta(t,s) \end{cases} \tag{9.12}$$

where $\theta(t,s)$ is defined implicitly by:

$$\frac{s}{t} = \int_0^\theta \sqrt{W^2(\theta) + W'^2(\theta)}\, d\theta = \int_0^{\nu(\theta)} \left[\gamma(\nu) + \gamma''(\nu)\right] d\nu \tag{9.13}$$

and $\nu(\theta)$ in turn is defined by

$$\theta = \nu + \tan^{-1}\left(\frac{\gamma'(\nu)}{\gamma(\nu)}\right),$$

where $W(\theta) = \gamma_*(\theta)$.

By the chain rule, we have

$$\begin{aligned} \frac{\partial \nu}{\partial t} &= \frac{\partial \nu}{\partial \tau} + \frac{\partial \nu}{\partial \xi}\frac{\partial \xi}{\partial t}, \\ \frac{\partial \nu}{\partial s} &= \frac{\partial \nu}{\partial \xi}\frac{\partial \xi}{\partial s}, \\ \frac{\partial \xi}{\partial t} &= \theta + t\frac{\partial \theta}{\partial t}, \\ \frac{\partial \xi}{\partial s} &= t\frac{\partial \theta}{\partial s}. \end{aligned}$$

We have the following

$$\begin{aligned}\frac{\partial\theta}{\partial t} &= -\frac{s}{t^2}\frac{\gamma(\nu)}{\gamma^2(\nu)+\gamma'^2(\nu)},\\ \frac{\partial\theta}{\partial s} &= \frac{1}{t}\frac{\gamma(\nu)}{\gamma^2(\nu)+\gamma'^2(\nu)}\end{aligned}$$

and thus

$$\begin{aligned}\frac{\partial\nu}{\partial t} &= \frac{\partial\nu}{\partial\tau}+\frac{\partial\nu}{\partial\xi}\left(\theta-\frac{s}{t}\frac{\gamma(\nu)}{\gamma^2(\nu)+\gamma'^2(\nu)}\right),\\ \frac{\partial\nu}{\partial s} &= \frac{\partial\nu}{\partial\xi}\frac{\gamma(\nu)}{\gamma^2(\nu)+\gamma'^2(\nu)}.\end{aligned}$$

Inserting these expression into equation (9.11), we get

$$\frac{\partial\nu}{\partial\tau}+\frac{\partial F(\nu)}{\partial\xi}=0,$$

where $F(\nu)=\frac{\nu^2}{2}+\int_0^\nu \tan^{-1}\frac{\gamma'(u)}{\gamma(u)}du$.

III. The Euler-Lagrange Equation of Surface Energy. The Lagrangian (7.3) when γ is a homogeneous function of degree 1 is of the following form

$$\mathcal{L}(\phi,\lambda)=\int\gamma(\nabla\phi)\delta(\phi)|\nabla\phi|dx-\lambda\int H(-\phi)dx. \tag{9.14}$$

Take $\psi\in C_0^\infty$, we have

$$\begin{aligned}<\frac{\delta\mathcal{L}}{\delta\phi},\psi> &\overset{def}{=} \lim_{\epsilon\to0}\frac{1}{\epsilon}[\mathcal{L}(\phi+\epsilon\psi,\lambda)-\mathcal{L}(\phi,\lambda)]\\ &= \int\{D\gamma(\nabla\phi)\cdot\nabla\psi\delta(\phi)+\gamma(\nabla\phi)\delta\prime(\phi)\psi+\delta(-\phi)\psi\}dx\\ &= \int\{-\nabla\cdot[\delta(\phi)D\gamma(\nabla)]+\gamma(\nabla\phi)\delta\prime(\phi)+\lambda\delta(\phi)\}\psi dx\\ &= \int\{-\delta\prime(\phi)\nabla\cdot D\gamma(\nabla\phi)-\delta(\phi)\nabla\cdot D\gamma(\nabla)+\gamma(\nabla\phi)\delta\prime(\phi)-\lambda\delta(\phi)\}\\ &= -\int\{\nabla\cdot D\gamma(\nabla\phi)-\lambda\}\delta(\phi)\psi dx\\ &= -\int\{\nabla\cdot D\gamma(\nabla\phi)-\lambda\}\frac{\psi}{|\nabla\phi|}\delta(\phi)|\nabla\phi|dx.\end{aligned}$$

Hence the Euler-Lagrange equation is

$$\nabla\cdot D\gamma(\nabla\phi)=\lambda, \tag{9.15}$$

where the Lagrange constant λ is chosen such that the volume enclosed is as given.

Noting that when γ is a homogeneous function of degree 1, then $D\gamma$ is homogeneous of degree 0 and hence $\gamma(\nabla\phi)=\gamma(\frac{\nabla\phi}{|\nabla\phi|})$.

For extensions of γ which are not necessarily homogeneous of degree 1, the Euler-Lagrange equation can be obtained through a similar but more involved calculation and is found to be

$$\nabla\cdot[\gamma(\hat{n})+\nabla_{\hat{n}}\gamma(\hat{n})-(\nabla_{\hat{n}}\gamma(\hat{n})\cdot\hat{n})\hat{n}]=0, \tag{9.16}$$

where $\hat{n}=\frac{\nabla\phi}{|\nabla\phi|}$ is function of space variable x, and the ∇ means gradient with respect to x, and $\nabla_{\hat{n}}$ means gradient with respect to the variables of (extended) γ.

For example, if we extend γ to be constant in the radial direction, then $\nabla_{\hat{n}}\gamma(\hat{n})\cdot\hat{n}=0$ and the Euler-Lagrange equation would be

$$\nabla\cdot[\gamma(\hat{n})+\nabla_{\hat{n}}\gamma(\hat{n})]=0, \tag{9.17}$$

which is different from equation (9.15). We will use the homogeneous extension of degree 1 next.

The gradient flow of the surface energy with the volume constraints is

$$\phi_t=|\nabla\phi|\left[\nabla D\gamma\left(\frac{\nabla\phi}{|\nabla\phi|}\right)-\lambda\right],$$

where

$$\lambda=\frac{\int\nabla\cdot\left[D\gamma(\frac{\nabla\phi}{|\nabla\phi|})\right]\delta(\phi)|\nabla\phi|dx}{\int\delta(\phi)|\nabla\phi|dx}. \tag{9.18}$$

The reason that we have included the extra term $|\nabla\phi|$ is to make the above equation rescaling invariant, i.e. ϕ can be replaced by $h(\phi)$ with $h'>0$ and $h(0)=0$.

The surface energy on the gradient flow is diminishing. To see this, let

$$\mathcal{F}=\nabla\cdot D\gamma(\nabla\phi)$$

and we have

$$\begin{aligned}\frac{dE}{dt} &= \frac{d}{dt}\int\gamma(\nabla\phi)\delta(phi)dx\\ &= -\int\mathcal{F}(\mathcal{F}-\lambda)\delta(phi)|\nabla\phi|dx\\ &= -\int_\gamma\mathcal{F}(\mathcal{F}-\lambda)dA\end{aligned}$$

where $dA=\delta(\phi)|\nabla\phi|dx$ is area element in 3D and arclength element in 2D. Using the Schwarz inequality

$$|\int_\gamma\mathcal{F}dA|^2\le\int_\gamma\mathcal{F}^2dA\int_\gamma dA, \tag{9.19}$$

one easily sees that

$$\frac{dE}{dt}\le 0. \tag{9.20}$$

References

[1] D. Adalsteinsson and J. A. Sethian, *A Fast Level Set Method for Propagating Interfaces*, J. Comput. Phys., v118, pp. 269-277, 1995

[2] J. E. Brothers and F. Morgan, *The Isoperimetric Theorem for General Integrands*, Michigan Math. J., 41 (1994), no. 3, pp. 419-431.

[3] J.W. Cahn, J. E. Taylor, and C. A. Handwerker, *Evolving Crystal Forms: Frank's Characteristics Revisited*, a chapter in Sir Charles Frank, OBE, FRS: An Eightieth Birthday Tribute, eds. R. G. Chambers, J. E. Enderby, A. Keller, A. R. Lang and J.W. Steeds, published by Adam Hilger Co., Bristol, England, 1991.

[4] A.A. Chernov, *The Kinetics of the Growth Form of Crystals*, Soviet Physics Crystallography, v7 (1963), pp.728-730, translated from Krystallografiya, v7 (1962) pp. 895-898.

[5] A.A. Chernov, *Modern Crystallography III, Crystal Growth*, Springer-Verlag, Berlin (1984).

[6] R. Courant and K.-O. Friedrichs, Supersonic Flow and Shock Waves, Wiley, Interscience, 1962.

[7] R. Caflisch, M. Gyurhe, B. Merriman, S. Osher, C. Ratsch, D. Vredensky and J. Zinck, *Island Dynamics and the Level Set Method for Epitaxial Growth*, Applied Math. Letters, 1999, to appear.

[8] M.G. Crandall and P.L. Lions, *Two approximations of solutions of Hamilton-Jacobi equations*, Math. Comput., v43, 1984, pp. 1-19.

[9] I. Fonseca, *The Wulff Theorem Revisited*, Proc. Royal Soc. London A, v432 (1991) pp. 125-145

[10] F.C. Frank, *The Geometrical Thermodynamics of Surfaces, in Metal Surfaces: Structure, Energies and Kinetics*, Am. Soc. Metals, Metals Park, Ohio, 1963.

[11] M. E. Gurtin, Thermomechanics of Evolving Phase Boundaries in the Plane, Clarendon Press, Oxford, 1993.

[12] M. Gage and R. S. Hamilton, *The Heat Equation Shrinking Convex Plane Curves*, J. Differental Geometry, 23(1986) 69-96.

[13] C. Herring, chapter in *The Physics of Powder Metallurgy*, edited by W. E. Kingston, McGraw-Hill Book Co., New York, 1951.

[14] G. S. Jiang and D. Peng, *WENO Schemes for Hamilton-Jacobi equations*, UCLA CAM report 97-29, 1997. To appear in SIAM J. Sci. Comput.

[15] S. Osher, *The Riemann Problem for Nonconvex Scalar Conservation Laws and Hamilton-Jacobi Equations*, Proc. Amer. Math. Soc., v 89, pp 641-645, 1983.

[16] S. Osher, *Riemann Solvers, the Entropy Condition and Difference Approximation*. SIAM J. Numer. Anal. 21(1984) pp217-235.

[17] S. Osher and J. Sethian, *Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formulations*, J. Comput. Phys., v79, 1988, pp. 12-49.

[18] S. Osher and C-W. Shu, *High-order essentially non-oscillatory schemes for Hamilton-Jacobi equations*, J. Numer. Anal., v28, 1991, pp. 907-922.

[19] S. Osher and B. Merriman, *The Wulff Shape as the Asymptotic Limit of a Growing Crytalline Interface*. Asian J. Math., v1, no. 3, pp560-571, Sept. 1997.

[20] D. Peng, B. Merriman, S. Osher, H-K Zhao and M. Kang *A PDE Based Fast Local Level Set Method*. UCLA CAM Report 98-25. Submitted to J. Comp. Phys.

[21] B. Riemann, *Über die Forteflanzung ebener Luftwellen von enolicher Schwingungsweite, Abhandlungen der Gessellschaft der Wissenschafter zu Göttingen*, Mathematish-Physikalische Klasse 8, v. 43, 1860.

[22] S. Ruuth and B. Merriman, *Convolution Generated Motion and Generalized Huygens' Principles for Interface Motion*. UCLA CAM Report 98-4.

[23] C-W. Shu, *Total-Variation-Diminishing Time Discretization*, SIAM J. Sci. Stat. Comput. v9, pp. 1073-1084, 1988.

[24] C-W. Shu and S. Osher, *Efficient implementation of essentially non-oscillatory shock-capturing schemes II*, J. Comput. Phys., v83, 1989, pp. 32-78.

[25] J. A. Sethian, *A Fast Marching Level Set Method for Monotonically Advancing Fronts* Proc. Nat. Acad. Sci., v93, 4, pp. 1591-1595, 1996.

[26] M. Sussman, P. Smereka and S. Osher, *A Level Set Approach for Computing Solutions to Incompressible Two-Phase Flow*, vol. 114, no. 1, Sept. 1994, pp. 146-159.

[27] J. E. Taylor, *Existence and Structure of Solutions to a Class of Non-elliptic Variational Problems*. Symposia Mathematica v 14, pp 499-508, 1974.

[28] G. Wulff, *Zur Frage der Geschwindigkeit des Wachsthums und der Auflösung der Krystallflachen*, Zeitschrift für Krystallographie und Minerologie, v 34, pp 449-530, 1901.

Department of Mathematics, University of California, Los Angeles, CA 90095-1555
E-mail address: dpeng@math.ucla.edu

Department of Mathematics, University of California, Los Angeles, CA 90095-1555
E-mail address: sjo@math.ucla.edu

Department of Mathematics, University of California, Los Angeles, CA 90095-1555
E-mail address: barry@math.ucla.edu

Department of Mathematics, University of California, Los Angeles, CA 90095-1555
Current address: Department of Mathematics, Stanford University, Stanford, CA 74305-2125.
E-mail address: zhao@math.stanford.edu