Abel Symposia 7

Helge Holden
Kenneth H. Karlsen *Editors*

Nonlinear Partial Differential Equations

The Abel Symposium 2010

Springer

ABEL SYMPOSIA

Edited by the Norwegian Mathematical Society

For further volumes:
www.springer.com/series/7462

Helge Holden • Kenneth H. Karlsen
Editors

Nonlinear Partial Differential Equations

The Abel Symposium 2010

ABEL PRISEN

Springer

Editors
Helge Holden
Dept. of Mathematics
Norwegian University of Science and Technology
Trondheim
Norway

Kenneth H. Karlsen
Centre of Mathematics for Applications
University of Oslo
Oslo
Norway

ISBN 978-3-642-25360-7 e-ISBN 978-3-642-25361-4
DOI 10.1007/978-3-642-25361-4
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012930729

Mathematics Subject Classification (2010): 35-XX, 34-XX, 76-XX, 65-XX, 81-XX, 70-XX

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface to the Series

The Niels Henrik Abel Memorial Fund was established by the Norwegian government on January 1, 2002. The main objective is to honor the great Norwegian mathematician Niels Henrik Abel by awarding an international prize for outstanding scientific work in the field of mathematics. The prize shall contribute towards raising the status of mathematics in society and stimulate the interest for science among school children and students. In keeping with this objective the Board of the Abel Fund has decided to finance an annual Abel Symposium. The topic may be selected broadly in the area of pure and applied mathematics. The Symposia should be at the highest international level, and serve to build bridges between the national and international research communities. The Norwegian Mathematical Society is responsible for the events. It has also been decided that the contributions from these Symposia should be presented in a series of proceedings, and Springer Verlag has enthusiastically agreed to publish the series. The board of the Niels Henrik Abel Memorial Fund is confident that the series will be a valuable contribution to the mathematical literature.

Helge Holden
Chairman of the board of the Niels Henrik Abel Memorial Fund

Preface

The topic of the 2010 Abel Symposium was *Nonlinear Partial Differential Equations*. The study of differential equations is of fundamental importance in mathematics and in almost all of the applications of mathematics in natural sciences, economics, and engineering. This area of mathematics is currently in the midst of an unprecedented development worldwide. Differential equations are used to model phenomena of increasing complexity, and in areas that have traditionally been outside the realm of mathematics. New analytical tools and mathematical theories, coupled with new numerical methods, are dramatically improving our understanding of nonlinear models. Nonlinearity gives raise to solutions having singularities, oscillations, or concentration effects, which in the real world are reflected in the appearance of shock waves, turbulence, material defects, etc. These effects frequently require new techniques, and offer challenging novel problems for mathematicians. On the other hand, new mathematical developments provide (numerical) solutions and new insight in many applications. The purpose of these Abel Symposium proceedings is to present a selection of the latest exciting results by world leading researchers in the area of nonlinear partial differential equations.

The Abel Symposium was hosted at the Norwegian Academy of Science and Letters, Oslo, from September 28 to October 2, 2010. Attendance was by invitation only, and the symposium had a total of 74 participants, out of which 32 were from Norwegian universities. The Scientific Committee consisted of Alberto Bressan (Penn State), Helge Holden (Trondheim), Kenneth H. Karlsen (Oslo), Sergiu Klainerman (Princeton), and Eitan Tadmor (Maryland).

Talks were presented by

Luigi Ambrosio (Pisa)
Alberto Bressan (Penn State)
Luis A. Caffarelli (Texas)
Gui-Qiang Chen (Oxford)
Camillo De Lellis (Zürich)
Maria J. Esteban (Paris)
Eduard Feireisl (Prague)

Gerhard Huisken (Golm)
Carlos Kenig (Chicago)
Alex Kiselev (Madison)
Sergiu Klainerman (Princeton)
Robert V. Kohn (New York)
Pierre-Louis Lions (Paris)
Andrew J. Majda (New York)
Igor Rodnianski (Princeton)
Laure Saint-Raymond (Paris)
Eitan Tadmor (Maryland)
Juan Luis Vazquez (Madrid)
Cédric Villani (Lyon)

These proceedings include most of the lectures presented at the Abel Symposium, and the editors appreciate the efforts made by the speakers to present their talks in these proceedings.

Participants were:

Andreianov, Boris
Berre, Inga
Borluk, Handan
Bouchut, François
Brenier, Yann
Carrillo, José A.
Celledoni, Elena
Christiansen, Snorre H.
Colombo, Rinaldo M.
Dafermos, Constantine
Engquist, Bjørn
Evje, Steinar
Fuchs, Franz
Golse, François
Grong, Erlend
Hanche-Olsen, Harald
Jakobsen, Espen R.
Kalisch, Henrik
Kalise, Dante
Karper, Trygve
Klausen, Runhild Aae
Kvamsdal, Trond
Kværnø, Anne
Langtangen, Hans Petter
Li, Qifan
Lie, Knut Andreas
Lindqvist, Peter
Lyubarskii, Yurii
Malinnikova, Eugenia
Marcati, Pierangelo
Markina, Irina
Munthe-Kaas, Hans
Nordbotten, Jan
Nørsett, Syvert P.
Owren, Brynjulf
Raynaud, Xavier
Risebro, Nils Henrik
Rønquist, Einar
Sande, Hilde
Selberg, Sigmund
Serre, Denis
Signahl, Mikael
Svanstedt, Nils
Süli, Endre
Titi, Edriss S.
Trivisa, Konstantina
Tveito, Aslak
Vasiliev, Alexander
Wasylewicz, Agnieszka
Winther, Ragnar
Xin, Zhouping
Yaguchi, Takaharu
Zanna, Antonella

We gratefully acknowledge financial support from the Niels Henrik Abel Memorial Fund, The Centre for Advanced Study at the Norwegian Academy of Science and Letters, as well as the Centre of Mathematics for Applications at the University of Oslo. The kind assistance of Ms Marit F. Strøm is greatly appreciated. Thanks go to Marius Thaule for his assistance in the preparation of these proceedings.

Trondheim, Norway — Helge Holden
Oslo, Norway — Kenneth H. Karlsen

Contents

Convergence of Wigner Transforms in a Semiclassical Limit

Luigi Ambrosio

Abstract We prove convergence of the Wigner transforms of solutions to the Schrödinger equation, in a semiclassical limit, to solutions to the Liouville equation. We are able to include in our convergence result rough or singular potentials (with Coulomb repulsive singularities), provided convergence is understood for "almost all" initial data. The rigorous statement involves a suitable extension of the DiPerna–Lions theory to the infinite-dimensional space of probability measure, where both the Wigner and the Liouville dynamics can be read.

1 Introduction

In this paper, which reflects with minor changes the talk given in Oslo, I would like to illustrate the content of the papers [4, 5]. The goal of these papers is a rigorous derivation of classical dynamics as a limit of quantum dynamics, based on Schrödinger's equation (semiclassical limit)

$$i\varepsilon\partial_t\psi_t^\varepsilon = -\frac{\varepsilon^2}{2}\Delta\psi_t^\varepsilon + U\psi_t^\varepsilon.$$

The main new ingredient, with respect to Gerard [11], Lions–Paul [12], is the introduction of physically relevant potentials as (here $n = 3M$ and $x_i \in \mathbb{R}^3$)

$$U(x_1,\ldots,x_M) = \sum_{1\le\alpha<\beta\le M}\frac{Z_\alpha Z_\beta}{|x_\alpha - x_\beta|} + U_b(x), \quad \text{with } U_b \text{ bounded, Lipschitz.}$$

In the first paper, assuming that U_b is also C^1, we show the validity of the Liouville equation in the semiclassical limit. In the second paper [5] we relax the assumption on U_b and study the problem of uniqueness of the limit, i.e. full convergence as $\varepsilon \to 0$. This requires an extension of the theory of Lagrangian flows to the case when the state space is not $\mathbb{R}^n_x \times \mathbb{R}^n_p$ but $\mathscr{P}(\mathbb{R}^n_x \times \mathbb{R}^n_p)$, see also the CRAS note [3], where this extension is briefly presented.

L. Ambrosio (✉)
Scuola Normale Superiore, Pisa, Italy
e-mail: l.ambrosio@sns.it

H. Holden, K.H. Karlsen (eds.), *Nonlinear Partial Differential Equations*,
Abel Symposia 7, DOI 10.1007/978-3-642-25361-4_1,

Let us consider solutions $\psi^\varepsilon(t,x)=\psi^\varepsilon_t(x)$ of the linear Schrödinger equations

$$\begin{cases} i\varepsilon\partial_t\psi^\varepsilon_t = -\dfrac{\varepsilon^2}{2}\Delta\psi^\varepsilon_t + U\psi^\varepsilon_t := H_\varepsilon\psi^\varepsilon_t, \\ \psi^\varepsilon_0 = \psi_{0,\varepsilon}. \end{cases} \tag{SE$_\varepsilon$}$$

Here U satisfies standard Kato conditions, hence H_ε is self-adjoint in $L^2(\mathbb{R}^n;\mathbb{C})$ with domain $H^2(\mathbb{R}^n;\mathbb{C})$, the Cauchy problem is well posed and norms and scalar products are preserved.

Assuming that

$$\int_{\mathbb{R}^n} |\psi_{0,\varepsilon}|^2(x)\,dx = 1$$

the goal is to describe the limit of ψ^ε_t (more precisely of their Wigner transforms) when $\varepsilon\to 0$.

The Wigner transform maps $L^2(\mathbb{R}^n;\mathbb{C})$ in $L^\infty(\mathbb{R}^{2n};\mathbb{C})$:

$$W_\varepsilon\psi(x,p) := \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n}\psi\left(x+\frac{\varepsilon}{2}y\right)\overline{\psi\left(x-\frac{\varepsilon}{2}y\right)}e^{-ipy}\,dy.$$

By symmetry in y, $W_\varepsilon\psi$ is real valued. In addition, its (formal, because W_ε need not be integrable) marginals are nonnegative, the so-called position and momentum densities:

$$\int_{\mathbb{R}^n} W_\varepsilon\psi(x,p)\,dp = |\psi|^2(x),$$

$$\int_{\mathbb{R}^n} W_\varepsilon\psi(x,p)\,dx = \left(\frac{1}{2\pi\varepsilon}\right)^n\left|\mathcal{F}\psi\left(\frac{p}{\varepsilon}\right)\right|^2,$$

where $\mathcal{F}$ stands for Fourier transform.

An elementary computation, going back to Wigner, shows that $W_\varepsilon\psi^\varepsilon_t$ solve, in the sense of distributions in $\mathbb{R}\times\mathbb{R}^{2n}$,

$$\partial_t W_\varepsilon\psi^\varepsilon_t + p\cdot\nabla_x W_\varepsilon\psi^\varepsilon_t = \mathscr{E}_\varepsilon(U,\psi^\varepsilon_t),$$

where $\mathscr{E}_\varepsilon(U,\psi)(x,p)$ is given by

$$-\frac{i}{(2\pi)^n}\int_{\mathbb{R}^n}\left[\frac{U(x+\frac{\varepsilon}{2}y)-U(x-\frac{\varepsilon}{2}y)}{\varepsilon}\right]\psi\left(x+\frac{\varepsilon}{2}y\right)\overline{\psi\left(x-\frac{\varepsilon}{2}y\right)}e^{-ipy}\,dy.$$

Adding and subtracting $\nabla U(x)\cdot y$ in the term between square brackets and using $ye^{-ip\cdot y}=i\nabla_p e^{-ip\cdot y}$ we get

$$\mathscr{E}_\varepsilon(U,\psi) = \nabla U(x)\cdot\nabla_p W_\varepsilon\psi + \mathscr{E}'_\varepsilon(U,\psi),$$

where $-i(2\pi)^n \mathscr{E}'_\varepsilon(U,\psi)(x,p)$ is given by

$$\int\left[\frac{U(x+\frac{\varepsilon}{2}y)-U(x-\frac{\varepsilon}{2}y)}{\varepsilon}-\langle\nabla U(x),y\rangle\right]\psi\left(x+\frac{\varepsilon}{2}y\right)\overline{\psi\left(x-\frac{\varepsilon}{2}y\right)}e^{-ipy}\,dy.$$

Hence, $W_\varepsilon\psi_t^\varepsilon$ solves the continuity equation with velocity (Hamiltonian, divergence-free)

$$\boldsymbol{b}(x,p)=\big(p,-\nabla U(x)\big)$$

and right hand side $\mathscr{E}'_\varepsilon(U,\psi_t^\varepsilon)$:

$$\partial_t W_\varepsilon\psi_t^\varepsilon+\nabla_{x,p}\cdot\big(\boldsymbol{b}W_\varepsilon\psi_t^\varepsilon\big)=\mathscr{E}'_\varepsilon(U,\psi_t^\varepsilon).$$

The goal is to show that the right hand side is infinitesimal, at least in the duality with nice test functions ϕ. In the study of $\mathscr{E}'_\varepsilon(U,\psi)$, Coulomb singularities are a source of difficulty. But, even if U were Lipschitz, the analysis would not be trivial. Indeed, denoting by $\Delta_\varepsilon(\cdot,z)$ the difference between ε-difference quotient of U along z and the partial derivative along z, it holds

$$\int\mathscr{E}'_\varepsilon(U,\psi)\phi\,dx\,dp=\int\int\Delta_\varepsilon(x,y)\psi\left(x+\frac{\varepsilon}{2}y\right)\overline{\psi\left(x-\frac{\varepsilon}{2}y\right)}\mathcal{F}_p\phi(x,y)\,dx\,dy.$$

The basic idea is to use the decay of $\mathcal{F}_p\phi(x,y)$ per y "large" ($|y|\sqrt{\varepsilon}>1$) and the differentiability of U for y "small". But, if $\Delta_\varepsilon(x,z)\to 0$ as $\varepsilon\downarrow 0$ *only* for $\mathscr{L}^n$-a.e. x, we need that $|\psi_t^\varepsilon|^2$ does not concentrate as $\varepsilon\to 0$!

On the other hand, this estimate is not compatible with families of initial conditions $\psi_{0,\varepsilon}$ in $(\mathrm{SE})_\varepsilon$ displaying concentration of positions and/or momentum (in some respect, the most natural when studying a semiclassical limit). An example

$$\psi_0^\varepsilon(x)=\varepsilon^{-n\alpha/2}\phi_0\left(\frac{x-w}{\varepsilon^\alpha}\right)e^{i(q\cdot x/\varepsilon)},\quad \alpha\in(0,1],\ w\in\mathbb{R}^n,\ q\in\mathbb{R}^n.$$

In this case

$$\lim_{\varepsilon\downarrow 0}W_\varepsilon\psi_0^\varepsilon\,dx\,dp=d\delta_{(w,q)}(x,p)\quad\forall\alpha\in(0,1),$$

$$\lim_{\varepsilon\downarrow 0}W_\varepsilon\psi_0^\varepsilon\,dx\,dp=d\delta_w(x)\times|\mathcal{F}\phi_0|^2(p-q)\,dp,\quad\alpha=1.$$

We would like to include potentials U of the form $U=U_{nn}+U_{ne}$, with

$$U_{nn}(x)=\sum_{1\le\alpha<\beta\le M}\frac{Z_\alpha Z_\beta}{|x_\alpha-x_\beta|},$$

and ($n=3M$, $M=$ number of atomic nuclei, $N=$ number of electrons)

$$U_{ne}(x):=\inf\left\{\langle\varphi,H_{ne}(x)\varphi\rangle:\int_{(\mathbb{R}^3\times\mathbb{Z}_2)^N}|\varphi|^2\,dy=1\right\}.$$

Here $H_{ne}(x)$ is the operator in $\mathbb{R}^{3N}$ given by

$$H_{ne}(x) := \sum_{i=1}^{N}\left(-\frac{1}{2}\Delta_{y_i} - \sum_{\alpha=1}^{M}\frac{Z_\alpha}{|y_i - x_\alpha|}\right) + \sum_{1\le i<j\le N}\frac{1}{|y_i - y_j|}.$$

Since U_{ne} is defined by a minimization problem, the best we can hope for is that U_{ne} is bounded and Lipschitz, and that $\nabla U_{ne} \in BV_{\mathrm{loc}}$. Indeed, at points where more than one minimizer occurs we expect a jump discontinuity in the gradient (this corresponds to the so-called eigenvalue crossings, whose structure can be in some cases studied in detail, see [8, 9]). The first convergence result reads as follows:

Theorem 1 (A, Friesecke, Giannoulis, [4]) *Assume that* $U = U_{nn} + U_b$, *with* U_b *bounded, Lipschitz and* C^1. *Let* $\psi_{0,\varepsilon}$ *be such that* $|\psi_{0,\varepsilon}|^2$ *is equi-tight and*

$$\sup_\varepsilon \int_{\mathbb{R}^n}\left|\left(-\frac{\varepsilon^2}{2}\Delta + U\right)\psi_{0,\varepsilon}\right|^2 dx < \infty.$$

Then the family of Wigner transforms has limit points (in a suitable weak topology in phase space, pointwise in time) as $\varepsilon \to 0$, *and any limit point* $\mu_t \in \mathscr{P}(\mathbb{R}^{2n})$ *is concentrated in* $\mathbb{R}^{2n} \setminus \Sigma$ *and satisfies*

$$\frac{d}{dt}\mu_t + \nabla\cdot\big((p, -\nabla U(x))\mu_t\big) = 0 \quad \text{in } \mathbb{R}\times\mathbb{R}^{2n}.$$

In the statement of the theorem

$$\Sigma := \bigcup_{\alpha<\beta}\{x_\alpha = x_\beta\}$$

is the set of Coulomb singularities. The key point is to show not only that μ_t do not charge Σ, but even the validity of the Liouville equation *up to* Σ. The proof uses the pointwise estimate $|\nabla U_{nn}| \le C U_{nn}^2$ and the uniform bound on $\int |H_\varepsilon \psi_{0,\varepsilon}|^2\, dx$ (propagated in time) to get

$$\sup_\varepsilon \sup_{t\in\mathbb{R}} \int U_{nn}^2 |\psi_t^\varepsilon|^2\, dx < \infty. \tag{$*$}$$

The estimate ($*$) is much stronger than the one given by energy conservation:

$$\frac{1}{2}\int_{\mathbb{R}^n}\varepsilon^2|\nabla\psi_t^\varepsilon|^2 + U|\psi_t^\varepsilon|^2\, dx = \frac{1}{2}\int_{\mathbb{R}^n}\varepsilon^2|\nabla\psi_{\varepsilon,0}|^2 + U|\psi_{\varepsilon,0}|^2\, dx$$

and it depends in a very specific way on the Coulomb structure of U_{nn}.

In the second paper [5] our goals have been the relaxation from C^1 to Lipschitz of the assumptions on U_b (in such a way to include more general potentials U_b as U_{en}) and a to achieve a *full* convergence as $\varepsilon \to 0$.

The second goal basically amounts to look for well-posedness results for *measure-valued* solutions μ_t to the Liouville equation

$$\frac{d}{dt}\mu_t + \nabla\cdot\big((p, -\nabla U(x))\mu_t\big) = 0 \quad \text{in } \mathbb{R}\times\mathbb{R}^{2n}$$

without assuming that ∇U is Lipschitz (in our model, at most we can hope for $\nabla U \in BV$ out of Coulomb singularities).

In the next sections we shall describe more carefully the ideas underlying the proof of these convergence results.

2 Well Posedness of the Liouville Equation and Flows

The first seminal results are due to DiPerna–Lions [7], under Sobolev regularity assumptions on ∇U.

Theorem 2 (Bouchut, [6]) *Assume* $\boldsymbol{b}(x,p) = (p, -\nabla U(x))$ *with* U *Lipschitz and* $\nabla U \in BV_{\mathrm{loc}}(\mathbb{R}^n;\mathbb{R}^n)$. *Then the Liouville equation is well posed in*

$$L^\infty\big([0,T]; L^1\cap L^\infty(\mathbb{R}^{2n})\big).$$

Under these assumptions on U it is hard to imagine well-posedness results in the class of measure-valued solutions. A basic difficulty is, for instance, the definition of moments $\boldsymbol{b}\mu_t$ when μ_t has a singular part with respect to $\mathscr{L}^{2n}$. On the other hand, thinking of the Liouville equation as an infinite-dimensional ODE in $\mathscr{P}(\mathbb{R}^{2n})$ (with constant, but rough coefficients), we can still try to obtain a unique *flow* of solutions in $\mathscr{P}(\mathbb{R}^{2n})$, using the finite-dimensional theory as an analogy. In this theory, due to DiPerna–Lions [7], one does not try to show that the solution to the ODE

$$\begin{cases} \dot{x}(t) = \boldsymbol{c}_t(x(t)), \\ x(0) = x_0, \end{cases} \qquad \boldsymbol{c}: [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$$

is unique and stable for a *specific* $x_0 \in \mathbb{R}^d$; rather one looks at the family of solutions as a whole, through the concept of flow. This point of view leads to very natural existence and uniqueness results, in fluid dynamics and in the theory of conservation laws, relating the “Eulerian” and “Lagrangian” viewpoints even when the velocity is not so smooth.

I will present the axiomatization of the theory given in [1] and [2], more flexible than the original one, based on the transport equation. The use of the continuity equation, instead, allows to deal (at least in some cases) with vector fields with unbounded divergence and highlights more the connections with the probabilistic viewpoint.

Definition 1 (ν-Regular flow) Let $\boldsymbol{X}(t,x): [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ and $\nu \in \mathscr{P}(\mathbb{R}^d)$ with $\nu \ll \mathscr{L}^d$ and bounded density. We say that $\boldsymbol{X}(t,x)$ is a ν-RF in $\mathbb{R}^d$ (relative a $\boldsymbol{c}$) if:

(i) for ν-a.e. x, the path $t \mapsto \boldsymbol{X}(t,x)$ is an integral absolutely continuous solution of the ODE $\dot\gamma(t) = \boldsymbol{c}_t(\gamma(t))$ in $[0,T]$ with $\boldsymbol{X}(0,x) = x$;
(ii) $\boldsymbol{X}(t,\cdot)_\sharp \nu \le C\mathscr{L}^d$ for $t \in [0,T]$, for some constant C independent of t.

The second condition, equivalent to

$$\int_{\mathbb{R}^d} \Phi(\boldsymbol{X}(t,x))\, d\nu(x) \le C \int_{\mathbb{R}^d} \Phi(z)\, dz, \quad \Phi \in C_c(\mathbb{R}^d),\ \Phi \ge 0$$

is crucial: it ensures, among other things, invariance of the concept with respect to modifications of $\boldsymbol{c}$ in $\mathscr{L}^{1+d}$-negligible sets. Indeed, if $\tilde{\boldsymbol{c}}$ is a modification of $\boldsymbol{c}$, thanks to Fubini's theorem we see that the set

$$\left\{x \in \mathbb{R}^d :\ \mathscr{L}^1\left(\{t \in [0,T] :\ \boldsymbol{X}(t,x) \in \{\boldsymbol{c} \ne \tilde{\boldsymbol{c}}\}\}\right) > 0\right\}$$

is ν-negligible; as a consequence the validity of (i) for $\boldsymbol{c}$ is transferred to (ii).

Theorem 3 [2] *Let $\boldsymbol{c} : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ be locally integrable and assume that the continuity equation $\partial_t w_t + \nabla \cdot (\boldsymbol{c} w_t) = 0$ is well posed in*

$$L^\infty_+\left([0,T]; L^1 \cap L^\infty(\mathbb{R}^{2n})\right).$$

Then, for all $\nu \ll \mathscr{L}^d$ with bounded density the ν-RF exists. It is even unique: if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are $f\mathscr{L}^d$-RF and $g\mathscr{L}^d$-RF respectively, then

$$\boldsymbol{X}(\cdot,x) = \boldsymbol{Y}(\cdot,x) \quad \textit{for } \mathscr{L}^d\textit{-a.e. } x \in \{f > 0\} \cap \{g > 0\}.$$

Thanks to this, by an exhaustion procedure we can even define a (unique) $\mathscr{L}^d$-RF. Let us now transpose these concepts from $\mathbb{R}^d$ to $\mathscr{P}(\mathbb{R}^d)$; here the main difficulty is the role played by $\mathscr{L}^d$, since no canonical measure in $\mathscr{P}(\mathbb{R}^d)$ exists.

In the particular case we are dealing with, namely the velocity is independent of the density, a satisfactory solution comes with the concept of regular measure. We say that $\boldsymbol{\nu} \in \mathscr{M}_+(\mathscr{P}(\mathbb{R}^d))$ is *regular* if its expected value $\mathbb{E}\boldsymbol{\nu} \in \mathscr{M}_+(\mathbb{R}^d)$ is absolutely continuous with respect to $\mathscr{L}^d$ and has a bounded density, i.e.

$$\int_{\mathscr{P}(\mathbb{R}^d)} \left(\int_{\mathbb{R}^d} \Phi\, d\mu\right) d\boldsymbol{\nu}(\mu) \le C \int_{\mathbb{R}^d} \Phi\, dx, \quad \Phi \in C_c(\mathbb{R}^d),\ \Phi \ge 0.$$

Examples

(1) The law $\boldsymbol{\nu}$ under $\mathscr{L}^d$ of the map $x \mapsto \delta_x$ is regular; it corresponds to a measure concentrated on Dirac masses;
(2) if $d = 2n$ and $f \in L^\infty(\mathbb{R}^n)$, the law under $\mathscr{L}^n$ of

$$x \mapsto \delta_x \times (f\mathscr{L}^n)$$

is regular, but not invariant under the Hamiltonian flow (it corresponds to concentration of position only).

The second example shows that the usual paradigm in dynamical systems, i.e. to consider measures $\boldsymbol{\nu}$ in $\mathscr{M}_+(\mathscr{P}(\mathbb{R}^d))$ invariant or quasi-invariant under the flow might be too restrictive. Another fact suggesting that quasi-invariance might be too restrictive is infinite-dimensionality: let γ be a Gaussian probability measure in a Hilbert space H, $\boldsymbol{b} = v \in H$, so that $\boldsymbol{X}(t,x) = x + tv$. Then $\boldsymbol{X}(t,\cdot)_\sharp\gamma \ll \gamma$ if and only if v belongs to the *Cameron–Martin* subspace $\mathcal{H}$, a much smaller subspace, since $\gamma(\mathcal{H}) = 0$ whenever $\dim(H) = \infty$.

Definition 2 Let $\boldsymbol{\mu} : [0,T] \times \mathscr{P}(\mathbb{R}^d) \to \mathscr{P}(\mathbb{R}^d)$ and $\boldsymbol{\nu} \in \mathscr{M}_+(\mathscr{P}(\mathbb{R}^d))$. We say that $\boldsymbol{\mu}$ is a $\boldsymbol{\nu}$-RLF in $\mathscr{P}(\mathbb{R}^d)$ (relative a $\boldsymbol{c}$) if

(i) for $\boldsymbol{\nu}$-a.e. μ, $t \mapsto \mu_t := \boldsymbol{\mu}(t,\mu)$ is a solution of the continuity equation with velocity $\boldsymbol{c}$ and $\boldsymbol{\mu}(0,\mu) = \mu$;
(ii) $\mathbb{E}(\boldsymbol{\mu}(t,\cdot)_\sharp\boldsymbol{\nu}) \le C\mathscr{L}^d$ for all $t \in [0,T]$, for some constant C independent of t.

Again, condition (ii) is crucial to relate the different ODE trajectories, to hope for an invariant theory and to hope for existence, uniqueness and stability.

Let us discuss now the relation between the flows in the "base space" $\mathbb{R}^d$ in the "lifted space" $\mathscr{P}(\mathbb{R}^d)$.

(Existence) If a ν-RF $\boldsymbol{X}$ in $\mathbb{R}^d$ exists, then a $\boldsymbol{\nu}$-RF $\boldsymbol{\mu}$ in $\mathscr{P}(\mathbb{R}^d)$ exists for all $\boldsymbol{\nu}$ such that $\mathbb{E}\boldsymbol{\nu} = \nu$, given by

$$\boldsymbol{\mu}(t,\mu) := \int_{\mathbb{R}^d} \delta_{\boldsymbol{X}(t,x)}\, d\mu(x). \tag{$**$}$$

(Uniqueness) Because of ($**$), uniqueness (and stability) results are *stronger* when stated at the level of $\boldsymbol{\nu}$-RF in $\mathscr{P}(\mathbb{R}^d)$, instead of ν-RF in $\mathbb{R}^d$.

Theorem 4 (A, Figalli, Friesecke, Giannoulis, Paul, [5]) *Let $\boldsymbol{c} : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ be locally integrable and let us assume that the continuity equation $\partial_t w_t + \nabla \cdot (\boldsymbol{c} w_t) = 0$ is well-posed in*

$$L^\infty_+\big([0,T]; L^1 \cap L^\infty(\mathbb{R}^d)\big).$$

Then, for all $\boldsymbol{\nu} \in \mathscr{M}_+(\mathscr{P}(\mathbb{R}^d))$ regular, the $\boldsymbol{\nu}$-RF $\boldsymbol{\mu}$ is unique, and therefore related to the unique $\mathbb{E}\boldsymbol{\nu}$-RF $\boldsymbol{X}$ as in ($**$).

We now discuss existence of the flow in the case of potentials $U_{nn} + U_b$.

Theorem 5 (A, Figalli, Friesecke, Giannoulis, Paul, [5]) *Assume $U = U_{nn} + U_b$, with U_b bounded, Lipschitz and $\nabla U_b \in BV_{\mathrm{loc}}(\mathbb{R}^n;\mathbb{R}^n)$. Then the continuity equation with velocity $\boldsymbol{c}(x,p) = (p, -\nabla U(x))$ is well posed in*

$$L^\infty_+\big([0,T]; L^1 \cap L^\infty(\mathbb{R}^{2n})\big).$$

As a consequence, both the ν-RF $\boldsymbol{X}$ and the $\boldsymbol{\nu}$-RF $\boldsymbol{\mu}$ exist and are unique.

The strategy of proof is a localization in phase space, using that the energy $E(x, p) := \frac{1}{2}|p|^2 + U(x)$ is formally preserved. Since the sublevels of the energy $\{E \le k\}$ are distant from Coulomb singularities, the "classical" theory with bounded BV vector fields is applicable. Notice that the argument would not work in the Coulomb attractive case!

3 Stability of the Flows

In order to discuss stability, let us assume to have $\boldsymbol{\nu}_n$-RF $\boldsymbol{\mu}_n$, where $\boldsymbol{\nu}_n \in \mathscr{M}_+(\mathscr{P}(\mathbb{R}^d))$ are generated in this way:

$$\boldsymbol{\nu}_n = (i_n)_\sharp \mathbb{P} \quad \text{with } i_n : W \to \mathscr{P}(\mathbb{R}^d),\ i_n \to i\ \mathbb{P}\text{-a.e.}$$

Here $(W, \mathcal{F}, \mathbb{P})$ is a given finite measure space and $\boldsymbol{\nu} := i_\sharp \mathbb{P}$ is the limit measure. This assumption is not particularly restrictive, since a classical result of Skorokhod shows that any weakly convergent sequence of measures has this representation (with $W = [0, 1]$ endowed with the standard probability structure).

In our model, namely the convergence of Wigner transforms, this assumption is natural as well. We may for instance consider:

$$i_\varepsilon(w) := W_\varepsilon \psi^\varepsilon_{0,w} \quad \text{with } \psi^\varepsilon_{0,w}(x) = \varepsilon^{-n/2}\phi_0\left(\frac{x-w}{\varepsilon}\right)e^{i(q\cdot x/\varepsilon)},$$

with q and ϕ_0 given. In this case $i = \lim_\varepsilon i_\varepsilon$ is given by

$$di(w)(x, p) = d\delta_w(x) \times |\mathcal{F}\phi_0|^2(p-q)\,dp.$$

We can now state the stability result for vector fields $(p, -\nabla U)$ of Hamiltonian type, under the same assumptions on U which ensure existence and uniqueness of generalized flows.

1. **(Uniform regularity)** For all $\phi \in C_c(\mathbb{R}^d) \ge 0$ it holds

$$\sup_{n\in\mathbb{N}} \int_W \int_{\mathbb{R}^d} \phi\, d\boldsymbol{\mu}_n(t, i_n(w))\, d\mathbb{P}(w) \le C \int_{\mathbb{R}^d} \phi\, dx;$$

2. **(Uniform decay near to Coulomb singularities)** For all $R > 0$ it holds

$$\sup_{\delta>0} \limsup_{n\to\infty} \int_W \int_0^T \int_{B_R} \frac{1}{\mathrm{dist}^2(x, \Sigma) + \delta}\, d\boldsymbol{\mu}_n(t, i_n(w))\, dt\, d\mathbb{P}(w) < \infty;$$

3. **(Compactness in space)** For all $\varepsilon > 0$ it holds

$$\lim_{R\uparrow\infty} \mathbb{P}\left(\left\{w \in W : \sup_{n\in\mathbb{N},\, t\in[0,T]} \boldsymbol{\mu}_n(t, i_n(w))(\mathbb{R}^d \setminus B_R) > \varepsilon\right\}\right) = 0;$$

4. **(Compactness in time)** For all $\phi \in C_c^\infty(\mathbb{R}^d)$ it holds

$$\lim_{M\uparrow\infty} \mathbb{P}\left(\left\{w \in W : \sup_n \int_0^T \left|\left(\int_{\mathbb{R}^d} \phi \, d\boldsymbol{\mu}_n(t, i_n(w))\right)'\right| dt > M\right\}\right) = 0;$$

5. **(Limit continuity equation)**

$$\int \left| \int_0^T \left[\varphi'(t) \int \phi \, d\boldsymbol{\mu}_n(t, i_n(w)) + \varphi(t) \int \langle \boldsymbol{b}, \nabla\phi \rangle \, d\boldsymbol{\mu}_n(t, i_n(w))\right] dt \right| d\mathbb{P}(w) = 0$$

is infinitesimal for all $\phi \in C_c^\infty(\mathbb{R}^d \setminus (\Sigma \times \mathbb{R}^n))$, $\varphi \in C_c^\infty(0, T)$.

Under these 5 assumptions it holds

$$\lim_{n\to\infty} \int_W \sup_{t\in[0,T]} d_{\mathscr{P}}(\boldsymbol{\mu}_n(t, i_n(w)), \boldsymbol{\mu}(t, i(w))) \, d\mathbb{P}(w) = 0.$$

Here $\boldsymbol{\mu}(t, \mu)$ is the $\boldsymbol{\nu}$-RF and $d_{\mathscr{P}}$ is any distance in $\mathscr{P}(\mathbb{R}^d)$ (e.g. an optimal transportation distance). Of course, since sequences converging in L^1 have subsequences converging a.e., the previous convergence result can also be seen as an almost sure convergence result.

4 Convergence of Wigner/Husimi Transforms

In order to apply the abstract stability result to the convergence of Wigner transforms we have to check that the assumptions are satisfied for suitable families of initial conditions. Let us consider, for instance, the family

$$\psi_{0,w}^\varepsilon(x) = \varepsilon^{-n\alpha/2} \phi_0\left(\frac{x - z}{\varepsilon^\alpha}\right) e^{i(q\cdot x/\varepsilon)},$$

corresponding to the choice of the "random" parameter $w = (z, q)$. We shall denote by $\psi_{t,w}^\varepsilon$ the solution of $(\mathrm{SE})_\varepsilon$ at time t, starting from $\psi_{0,w}^\varepsilon$ at time 0.

In order to work with genuine probability measures even on the "scale" ε we consider, as in Lions–Paul [12], the *Husimi* transforms of $\psi_{t,w}^\varepsilon$:

$$\tilde{W}_\varepsilon \psi = (W_\varepsilon \psi) * G_\varepsilon^{(2n)} \quad \text{with } G_\varepsilon^{(2n)}(x, p) := \frac{e^{-(|x|^2+|p|^2)/\varepsilon}}{(\pi\varepsilon)^n}.$$

Notice that the asymptotic behavior as $\varepsilon \to 0$ of the two transforms (in weak topologies) is the same.

The verification of the compactness in space/time is standard and suffices to integrate in $d\mathbb{P}$ estimates with ω fixed. The same happens for the verification of the uniform decay out of Coulomb singularities, a byproduct of the "deterministic" estimate

$$\sup_{t\in\mathbb{R}}\sup_{\varepsilon>0}\int_{\mathbb{R}^n} U_{nn}^2|\psi_{t,w}^{\varepsilon}|^2\,dx<\infty$$

of the first paper.

On the other hand, both the verification of uniform regularity and of the limit continuity equation depend on two a priori estimates on the averages of $\psi_{t,w}^{\varepsilon}$ with respect to w. The first one is

$$\sup_{\varepsilon>0}\sup_{t\in\mathbb{R}}\left\|\int_W (W_\varepsilon\psi_{t,w}^{\varepsilon})*G_{\varepsilon}^{(2n)}(x,p)\,d\mathbb{P}(w)\right\|_{L^\infty(\mathbb{R}^{2n})}<\infty,$$

which is nothing but a uniform L^∞ bound on $\int_W \tilde{W}_\varepsilon\psi_{t,w}^{\varepsilon}(x,p)\,d\mathbb{P}(w)$.

The second, instead, requires convolutions on scale ε^2:

$$\sup_{\varepsilon}\sup_{t\in\mathbb{R}}\left\|\int_W |\psi_{t,w}^{\varepsilon}*G_{\varepsilon^2}^{(2n)}|^2(x,p)\,d\mathbb{P}(w)\right\|_{L^\infty(\mathbb{R}^{2n})}<\infty.$$

Both the first and the second can be derived by inserting suitable "test" $\phi=\phi_{x,p}$ in the operator inequality in $\mathcal{L}(L^2(\mathbb{R}^n);L^2(\mathbb{R}^n))$:

$$\sup_{\varepsilon>0,\,t\in\mathbb{R}}\varepsilon^{-n}\int_W \rho^{\psi_{t,w}^{\varepsilon}}\,d\mathbb{P}(w)\le C\mathrm{Id}\quad\big(\rho^{\psi}\phi:=\langle\psi,\phi\rangle\psi\big).$$

This operator inequality (thanks to the unitary structure of the Schrödinger evolution) has the nice feature of being propagated in time; so, suffices to impose it just on the initial conditions (see also [10] for a more detailed discussion).

In conclusion, given $\alpha\in(0,1)$ and $g\in L^1\cap L^\infty(\mathbb{R}^{2n})$ nonnegative and the initial conditions (but other choices are possible and compatible with the uniform operator inequality)

$$\psi_{0,w}^{\varepsilon}(x)=\varepsilon^{-n\alpha/2}\phi_0\left(\frac{x-z}{\varepsilon^\alpha}\right)e^{i(q\cdot x/\varepsilon)},\qquad w=(z,q)$$

the stability theorem with $\mathbb{P}=g\mathscr{L}^{2n}$ gives

$$\lim_{\varepsilon\to 0}\int_{\mathbb{R}^{2n}}\sup_{t\in[-T,T]} d_{\mathscr{P}}\big(\tilde{W}_\varepsilon\psi_{t,w}^{\varepsilon},\delta_{\boldsymbol{X}(t,w)}\big)g(w)\,dw=0\quad\forall T>0,$$

where $\boldsymbol{X}(t,w)$ is the flow in $\mathbb{R}^{2n}$ induced by $(p,-\nabla U(x))$. In general, we should replace $\delta_{\boldsymbol{X}}(t,w)$ by the superposition flow induced by the limiting initial conditions, i.e. $\boldsymbol{\mu}(t,w)=\int\delta_{\boldsymbol{X}(t,x)}\,di(w)(x)$.

This way, we get *full* convergence as $\varepsilon\to 0$ even for fields not of class C^2.

5 Conclusions

1. The "flow" viewpoint is very common in Probability, and the regularizing effect due to the addition of noise has been studied by many authors and in many contexts. In our case the equation *is not* stochastically perturbed, but convergence is studied with analytic and probabilistic tools.
2. The case of *attractive* potentials U seems for the moment to be completely out of reach.
3. The transfer mechanisms from flows in $\mathbb{R}^d$ to flows in $\mathscr{P}(\mathbb{R}^d)$ rely on the fact that the equation is linear (in the abstract perspective, on the fact that the ODE in $\mathscr{P}(\mathbb{R}^d)$ has constant coefficients). The extension to non-linear equations, as for instance Vlasov–Poisson (where density is coupled to the velocity by Poisson's equation)

$$-\Delta U_t(x) = \int f_t(x, p)\, dp$$

also seems at this moment to be out of reach.

References

1. Ambrosio, L.: Transport equation and Cauchy problem for BV vector fields. Invent. Math. **158**, 227–260 (2004)
2. Ambrosio, L.: Transport equation and Cauchy problem for non-smooth vector fields. In: Dacorogna, B., Marcellini, P. (eds.) Calculus of Variations and Non-Linear Partial Differential Equations. CIME Series, Cetraro, 2005. Lecture Notes in Mathematics, vol. 1927, pp. 2–41 (2008)
3. Ambrosio, L., Figalli, A.: Almost everywhere well-posedness of continuity equations with measure initial data. CRAS **348**, 249–252 (2010)
4. Ambrosio, L., Friesecke, G., Giannoulis, J.: Passage from quantum to classical molecular dynamics in the presence of Coulomb interactions. Commun. Partial Differ. Equ. **35**, 1490–1515 (2010)
5. Ambrosio, L., Figalli, A., Friesecke, G., Giannoulis, J., Paul, T.: Semiclassical limit of quantum dynamics with rough potentials and well posedness of transport equations with measure initial data. Commun. Pure Appl. Math. **64**, 1199–1242 (2011)
6. Bouchut, F.: Renormalized solutions to the Vlasov equation with coefficients of bounded variation. Arch. Ration. Mech. Anal. **157**, 75–90 (2001)
7. DiPerna, R.J., Lions, P.L.: Ordinary differential equations, transport theory and Sobolev spaces. Invent. Math. **98**, 511–547 (1989)
8. Fermanian-Kammerer, C., Gérard, P.: Mesures semi-classiques et croisement de modes. Bull. Soc. Math. Fr. **130**, 123–168 (2002)
9. Fermanian-Kammerer, C., Gérard, P.: A Landau–Zener formula for non-degenerated involutive codimension 3 crossings. Ann. Henri Poincaré **4**, 513–552 (2003)
10. Figalli, A., Ligabó, M., Paul, T.: Work in progress
11. Gérard, P.: Mesures semi-classiques et ondes de Bloch. In: Seminaire sur les Équations aux Dérivées Partielles, 1990–1991. Exp. No. XVI, 19 pp. École Polytechnique, Palaiseau (1991)
12. Lions, P.L., Paul, T.: Sur les mesures de Wigner. Rev. Mat. Iberoam. **9**, 553–618 (1993)

Contractive Metrics for Nonsmooth Evolutions

Alberto Bressan

Abstract Given an evolution equation, a standard way to prove the well posedness of the Cauchy problem is to establish a Gronwall type estimate, bounding the distance between any two trajectories. There are important cases, however, where such estimates cannot hold, in the usual distance determined by the Euclidean norm or by a Banach space norm.

In alternative, one can construct different distance functions, related to a Riemannian structure or to an optimal transportation problem. This paper reviews various cases where this approach can be implemented, in connection with discontinuous ODEs on $\mathbb{R}^n$, nonlinear wave equations, and systems of conservation laws. For all the evolution equations considered here, a metric can be constructed such that the distance between any two solutions satisfies a Gronwall type estimate. This yields the uniqueness of solutions, and estimates on their continuous dependence on the initial data.

1 Introduction

Consider an abstract evolution equation in a Banach space

$$\frac{d}{dt}u(t) = F(u(t)). \tag{1}$$

If F is a continuous vector field with Lipschitz constant L, the classical Cauchy–Lipschitz theory applies. For any given initial data

$$u(0) = \bar{u}, \tag{2}$$

the solution of (1) is thus unique, and depends continuously on $\bar{u}$. Indeed, the distance between any two solutions grows at a controlled rate:

$$\frac{d}{dt}\|u(t) - v(t)\| \leq L\|u(t) - v(t)\|. \tag{P1}$$

A. Bressan (✉)
Department of Mathematics, Penn State University, University Park, PA 16802, USA
e-mail: bressan@math.psu.edu

H. Holden, K.H. Karlsen (eds.), *Nonlinear Partial Differential Equations*, Abel Symposia 7, DOI 10.1007/978-3-642-25361-4_2,

In this case, the classical Gronwall's estimate yields

$$\|u(t) - v(t)\| \le C(t)\|u(0) - v(0)\|, \tag{P2}$$

with $C(t) = e^{Lt}$. Semigroup theory has extended the validity of estimates such as (P1), (P2) to a wide class of right hand sides, including differential operators, which generate a continuous flow [18, 24, 29, 34].

On the other hand, there are cases (such as the Camassa–Holm equation) where the flow generated by (1) is not Lipschitz continuous w.r.t. the initial data, in any standard Hölder or Sobolev norm. In other cases, such as hyperbolic systems of conservation laws, the generated semigroup is globally Lipschitz continuous w.r.t. the $\mathbf{L}^1$ norm but does not satisfy an estimate of the form (P1), for any constant L. In all these situations, a natural problem is to seek an alternative distance $d^\diamondsuit(\cdot,\cdot)$, possibly not equivalent to any of the usual norm distances, for which (P1) or (P2) still hold.

Aim of this note is to discuss a few examples where this goal can be achieved. Typically, the distance $d^\diamondsuit$ is defined as a Riemann type distance. In other words, one starts with a Banach space E and a family Σ of sufficiently regular paths $\gamma : [0,1] \mapsto E$, for which some kind of "weighted length" $|\gamma|_*$ can be defined. This needs not be equivalent to the length derived from the norm distance. Given two elements $u, v \in E$, one first defines

$$d^*(u,v) = \inf\big\{|\gamma|_*;\ \gamma \in \Sigma,\ \gamma(0) = u,\ \gamma(1) = v\big\}, \tag{3}$$

and then takes the lower semicontinuous envelope (w.r.t. convergence in norm):

$$d^\diamondsuit(u,v) \doteq \liminf_{u' \to u,\, v' \to v} d^*(u',v'). \tag{4}$$

Besides achieving a proof of uniqueness and continuous dependence, estimates of the form (P2) are useful for establishing error estimates. Indeed, adopting a semigroup notation, call $t \mapsto S_t\bar u$ the solution to the Cauchy problem

$$\frac{d}{dt}u = F(u), \qquad u(0) = \bar u. \tag{5}$$

Assume that, for any couple of initial data $\bar u, \bar v$, there holds

$$d^\diamondsuit(S_t\bar u,\ S_t\bar v) \le C d^\diamondsuit(\bar u,\ \bar v), \qquad t \in [0,T]. \tag{6}$$

Then for any Lipschitz continuous trajectory $t \mapsto w(t)$ one can deduce the error estimate [4]

$$d^\diamondsuit\big(w(T), S_T w(0)\big) \le C \cdot \int_0^T \left\{ \liminf_{h\to 0+} \frac{d^\diamondsuit(w(t+h), S_h w(t))}{h} \right\} dt. \tag{7}$$

Here the left hand side is the distance at time T between the approximate solution $w(\cdot)$ and the exact solution of (1) with the same initial data $w(0)$. The right hand side is the integral of an instantaneous error rate.

The estimate (P1) can be also useful in order to understand which kind of Lipschitz perturbations preserve the well-posedness property.

In the following sections we shall review three different settings where these ideas can be implemented. Section 2 is devoted to discontinuous ODEs in a finite dimensional space [21]. Following [5], for a vector field $F = F(t, x)$ having finite directional variation, a general formula yielding a time-dependent contractive Riemann metric can here be given. Section 3 reviews two different constructions of a distance functional which satisfies an estimate of the form (P1), in connection with the Camassa–Holm equation [10, 23]. Finally, in Sect. 4 we discuss distance functionals which are contractive for the flow generated by a hyperbolic system of conservation laws [3, 12].

2 Discontinuous ODEs

To motivate the search for contractive metrics, we start with two elementary examples.

Example 1 The ODE $\dot{x} = |x|^{1/2}$ yields a textbook case of a Cauchy problem with multiple solutions. Yet, there is a simple way to select a unique solution for each initial data $x(0) = \bar{x}$. Let us define a solution $t \mapsto x(t)$ to be "admissible" if and only if it is strictly increasing. These admissible solutions are then unique, and depend continuously on the initial data. For $\bar{x} = 0$, the corresponding admissible solution is $t \mapsto S_t\bar{x} = (\text{sign}\, t)\, t^2/4$. Notice that the trajectory $w(t) \equiv 0$ is not an admissible solution, but the error $|w(t) - S_t w(0)| = t^2/4$ cannot be estimated integrating the "instantaneous error rate" $|\dot{w}(t) - |w(t)|^{1/2}| \equiv 0$.

On the other hand, a direct computation shows that the Riemann distance

$$d^{\diamond}(x, y) = \left| \int_x^y \frac{ds}{|s|^{1/2}} \right|$$

is invariant w.r.t. the flow of admissible solutions. Namely $d^{\diamond}(S_t\bar{x},\ S_t\bar{y}) = d^{\diamond}(\bar{x}, \bar{y})$ for every $\bar{x}, \bar{y}, t$. Using this distance, the error estimate (7) retains its validity. Indeed

$$d^{\diamond}\big(w(T), S_T w(0)\big) = d^{\diamond}(0, T^2/4) = \int_0^{T^2/4} \frac{ds}{s^{1/2}} = T,$$

$$\int_0^T \left\{ \liminf_{h \to 0+} \frac{d^{\diamond}(w(t+h), S_h w(t))}{h} \right\} dt = \int_0^T \left\{ \liminf_{h \to 0+} \frac{d^{\diamond}(0, h^2/4)}{h} \right\} dt$$
$$= \int_0^T 1\, dt = T.$$

Example 2 Consider the discontinuous ODE

$$\dot{x} = f(x) = \begin{cases} 1 & \text{if } x < 0, \\ 3 & \text{if } x \geq 0. \end{cases} \tag{8}$$

For any initial data $x(0) = \bar{x}$, the Cauchy problem is well posed. Indeed, any two solutions satisfy

$$|x_1(t) - x_2(t)| \leq 3|x_1(0) - x_2(0)|. \tag{9}$$

The estimate (9) alone, however, does not tell for which vector fields $g(\cdot)$ the ODE

$$\dot{x} = f(x) + g(x)$$

generates a continuous semigroup. For example, taking $g(x) \equiv 2$, the Cauchy problem is well posed, while taking $g(x) \equiv -2$ it is not. The difference between the two above cases becomes apparent by introducing the equivalent distance

$$d^{\diamondsuit}(x, y) = \begin{cases} 3|y - x| & \text{if } x \leq y \leq 0, \\ |y - x| & \text{if } 0 \leq x \leq y, \\ 3|x| + |y| & \text{if } x \leq 0 \leq y. \end{cases}$$

Notice that $d^{\diamondsuit}$ is invariant w.r.t. the flow generated by (8). Denote by S_t^g and S_t^{-g} the semigroups generated by the ODEs $\dot{x} = g(x)$ and $\dot{x} = -g(x)$, respectively. Then for every $\bar{x}, \bar{y} \in \mathbb{R}$ and $t \geq 0$ we have

$$d^{\diamondsuit}(S_t^g \bar{x}, S_t^g \bar{y}) \leq d^{\diamondsuit}(\bar{x}, \bar{y}). \tag{10}$$

On the other hand, taking $\bar{x} \leq 0$, $\bar{y} > 0$, one has

$$\lim_{h \to 0+} \frac{d^{\diamondsuit}(S_t^{-g} \bar{x}, S_t^{-g} \bar{y}) - d^{\diamondsuit}(\bar{x}, \bar{y})}{h} = 2. \tag{11}$$

Comparing (10) with (11), we see that the flow generated by g contracts the distance $d^{\diamondsuit}$, while the flow generated by $-g$ can increase it, at a rate which does not approach zero as $d^{\diamondsuit}(\bar{x}, \bar{y}) \to 0$.

Next, consider a general ODE with bounded, possibly discontinuous right hand side

$$\dot{x} = f(t, x), \qquad x \in \mathbb{R}^n. \tag{12}$$

In the Euclidean space $\mathbb{R}^{1+n}$, consider the cone with opening M:

$$\Gamma^M \doteq \big\{(\tau, y);\ |y| \leq M\tau\big\}.$$

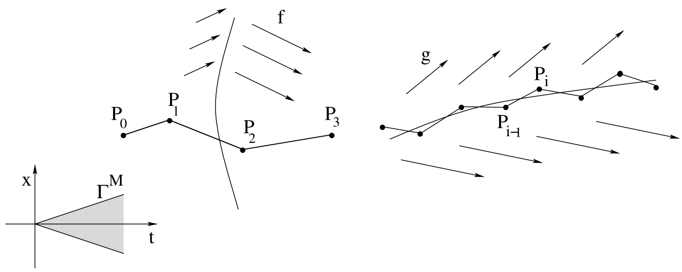


Fig. 1 *Left*: the vector field f is transversal to the surface where it is discontinuous. The directional variation V^M is a bounded function. *Right*: the vector field g is not transversal to the surface where it has a discontinuity. Its directional variation is thus unbounded

Following [2], the total directional variation of the vector field f up to the point (t, x) is defined as

$$V^M(t, x) \doteq \sup\left\{ \sum_{i=1}^{N} |f(P_i) - f(P_{i-1})|;\ N \geq 1,\ P_i - P_{i-1} \in \Gamma^M,\ P_N = (t, x) \right\}. \tag{13}$$

Notice that, in order that V^M be bounded the jumps in f must be located along hypersurfaces which are transversal to the directions in the cone Γ^M. Otherwise, one can choose a large number of points P_i, alternatively on opposite sides of the discontinuity, and render the sum in (13) arbitrarily large (see Fig. 1).

Given two constants $0 < L < M$, we define the weighted length of a Lipschitz continuous path $\gamma : [0, 1] \mapsto \mathbb{R}^n$ at time t as

$$\|\gamma\|_t \doteq \int_0^1 \exp\left(-\frac{V^M(t, \gamma(s))}{M - L} \right) |\dot{\gamma}(s)|\, ds. \tag{14}$$

The weighted distance between two points $x, y \in \mathbb{R}^n$ at time t is defined as

$$d_t(x, y) \doteq \inf\left\{ \|\gamma\|_t;\ \gamma \text{ is a Lipschitz path joining } x \text{ with } y \right\}. \tag{15}$$

Notice that, as t increases, the directional variation $V^M(t, \cdot)$ also increases. Hence, by (14), the weighted length of the path γ becomes smaller.

We recall that a Carathéodory solution to the (possibly discontinuous) ODE (12) is an absolutely continuous function $t \mapsto x(t)$ that satisfies (12) for a.e. time t. The main result proved in [5] is as follows.

Theorem 1 *Let $f = f(t, x)$ be a time dependent vector field on $\mathbb{R}^n$. Assume that there exist constants $L < M$ such that $|f(t, x)| \leq L$ for all t, x, and the directional variation V^M of f defined at* (13) *is locally bounded.*

Then, for every initial data $x(t_0) = x_0$, the ODE (12) *has a unique, globally defined Carathéodory solution. Any two solutions $x(\cdot)$, $y(\cdot)$ of* (12) *satisfy*

$$d_\tau(x(\tau),\ y(\tau)) \leq d_t(x(t),\ y(t)) \quad \textit{for all } t \leq \tau. \tag{16}$$

Remark 1 Since the Euclidean distance between two nearby trajectories can rapidly increase across a surface where f is discontinuous, to achieve the contractive property (16) this must be compensated by the decrease of the exponential weight inside the integral in (14).

3 Nonlinear Wave Equations

The Camassa–Holm equation can be written as a scalar conservation law with an additional integro-differential term:

$$u_t + (u^2/2)_x + P_x = 0, \tag{17}$$

where P is defined as the convolution

$$P \doteq \frac{1}{2} e^{-|x|} * \left(u^2 + \frac{u_x^2}{2} \right). \tag{18}$$

For the physical motivations of this equation we refer to [14–16]. One can regard (17) as an evolution equation on a space of absolutely continuous functions with derivatives $u_x \in \mathbf{L}^2$. In the smooth case, differentiating (17) w.r.t. x one obtains

$$u_{xt} + uu_{xx} + u_x^2 - \left(u^2 + \frac{u_x^2}{2} \right) + P = 0. \tag{19}$$

Multiplying (17) by u and (19) by u_x one obtains the two balance laws

$$\left(\frac{u^2}{2} \right)_t + \left(\frac{u^3}{3} + uP \right)_x = u_x P, \qquad \left(\frac{u_x^2}{2} \right)_t + \left(\frac{uu_x^2}{2} - \frac{u^3}{3} \right)_x = -u_x P. \tag{20}$$

As a consequence, for regular solutions the total energy

$$E(t) \doteq \int \left[u^2(t,x) + u_x^2(t,x) \right] dx \tag{21}$$

remains constant in time.

As in the case of conservation laws, because of the strong nonlinearity of the equations, solutions with smooth initial data can lose regularity in finite time. For the Camassa–Holm equation (17), however, the uniform bound on $\|u_x\|_{\mathbf{L}^2}$ guarantees that only the $\mathbf{L}^\infty$ norm of the gradient can blow up, while the solution u itself remains Hölder continuous at all times.

The equation (17) admits *multi-peakon* solutions, depending on finitely many parameters. These have the form

$$u(t,x) = \sum_{i=1}^{N} p_i(t) e^{-|x - q_i(t)|}, \tag{22}$$

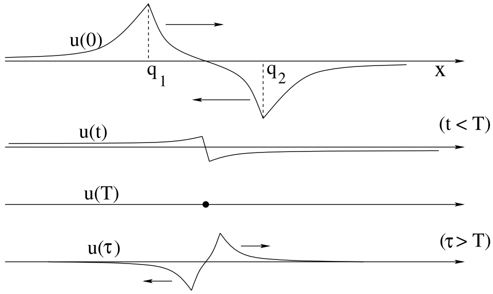


Fig. 2 A solution consisting of two peakons with opposite strengths

where the coefficients p_i, q_i are obtained by solving the Hamiltonian system of ODEs

$$\begin{cases} \dot{q}_i = \dfrac{\partial}{\partial p_i} H(p,q), \\ \dot{p}_i = -\dfrac{\partial}{\partial q_i} H(p,q), \end{cases} \qquad H(p,q) \doteq \frac{1}{2}\sum_{i,j=1}^{N} p_i p_j e^{-|q_i - q_j|}. \tag{23}$$

According to (22), the coefficient p_i determines the amplitude of the i-th peakon, while q_i describes its location.

By (20), the H^1 norm is constant in time along regular solutions. The space $H^1(\mathbb{R}) = W^{1,2}(\mathbb{R})$ thus provides a natural domain where to construct global solutions to the Camassa–Holm equation. Observe that, for $u \in W^{1,p}$ with $p < 2$, the convolution (18) may not be well defined. On the other hand, if $p > 2$, the Sobolev norm $\|u(t)\|_{W^{1,p}}$ of a solution can blow up in finite time.

Given an initial condition

$$u(0) = u_0 \in H^1(\mathbb{R}), \tag{24}$$

by a solution of the Cauchy problem (17)–(24) on $[0, T]$ we mean a Hölder continuous function $u = u(t, x)$ defined on $[0, T] \times \mathbb{R}$ with the following properties. At each fixed t one has $u(t, \cdot) \in H^1(\mathbb{R})$. Moreover, the map $t \mapsto u(t, \cdot)$ is Lipschitz continuous from $[0, T]$ into $\mathbf{L}^2(\mathbb{R})$, satisfying the initial condition (24) together with

$$\frac{d}{dt} u = -u u_x - P_x \tag{25}$$

for a.e. t. Here (25) is understood as an equality between functions in $\mathbf{L}^2(\mathbb{R})$. The solution is called *conservative* if the corresponding energy $E(t)$ in (21) coincides a.e. with a constant function.

A globally defined flow of conservative solutions was constructed in [8]. One should be aware, however, that the Cauchy problem for the Camassa–Holm equation is not well posed, even in the "natural" space $H^1(\mathbb{R})$. Failure of continuous dependence on initial data can be seen by looking at special solutions with two opposite peakons (Fig. 2). In this case we have $p_1(t) + p_2(t) \equiv 0$, $q_1(t) + q_2(t) \equiv 0$.

Let T be the interaction time, so that $q_1(T) = q_2(T) = 0$. As $t \to T-$, one has

$$p_1(t) \to +\infty, \qquad p_2(t) \to -\infty, \qquad q_1(t) = 0, \qquad q_2(t) \to 0, \tag{26}$$

$$\|u(t)\|_{\mathscr{C}^0} \to 0, \qquad \|u(t)\|_{\mathbf{L}^2} \to 0, \qquad \int_{q_1(t)}^{q_2(t)} u_x^2(t,x)\,dx \to E_0, \tag{27}$$

where E_0 is the energy of the solution, which is a constant (except at $t = T$). For detailed computations we refer to Sect. 5 in [8]. The last limit in (27) shows that, as $t \to T-$, nearly all the energy is concentrated within the small interval $[q_1(t),\ q_2(t)]$ between the two peakons.

Next, in addition to this special solution u, consider a family of solutions defined as $u^\varepsilon(t,x) \doteq u(t-\varepsilon, x)$. At time $t = 0$, as $\varepsilon \to 0$ we have $\|u^\varepsilon(0,\cdot) - u(0,\cdot)\|_{H^1} \to 0$. However, for any $\varepsilon > 0$, as $t \to T-$ one has

$$\begin{aligned}
\|u(t) - u^\varepsilon(t)\|_{H^1}^2 &= \int_{\mathbb{R}} |u(t,x) - u^\varepsilon(t,x)|^2\,dx \\
&\quad + \left(\int_{\mathbb{R}\setminus[q_1(t),\,q_2(t)]} + \int_{[q_1(t),\,q_2(t)]}\right)|u_x(t,x) - u_x^\varepsilon(t,x)|^2\,dx \\
&\to \int_{\mathbb{R}} |u^\varepsilon(T,x)|^2\,dx + \int_{\mathbb{R}} |u_x^\varepsilon(T,x)|^2\,dx + E_0^2 = 2E_0^2.
\end{aligned} \tag{28}$$

According to (28), solutions u^ε which initially start arbitrarily close to u, within finite time split apart at a uniformly positive distance $\sqrt{2}E_0$.

3.1 A Metric Induced by Optimal Transportation

In order to analyze uniqueness questions for solutions to the Camassa–Holm equations, it is of interest to construct an alternative distance functional $J(\cdot,\cdot)$ on H^1. For any two solutions of (17), this distance should satisfy an inequality of the form

$$\frac{d}{dt} J(u(t), v(t)) \le \kappa \cdot J(u(t), v(t)), \tag{29}$$

with a constant κ depending only on the maximum of the two norms $\|u(t)\|_{H^1}$, $\|v(t)\|_{H^1}$ (which remain a.e. constant in time). For spatially periodic conservative solutions to the Camassa–Holm equation, this goal was achieved in [10], building on insight gained in [7]. We describe here the key steps of this construction.

Consider the unit circle $\mathbf{T} = [0, 2\pi]$ with endpoints identified. The distance between two angles $\theta, \tilde\theta \in \mathbf{T}$ will be denoted as $|\theta - \tilde\theta|_*$. Consider the manifold $X \doteq \mathbb{R}\times\mathbb{R}\times\mathbf{T}$ with distance

$$d^*\big((x,u,\theta),(\tilde x,\tilde u,\tilde\theta)\big) \doteq \big(|x-\tilde x| + |u - \tilde u| + |\theta - \tilde\theta|_*\big) \wedge 1, \tag{30}$$

where $a \wedge b \doteq \min\{a, b\}$. Let H^1_{per} be the space of absolutely continuous periodic functions u, with $u(x) = u(x+1)$ for every $x \in \mathbb{R}$, and such that

$$\|u\|_{H^1_{per}} \doteq \left(\int_0^1 [u^2(x) + u_x^2(x)]\, dx \right)^{1/2} < \infty.$$

Given $u \in H^1_{per}(\mathbb{R})$, define its extended graph

$$\text{Graph}(u) = \Big\{ \big(x, u(x), 2\arctan u_x(x)\big);\, x \in \mathbb{R} \Big\} \subset X.$$

Moreover, let μ^u be the measure supported on Graph(u), whose projection on the x-axis has density $1 + u_x^2$ w.r.t. Lebesgue measure. In other words, for every open set $A \subset X$ we require

$$\mu^u(A) = \int_{\{x;\, (x, u(x), 2\arctan u_x(x)) \in A\}} \big(1 + u_x^2(x)\big)\, dx.$$

The distance $J(u, v)$ between two functions $u, v \in H^1_{per}$ is determined as the minimum cost for a constrained optimal transportation problem. More precisely, consider the measures μ^u, μ^v, supported on Graph(u) and on Graph(v), respectively. An absolutely continuous strictly increasing map $\psi : \mathbb{R} \mapsto \mathbb{R}$ satisfying the periodicity condition

$$\psi(x+1) = \psi(x) + 1 \quad \forall x \in \mathbb{R}$$

will be called an *admissible transportation plan* (see Fig. 3). Given ψ, we can move the mass μ^u to μ^v, from the point $(x, u(x), 2\arctan u_x(x))$ to the point $(\psi(x), v(\psi(x)), 2\arctan v_x(\psi(x)))$. In general, however, the measure μ^v is not equal to the push forward of the measure μ^u determined by the map ψ. We thus need to introduce an additional cost, penalizing this discrepancy. Using the function

$$\phi_1(x) \doteq \sup\Big\{ \theta \in [0,1];\, \theta \cdot \big(1 + u_x^2(x)\big) \le \big(1 + \tilde{u}_x^2(\psi(x))\big)\psi'(x) \Big\},$$

the cost associated to the transportation plan ψ is now defined as

$$\begin{aligned}
J^\psi(u, v) &= \int_0^1 [\text{distance}] \cdot [\text{transported mass}] + \int_0^1 [\text{excess mass}] \\
&\doteq \int_0^1 d^*\Big(\big(x, u(x), 2\arctan u_x(x)\big), \big(\psi(x), \tilde{u}(\psi(x)), 2\arctan \tilde{u}_x(\psi(x))\big)\Big) \\
&\quad \cdot \phi_1(x)(1 + u_x^2(x))\, dx \\
&\quad + \int_0^1 \Big| \big(1 + u_x^2(x)\big) - \big(1 + \tilde{u}_x^2(\psi(x))\big)\psi'(x) \Big|\, dx. \qquad (31)
\end{aligned}$$

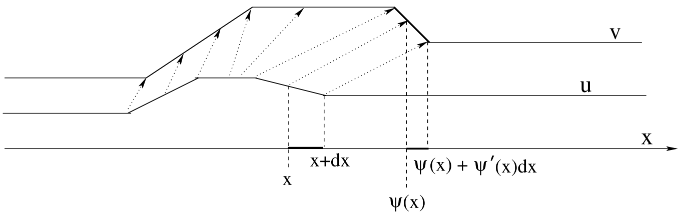


Fig. 3 Transporting the mass from the graph of u to the graph of v

Minimizing over all admissible transportation plans, one obtains a distance functional:

$$J(u, v) \doteq \inf_{\psi} J^{\psi}(u, v). \tag{32}$$

The analysis in [10] shows that this functional is indeed a distance on H^1_{per}, and grows at the controlled rate (29) along any couple of conservative solutions to the Camassa–Holm equation. In turn, this yields the uniqueness of conservative solutions, and a sharp estimate on their continuous dependence on initial data.

Remark 2 In the definition of the distance functional J, the requirement that the transportation must be achieved in terms of a non-decreasing function ψ plays an essential role. Indeed, the topology generated by the distance J is different from the topology of weak convergence of measures, corresponding to the standard transportation distance

$$d(\mu^u, \mu^v) = \sup\left\{\left|\int f\, d\mu^u - \int f\, d\mu^v\right|, \|f\|_{Lip} \le 1\right\}, \tag{33}$$

where the supremum is taken over all Lipschitz continuous functions with Lipschitz constant 1.

For example, consider the sequence of saw-tooth functions as in Fig. 4, where u_m is defined as the unique function of period 2^{-m} such that

$$u_m(x) \doteq \min\{x, 2^{-m} - x\}, \qquad x \in [0, 2^{-m}].$$

Observe that μ^{u_m} is a measure supported on Graph(u_m), whose projection on the x-axis has constant density $1 + (u_m)_x^2 \equiv 2$ w.r.t. Lebesgue measure. As $m \to \infty$, one has the weak convergence $\mu^{u_m} \rightharpoonup \mu$, where μ is the sum of two copies of Lebesgue measure, one on the line $\{(x, 0, \pi);\ x \in \mathbb{R}\}$, and one on the line $\{(x, 0, -\pi);\ x \in \mathbb{R}\}$.

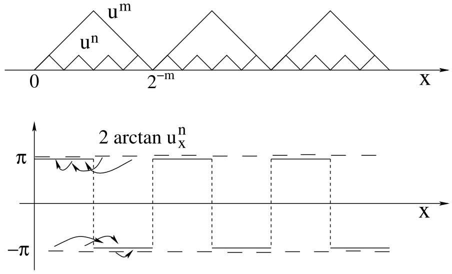


Fig. 4 If one allows transportation plans $x \mapsto \psi(x)$ which are not monotone, the optimal transportation of the measure μ^{u_n} to the measure μ^{u_m} can be achieved with a much smaller cost. However, such plans are not allowed by the definition (32)

In particular, the sequence $(\mu^{u_m})_{m\geq 1}$ is a Cauchy sequence w.r.t. the distance (33). However, $(u_m)_{m\geq 1}$ is not a Cauchy sequence w.r.t. the distance (32).

3.2 A Metric Induced by Relabeling Equivalence

Next, we discuss an alternative approach to the construction of a distance functional J, having the controlled growth property (29) along solutions to the Camassa–Holm equation. Here the starting point is the representation of solutions in terms of new variables, introduced in [8].

As independent variables we use time t and an "energy" variable $\xi \in \mathbb{R}$, which is constant along characteristics (see Fig. 5). This means that, in the t-x plane, for each fixed ξ the curve $t \mapsto y(t,\xi)$ provides a solution to the Cauchy problem

$$\frac{d}{dt}y(t) = u(t, y(t)), \qquad y(0,\xi) = \bar{y}(\xi).$$

In addition, we use the three dependent variables

$$U = u, \qquad v = 2\arctan u_x, \qquad q = (1+u_x^2)\cdot\frac{\partial y}{\partial \xi}.$$

There is considerable freedom in the parameterization of characteristics. A natural way to choose the function $\xi \mapsto \bar{y}(\xi)$ is to require that

$$\int_0^{\bar{y}(\xi)} (1+\bar{u}_x^2)\, dx = \xi.$$

At time $t = 0$, this achieves the identity $q(0,\xi) \equiv 1$.

As proved in [8], for a given initial data $u(0) = \bar{u} \in H^1(\mathbb{R})$, a conservative solution to the Camassa–Holm equation (17) can be constructed as follows. As a first

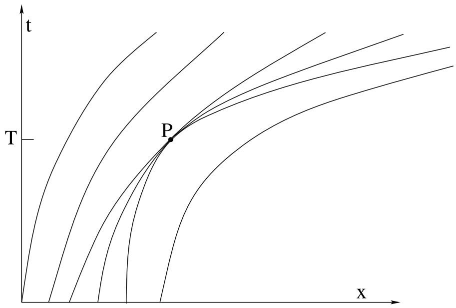


Fig. 5 Characteristic curves, for a solution to the Camassa–Holm equation. It is quite possible that characteristics join together at an isolated time T. This happens, for example, when two peakons cross each other as in Fig. 2. In this case, as $t \to T-$, the measure with density $1+u_x^2$ approaches a point mass at P. However, in the variables (U, v, q), the solution of (34) remains smooth. As $u_x \to \pm\infty$ we simply have $2\arctan u_x \to \pm\pi$, and the singularity is completely resolved by the variable transformation

step, we solve the Cauchy problem

$$\begin{cases} \dfrac{\partial U}{\partial t} = -P_x, \\ \dfrac{\partial v}{\partial t} = 2(U^2 - P)\cos^2\dfrac{v}{2} - \dfrac{1}{2}\sin^2\dfrac{v}{2}, \\ \dfrac{\partial q}{\partial t} = \left(U - \dfrac{1}{2} - P\right)\sin v \cdot q, \end{cases} \qquad \begin{cases} U(0,\xi) = \bar{u}\big(\bar{y}(\xi)\big), \\ v(0,\xi) = 2\arctan \bar{u}_x\big(\bar{y}(\xi)\big), \\ q(0,\xi) = 1. \end{cases} \tag{34}$$

$$P(t,\xi) = \frac{1}{2}\int_{-\infty}^{\infty} \exp\left\{-\left|\int_{\xi}^{\xi'} \cos^2\frac{v(s)}{2}\cdot q(s)\,ds\right|\right\} \cdot \left[U^2(\xi')\cos^2\frac{v(\xi')}{2} + \frac{1}{2}\sin^2\frac{v(\xi')}{2}\right] q(\xi')\,d\xi'.$$

This can be regarded as a Cauchy problem for an ODE on the Banach space

$$E \doteq H^1 \oplus \mathbf{L}^\infty \oplus \mathbf{L}^\infty.$$

By a fixed point argument, one obtains a unique solution, globally defined for all $t \in \mathbb{R}$. In turn, from a solution $(U, v, q)(t, \xi)$ of (34) one recovers a solution $u(t, x)$ of the Camassa–Holm equation (17) by setting

$$y(t,\xi) \doteq \bar{y}(\xi) + \int_0^t U(\tau,\xi)\,d\tau$$

and then defining

$$u(t,x) = U(t,\xi) \quad \text{if } x = y(t,\xi). \tag{35}$$

As proved in [8], this procedure yields a group of solutions continuously depending on the initial data. Namely, given a sequence of initial data such that

$\|\bar{u}_n - \bar{u}\|_{H^1} \to 0$, the corresponding solutions $u_n(t,x)$ converge to $u(t,x)$ uniformly for t, x in bounded sets.

By itself, this result does not guarantee the uniqueness of conservative solutions. In principle one may use a completely different construction procedure (say, by vanishing viscosity approximations as in [32, 33]) and generate different solutions.

To construct a distance functional providing precise information on the continuous dependence of solutions, the approach developed in [23] is based on a relabeling technique. See also [13] for an earlier result in connection with the Hunter–Saxton equation.

As motivation, observe that the same solution $u(t,x)$ of the Camassa–Holm equation (17) corresponds to infinitely many equivalent solutions $(U, v, q)(t, \xi)$ of the system (34). Indeed, here the variable ξ is simply used as a label to identify different characteristics. A smooth relabeling $\xi \mapsto \zeta(\xi)$ would produce a different solution $(\widetilde{U}, \tilde{v}, \tilde{q})$ of (34), with

$$\widetilde{U}(t,\zeta) = U(t,\xi), \qquad \tilde{v}(t,\zeta) = v(t,\xi), \qquad \tilde{q}(t,\zeta) = q(t,\xi) \cdot \frac{\partial \xi}{\partial \zeta}.$$

However, the corresponding solution $u(t,x)$ would be the same.

Given $u \in H^1$, consider the set of triples

$$\mathscr{F}(u) = \Big\{(U, v, q);\ \text{there exists } y(\cdot) \text{ such that } U(\xi) = u(y(\xi)),$$
$$v(\xi) = 2\arctan u_x(y(\xi)),\ q(\xi) = (1 + u_x^2(y(\xi))) \cdot \frac{\partial}{\partial \xi} y(\xi)\Big\}.$$

One can define the functional

$$J^\sharp(u, \tilde{u}) \doteq \inf \big\|(U, v, q) - (\widetilde{U}, \tilde{v}, \tilde{q})\big\|_E,$$

where the infimum is taken over all triples such that $(U, v, q) \in \mathscr{F}(u)$, $(\widetilde{U}, \tilde{v}, \tilde{q}) \in \mathscr{F}(\tilde{u})$. To achieve the triangle inequality, one needs to introduce a further functional

$$J(u, \tilde{u}) \doteq \inf \left\{ \sum_{i=1}^{N} J^\sharp(u_i, u_{i-1});\ u_0 = u,\ u_N = \tilde{u} \right\}. \tag{36}$$

As shown in [23], in connection with spatially periodic solutions to the Camassa–Holm equation, this approach yields an alternative construction of a distance functional which satisfies the crucial property (29).

Remark 3 While the well-posedness issue for the Camassa–Holm equation is now well understood, it remains a challenging open problem to establish similar results for the nonlinear wave equation

$$u_{tt} - c(u)\big(c(u)u_x\big)_x = 0. \tag{37}$$

In this case, for each given data $(u, u_t) \in H^1 \times \mathbf{L}^2$, one can define not one but two measures μ_+^u, μ_-^u, accounting for the energy transported by forward and by backward moving waves. Given two couples (u, u_t), (v, v_t), it is not clear how to extend a functional of the form (31) to a "double transportation problem", relating the two couples of measures (μ_+^u, μ_-^u) and (μ_+^v, μ_-^v).

As proved in [11], global conservative solutions to Eq. (37), continuously depending on the initial data, can also be obtained by a nonlinear transformation of independent and dependent variables. However, a relabeling technique here is hard to implement. Indeed, in (35) the independent variables are related by $(t, x) = (t, y(t, \xi))$. On the other hand, for Eq. (37), it is convenient to use independent variables X, Y which are constant along forward and backward characteristics, respectively. This yields a transformation $(X, Y) \mapsto (t(X, Y), x(X, Y))$ where the time variable has no preferred status. In general, for any constant c the set $\{(X, Y); t(X, Y) = c\}$ has an awkward structure.

4 Hyperbolic Conservation Laws

In this last section we discuss the construction of a contractive metric for the system of conservation laws

$$u_t + f(u)_x = 0. \tag{38}$$

Here $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$ is the vector of conserved quantities and $f = (f_1, \ldots, f_n) : \mathbb{R}^n \mapsto \mathbb{R}^n$ is the flux function [4, 19, 25, 30, 31]. For smooth solutions, this can be written in quasilinear form

$$u_t + A(u)u_x = 0, \qquad A(u) = Df(u).$$

We recall that the system is *strictly hyperbolic* if each Jacobian matrix $A(u) = Df(u)$ has real distinct eigenvalues $\lambda_1(u) < \lambda_2(u) < \cdots < \lambda_n(u)$. In this case, one can find dual bases of right and left eigenvectors $r_i(u), l_j(u)$, normalized so that

$$|r_i(u)| \equiv 1, \qquad l_j(u) \cdot r_i(u) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{39}$$

The existence and uniqueness of entropy admissible weak solutions to (38) was initially developed relying on the following assumption, stating that the directional derivative of eigenvalue in the direction of the corresponding eigenvector is identically zero, or has always the same sign [22, 27].

Lax Conditions For each $i \in \{1, \ldots, n\}$, the i-th characteristic field is either *linearly degenerate*, so that $D\lambda_i \cdot r_i \equiv 0$, or *genuinely nonlinear*, so that $D\lambda_i \cdot r_i > 0$ at every $u \in \mathbb{R}^n$.

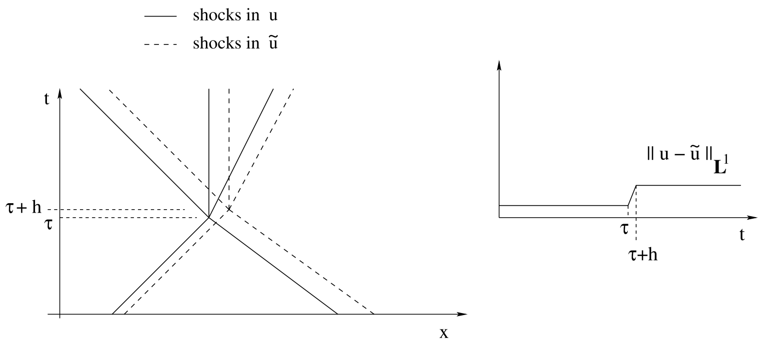


Fig. 6 The $\mathbf{L}^1$ distance between two nearby solutions can increase rapidly, during a short interval of time

In the case of a scalar conservation laws, a fundamental result of Kruzhkov [26] valid also in several space dimensions shows that the $\mathbf{L}^1$ distance between solutions does not increase in time. Indeed,

$$\|u(t,\cdot)-\tilde{u}(t,\cdot)\|_{\mathbf{L}^1} \le \|u(0,\cdot)-\tilde{u}(0,\cdot)\|_{\mathbf{L}^1}, \tag{40}$$

where $u(t,x)$, $\tilde{u}(t,x)$ are any two bounded, entropy-admissible solutions of (38). Thanks to this property, solutions to a scalar conservation law can also be constructed relying on the abstract theory of contractive semigroups [17].

For systems of two or more conservation laws, however, this contractive property fails. In general one cannot even find any constant L for which the property (P1) holds. For example, consider a solution $u = u(t,x)$ which initially contains two shocks, interacting at time τ and producing a third outgoing shock (see Fig. 6, left). Let $\tilde{u}$ be a perturbed solution, containing the same shocks, but slightly shifted in space. As a result, the interaction occur a bit later, say at time $\tau + h$. In this case, the $\mathbf{L}^1$ distance between the two solutions remains constant, except during the short interval $[\tau, \tau + h]$ where it increases very rapidly (Fig. 6, right).

Under the Lax conditions, two approaches are now available, in order to construct a distance functional on a domain d of functions with small total variation.

4.1 An Explicit Functional

In [12] an explicit formula was introduced, providing a functional Φ such that

$$\|u-v\|_{\mathbf{L}^1} \le \Phi(u,v) \le C\|u-v\|_{\mathbf{L}^1}, \tag{41}$$

and satisfying

$$\Phi(u(t',\cdot), v(t',\cdot)) \le \Phi(u(t,\cdot), v(t,\cdot)), \quad t < t', \tag{42}$$

for any couple of entropy-admissible weak solutions u, v to (38), with sufficiently small total variation. We review here the basic step of this construction.

1. Measuring the Strength of Shock and Rarefaction Waves Fix a state $u_0 \in \mathbb{R}^n$ and an index $i \in \{1, \ldots, n\}$. As before, let $r_1(u), \ldots, r_n(u)$ be the right eigenvectors of the Jacobian matrix $A(u) = Df(u)$, normalized as in (39). The integral curve of the vector field r_i through the point u_0 is called the *i-rarefaction curve* through u_0. It is obtained by solving the Cauchy problem in state space:

$$\frac{du}{d\sigma} = r_i(u), \qquad u(0) = u_0. \tag{43}$$

This curve, parameterized by arc-length, will be denoted as

$$\sigma \mapsto R_i(\sigma)(u_0). \tag{44}$$

Next, for a fixed $u_0 \in \mathbb{R}^n$ and $i \in \{1, \ldots, n\}$, one can show that there exists (locally, in a neighborhood of u_0) a unique smooth curve of states u which can be connected to the right of u_0 by an i-shock, satisfying the Rankine–Hugoniot equations

$$\lambda(u - u_0) = f(u) - f(u_0), \tag{45}$$

for some scalar speed λ, with $\lambda \to \lambda_i(u_0)$ as $u \to u_0$. This will be called the *i-shock curve* through the point u_0 and parameterized by arc-length:

$$\sigma \mapsto S_i(\sigma)(u_0). \tag{46}$$

It is well known that the two curves R_i, S_i have a second order contact at the point u_0. More precisely, the following estimates hold,

$$\begin{cases} R_i(\sigma)(u_0) = u_0 + \sigma r_i(u_0) + \mathcal{O}(1) \cdot \sigma^2, \\ S_i(\sigma)(u_0) = u_0 + \sigma r_i(u_0) + \mathcal{O}(1) \cdot \sigma^2. \end{cases} \tag{47}$$

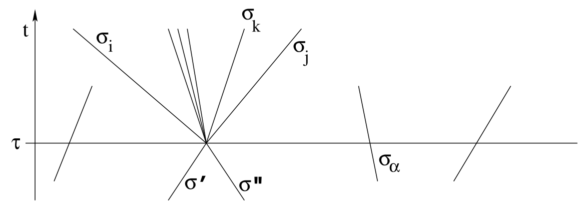
$$\big|R_i(\sigma)(u_0) - S_i(\sigma)(u_0)\big| = \mathcal{O}(1) \cdot \sigma^3. \tag{48}$$

Here and throughout the following, the Landau symbol $\mathcal{O}(1)$ denotes a quantity whose absolute value satisfies a uniform bound, depending only on the system (30).

Notice that the orientation of the unit vector $r_i(u_0)$ determines an orientation of the curves R_i, S_i. Recalling the Lax conditions, if the i-th characteristic field is genuinely nonlinear, the orientation is chosen so that the characteristic speed λ_i increases along the curves, as the parameter σ increases. On the other hand, if the i-th field is linearly degenerate, one can prove that $R_i(\sigma) = S_i(\sigma)$ for every σ, and that λ_i is constant along these curves. In this case, the orientation can be chosen arbitrarily.

2. The Interaction Potential It will be convenient to work within a special class of functions, which we call $\mathcal{PCS}$, consisting of all piecewise constant functions $u : \mathbb{R} \mapsto \mathbb{R}^n$, with simple jumps. We say that the jump at x is *simple* if either $u(x+) = R_i(\sigma)(u(x-))$ for some $\sigma > 0$ or $u(x+) = S_i(\sigma)(u(x-))$ for some $\sigma < 0$. In both cases, we regard $|\sigma|$ as the *strength* of the jump at x.

Fig. 7 Estimating the change in the total variation at a time where two fronts interact



For a piecewise constant function $u \in \mathscr{PCS}$, let x_α, $\alpha = 1, \dots, N$, be the locations of the jumps in $u(\cdot)$. Moreover, let $|\sigma_\alpha|$ be the strength of the wave-front at x_α, say of the family $k_\alpha \in \{1, \dots, n\}$. Following [22], we consider the two functionals

$$V(u) \doteq \sum_{\alpha=1}^{N} |\sigma_\alpha|, \tag{49}$$

measuring the *total strength of waves* in u, and

$$Q(u) \doteq \sum_{(\alpha,\beta)\in\mathscr{A}} |\sigma_\alpha \sigma_\beta|, \tag{50}$$

measuring the *wave interaction potential*. In (50), the summation ranges over the set $\mathscr{A}$ of all couples of approaching wave-fronts. More precisely, two fronts, located at points $x_\alpha < x_\beta$ and belonging to the characteristic families $k_\alpha, k_\beta \in \{1, \dots, n\}$ respectively, are *approaching* if $k_\alpha > k_\beta$ or else if $k_\alpha = k_\beta$ and at least one of the wave-fronts is a shock of a genuinely nonlinear family. Roughly speaking, two fronts are approaching if the one behind has the larger speed (and hence it can collide with the other, at a future time).

If now $u = u(t, x)$ is a piecewise constant approximate solution, a key observation is that the total strength of waves $V(u(t))$ can increase in time, but the interaction potential is monotone decreasing. Indeed, consider a time τ where two fronts of strength $|\sigma'|$, $|\sigma''|$ collide. Then the changes in V, Q are estimated by

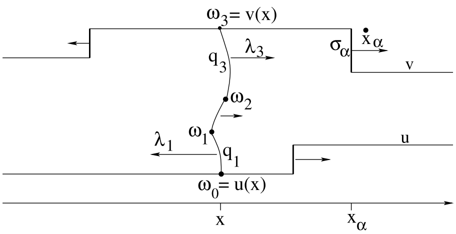
$$\Delta V(\tau) \doteq V(\tau+) - V(\tau-) = \mathscr{O}(1) \cdot |\sigma'\sigma''|, \tag{51}$$

$$\begin{aligned} \Delta Q(\tau) \doteq Q(\tau+) - Q(\tau-) &= -|\sigma'\sigma''| + \mathscr{O}(1) \cdot |\sigma'\sigma''| \cdot V(\tau-) \\ &\leq -\frac{|\sigma'\sigma''|}{2}, \end{aligned} \tag{52}$$

provided that $V(\tau-)$ is sufficiently small. Indeed (see Fig. 7), after time τ the two colliding fronts σ', σ'' are no longer approaching. Hence the product $|\sigma'\sigma''|$ is no longer counted within the summation (50). On the other hand, the new waves σ_k emerging from the interaction (having strength $\mathscr{O}(1) \cdot |\sigma'\sigma''|$) can approach all the other fronts not involved in the interaction (which have total strength $\leq V(\tau-)$).

By (51) and (52) we can thus choose a constant C_0 large enough so that the quantity $V(u(t)) + C_0 Q(u(t))$ is monotone decreasing, provided that V remains sufficiently small.

Fig. 8 Decomposing a jump $(u(x), v(x))$ in terms of n (possibly non-admissible) shocks



In turn, this yields an estimate on the total variation, globally in time:

$$\text{Tot.Var.}\{u(t)\} \leq V\big(u(t)\big) \leq V\big(u(0)\big) + C_0 Q\big(u(0)\big). \tag{53}$$

By Helly's theorem, this provides a crucial compactness property, toward a proof of the existence of globally defined weak solutions [22].

3. A Weighted Distance Functional Relying on the concepts and notations developed above, we can now describe the construction of a functional $\Phi(u, v)$, measuring the distance between solutions to the hyperbolic system (38) and satisfying the key properties (41)–(42).

Given two piecewise constant functions with simple jumps $u, v : \mathbb{R} \mapsto R^n$, recalling the construction of shock curves at (46), consider the scalar functions q_i defined implicitly by

$$v(x) = S_n\big(q_n(x)\big) \circ \cdots \circ S_1\big(q_1(x)\big)\big(u(x)\big). \tag{54}$$

Defining the intermediate states (see Fig. 8)

$$\omega_i \doteq S_i\big(q_i(x)\big) \circ \cdots \circ S_1\big(q_1(x)\big)\big(u(x)\big), \quad i = 0, 1, 2, \ldots, n,$$

this means that each couple of states ω_{i-1}, ω_i is connected by an i-shock of size $q_i(x)$. We regard $|q_i(x)|$ as the *strength* of the i-th component in the jump $v(x) - u(x)$, measured along shock curves. Since these curves are parameterized by arc length, as long as $u(x)$, $v(x)$ vary in a small neighborhood of the origin one clearly has

$$\big|v(x) - u(x)\big| \leq \sum_{i=1}^{n} \big|q_i(x)\big| \leq C_1 \cdot \big|v(x) - u(x)\big| \tag{55}$$

for some constant C_1. We can now define the functional

$$\Phi(u, v) \doteq \sum_{i=1}^{n} \int_{-\infty}^{\infty} \big|q_i(x)\big| W_i(x)\, dx, \tag{56}$$

where the weights W_i are defined by setting:

$$W_i(x) \doteq 1 + \kappa_1 \cdot \big[\text{total strength of waves in } u \text{ and in}$$

$$v \text{ which approach the } i\text{-wave } q_i(x)\big]$$
$$+\kappa_2 \cdot \big[\text{wave interaction potentials of } u \text{ and of } v\big]$$
$$\doteq 1 + \kappa_1 A_i(x) + \kappa_2\big[Q(u) + Q(v)\big]. \tag{57}$$

The quantity $A_i(x)$, accounting for the total strength of waves approaching an i-wave located at x, is defined as follows. If the i-th characteristic field is linearly degenerate, we simply take

$$A_i(x) \doteq \Bigg[\sum_{x_\alpha < x,\, i < k_\alpha \le n} + \sum_{x_\alpha > x,\, 1 \le k_\alpha < i}\Bigg] |\sigma_\alpha|. \tag{58}$$

The summations here extend to waves both of u and of v. Here $k_\alpha \in \{1, \ldots, n\}$ is the family of the jump located at x_α with size σ_α. On the other hand, if the i-th field is genuinely nonlinear, the definition of A_i contains an additional term, accounting for waves in u and in v of the same i-th family:

$$\begin{aligned} A_i(x) \doteq & \Bigg[\sum_{\substack{\alpha \in \mathcal{J}(u) \cup \mathcal{J}(v) \\ x_\alpha < x,\, i < k_\alpha \le n}} + \sum_{\substack{\alpha \in \mathcal{J}(u) \cup \mathcal{J}(v) \\ x_\alpha > x,\, 1 \le k_\alpha < i}}\Bigg] |\sigma_\alpha| \\ & + \begin{cases} \Bigg[\sum_{\substack{k_\alpha = i \\ \alpha \in \mathcal{J}(u),\, x_\alpha < x}} + \sum_{\substack{k_\alpha = i \\ \alpha \in \mathcal{J}(v),\, x_\alpha > x}}\Bigg] |\sigma_\alpha| & \text{if } q_i(x) < 0, \\ \Bigg[\sum_{\substack{k_\alpha = i \\ \alpha \in \mathcal{J}(v),\, x_\alpha < x}} + \sum_{\substack{k_\alpha = i \\ \alpha \in \mathcal{J}(u),\, x_\alpha > x}}\Bigg] |\sigma_\alpha| & \text{if } q_i(x) > 0. \end{cases} \end{aligned} \tag{59}$$

Here $\mathcal{J}(u)$ and $\mathcal{J}(v)$ denote the sets of all jumps in u and in v, while $\mathcal{J} \doteq \mathcal{J}(u) \cup \mathcal{J}(v)$.

As soon as the functional Φ is defined for piecewise constant functions, it can be extended to all functions $u \in \mathbf{L}^1(\mathbb{R}\mathbb{R}^n)$ having suitably small total variation, by taking the lower semicontinuous envelope:

$$\Phi_*(u, v) \doteq \liminf_{\substack{u' \to u,\, v' \to v \\ u', v' \in PCS}} \Phi(u', v'). \tag{60}$$

By choosing the constants $\kappa_2 \gg \kappa_1 \gg 1$ in (57) sufficiently large, if the total variation of the functions u, v remains small, the analysis in [12] shows that this functional is equivalent to the $\mathbf{L}^1$ distance and is non-increasing in time along couples of entropy-weak solutions to the system (38).

We remark that the functional $\Phi_*(\cdot, \cdot)$ in (60) is still not a distance, because it may not satisfy the triangle inequality. To achieve a distance, as in (36) one should define

$$d^{\Phi}(u, v) \doteq \inf\left\{ \sum_{i=1}^{N} \Phi(u_i, u_{i-1});\, u_0 = u,\, u_N = \tilde{u} \right\}.$$

In practice, it is more convenient to work out all the estimates on piecewise constant approximate solutions, using the explicit formula (57). The limits of approximate solutions, providing exact solutions, are taken only at the end.

An extension of these ideas to the initial-boundary value problem can be found in [20].

4.2 A Riemann Type Distance

With this approach, introduced in [3], one considers a family of sufficiently regular paths $\gamma : [0, 1] \mapsto \mathbf{L}^1$, for which a weighted length can be defined. For any couple of functions $u, \tilde{u}$, the weighted distance $d^{\diamond}(u, \tilde{u})$ is then defined as the infimum of lengths of all paths connecting u with $\tilde{u}$.

In connection with the system (38) we say that a function $u : \mathbb{R} \mapsto \mathbb{R}^n$ is in the class PLSD (*Piecewise Lipschitz with Simple Discontinuities*) if u is piecewise Lipschitz continuous with finitely many jumps, each jump consisting of a single, entropy admissible shock. In other words, at each point x_α where u has a jump, the left and right stated are related by

$$u(x_\alpha+) = S_i(\sigma_\alpha)(u(x_\alpha-)), \tag{61}$$

for some genuinely characteristic field i, and for some amplitude $\sigma_\alpha < 0$. The condition on the sign of σ_α guarantees that the shock is admissible.

If u is in PLSD and has N discontinuities at the points $x_1 < \cdots < x_N$, the space of *generalized tangent vectors* at u is defined as $T_u \doteq \mathbf{L}^1 \times \mathbb{R}^N$. Adopting the point of view of differential geometry, elements in T_u can be interpreted as first order tangent vectors as follows. On the family Σ_u of all continuous paths $\gamma : [0, \varepsilon_0] \to \mathbf{L}^1$ with $\gamma(0) = u$, define the equivalence relation

$$\gamma \sim \gamma' \quad \text{iff} \quad \lim_{\varepsilon \to 0+} \frac{1}{\varepsilon} \big\| \gamma(\varepsilon) - \gamma'(\varepsilon) \big\|_{\mathbf{L}^1} = 0. \tag{62}$$

We say that a continuous path $\gamma \in \Sigma_u$ *generates the tangent vector* $(v, \xi) \in T_u$ if γ is equivalent to the path $\gamma_{(v,\xi;u)}$ defined as

$$\begin{aligned} \gamma_{(v,\xi;u)}(\varepsilon) \doteq u + \varepsilon v &+ \sum_{\xi_\alpha < 0} \big[u(x_\alpha^+) - u(x_\alpha^-) \big] \chi_{[x_\alpha + \varepsilon \xi_\alpha,\, x_\alpha]} \\ &- \sum_{\xi_\alpha > 0} \big[u(x_\alpha^+) - u(x_\alpha^-) \big] \chi_{[x_\alpha,\, x_\alpha + \varepsilon \xi_\alpha]}, \end{aligned} \tag{63}$$

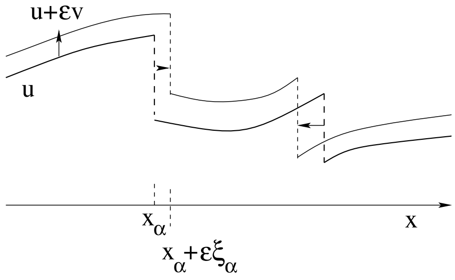


Fig. 9 A piecewise Lipschitz continuous function u and a perturbation $\gamma(\varepsilon)$ described by the tangent vector $(v,\xi)\in T_u$

where χ_I denotes the characteristic function of the interval I. Up to higher order terms, the perturbation $\gamma(\varepsilon)$ is thus obtained from u by adding εv and by shifting the points x_α, where the discontinuities of u occur, by $\varepsilon\xi_\alpha$ (see Fig. 9).

To define a norm on each tangent space T_u, we proceed as follows. Let u be a function in the class PLSD, with jumps at the points $x_1 < x_2 < \cdots < x_N$. For any $(v,\xi)\in T_u \doteq \mathbf{L}^1\times\mathbb{R}^N$, define the scalar components

$$v_i(x) \doteq l_i\big(u(x)\big)\cdot v(x). \tag{64}$$

Here $l_1,\ldots,l_n$ are the left eigenvectors of the Jacobian matrix $A(u)=Df(u)$, normalized as in (39). Following [3], the weighted norm is defined as

$$\big\|(v,\xi)\big\|_u \doteq \sum_{\alpha=1}^{N}|\sigma_\alpha||\xi_\alpha|W^u_{k_\alpha}(x_\alpha)+\sum_{i=1}^{m}\int_{-\infty}^{\infty}\big|v_i(x)\big|W_i^u(x)\,dx, \tag{65}$$

where $W_i^u(x)$ is the weight given to an i-wave located at x. More precisely:

$$W_i^u(x)\doteq 1+\kappa_1\,A_i^u(x)+\kappa_2 Q(u), \tag{66}$$

where

$$A_i^u(x)\doteq\left[\sum_{j\le i}\int_x^\infty+\sum_{j\ge i}\int_{-\infty}^{x}\right]|u_x^j(y)|\,dy+\left[\sum_{\substack{k_\alpha\le i\\ x_\alpha>x}}+\sum_{\substack{k_\alpha\ge i\\ x_\alpha<x}}\right]|\sigma_\alpha|$$

measures the total amount of waves in u approaching an i-wave located at x, while $Q(u)$ is the interaction potential, introduced at (50), and $1\ll\kappa_1\ll\kappa_2$ are suitable constants.

Let now $\gamma:[0,1]\mapsto\mathbf{L}^1$ be a Lipschitz continuous curve such that, for all but finitely many values of θ, the functions $\gamma(\theta)$ is in PLSD and the tangent vector $\dot\gamma=(v(\theta),\xi(\theta))\in T_{\gamma(\theta)}$ is well defined. One can then define the weighted length of γ by integrating the weighted norm of its tangent vector:

$$\|\gamma\|_*\doteq\int_0^1\|\dot\gamma(\theta)\|_{\gamma(\theta)}\,d\theta. \tag{67}$$

In turn, this provides a notion of distance between two functions $u, \tilde{u}$, as in (3)–(4).

By choosing the constants κ_1, κ_2 large enough, this distance is non-increasing along any couple of entropy-weak solutions to the hyperbolic system (38), having suitably small total variation. See [6, 9] for two implementations of this approach.

Remark 4 All of the previous analysis dealt with solutions having small total variation. An extension to large BV data has been achieved in [28]. In this case, a contractive metric can be constructed on a domain of functions consisting of small BV perturbations of a (possibly large) Riemann solution. While the total strength of waves $V(u)$ here can be large, the interaction potential $Q(u)$ must remain sufficiently small.

We remark that, for general initial data with large interaction potential, the a priori BV estimates in [22] do not apply and even the global existence of weak solutions remains an open problem.

Remark 5 For strictly hyperbolic systems which do not satisfy the Lax conditions, a Lipschitz semigroup of globally defined, entropy weak solutions was constructed in [1], taking limits of vanishing viscosity approximations. In this general case, a distance which is contractive w.r.t. the flow generated by (38) has not yet been constructed. Extending the explicit definition (56)–(57) appears to be a very difficult task. On the other hand, since the continuous dependence of viscous approximations was proved in [1] by studying the weighted length of smooth paths of solutions, constructing a Riemann type metric as in (65)–(67) may be a more promising approach.

References

1. Bianchini, S., Bressan, A.: Vanishing viscosity solutions of nonlinear hyperbolic systems. Ann. Math. **161**, 223–342 (2005)
2. Bressan, A.: Unique solutions for a class of discontinuous differential equations. Proc. Am. Math. Soc. **104**, 772–778 (1988)
3. Bressan, A.: A locally contractive metric for systems of conservation laws. Ann. Sc. Norm. Super. Pisa, Cl. Sci. (IV) **XXII**, 109–135 (1995)
4. Bressan, A.: Hyperbolic Systems of Conservation Laws. The One Dimensional Cauchy Problem. Oxford University Press, London (2000)
5. Bressan, A., Colombo, G.: Existence and continuous dependence for discontinuous O.D.E.'s. Boll. Unione Mat. Ital. B **4**, 295–311 (1990)
6. Bressan, A., Colombo, R.M.: The semigroup generated by 2×2 conservation laws. Arch. Ration. Mech. Anal. **133**, 1–75 (1995)
7. Bressan, A., Constantin, A.: Global solutions to the Hunter–Saxton equations. SIAM J. Math. Anal. **37**, 996–1026 (2005)
8. Bressan, A., Constantin, A.: Global conservative solutions to the Camassa–Holm equation. Arch. Ration. Mech. Anal. **183**, 215–239 (2007)
9. Bressan, A., Crasta, G., Piccoli, B.: Well posedness of the Cauchy problem for $n \times n$ systems of conservation laws. Mem. Am. Math. Soc. **694** (2000)
10. Bressan, A., Fonte, M.: An optimal transportation metric for solutions of the Camassa–Holm equation. Methods Appl. Anal. **12**, 191–220 (2005)

11. Bressan, A., Zheng, Y.: Conservative solutions to a nonlinear variational wave equation. Commun. Math. Phys. **266**, 471–497 (2006)
12. Bressan, A., Liu, T.P., Yang, T.: L^1 stability estimates for $n \times n$ conservation laws. Arch. Ration. Mech. Anal. **149**, 1–22 (1999)
13. Bressan, A., Holden, H., Raynaud, X.: Lipschitz metric for the Hunter–Saxton equation. J. Math. Pures Appl. **94**, 68–92 (2010)
14. Camassa, R., Holm, D.D.: An integrable shallow water equation with peaked solitons. Phys. Rev. Lett. **71**, 1661–1664 (1993)
15. Constantin, A., McKean, H.P.: A shallow water equation on the circle. Commun. Pure Appl. Math. **52**, 949–982 (1999)
16. Constantin, A., Molinet, L.: Global weak solutions for a shallow water equation. Commun. Math. Phys. **211**, 45–61 (2000)
17. Crandall, M.G.: The semigroup approach to first-order quasilinear equations in several space variables. Isr. J. Math. **12**, 108–132 (1972)
18. Crandall, M.G., Liggett, T.M.: Generation of semi-groups of nonlinear transformations on general Banach spaces. Am. J. Math. **93**, 265–298 (1971)
19. Dafermos, C.: Hyperbolic Conservation Laws in Continuum Physics, 2nd edn. Springer, Berlin (2005)
20. Donadello, C., Marson, A.: Stability of front tracking solutions to the initial and boundary value problem for systems of conservation laws. Nonlinear Differ. Equ. Appl. **14**, 569–592 (2007)
21. Filippov, A.F.: Differential Equations with Discontinuous Righthand Sides. Kluwer Academic, Dordrecht (1988)
22. Glimm, J.: Solutions in the large for nonlinear hyperbolic systems of equations. Commun. Pure Appl. Math. **18**, 697–715 (1965)
23. Grunert, K., Holden, H., Raynaud, X.: Lipschitz metric for the periodic Camassa–Holm equation. Preprint (2010)
24. Henry, D.: Geometric Theory of Semilinear Parabolic Equations. Lecture Notes in Mathematics, vol. 840. Springer, Berlin (1981)
25. Holden, H., Risebro, N.H.: Front Tracking for Hyperbolic Conservation Laws. Springer, New York (2002)
26. Kruzhkov, S.: First order quasilinear equations with several space variables. Math. USSR Sb. **10**, 217–243 (1970)
27. Lax, P.: Hyperbolic systems of conservation laws II. Commun. Pure Appl. Math. **10**, 537–566 (1957)
28. Lewicka, M.: Well-posedness for hyperbolic systems of conservation laws with large BV data. Arch. Ration. Mech. Anal. **173**, 415–445 (2004)
29. Martin, R.H. Jr.: Nonlinear Operators and Differential Equations in Banach Spaces. Wiley-Interscience, New York (1976)
30. Serre, D.: Systems of Conservation Laws I, II. Cambridge University Press, Cambridge (2000)
31. Smoller, J.: Shock Waves and Reaction-Diffusion Equations. Springer, New York (1983)
32. Xin, Z., Zhang, P.: On the weak solutions to a shallow water equation. Commun. Pure Appl. Math. **53**, 1411–1433 (2000)
33. Xin, Z., Zhang, P.: On the uniqueness and large time behavior of the weak solutions to a shallow water equation. Commun. Partial Differ. Equ. **27**, 1815–1844 (2002)
34. Yosida, K.: Functional Analysis. Classics in Mathematics. Springer, Berlin (1995). Reprint of the sixth (1980) edition

Non-local Diffusions, Drifts and Games

Luis Caffarelli

Abstract This is a brief discussion of the properties of solutions to several non-linear elliptic equations involving diffusive processes of non-local nature. These equation arise in several contexts: from continuum mechanics and phase transition, from population dynamics, from optimal control and game theory. The equations coming from continuum mechanics exhibit a variational structure and a theory parallel to the De Giorgi–Nash–Moser was necessary to show existence of regular solutions. Population dynamics suggests "porous media like equations" with a non-local pressure, and from optimal control we obtain fully non-linear equations that require methods of the type of the Krylov–Safonov–Evans theory. Finally, we discuss some non-local p and infinite Laplacian models coming from game theory.

1 Introduction

We are interested in integral diffusion equations:

$$u_t(x,t) = [L(u)](x,t)$$

where the operator L takes the form

$$L(u(x,t)) = \int [u(y) - u(x)] K(x,y)\, dy$$

for some positive kernel (or measure) $K(x,y)$ (or $K_x(y)$).

We call the equation a diffusion equation because solutions try to revert to some sort of "integral average" of u.

Indeed, if $u(x_0)$ is "smaller than" its surrounding values, as weighted by $K(x,y)$, $u(x_0,t)$ will tend to increase, if "bigger", to decrease (i.e., $u_t > 0$ or $u_t < 0$).

We may think of the heat equation as an infinitesimal version of this process.

L. Caffarelli (✉)
Department of Mathematics, The University of Texas at Austin, 1 University Station—C1200, Austin, TX 78712, USA
e-mail: caffarel@math.utexas.edu

H. Holden, K.H. Karlsen (eds.), *Nonlinear Partial Differential Equations*, Abel Symposia 7, DOI 10.1007/978-3-642-25361-4_3,

Indeed the Laplacian, Δu, is the limit of

$$\Delta u(x_0) = \lim_{r\to 0} \frac{1}{r^2} \fint_{B_r(x_0)} u(x) - u(x_0)\, dx.$$

In fact, if

$$K_\varepsilon(x, y) = \frac{1}{\varepsilon^2}\left(\frac{1}{\varepsilon^n}\varphi\left(\frac{x-y}{\varepsilon}\right)\right) = \frac{1}{\varepsilon^2}\varphi_\varepsilon(x-y)$$

for φ a probability density (a mollifier), the corresponding solutions u_ε should converge to a solution $u_0(x,t)$ of

$$(u_0)_t = a_{ij} D_{ij} u_0$$

where a_{ij} are the second moments of φ.

These types of equations (and the associated non-linear ones that we will discuss shortly) have roots in different phenomena and, as their second order counterpart, they naturally divide between those with variational structure and those coming from probabilistic considerations.

A familiar example for the first case is prescribing Neumann boundary data (for instance zero). Insulating a wall implies some temperature diffusivity along the surface, expressed by inverting the Dirichlet to Neumann map. A non-local related equation is the quasi-geostrophic equation that describes the evolution of temperature on the ocean surface, due to the (one-side) atmospheric conditions.

On the probabilistic side let us recall the Levy–Khintchine formula. In an informal "black-box" approach suppose we can observe the transition probability of a distribution of particles, for any sequence of times t_k, and we realize that the transition from t_1 to t_2 only depends on t_1 and t_2, in fact on $t_2 - t_1$.

Then, for any k, we can write the transition probability from t_1 to t_2 as the composition (convolution) k times of the transition from $t = 0$ to $t = \frac{1}{k}(t_2 - t_1)$. This suggests the possibility, as δt goes to zero, of describing the process through a "heat equation"—as a properly scaled infinitesimal limit of the δt transition.

This is what the Levy–Khintchine formula asserts: That the probability density evolves according to a heat equation

$$u_t = \cdots$$

consisting of a continuous part

$$\cdots = a_{ij} D_{ij} u + b_j \nabla u + \cdots$$

a symmetric jump process

$$\int [u(x+y) + u(x-y) - 2u(x)]\, d\mu(y)$$

$+\cdots$ an asymmetric part that we will discuss later

$$d\mu \text{``}=\text{''} K(y)\, dy.$$

Here, it is required to make sense for a C^2, bounded u, i.e.,

$$\int_{B_1(0)} \|y\|^2 \, d\mu(y) < \infty$$

and

$$\int_{\mathscr{C}(B_1(0))} d\mu(y) < \infty.$$

Between divergence and non-divergence lie the equations invariant under translations, i.e., where the kernel $K(x, y) = \widetilde{K}(x - y)$. In this case, the equation can be thought of as having both divergence and non-divergence structure and also, being of convolution type, they enjoy the advantage of allowing for methods of harmonic analysis.

That is the case, for instance, with the family of fractional Laplacians: For $0 < \alpha < 1$

$$\text{“}\Delta^{\alpha}\text{”}(x) = C(\alpha) \int [u(y) - u(x)] \frac{1}{|x - y|^{n+2\alpha}}, dy = (\hat{u}(\xi)|\xi|^{2\alpha})^{v}.$$

The constant $C(\alpha) \sim (1 - \alpha)$ to recuperate the standard Δu, as α goes to one.

Notice that the range of α's is such that it makes these kernels satisfy the Levy–Khintchine condition to be an infinite divisible distribution.

In fact the fractional Laplacians are also called “stable processes”.

On the other hand, the fractional Laplacians is what we obtain as an Euler–Lagrange equation for the energy integral corresponding to the $W^{\alpha,2}$ (the L^2 norm of the “alpha” fractional derivatives of u):

$$\text{“}D^{\alpha} u(x) = \int [u(y) - u(x)] \frac{1}{|x - y|^{n+\alpha}} \, dy\text{”}.$$

And finally, convolution with the Δ^{α} kernel corresponds after Fourier transform to the multiplier

$$(\widehat{\Delta^2}) = -|\xi|^{2\alpha}.$$

In that sense, the fractional Laplacian serves as a basic model for the three classical methods of second order PDE's.

- Superposition (potential theory, harmonic analysis)
- Energy method (calculus of variations, DeGiorgi–Nash–Moser)
- Probabilistic (optimal control-Krylov–Safonov)

Since we are interested in regularity properties of solutions to such an “elliptic” or “parabolic” equation, the kernel $K(x)$ should be singular at the origin to force u to be somewhat “special” in order to satisfy the equation: To know that after convolution with a smooth function u is smooth does not reflect so much on the regularity of u, at least at first glance.

In that sense, the fractional Laplacians provide a natural comparison scale of “order of differentiation” of the operator to help us develop a general setting.

2 Divergence Structure

Equations with "divergence" structure arise from continuum mechanics and calculus of variations.

A rough characterization would be that the kernel $K(x, y)$ is symmetric.

That makes the equation

$$\int [u(y) - u(x)]K(x, y)\,dy = 0$$

the Euler–Lagrange equation of

$$E(u)T(u) = \iint [u(x) - u(y)]^2 K(x, y)\,dx\,dy$$

and thus puts the problem in the framework of weak variational solutions test functions methods, etc.:

For a test function $\varphi(x)$, the bilinear form

$$B(u, \varphi) = \iint [u(y) - u(x)]K(x, y)[\varphi(y) - \varphi(x)],$$

depending on the problem at hand, must be zero, or prescribed or equal to

$$\int \varphi(y)u_t(x, t)$$

in the parabolic setting.

The general "non-linear calculus of variations" framework becomes then the study of the minimizers of the form:

$$\int \phi(u(x) - u(y))K(x - y)\,dx\,dy$$

with ϕ convex (quadratic for "uniform" fractional ellipticity).

The first, natural problem to study is that of regularity of local minimizers (the equivalent of the DeGiorgi solution of the Hilbert problem and the development of the DeGiorgi–Nash–Moser theory of regularity of solutions). Let us recall that in the second order case, the theory proceeds as follows:

A local minimizer, u, of the functional

$$E(w) = \int F(\nabla w)\,dx$$

satisfies the Euler–Lagrange equation

$$D_{x_i} F_i(\nabla u) = 0$$

or, in non-divergence form:

$$F_{ij}(\nabla u)D_{x_i x_j} u = 0.$$

If we would known that ∇u is continuous Shauder estimates would allow us to bootstrap the solution to higher regularity. In turn, first derivatives $D_e u = w$ satisfy

$$D_{x_i} F_{ij}(\nabla u) D_{x_j} w = 0.$$

But at this point we only know that ∇u is in L^2 and, from the uniform convexity of F, that the matrix $F_{ij}(\cdot) = A_{ij}(x)$ is strictly positive:

$$\lambda I \leq F_{ij}(\cdot) \leq \Lambda I.$$

But then, the celebrated DeGiorgi theorem establishes that solutions of an elliptic equation

$$D_i a_{ij}(x) D_j w = 0$$

with *no regularity* assumption on a_{ij} are Hölder continuous.

In particular, ∇u is Hölder continuous and higher regularity follows.

In this context, with Chan and Vasseur [9], we develop the DeGiorgi regularity theory for the parabolic case:

Let $u(x,t)$ be the solution of

$$u_t(x,t) = \int \phi'(u(x) - u(y)) K(x-y)$$

with "ϕ symmetric and quadratic" (i.e., $\lambda \leq \phi'' \leq \Lambda$) and

$$(1-\alpha) m |z|^{-(n+2\alpha)} \leq K(z) \leq (1-\alpha) M |z|^{-(n+2\alpha)}.$$

Then u becomes instantaneously smooth.

As in the second order case, the central step is to prove that first derivatives, $w = D_x u$, satisfy a "rough equation" and are Hölder continuous:

$$w_t(x,t) = \int [w(y,t) - w(x,t)] \underbrace{\phi''(u(y,t) - u(x,t)) K(x-y)}_{\substack{\text{"symmetric, measurable} \\ \text{fractional Laplacian like} \\ \text{kernel" } K(x,y,t)}} \, dy$$

(see also related articles by Barlow, Bass, Chen, Kassman, and of Komatsu [1, 3, 14, 15]).

The study of non-local, non-linear equations with "variational structure" has several motivations:

- What we could call surface diffusion: the quasigeostrophic equation that models ocean atmosphere interaction, the theory of semi-permeable membranes, planar fracture dynamics (see [5, 11]).
- Problems in statistical mechanics, like phase transition problems with long range interactions (as opposed to neighbor to neighbor). See for instance the work of Giacomin–Lebowitz and of Presutti [12].
- Material sciences, for instances polymers where many scales interact.
- Image processing, see for instance the work of Gilboa and Osher [13].

3 Non-divergence Equations

"Non-divergence" equations arise instead from probability (Levy processes), optimal control and game theory.

Suppose for instance that particles generate at some point x_0 of a domain Ω and bounce randomly until they exit Ω.

At that moment they release an amount of energy $u(y)$ depending on the point y where they land.

In principle to find out the expectation for future released energy $u(x_0)$ when starting at x_0, we should just solve $Lu = 0$ in Ω with external data $u(y)$ and the diffusion associated to the process.

In the case of optimal control we are able to "design" the jump process (the media) to maximize the expected value $u(x_0)$.

That is: We have a family of possible diffusion processes given by the kernels

$$L_\alpha u(x) = \int [u(x+y) - u(x)] K_\alpha(y)\, dy$$

and at each x we want to chose the optimal jump distribution

$$L_{\alpha(x)} = \int [u(x+y) - u(x)] K_{\alpha(x)}(y)\, dy.$$

In order to achieve that we have to find a solution u_0 of the equation

$$F(u_0) = \sup_\alpha L_\alpha u_0 = 0$$

with exterior data $u(y)$.

Indeed, this equation means that "u_0 is a supersolution of all the admissible operators, and at each point is the solution of at least one of the L_α." Therefore on one hand it is better than any choice and at the same time is an admissible distribution.

In the case of second order equations, the central result of the theory is the Evans–Krylov theorem:

In that case, the family of operators are second order

$$L_\alpha(u) = \sum a_{ij}^\alpha D_{ij} u,$$

the non-linear equation is

$$F(D^2 u) = \sup_\alpha \sum a_{ij}^\alpha D_{ij} u$$

and the Evans–Krylov theorem asserts that solutions to $F(D^2 u) = 0$ are $C^{2,\beta}$ and thus **classical** (i.e., the derivatives involved are continuous).

In collaboration with Silvestre, we reproduce their theory for the corresponding non-local equations [6–8].

If the kernels K_α are all comparable to the s-Laplacian:

$$\lambda(1-s)|y|^{-(n+2s)} \le K_\alpha(y) \le \Lambda(1-s)|y|^{-(n+2s)}$$

and they are **symmetric** in y (no "drift"), then solutions to $F(u)=0$ are $C^{2s+\beta}$ that makes the corresponding integrals convergent and the solutions "classical".

One of the main features of the work is the proof of a theorem equivalent to the Krylov–Safonov Harnack inequality for "bounded measurable" kernels:

If w is for every x a solution of a different equation

$$L_x(w) = \int [w(x+y) - w(x)] K_x(y)\, dy = 0$$

with K_x changing discontinuously with x "bounded measurable coefficients", w is still Hölder continuous.

4 Drifts

What I want to discuss now is the relation, or interaction between diffusion and drift in the optimal control context:

For second order equations, when addressing gradient dependence of an equation, we have two different issues. On one hand semilinear equations, say, for instance

$$\Delta u = g(u, \nabla u)$$

with an associated idea of drift or transport and on the other quasilinear equations:

$$a_{ij}(\nabla u) D_{ij} u = 0$$

for instance those coming from the calculus of variations.

Semilinear equations with fractional diffusions arise for instance in the case of the quasigeostrophic equation:

$$\text{“}u_t - \Delta^s u = g(u, \nabla u)\text{”}$$

and assuming nice dependence on u, there is here a clear competition between diffusion and transport that becomes critical where $s = 1/2$.

But there is a second, implicit form of drift in the asymmetry of the kernel for a Levy process:

The most general "heat equation" for an infinite divisible distribution, leaving aside the continuous part and the standard drift is

$$\begin{aligned} u_t &= \frac{1}{2}\int [u(x+y) + u(x-y) - 2u(x)]\, d\mu(y) \\ &\quad + \frac{1}{2}\int \big([u(x+y) - u(x-y)] - 2(\nabla u(x), y)\chi_{B_1}\, d\mu\big)\, d\mu \\ &= \text{symmetric} + \text{antisymmetric}. \end{aligned}$$

Note that the antisymmetric part has in it an extra cancellation to ensure that the process does not drift to infinity.

For quasilinear equations, one equivalent framework to the second order case is, of course, through the calculus of variations.

For instance, one defines the p- (s-Laplacian), i.e., s-derivatives in L^p, as the Euler–Lagrange equation of the L^p norm of the s-derivatives of a function

$$\|u\|^p_{W^{s,p}} = \iint \frac{[u(x) - u(y)]^p}{|x-y|^{n+sp}}\,dx\,dy.$$

This p-fractional Laplacian is naturally studied through "energy" and "test functions methods" (see [10]). But the p-Laplacian also can also be written in non-divergence form as

$$(p - \Delta)u = |\nabla u|^{p-2}(\Delta u + (p-2)u_{nn})$$

where u_{nn} denotes the second derivative in the direction of the gradient of u.

And this has a game-theoretical interpretation (Peres–Sheffield [16]): Let us go back to the example of expected energy release $u(x)$ of the random particle.

Suppose that as before the random process has the ("almost continuous") diffusion equation ($\delta_t \sim \varepsilon^2$)

$$\delta_t u(x,t) = \int [u(y+x)(y,t) - u(x,t)]\frac{1}{\varepsilon^2}\varphi_\varepsilon(y)\,dy$$

i.e., the particle at position x at time t, jumps, by time $t + \varepsilon^2$, to a position epsilon-away, according to the radially symmetric probability density $\varphi_\varepsilon(y) = \frac{1}{\varepsilon^n}\varphi(y/\varepsilon)$.

Then, as discussed before, when ε goes to zero, we would get the standard "heat" equation.

But, assume now that competing players P_1, P_2 are able to impose on the jump an epsilon-drift in their preferred direction, randomly in time, trying to maximize, respectively minimize, the expected value u.

That is, depending on which player has the input, the particle at x will jump to the position $(x+y)$, with probability density ($\tau_i = \tau_1$ or τ_2, a unit vector)

$$\varphi_\varepsilon(y + \lambda\tau_i) = \frac{1}{\varepsilon^n}\left(\varphi\left(\frac{y+\lambda\tau_i}{\varepsilon}\right)\right).$$

As a consequence, the jump probability density φ_ε has drifted in the direction τ_1 or τ_2 depending on which player imposed the drift. Here λ is the intensity of the drift and the expected value u will then satisfy the Isaac's equation

$$\inf_{\tau_1\in S^1}\sup_{\tau_2\in S^1}\left[\frac{1}{2\varepsilon^2}\int [u(x+y) - u(x)]\varphi_\varepsilon(y+\lambda\tau_1) + \varphi_\varepsilon(y+\lambda\tau_2)\,dy\right] = 0.$$

The natural choice for τ_2 is to push the drift in the direction of ∇u, and for τ_1 in that of $-\nabla u$. Therefore, if both players use the optimal strategy, the combination of

$$\varphi_\varepsilon(y+\lambda\tau_1) + \varphi_\varepsilon(y+\lambda\tau_2)$$

will shift the mass of φ_ε symmetrically in the directions of $\pm\nabla u$, increasing the second moment in that direction so that the limiting equation, as epsilon goes to zero becomes

$$\Delta u + C(\lambda)u_{nn}$$

i.e., the non-divergence form of the fractional Laplacian.

A similar argument can be made for jump processes:

In work with Bjorland and Figalli, we have studied existence and regularity properties of this "tug of war" game for jump processes. Let me start by pointing out that there are different ways to "influence the drift" that give rise to structurally different mathematical problems. A possible one is for shifted kernels:

That is, for kernels of the form

$$K_{e_1}(y) = K_0(y)[1 + A(y_1)]$$

with $K_0(y)$ a symmetric kernel of the size of a fractional Laplacian, and $A(y_1)$ a smooth odd function, $|A(y_1)| \le 1 - \delta$.

That is, we look at the Isaac's equation:

$$\inf_{\nu_1} \sup_{\nu_2} \frac{1}{2} \int [u(x+y) - u(x)] K_0(y)[2 + A(y \cdot \nu_1) + A(y \cdot \nu_2)]\, dy$$

(i.e., each player adds the implicit drift $A(y \cdot \nu)$ in his optimal direction ν_1 or ν_2).

Another possible way is that the player chooses a direction and it is this direction that suffers a random deviation (an "unsteady hand"). In that case the corresponding basic kernel $K_{e_1}(y)$ should be of the form

$$K_{e_1}(y) = K_0(y)\eta(\sigma \cdot e_1)$$

where η may vanish outside a neighborhood of e_1.

The final operator is as before, the inf sup over all rotations of $K_{e_1}(y)$.

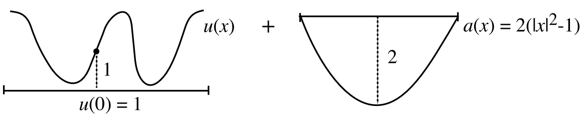
In both cases, it follows from the non-local Harnack inequality and ABP theorem [6] that solutions are C^α for some α.

In fact, let me take this opportunity to discuss informally the non-local ABP theorem, that is central to many of the developments for non-local optimal control.

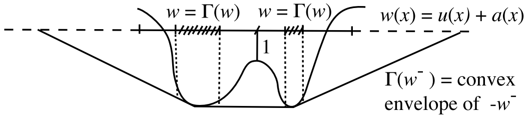
The local version of the ABP theorem needed for the Harnack inequality (as presented in [4]) is the following:

Theorem 1 *$u \ge 0$ in B_1, $Lu = a_{ij}(x)D_{ij}u \le 0$, $u(0) \le 1$. Then, $\exists \varepsilon_0$, such that $|\{u < 2\}| \ge \varepsilon_0(\lambda, \Lambda) > 0$.*

Proof We add to u a negative paraboloid in B_1:



and construct its convex envelope in B_1:

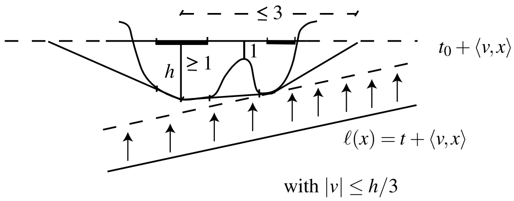


We will estimate $|\{w=\Gamma(w)\}|$ by below.

Indeed, in this set w is negative and

$$u = w - a \le 0 + 2.$$

For this purpose, we use the classical A-B-P argument, i.e., we estimate the volume of the image of the gradient map: $\nabla\Gamma : B_1 \to \mathbb{R}^n$. To do that, we lift from minus infinity a plane with generic slope v:



If $\ell(x) = t + \langle v, x\rangle$ with $|v| \le h/3$, for some value t_0, ℓ is a supporting plane of $\Gamma(w^-)$ at some interior point $x_0 \in \{\Gamma = w\}$. Thus "the image of $\{w^- = \Gamma(w)\}$ by the map: $x \to \nabla\Gamma(x)$ contains the ball of "v's" of radius $\frac{h}{3} = \frac{\sup w^-}{3} \ge \frac{1}{3}$, i.e.,

$$\left(\frac{1}{3}\right)^n \le C\,\mathrm{Vol}[\nabla\Gamma(\{w^- = \Gamma(w)\})].$$

We now "change variables", from v to x

$$1 \le C\,\mathrm{Vol}[\nabla\Gamma(\{w^- = \Gamma w\})] = \int 1\,dv = \int_{\{w=\Gamma(w)\}} |\det D\nabla\Gamma|\,dx.$$

But $D\nabla\Gamma = D^2\Gamma$, a non-negative matrix, since Γ is convex, so

$$\left(\frac{1}{3}\right)^n \le \int_{\{w=\Gamma(w)\}} \det D^2\Gamma \le \int_{\{w=\Gamma(w)\}} \det D^2 w \le \int_{\{w=\Gamma w\}} [\mu_{\max}(D^2(w))]^n$$

with μ the largest eigenvalue of D^2w (at a contact point $w=\Gamma(w)$, $D^2w \ge D^2\Gamma \ge 0$). Since all max eigenvalues of $D^2w \ge 0$,

$$Lw \cong \lambda\mu_{\max}, \quad \text{but also} \quad Lw \le Lh \le 2n\Lambda.$$

We then get

$$1 \le C\left(\frac{\Lambda}{\lambda}\right)^n \int_{\{w=\Gamma(w)\}} 1 = C\left(\frac{\Lambda}{\lambda}\right)^n |\{w=\Gamma(w)\}|.$$

This is "almost" the proof of the ABP version we need for the Harnack inequality.

What we are missing is the localization property:

$$\text{“}\{w=\Gamma(w)\}\cap Q_{1/4}(y) \quad \text{for } |y|\le 1/4\text{”} \quad \text{instead of} \quad \{w=\Gamma(w)\},$$

i.e., we need the extra fact that we can get the contact set to be inside of any cube of size $1/4$ close to the origin in order to make a C-Z decomposition. For that, all we need is to change h by an h' with: $Lh'\le 0$ outside $Q_{1/4}$ (so that $Lw\le 0$ outside $Q_{1/4}$ and contact cannot occur), $h'(0)\le -2$ so $\inf w\le -1$, and Lh' still bounded above, so Lw is bounded above. □

5 The Corresponding ABP for Integral Diffusions [6]

As before, we assume $u\ge 0$ in B_3, $Lu\le 0$, $u(0)\le 1$. Now

$$Lu(x)=\int [u(x+y)+u(x-y)-2u(x)]K_x(y)\,dy=\int \delta_2 u(x,y)K_x.$$

For simplicity we will truncate K_x:

$$\lambda(2-s)|y|^{-n+s}\chi_{B_1}(y)\le K_x(y)\le \Lambda(2-s)|y|^{-(n+s)}\chi_{B_1}(y)$$

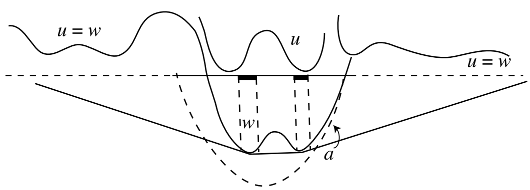
and restrict ourselves to $x\in B_1(0)$, so L is well defined.

We want to show:

$$\text{“}\exists M,\varepsilon>0,\ M,\varepsilon(\lambda,\Lambda,s), \quad \text{such that} \quad |\{u<M\}\cap B_1|\ge\varepsilon\text{”}$$

$$M,\varepsilon \text{ deteriorate with } s \text{ only for } \boxed{s\to 0}.$$

We proceed as before. Consider $w=u+a$, with $a=2(|x|^2-1)\wedge 0$ and construct the convex envelope $\Gamma(w^-)$ in B_3

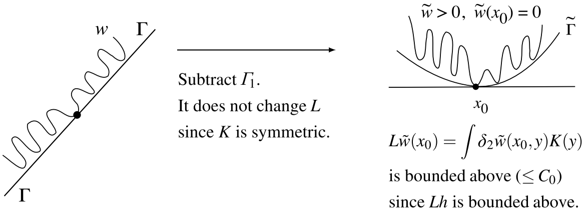


As before

$$\text{Vol } \nabla\Gamma(\{w=\Gamma(2)\})\ge\left(\frac{|\inf w|}{4}\right)^n.$$

The problem is how to relate $|\{w=\Gamma(w)\}|$ with its image (no good change of variables formula).

Consider a point x_0 in $w = \Gamma(w)$. We have the following local geometry



We are going to prove the following family of *steps*. Consider x_0 as above.

(a) For some small diadic ring $\mathbb{R}_k = B_{2^{-k}} \setminus B_{2^{-(k+1)}}$ $\tilde{w}$ "grows quadratically on average" in the sense that

$$\fint_{R_k} \tilde{w} \leq C_1 (r_k)^2 \quad (r_k = 2^{-k}).$$

(b) Of course, this does not imply that $\tilde{w} \leq c_1 r_k^2$, but since $0 \leq \tilde{\Gamma}(\tilde{w}) \leq \tilde{w}$, and $\tilde{\Gamma}$ is convex. (a) does imply that

$$\tilde{\Gamma}|_{B_{2^{-(k+1)}}} \leq C_2 2^{-2k} \quad \text{and} \quad \nabla \tilde{\Gamma}|_{B_{2^{-(k+2)}}} \leq C_2 2^{-k}$$

($k = k(x_0)$ of course).

(c) In particular:

$$\operatorname{Vol} \nabla \Gamma(B \ldots) = \operatorname{Vol} \nabla \tilde{\Gamma}(B_{2^{-(k+2)}}(x_0)) \leq C |B_{2^{-(k+2)}}(x_0)|.$$

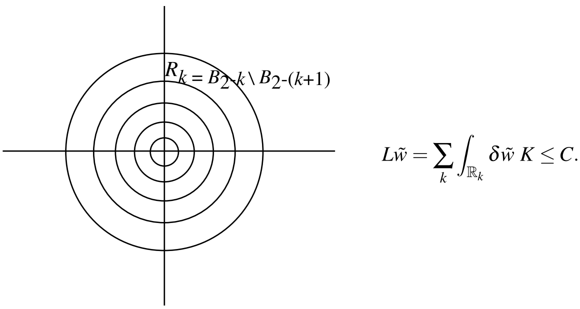
We now extract a covering of $\{w = \Gamma(w)\}$ with the family of these balls $B_{r(x)}(x)$ and we have

$$1 \leq \operatorname{Vol} \nabla \Gamma(\{w = \Gamma\}) \leq C \sum |B_{r(x_j)}(x_j)|.$$

But in each B_{r_j}, $(\cdot)\tilde{w}$ differs from Γ by at most $(r_j)^2$ in a large portion of $B_{r(x_j)}$ since

$$u \leq w + 2; \quad |\{u \leq 3\}| \geq C \sum |B_{r(x_j)}| \geq 1.$$

We divide the integration in diadic rings around x_0



Since $\tilde{w}(x_0) = 0$, and $\tilde{w} \ge 0$, the integrand in all of the rings is positive and

$$Lw \sim (2-s) \sum (r_k)^{-(n+s)} \int_{\mathbb{R}_k} \tilde{w} \le C.$$

(a) We first show that if $C_1 = MC$ is a large multiple of C there is at least one ring where

$$\fint_{\mathbb{R}_k} \tilde{w} \le C_1 r_k^2.$$

If not

$$C \ge L\tilde{w} = (2-s) \sum r_k^{-(n+s)} \int \tilde{w} \ge (2-s) \sum r_k^{-s} \fint \tilde{w} \ge C_1 \frac{2-s}{1-2^{(s-2)}}$$

$\sim C_1$, a contradiction.

In fact, if M is large, we can start the sum from $k = k_0$ and we get

$$C \ge C_1 \frac{(2-s)}{1-2^{(s-2)}} \cdot 2^{(2-s)k_0}$$

still a contradiction.

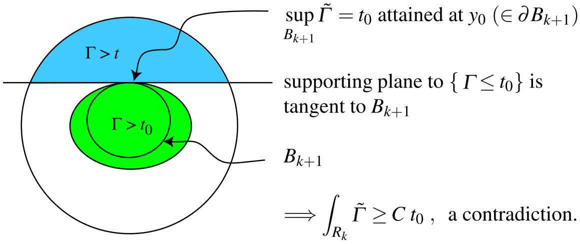
Of course, $\tilde{w}$ may still be highly oscillatory but

(a) In 99% of the rings, $\tilde{w} \le 100 C_1 r_k^2$, that is, in the original configuration w stays close to its convex envelope Γ.
(b) Further, since $0 \le \tilde{\Gamma} \le \tilde{w}$

$$\fint_{\mathbb{R}_k} \tilde{\Gamma} \le C_1 r_k^2.$$

But $\tilde{\Gamma}$ is convex, so this implies a bound $\tilde{\Gamma} \le C_2 r_k^2$ in B_{k+1} and $\nabla\Gamma \le r_k$ in B_{k+2}.

Let $t_0 = r_k^2$

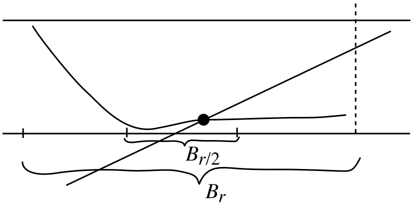


In turn, this implies that $\sup_{B_{k+2}} \nabla\tilde{\Gamma} < C\, r_k^2$.

Finally, it is a general fact that

If Γ is convex in B_r,

$$\nabla\Gamma\Big|_{B_{r/2}} \leq \left(\underset{B_r}{\operatorname{osc}\Gamma}\right) 2r^{-1}$$



$$\left(\text{slope of } \ell \leq \frac{\operatorname{osc}\Gamma}{(r/2)}\right)$$

A covering lemma completes the proof.

This proof, of course, requires in principle that the kernels be symmetric (some asymmetry is "tolerated" by the fact that the gradient of Γ is bounded, as in the second order case).

But the nature of the "game" symmetrizes the kernel:

From the "inf sup" property, for any x_0, there exists a direction ν^+ so that

$$0 \leq \int [u(x+y) + u(x-y) - 2(u(x))] K_{\nu^+} + K_\mu$$

for any μ (in particular $-\nu^+$) and vice versa, a ν^- so that

$$0 \geq \int [u(x+y) + u(x-y) - 2u(x)] K_{\nu^-} + K_\mu$$

for any μ, and this property is all that's needed.

Going back to the two possible "integral drifts", in the first case it is also possible to prove that solutions are in fact $C^{2s+\sigma}$, i.e., the integrals converge and the solution is classical (see [7]).

This is because the nature of the drift is such that, as the problem is rescaled the perturbation term $A(x_1)$ drifts to infinity.

6 Non-local Infinite Laplacian

Finally, I would like to discuss briefly the "tug of war" non-local "infinite Laplacian".

The Infinite Laplacian appears in the case when there is no diffusion left, i.e., when it is just the players taking random turns in choosing the direction of the drift (tug of war).

For the infinitesimal case, when the length of the jump is predetermined you formally get "$u_{nn} = 0$", n the direction of the gradient (Peres–Schramm–Sheffield–Wilson [17]).

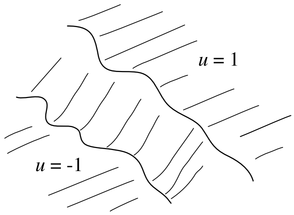
Also, in collaboration with Bjorland and Figalli [2], we consider the case in which the jump of the particle follows the distribution of the s-Laplacian

$$\inf_{\nu_1 \in s^1} \sup_{\nu_2 \in s^2} \int \frac{u(x+\nu_1 t) + u(x+\nu_2 t) - 2u(x)}{t^{1+2s}}\, dt = 0.$$

(That is, each player pulls in the directions ν_1 and ν_2.)

Formally, for $s > 1/2$, the direction of the jump is given by ∇u: Since the integrals diverge, each players is "forced" to take that choice.

We prove existence, uniqueness and (some) regularity, under a monotone geometry, for $s > 1/2$.

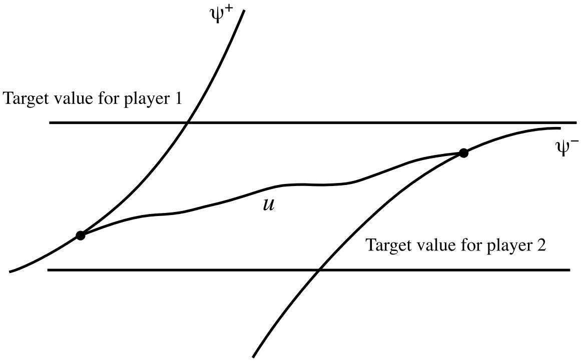


We assume that the domain Ω is "strip like", i.e., between bounded Lipschitz graphs with uniform separation and pay off is respectively 1 and -1. We show that there exists a unique viscosity solution (the least supersolution and larger subsolution coincide), and it is C^{2s-1}.

($|x|^{2s-1}$ are the "cones" for this problem, note that for $s = 1/2$ the theory breaks down.)

We end up with some comments:

- The case $s < 1/2$ seems very interesting since "∇u" does not fix the direction of the jump any more, and players will choose to jump in "non-opposite directions" most of the time.
- Instead of prescribing boundary values, it seems more natural to prescribe upper and lower obstacles where it would be optimal for one of the players to stop playing (execute an option).



We can prove in this case similar results as to the boundary value problem discussed before [2].

Acknowledgements The author was partially supported by National Science Foundation Grant DMS-0654267.

References

1. Barlow, M.T., Bass, R.F., Chen, Zh.Q., Kassmann, M.: Non-local Dirichlet forms and symmetric jump processes. Trans. Am. Math. Soc. **361**(4), 1963–1999 (2009)
2. Bjorland, C., Caffarelli, L., Figalli, A.: Non-local tug-of-war and the infinity fractional Laplacian. arXiv:1011.1966
3. Bass, R.F., Kassmann, M.: Harnack inequalities for non-local operators of variable order. Trans. Am. Math. Soc. **357**(2), 837–850 (2005)
4. Caffarelli, L., Cabre, X.: Fully Nonlinear Elliptic Equations. Am. Math. Soc. Colloq. Publ., vol. 43. Am. Math. Soc., Providence (1995)
5. Caffarelli, L., Vasseur, A.: Drift diffusion equations with fractional diffusion and the quasi-geostrophic equation. Ann. Math. (2011, in press)
6. Caffarelli, L.A., Silvestre, L.: Regularity theory for fully nonlinear integro-differential equations. Commun. Pure Appl. Math. **62**(5), 597–638 (2009)
7. Caffarelli, L.A., Silvestre, L.: Regularity results for nonlocal equations by approximation. Arch. Ration. Mech. Anal. (2011, in press). arXiv:0902.4030
8. Caffarelli, L.A., Silvestre, L.: The Evans-Krylov theorem for non local fully non linear equations. Ann. Math. (2011, in press). arXiv:0905.1339
9. Caffarelli, L.A., Chan, C.H., Vasseur, A.: Regularity theory for nonlinear integral operators. arXiv:1003.1699
10. Chambolle, A., Lindgren, E., Monneau, R.: The Hölder infinite Laplacian and Hölder extensions. Preprint (2010)
11. Duvant, G., Lions, J.-L.: Les inéquations en mécanique et en physique. Travaux et Recherches Mathématiques, vol. 21. Dunod, Paris (1972) (in French)
12. Giacomin, G., Lebowitz, J.L., Presutti, E.: Deterministic and stochastic hydrodynamic equations arising from simple microscopic model systems. In: Stochastic Partial Differential Equations: Six Perspectives. Math. Surveys Monogr., vol. 64, pp. 107–152. Am. Math. Soc., Providence (1999)
13. Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. Multiscale Model. Simul. **7**(3), 1005–1028 (2008)
14. Kassmann, M.: A priori estimates for integro-differential operators with measurable kernels. Calc. Var. Partial Differ. Equ. **34**(1), 1–21 (2009)
15. Komatsu, T.: Uniform estimates for fundamental solutions associated with non-local Dirichlet forms. Osaka J. Math. **32**(4), 833–860 (1995)
16. Peres, Y., Sheffield, S.: Tug-of-war with noise: a game-theoretic view of the p-Laplacian. Duke Math. J. **145**(1), 91–120 (2008)
17. Peres, Y., Schramm, O., Sheffield, S., Wilson, D.B.: Tug-of-war and the infinity Laplacian. J. Am. Math. Soc. **22**(1), 167–210 (2009)

Characteristic Discontinuities and Free Boundary Problems for Hyperbolic Conservation Laws

Gui-Qiang Chen and Ya-Guang Wang

Abstract We are concerned with entropy solutions of hyperbolic systems of conservation laws in several space variables. The Euler equations of gas dynamics and magnetohydrodynamics (MHD) are prototypes of hyperbolic conservation laws. In general, there are two types of discontinuities in the entropy solutions: shock waves and characteristic discontinuities, in which characteristic discontinuities can be either vortex sheets or entropy waves. In gas dynamics and MHD, across a vortex sheet, the tangential velocity field has a jump while the normal velocity is continuous; across an entropy wave, the entropy has a jump while the velocity field is continuous. A vortex sheet or entropy wave front is a part of the unknowns, which is a free boundary. Compressible vortex sheets and entropy waves, along with shock and rarefaction waves, occur ubiquitously in nature and are fundamental waves in the entropy solutions to multidimensional hyperbolic conservation laws. The local stability of shock and rarefaction waves has been relatively better understood. In this paper we discuss the stability issues for vortex sheets/entropy waves and present some recent developments and further open problems in this direction. First we discuss vortex sheets and entropy waves for the Euler equations in gas dynamics and some recent developments for a rigorous mathematical theory on their nonlinear stability/instability. Then we review our recent study and present a supplement to the proof on the nonlinear stability of compressible vortex sheets under the magnetic effect in three-dimensional MHD. The compressible vortex sheets in three dimensions are unstable in the regime of pure gas dynamics. Our main concern is

G.-Q. Chen (✉)
Mathematical Institute, University of Oxford, Oxford, OX1 3LB, UK
e-mail: chengq@maths.ox.ac.uk

G.-Q. Chen
School of Mathematical Sciences, Fudan University, Shanghai 200433, China

G.-Q. Chen
Department of Mathematics, Northwestern University, Evanston, IL 60208, USA

Y.-G. Wang
Department of Mathematics, Shanghai Jiao Tong University, Shanghai 200240, China
e-mail: ygwang@sjtu.edu.cn

H. Holden, K.H. Karlsen (eds.), *Nonlinear Partial Differential Equations*,
Abel Symposia 7, DOI 10.1007/978-3-642-25361-4_4,

whether such vortex sheets can be nonlinearly stabilized under the magnetic fields. To achieve this, we first set up the current-vortex sheet problem as a free boundary problem; then we establish high-order energy estimates of the solutions to the linearized problem, which shows that the current-vortex sheets are linearly stable when the jump of the tangential velocity is dominated by the jump of the non-paralleled tangential magnetic fields; and finally we develop a suitable iteration scheme of the Nash–Moser–Hörmander type to obtain the existence and nonlinear stability of compressible current-vortex sheets, locally in time. Some further open problems and several related remarks are also presented.

1 Introduction

We are concerned with entropy solutions of hyperbolic systems of conservation laws in several space variables:

$$\partial_t \mathbf{U} + \sum_{j=1}^{d} \partial_{x_j} \mathbf{f}_j(\mathbf{U}) = 0, \tag{1}$$

where $\mathbf{U} = (U_1, \dots, U_m)^\top$ and $\mathbf{f}_j : \mathbb{R}^m \to \mathbb{R}^m$, $j = 1, \dots, d$, are nonlinear smooth functions. System (1) consists of m quasilinear hyperbolic equations in the d-dimensional space variables $\mathbf{x} = (x_1, \dots, x_d)$. The prototypes of hyperbolic conservation laws include the Euler equations of gas dynamics and magnetohydrodynamics (MHD).

Let the level set surface $\Gamma := \{\Phi(t, \mathbf{x}) = 0\}$ of $\Phi(t, \mathbf{x})$ be a discontinuity of a piecewise smooth entropy solution:

$$\mathbf{U}(t, \mathbf{x}) = \begin{cases} \mathbf{U}^-(t, \mathbf{x}) & \text{for } \Phi(t, \mathbf{x}) < 0, \\ \mathbf{U}^+(t, \mathbf{x}) & \text{for } \Phi(t, \mathbf{x}) > 0, \end{cases} \tag{2}$$

where $\mathbf{U}^\pm(t, \mathbf{x})$ are smooth solutions of (1) in the respective domains separated by Γ. Then $\mathbf{U}^\pm|_\Gamma$ and Φ must satisfy the Rankine–Hugoniot jump conditions across Γ:

$$\partial_t \Phi [\mathbf{U}] + \sum_{j=1}^{d} \partial_{x_j} \Phi [\mathbf{f}_j(\mathbf{U})] = 0, \tag{3}$$

where the bracket $[\cdot]$ stands for the jump of the associated function across Γ, that is,

$$[\mathbf{U}] = \mathbf{U}^+|_\Gamma - \mathbf{U}^-|_\Gamma,$$

with $\mathbf{U}^\pm|_\Gamma$ as the traces of $\mathbf{U}^\pm$ taken on the respective sides of Γ.

In general, there are two types of discontinuities in the entropy solutions of (1). The first type of discontinuities is called shock waves (or fronts), across which the strict Lax entropy inequality holds for at least one convex entropy-entropy flux pair $(\eta, \mathbf{q}) = (\eta, q_1, \dots, q_d)$, $\nabla^2 \eta(\mathbf{U}) \geq 0$:

$$\partial_t \Phi [\eta(\mathbf{U})] + \sum_{j=1}^{d} \partial_{x_j} \Phi [q_j(\mathbf{U})] > 0. \tag{4}$$

The second type of discontinuities is called characteristic discontinuities, which are characteristic surfaces of the hyperbolic system (1). That is, for this case, the function $\Phi(t, \mathbf{x})$ satisfies the eikonal equation on $\Gamma = \{\Phi(t, \mathbf{x}) = 0\}$:

$$\partial_t \Phi + \lambda(\mathbf{U}^{\pm}; \nabla_{\mathbf{x}} \Phi) = 0, \tag{5}$$

where $\lambda(\mathbf{U}; \vec{\xi})$ is an eigenvalue of the matrix $\sum_{j=1}^{d} \xi_j \mathbf{f}_j'(\mathbf{U})$ for $\vec{\xi} = (\xi_1, \ldots, \xi_d)$. Usually, there are two different kinds of characteristic discontinuities: vortex sheets and entropy waves. In gas dynamics and MHD, across a vortex sheet, the tangential velocity field has a jump, while the normal velocity is continuous; across an entropy wave, the entropy has a jump while the velocity field is continuous. A vortex sheet or entropy wave front Γ is a part of the unknowns, which is a free boundary. This free boundary is a characteristic surface with respect to either side of Γ.

Compressible vortex sheets and entropy waves, along with shock and rarefaction waves, are fundamental waves in the entropy solutions to multidimensional hyperbolic systems of conservation laws. They occur ubiquitously in nature including slip-stream interfaces, lifting of aircrafts, galactic jets, tornadoes, Mach configurations in the shock reflection-diffraction patterns, and interactions among nonlinear waves; see [1, 3–14, 24, 25, 28, 34–36, 40, 41] and the references cited therein. The stability of shock and rarefaction waves has been studied in Majda [32], Métivier [33], and Alinhac [2]; also see [30].

In this paper, we discuss the stability issues for vortex sheets/entropy waves, present some recent developments, and address further open problems in this direction.

In Sect. 2, we discuss vortex sheets and entropy waves for the Euler equations in gas dynamics. It was observed in Miles [34, 35], by mode analysis, that the vortex sheets in two-dimensional isentropic gas dynamics are linearly stable when the Mach number is larger than $\sqrt{2}$ and are violently unstable when the Mach number is less than $\sqrt{2}$, while they are always unstable in three space variables no matter how large the Mach number is. A rigorous mathematical theory on the nonlinear stability of the two-dimensional vortex sheets with the Mach number larger than $\sqrt{2}$ locally in time was obtained recently by Coulombel–Secchi [21, 22] when the initial data is in a class of small perturbation functions of a planar vortex sheet.

In Sects. 3–5, we review our recent study and present a supplement to the proof on the stability of three-dimensional compressible vortex sheets under the magnetic effect, that is, the nonlinear stability of current-vortex sheets in three-dimensional MHD in Chen–Wang [17]. As we mentioned above, the compressible vortex sheets in three dimensions are unstable in the regime of pure gas dynamics. Our main concern is whether such vortex sheets can be nonlinearly stabilized under the magnetic fields. In Sect. 3, we first set up the current-vortex sheet problem as a free boundary problem and state the main results. In Sect. 4, we establish high-order energy estimates of the solutions to the linearized problem, which shows that the current-vortex sheets are linearly stable when the jump of the tangential velocity is dominated by the jump of the non-paralleled tangential magnetic fields in the sense that $\lambda^{\pm}$ determined by (39) satisfy condition (38), as observed in [37]. To achieve this, our key observation is that the linearized problem (41), equivalently (43), for current-vortex sheets is endowed with a well-structured decoupled formulation so that the linear problem is decoupled into one standard initial-boundary value problem (48) for a symmetric hyperbolic system and the other problem (52) for an ordinary differential equation for the front. This decoupled formulation is essential for us to establish our desired high-order energy estimates of solutions, which is one of the key ingredients for developing our nonlinear approach for the stability problem. Also see Trakhinin [38] for a different approach to make related estimates. The energy estimates of the linearized problems have a loss of regularity with respect to the nonhomogeneous terms and initial data, mainly due to that the front is characteristic in the current-vortex sheets. As in [2, 22], this has inspired us to develop a suitable iteration scheme of the Nash–Moser–Hörmander type to obtain the existence and structural stability of compressible current-vortex sheets, locally in time, in the three-dimensional MHD. This is done in Sect. 5. In Sect. 6, we present further open problems and several related remarks.

2 Characteristic Discontinuities for the Euler Equations in Gas Dynamics

In this section we discuss vortex sheets and entropy waves for the Euler equations in gas dynamics.

2.1 *Isentropic Euler Equations*

The isentropic Euler equations in gas dynamics in $\mathbb{R}^d$ describing the motion of inviscid gases take the following form:

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = 0, \\ \partial_t (\rho \mathbf{v}) + \nabla \cdot (\rho \mathbf{v} \otimes \mathbf{v}) + \nabla p = 0, \end{cases} \tag{6}$$

where ρ and $\mathbf{v} = (v_1, \dots, v_d)^\top \in \mathbb{R}^d$ are the density and velocity, respectively; the pressure p is a function of the density ρ:

$$p = p(\rho) \tag{7}$$

with $p'(\rho) > 0$ when $\rho > 0$.

For a piecewise smooth weak solution $\mathbf{U}(t, \mathbf{x})$ of (6):

$$\mathbf{U}(t, \mathbf{x}) = \begin{cases} \mathbf{U}^-(t, \mathbf{x}) & \text{for } \Phi(t, \mathbf{x}) < 0, \\ \mathbf{U}^+(t, \mathbf{x}) & \text{for } \Phi(t, \mathbf{x}) > 0 \end{cases} \tag{8}$$

on the front $\Gamma := \{\Phi(t, \mathbf{x}) = 0\}$, the Rankine–Hugoniot conditions must be satisfied:

$$\begin{cases} [m_N] = 0, \\ m_N [v_N] + |\nabla_{\mathbf{x}} \Phi|^2 [p] = 0, \\ m_N [\mathbf{v}_\tau] = 0, \end{cases} \tag{9}$$

where $v_N := \mathbf{v}\cdot\nabla_{\mathbf{x}}\Phi$ and $\mathbf{v}_\tau \in \mathbb{R}^{d-1}$ are the normal and tangential components of $\mathbf{v}$ on Γ, and $m_N = \rho(v_N + \Phi_t)$ is the mass transfer flux.

Suppose that $m_N = 0$ on Γ, i.e., no mass transfer flux across the front, so $(\mathbf{U}^{\pm}, \Gamma)$ is a characteristic discontinuity for (6). Then, on Γ,

$$[p] = [v_N] = 0. \tag{10}$$

In this case, there is only one kind of characteristic discontinuities, vortex sheets, since the tangential velocity field (with respect to the interface Γ) is the only quantity that experiences a jump across Γ:

$$[\mathbf{v}_\tau] \neq 0, \qquad [p] = [v_N] = 0. \tag{11}$$

2.2 Full Euler Equations

The full Euler equations for gas dynamics in $\mathbb{R}^d$ take the following form:

$$\begin{cases} \partial_t\rho + \nabla\cdot(\rho\mathbf{v}) = 0, \\ \partial_t(\rho\mathbf{v}) + \nabla\cdot(\rho\mathbf{v}\otimes\mathbf{v}) + \nabla p = 0, \\ \partial_t\left(\frac{1}{2}\rho|\mathbf{v}|^2 + e\right) + \nabla\cdot\left(\left(\frac{1}{2}\rho|\mathbf{v}|^2 + \rho e + p\right)\mathbf{v}\right) = 0, \end{cases} \tag{12}$$

where $p = p(\rho, S)$ and $e = e(\rho, S)$ are the pressure and internal energy with the entropy S, respectively.

Let a piecewise smooth function $\mathbf{U}(t,\mathbf{x})$:

$$\mathbf{U}(t,\mathbf{x}) = \begin{cases} \mathbf{U}^-(t,\mathbf{x}) & \text{for } \Phi(t,\mathbf{x}) < 0, \\ \mathbf{U}^+(t,\mathbf{x}) & \text{for } \Phi(t,\mathbf{x}) > 0 \end{cases} \tag{13}$$

be a weak solution to (12). Then, on the front $\Gamma := \{\Phi(t,\mathbf{x}) = 0\}$, $\mathbf{U}(t,\mathbf{x})$ must satisfy the Rankine–Hugoniot conditions:

$$\begin{cases} [m_N] = 0, \\ m_N[v_N] + |\nabla_{\mathbf{x}}\Phi|^2[p] = 0, \\ m_N[\mathbf{v}_\tau] = 0, \\ m_N\left[e + \frac{1}{2}|\mathbf{v}|^2\right] + [pv_N] = 0, \end{cases} \tag{14}$$

where $v_N := \mathbf{v}\cdot\nabla_{\mathbf{x}}\Phi|_\Gamma$ and $\mathbf{v}_\tau$ are the normal and tangential components of $\mathbf{v}$ on Γ, and $m_N = \rho(v_N + \Phi_t)$ is the mass transfer flux.

As above, we consider the case that $m_N = 0$ on Γ, i.e., no mass transfer flux across the front; so $(\mathbf{U}^{\pm}, \Gamma)$ is a characteristic discontinuity for (12). Then, on Γ,

$$[p] = [v_N] = 0. \tag{15}$$

Different from the isentropic case, there are two different kinds of characteristic discontinuities on which $[p] = [v_N] = m_N = 0$:

(i) Vortex sheets:

$$[\mathbf{v}_\tau] \neq 0; \tag{16}$$

(ii) Entropy waves:

$$[\mathbf{v}_\tau] = 0, \qquad [\rho] \neq 0, \qquad [S] \neq 0. \tag{17}$$

2.3 Stability of Vortex Sheets for the Two-Dimensional Isentropic Euler Equations

Choose

$$\Phi(t, x_1, x_2) = x_1 - \varphi(t, x_2).$$

Then

$$\Gamma = \{x_1 = \varphi(t, x_2) : t > 0, x_2 \in \mathbb{R}\}.$$

The vortex sheet Γ satisfies that

(i). The isentropic Euler equations (6) are satisfied on either side of Γ;
(ii). The Rankine–Hugoniot jump relations are satisfied on Γ:

$$\partial_t \varphi = \mathbf{v}^+ \cdot (1, -\partial_{x_2}\varphi) = \mathbf{v}^- \cdot (1, -\partial_{x_2}\varphi), \qquad \rho^- = \rho^+.$$

As usual, $\rho^\pm, \mathbf{v}^\pm$ denote the traces of $\rho, \mathbf{v}$ taken on either side of Γ. The vortex sheet Γ is a part of the unknowns, which is a free boundary. This free boundary is a characteristic with respect to either side of Γ.

Consider a planar vortex sheet Γ_0 with constant states on either side. Then, by the Galilean invariance of frame, such a vortex sheet can be always reformulated as the following form:

$$\mathbf{U}^\pm = (\bar{\rho}, 0, \pm\bar{\rho}\bar{v})^\top, \qquad \pm x_1 > 0, \tag{18}$$

where $\bar{\rho} > 0$ is a fixed density, $\bar{v} > 0$ is a fixed tangential velocity, while the normal velocity vanishes. The sonic speed on Γ_0 is $\bar{c} = \sqrt{p'(\bar{\rho})}$, and the relative Mach number $\bar{M}$ is

$$\bar{M} = \frac{\bar{v}}{\bar{c}}.$$

By mode analysis, it was observed by Miles in [34, 35] that the vortex sheets in two-dimensional isentropic gas dynamics are linearly stable when the Mach number $\bar{M} > \sqrt{2}$ and violently unstable when $\bar{M} < \sqrt{2}$.

A rigorous mathematical theory on the nonlinear stability of the two-dimensional vortex sheets with $\bar{M} > \sqrt{2}$ locally in time was obtained recently by Coulombel–Secchi [21, 22] when the initial data function is in a class of small perturbation functions of a planar vortex sheet Γ_0. On the other hand, the seminal work by Artola–Majda [4–6] indicates that the stability of compressible vortex sheets depends on the class of initial perturbation functions, even when $\bar{M} > \sqrt{2}$.

For the two-dimensional full Euler equations, as indicated in Sect. 2.2, there is an additional type of characteristic discontinuities, called entropy waves. Across an entropy wave, the velocity and pressure are continuous, though the entropy, equivalently the density, has a jump. It would be interesting to analyze the stability of entropy waves to understand fundamental features of entropy solutions.

For the Euler equations in three space-dimensions, every compressible vortex sheet is violently unstable, and this violent instability is the analogue of the Kelvin–Helmholtz instability for incompressible fluids (cf. Fejer–Miles [25]). In the next sections, Sects. 3–5, we analyze whether compressible vortex sheets in three dimensions (which are unstable in the regime of pure gas dynamics) become stable under the magnetic effect in three-dimensional MHD.

3 Compressible Current-Vortex Sheets in MHD: Main Theorem

The equations for three-dimensional MHD describing the motion of inviscid MHD fluids take the following form:

$$\begin{cases} \partial_t \rho + \nabla\cdot(\rho\mathbf{v}) = 0, \\ \partial_t(\rho\mathbf{v}) + \nabla\cdot(\rho\mathbf{v}\otimes\mathbf{v} - \mathbf{H}\otimes\mathbf{H}) + \nabla\Big(p + \dfrac{1}{2}|\mathbf{H}|^2\Big) = 0, \\ \partial_t\mathbf{H} - \nabla\times(\mathbf{v}\times\mathbf{H}) = 0, \\ \partial_t\Big(\dfrac{1}{2}\rho|\mathbf{v}|^2 + \rho e + |\mathbf{H}|^2\Big) + \nabla\cdot\Big(\Big(\dfrac{1}{2}\rho|\mathbf{v}|^2 + \rho e + p\Big)\mathbf{v} + \mathbf{H}\times(\mathbf{v}\times\mathbf{H})\Big) = 0, \end{cases} \tag{19}$$

and

$$\nabla\cdot\mathbf{H} = 0, \tag{20}$$

where $\rho, \mathbf{v} = (v_1, v_2, v_3)^\top$, $\mathbf{H} = (H_1, H_2, H_3)^\top$, and $p = p(\rho, S)$ are the density, velocity, magnetic field, and pressure, respectively; $e = e(\rho, S)$ is the internal energy; and S is the entropy.

For smooth solutions, the equations in (19) are equivalent to

$$\begin{cases} (\partial_t + \mathbf{v}\cdot\nabla)p + \rho c^2\nabla\cdot\mathbf{v} = 0, \\ \rho(\partial_t + \mathbf{v}\cdot\nabla)\mathbf{v} + \nabla p - (\nabla\times\mathbf{H})\times\mathbf{H} = 0, \\ (\partial_t + \mathbf{v}\cdot\nabla)\mathbf{H} - (\mathbf{H}\cdot\nabla)\mathbf{v} + \mathbf{H}\nabla\cdot\mathbf{v} = 0, \\ (\partial_t + \mathbf{v}\cdot\nabla)S = 0, \end{cases} \tag{21}$$

where $c = \sqrt{p_\rho(\rho, S)}$ is the sonic speed of the fluid. The equations in (21) can be written as a 8×8 symmetric hyperbolic system for $\mathbf{U} = (p, \mathbf{v}, \mathbf{H}, S)^\top$ of the form:

$$\mathbf{B}_0(\mathbf{U})\partial_t\mathbf{U} + \sum_{j=1}^{3}\mathbf{B}_j(\mathbf{U})\partial_{x_j}\mathbf{U} = 0. \tag{22}$$

Let a piecewise smooth function $\mathbf{U}(t, \mathbf{x})$:

$$\mathbf{U}(t,\mathbf{x}) = \begin{cases} \mathbf{U}^-(t,\mathbf{x}) & \text{for } x_1 < \psi(t, x_2, x_3), \\ \mathbf{U}^+(t,\mathbf{x}) & \text{for } x_1 > \psi(t, x_2, x_3) \end{cases} \tag{23}$$

be a weak solution to (19). Then, on the front $\Gamma := \{x_1 = \psi(t, x_2, x_3)\}$, $\mathbf{U}(t, \mathbf{x})$ must satisfy the Rankine–Hugoniot conditions:

$$\begin{cases} [m_N] = [H_N] = 0, \\ m_N[v_N] + (1 + \psi_{x_2}^2 + \psi_{x_3}^2)[q] = 0, \\ m_N[\mathbf{v}_\tau] = H_N[\mathbf{H}_\tau], \\ m_N\left[\dfrac{\mathbf{H}_\tau}{\rho}\right] = H_N[\mathbf{v}_\tau], \\ m_N\left[e + \dfrac{1}{2}\left(|\mathbf{v}|^2 + \dfrac{|\mathbf{H}|^2}{\rho}\right)\right] + [qv_N - H_N(\mathbf{H}\cdot\mathbf{v})] = 0, \end{cases} \tag{24}$$

where $(v_N, \mathbf{v}_\tau)$ (resp. $(H_N, \mathbf{H}_\tau)$) are the normal and tangential components of $\mathbf{v}$ (resp. $\mathbf{H}$) on Γ, i.e.,

$$\begin{aligned} &v_N := v_1 - \psi_{x_2}v_2 - \psi_{x_3}v_3, \\ &\mathbf{v}_\tau = (v_{\tau_1}, v_{\tau_2})^\top := (\psi_{x_2}v_1 + v_2, \psi_{x_3}v_1 + v_3)^\top, \\ &H_N := H_1 - \psi_{x_2}H_2 - \psi_{x_3}H_3, \\ &\mathbf{H}_\tau = (H_{\tau_1}, H_{\tau_2})^\top := (\psi_{x_2}H_1 + H_2, \psi_{x_3}H_1 + H_3)^\top, \end{aligned}$$

$m_N = \rho(v_N - \psi_t)$ is the mass transfer flux, and $q = p + \frac{|\mathbf{H}|^2}{2}$ is the total pressure.

As in Sect. 2.2, we consider the case that $m_N = 0$ on Γ, i.e., no mass transfer flux across the front, so $(\mathbf{U}^\pm, \Gamma)$ is a characteristic discontinuity for (19). We now focus on the current-vortex sheets:

$$H_N^+ = H_N^- = 0, \qquad \mathbf{H}_\tau^+ \nparallel \mathbf{H}_\tau^-. \tag{25}$$

Then the Rankine–Hugoniot conditions are equivalent to

$$\psi_t = v_N^+ = v_N^-, \qquad \left[p + \frac{|\mathbf{H}|^2}{2}\right] = 0 \quad \text{on } \Gamma \tag{26}$$

and generically $([\rho], [\mathbf{v}_\tau], [S]) \neq 0$.

First, we have

Lemma 1 *Let* $(\mathbf{U}^\pm, \psi)$ *be a current-vortex sheet defined as above for* $0 \le t < T$. *Then, if*

$$H_N^\pm|_{\Gamma\cap\{t=0\}} = 0, \qquad \nabla\cdot\mathbf{H}^\pm(0, \mathbf{x}) = 0,$$

we have

$$H_N^\pm|_\Gamma = 0, \qquad \nabla\cdot\mathbf{H}^\pm(t, \mathbf{x}) = 0, \quad \textit{for all } t \in [0, T). \tag{27}$$

By a direct calculation, one knows that both $H_N^\pm$ and $\nabla\cdot\mathbf{H}^\pm$ satisfy a homogeneous transport equation tangential to Γ, so assertion (27) follows immediately if it holds initially.

This lemma shows that both the divergence-free condition (20) and the condition $H_N^\pm|_\Gamma = 0$ are only the constraints on the initial data.

Set

$$\mathbf{D}(\lambda,\mathbf{U}) := \begin{pmatrix} \tilde{\mathbf{D}}(\lambda,\mathbf{U}) & 0 \\ 0 & 1 \end{pmatrix} \quad \text{with } \tilde{\mathbf{D}}(\lambda,\mathbf{U}) := \begin{pmatrix} 1 & \frac{\lambda}{\rho c^2}\mathbf{H}^\top & \mathbf{O}_{1\times 3} \\ \lambda\rho\mathbf{H} & I_3 & -\rho\lambda I_3 \\ \mathbf{O}_{3\times 1} & -\lambda I_3 & I_3 \end{pmatrix}.$$

As in Trakhinin [37], we know from Lemma 1 that system (19)–(20) is equivalent to the following system on both sides of $\varGamma$:

$$\mathbf{D}(\lambda^\pm,\mathbf{U}^\pm)\left(\mathbf{B}_0(\mathbf{U}^\pm)\partial_t\mathbf{U}^\pm + \sum_{j=1}^{3}\mathbf{B}_j(\mathbf{U}^\pm)\partial_{x_j}\mathbf{U}^\pm\right) + \lambda^\pm\mathbf{G}^\pm\nabla\cdot\mathbf{H}^\pm = 0, \tag{28}$$

provided $\nabla\cdot\mathbf{H}^\pm(0,\mathbf{x}) = 0$, where

$$\mathbf{G}^\pm = -(1,0,0,0,\mathbf{H}^\pm,0)^\top,$$

and $\lambda^\pm = \lambda^\pm(\mathbf{U}^+,\mathbf{U}^-)$ will be determined later.

System (28) can be rewritten as the following symmetric form

$$\mathbf{A}_0(\mathbf{U}^\pm)\partial_t\mathbf{U}^\pm + \sum_{j=1}^{3}\mathbf{A}_j(\mathbf{U}^\pm)\partial_{x_j}\mathbf{U}^\pm = 0. \tag{29}$$

System (29) is still hyperbolic, provided that

$$(\lambda^\pm)^2 < \frac{1}{\rho^\pm + |\mathbf{H}^\pm|^2/(c^\pm)^2}. \tag{30}$$

The main task of this section and Sects. 4–5 is to study the existence and stability of the states $\mathbf{U}^\pm(t,\mathbf{x})$ and a free boundary $\varGamma = \{x_1 = \psi(t,x_2,x_3)\}$ for $0 \le t < T$ such that

$$\begin{cases} \mathbf{A}_0(\mathbf{U}^\pm)\partial_t\mathbf{U}^\pm + \sum_{j=1}^{3}\mathbf{A}_j(\mathbf{U}^\pm)\partial_{x_j}\mathbf{U}^\pm = 0 \quad \text{for } \pm(x_1-\psi(t,x_2,x_3)) > 0, \\ \mathbf{U}|_{t=0} = \begin{cases} \mathbf{U}_0^+(\mathbf{x}) & \text{for } x_1 > \psi_0(x_2,x_3), \\ \mathbf{U}_0^-(\mathbf{x}) & \text{for } x_1 < \psi_0(x_2,x_3) \end{cases} \end{cases} \tag{31}$$

with the transmission conditions on $\varGamma$:

$$\psi_t = v_N^+ = v_N^-, \qquad \left[p + \frac{|\mathbf{H}|^2}{2}\right] = 0, \tag{32}$$

provided that $H_N^\pm|_{\varGamma\cap\{t=0\}} = 0$, $\nabla\cdot\mathbf{H}_0^\pm(\mathbf{x}) = 0$, and $\mathbf{H}_\tau^+ \nparallel \mathbf{H}_\tau^-$ hold at $t=0$, where $\psi_0(x_2,x_3) = \psi(0,x_1,x_2)$.

In the above problem, the front $\varGamma$ is unknown. To deal with such a free boundary problem, it is convenient to use the following standard transformation:

$$\begin{cases} t = \tilde{t}, \quad x_2 = \tilde{x}_2, \quad x_3 = \tilde{x}_3, \\ x_1 = \varPsi^\pm(\tilde{t},\tilde{x}_1,\tilde{x}_2,\tilde{x}_3) \end{cases} \tag{33}$$

with $\varPsi^\pm$ satisfying

$$\begin{cases} \pm(\varPsi^\pm)_{\tilde{x}_1} \ge \kappa > 0, \\ \varPsi^+|_{\tilde{x}_1=0} = \varPsi^-|_{\tilde{x}_1=0} = \psi(\tilde{t},\tilde{x}_2,\tilde{x}_3) \end{cases} \tag{34}$$

for some constant $\kappa > 0$. Under (33), the domains $\Omega^{\pm} := \{\pm(x_1 - \psi(t, x_2, x_3)) > 0\}$ are transformed into $\{\tilde{x}_1 > 0\}$ and the free boundary Γ into the fixed boundary $\{\tilde{x}_1 = 0\}$.

The natural candidates of $\Psi^{\pm}$ can be proper extensions of $\psi(\tilde{t}, \tilde{x}_2, \tilde{x}_3)$ in $\{\tilde{x}_1 > 0\}$ satisfying the first non-degenerate condition $(34)_1$. With this in mind, we choose $\Psi^{\pm}$ to be the solutions to the following problem:

$$\begin{cases} \partial_t \Psi^{\pm} - v_1^{\pm} + v_2^{\pm}\partial_{x_2}\Psi^{\pm} + v_3^{\pm}\partial_{x_3}\Psi^{\pm} = 0, \quad t, x_1 > 0, \\ \Psi^{\pm}|_{t=0} = \Psi_0^{\pm}(\mathbf{x}) := \pm x_1 + \chi(\pm x_1)\psi_0(x_2, x_3), \end{cases} \tag{35}$$

where we drop the tildes in the formula for simplicity, $\chi(s)$ is a smooth cut-off function that is 1 for $|s| \le 1$ and 0 for $|s| > 2$ such that $\pm(\Psi_0^{\pm})_{x_1} \ge \kappa > 0$ in $\{x_1 > 0\}$.

Under transformation (33), it is easy to know that problem (31)–(32) is equivalent to that $\tilde{\mathbf{U}}^{\pm}(\tilde{t}, \tilde{\mathbf{x}}) = \mathbf{U}^{\pm}(t, \mathbf{x})$ satisfy the following problem with a fixed boundary $\{x_1 = 0\}$:

$$\begin{cases} \mathbf{L}(\mathbf{U}^{\pm}, \Psi^{\pm})\mathbf{U}^{\pm} = 0 \quad \text{in } \{x_1 > 0\}, \\ \mathbf{B}(\mathbf{U}^{+}, \mathbf{U}^{-}, \psi)|_{x_1=0} = 0, \\ (\mathbf{U}^{\pm}, \psi)|_{t=0} = (\mathbf{U}_0^{\pm}(\mathbf{x}), \psi_0(x_2, x_3)), \end{cases} \tag{36}$$

where the tildes have also been dropped,

$$\mathbf{L}(\mathbf{U}, \Psi)\mathbf{V} = \mathbf{A}_0(\mathbf{U})\partial_t \mathbf{V} + \bar{\mathbf{A}}_1(\mathbf{U}, \Psi)\partial_{x_1}\mathbf{V} + \sum_{j=2}^{3} \mathbf{A}_j(\mathbf{U})\partial_{x_j}\mathbf{V}$$

with $\bar{\mathbf{A}}_1(\mathbf{U}, \Psi) = \frac{1}{\Psi_{x_1}}(\mathbf{A}_1(U) - \Psi_t \mathbf{A}_0(U) - \sum_{j=2}^{3} \Psi_{x_j}\mathbf{A}_j(U))$, and

$$\mathbf{B}(\mathbf{U}^{+}, \mathbf{U}^{-}, \psi) = (\psi_t - U_{v,N}^{\pm}, q^{+} - q^{-})^{\top}$$

with $U_{v,N}^{\pm} = U_2^{\pm} - \psi_{x_2}U_3^{\pm} - \psi_{x_3}U_4^{\pm}$ and $q = U_1 + \frac{1}{2}|\mathbf{U}_H|^2$ for $\mathbf{U}_H = (U_5, U_6, U_7)^{\top}$, under the constraints that

$$\begin{cases} \mathbf{U}_{\mathbf{H},N}^{\pm} = U_5^{\pm} - \psi_{x_2}U_6^{\pm} - \psi_{x_3}U_7^{\pm} = 0 \quad \text{on } \{x_1 = 0\}, \\ \widetilde{\nabla} \cdot \mathbf{H}^{\pm} := \partial_{x_1}U_5^{\pm} + (\partial_{x_1}\Psi^{\pm}\partial_{x_2} - \partial_{x_2}\Psi^{\pm}\partial_{x_1})U_6^{\pm} + (\partial_{x_1}\Psi^{\pm}\partial_{x_3} - \partial_{x_3}\Psi^{\pm}\partial_{x_1})U_7^{\pm} \\ \qquad = 0 \quad \text{in } \{x_1 > 0\} \end{cases} \tag{37}$$

hold at $\{t = 0\}$.

The main feature of problem (36) is that the fixed boundary $\{x_1 = 0\}$ is a characteristic plane of constant multiplicity. To solve (36), as in [2, 18, 29], it is natural to introduce the weighted anisotropic Sobolev spaces defined on $\Omega_T := \{(t, \mathbf{x}) \in [0, T] \times \mathbb{R}^3 : x_1 > 0\}$:

$$B_{\mu}^{s}(\Omega_T) := \{u \in L^2(\Omega_T) : e^{-\mu t} M^{\alpha}\partial_{x_1}^{k} u \in L^2(\Omega_T) \text{ for } |\alpha| + 2k \le s\}$$

for all $s \in \mathbb{N}$ and $\mu > 0$, where the tangential vectors $M = (M_0, M_1, M_2, M_3)$ of

$\{x_1 = 0\}$ are given by

$$M_0 = \partial_t, \qquad M_1 = \sigma(x_1)\partial_{x_1}, \qquad M_2 = \partial_{x_2}, \qquad M_3 = \partial_{x_3},$$

with

$$\sigma(x_1) := \begin{cases} x_1 & \text{for } 0 \le x_1 \le 1, \\ 2 & \text{for } x_1 \ge 2. \end{cases}$$

The norms in $B^s_\mu(\Omega_T)$ are as usual:

$$\|u\|_{s,\mu,T} := \left(\int_0^T \|u(t,\cdot)\|^2_{s,\mu} dt \right)^{1/2},$$

with

$$\|u(t,\cdot)\|^2_{s,\mu} := \sum_{|\alpha|+2k\le s} \mu^{2(s-|\alpha|-2k)} \|e^{-\mu t} M^\alpha \partial^k_{x_1} u(t,\cdot)\|^2_{L^2}.$$

We will also use a similar notation for the spaces with $\mu = 0$, $B^s(\Omega_T)$, with norm:

$$\|u\|_{s,T} := \left(\sum_{|\alpha|+2k\le s} \|M^\alpha \partial^k_{x_1} u\|^2_{L^2(\Omega_T)} \right)^{1/2}.$$

Also denote $b\Omega_T := \{(t, x_2, x_3) : t \in [0,T], (x_2, x_3) \in \mathbb{R}^2\}$, and denote by $|u|_{s,T}$ the norm of u in $H^s(b\Omega_T)$.

Consider the initial data functions $\mathbf{U}_0^\pm = (\hat{\rho}^\pm, \hat{\mathbf{v}}^\pm, \hat{\mathbf{H}}^\pm, \hat{S}^\pm)$ and ψ_0 that are a small perturbation of a planar current-vortex sheet $(\bar{\mathbf{U}}^\pm, \bar{\psi})$ for constant states $\bar{\mathbf{U}}^\pm$ and $\bar{\psi} = 0$ with (25)–(26) so that the following stability condition holds:

$$\left(\hat{\lambda}^\pm\right)^2 < \frac{1}{\hat{\rho}^\pm + |\hat{\mathbf{H}}^\pm|^2/(\hat{c}^\pm)^2}, \tag{38}$$

for $\hat{\lambda}^\pm$ uniquely determined by

$$\begin{pmatrix} \hat{H}_2^+ & -\hat{H}_2^- \\ \hat{H}_3^+ & -\hat{H}_3^- \end{pmatrix} \begin{pmatrix} \hat{\lambda}^+ \\ \hat{\lambda}^- \end{pmatrix} = \begin{pmatrix} \hat{v}_2^+ - \hat{v}_2^- \\ \hat{v}_3^+ - \hat{v}_3^- \end{pmatrix} \quad \text{on } \{x_1 = 0\}. \tag{39}$$

Then we have the following main result.

Theorem 1 (Chen–Wang [17]) *Assume that, for any fixed* $\alpha \ge 15$ *and* $s \in [\alpha + 7, 2\alpha - 5]$, *the initial data functions* $\psi_0 \in H^{2s+3}(\mathbb{R}^2)$ *and* $\mathbf{U}_0^\pm - \bar{\mathbf{U}}^\pm \in B^{2(s+2)}(\mathbb{R}^3_+)$ *satisfy constraints* (37), *the compatibility conditions of problem* (36) *up to order* $s + 2$, *and the stability condition* (38)–(39). *Then there exists a solution* $(\mathbf{U}^\pm, \psi)$ *of the initial-boundary value problem* (36) *such that*

$$\mathbf{U}^\pm - \bar{\mathbf{U}}^\pm \in B^\alpha(\Omega_T) \quad \textit{and} \quad \psi \in H^{\alpha-1}(b\Omega_T).$$

Remark 1 The stability conditions (38) and (39) for the initial data functions $\mathbf{U}_0^\pm = (\hat{\rho}^\pm, \hat{\mathbf{v}}^\pm, \hat{\mathbf{H}}^\pm, \hat{S}^\pm)$ and ψ_0 are equivalent to

$$\max\{|\hat{H}_2^-(\hat{v}_3^+ - \hat{v}_3^-) - \hat{H}_3^-(\hat{v}_2^+ - \hat{v}_2^-)|, |\hat{H}_2^+(\hat{v}_3^+ - \hat{v}_3^-) - \hat{H}_3^+(\hat{v}_2^+ - \hat{v}_2^-)|\}$$
$$\leq \frac{|\hat{H}_2^+\hat{H}_3^- - \hat{H}_2^-\hat{H}_3^+|}{\sqrt{\hat{\rho}^\pm + |\hat{H}^\pm|^2/(\hat{c}^\pm)^2}}. \tag{40}$$

Also see Trakhinin [37] for another equivalent form and [38] for a similar result but different proof independently.

Remark 2 Using the same argument as in Coulombel–Secchi [23], we conclude that the above current-vortex sheet solution to system (19) is also uniquely determined by its initial data.

To establish Theorem 1, in Chen–Wang [17], we developed an approach by combining a well-structured decoupled formulation of a linearized problem derived from the current-vortex sheet and careful high-order energy estimates of the solutions of the linearized problem with a suitable iteration scheme of the Nash–Moser–Hörmander type. In order to establish the energy estimates, especially high-order energy estimates, of solutions to the linearized problem, one of the main contributions in [17] is to identify the *well-structured decoupled formulation* so that the linear problem decoupled into *one standard initial-boundary value problem* (48) *for a symmetric hyperbolic system and the other problem* (52) *for a transport equation for the front*. This decoupled formulation is essential for us in a much more convenient way to establish the desired high-order energy estimates of solutions and to develop the suitable iteration scheme of the Nash–Moser–Hörmander type that converges. On the other hand, in [17], there is a gap in the presentation for constructing the iteration scheme of Nash–Moser–Hörmander type for the nonlinear problem (36). In Sects. 4–5, we provide a supplement and describe the complete arguments of the proof of Theorem 1 here, i.e., Theorem 2.1 in [17].

4 Compressible Current-Vortex Sheets in MHD: Linear Stability

To study the linear stability of current-vortex sheets, we first derive a linearized problem from the nonlinear problem (36). By a direct calculation, we have

$$\frac{d}{ds}\big(\mathbf{L}(\mathbf{U}+s\mathbf{V}, \Psi + s\Phi)(\mathbf{U}+s\mathbf{V})\big)|_{s=0}$$
$$= \mathbf{L}(\mathbf{U}, \Psi)\mathbf{W} + \mathbf{E}(\mathbf{U}, \Psi)\mathbf{W} + \frac{\Phi}{\Psi_{x_1}}(\mathbf{L}(\mathbf{U}, \Psi)\mathbf{U})_{x_1},$$

where $\mathbf{W} = \mathbf{V} - \frac{\Phi}{\Psi_{x_1}}\mathbf{U}_{x_1}$ is the good unknown as introduced in [2] (see also [26, 33]), and

$$\mathbf{E}(\mathbf{U}, \Psi)\mathbf{W} = \mathbf{W}\cdot\nabla_{\mathbf{U}}(\bar{\mathbf{A}}_1(\mathbf{U}, \Psi))\mathbf{U}_{x_1} + \sum_{j=2}^{3}\mathbf{W}\cdot\nabla_{\mathbf{U}}\mathbf{A}_j(\mathbf{U})\mathbf{U}_{x_j}.$$

Then we obtain the following linearized problem of (36):

$$\begin{cases} \mathbf{L}(\mathbf{U}^{\pm}, \Psi^{\pm})\mathbf{W}^{\pm} + \mathbf{E}(\mathbf{U}^{\pm}, \Psi^{\pm})\mathbf{W}^{\pm} = \mathbf{F}^{\pm} & \text{in } \{x_1 > 0\}, \\ \phi_t - (W_2^{\pm} - \psi_{x_2} W_3^{\pm} - \psi_{x_3} W_4^{\pm}) + U_3^{\pm}\phi_{x_2} + U_4^{\pm}\phi_{x_3} = h_1^{\pm} & \text{on } \{x_1 = 0\}, \\ W_1^{+} - W_1^{-} + \sum_{j=5}^{7}(U_j^{+}W_j^{+} - U_j^{-}W_j^{-}) = h_2 & \text{on } \{x_1 = 0\}, \\ (\mathbf{W}^{\pm}, \phi)|_{t=0} = 0, \end{cases} \tag{41}$$

for some functions $\mathbf{F}^{\pm}$, $h_1^{\pm}$, and h_2, where $\Psi^{\pm}(t, \mathbf{x})$ are proper extensions of $\psi(t, x_2, x_3)$ in $\{x_1 > 0\}$ satisfying (34).

To simplify problem (41), we introduce $\mathbf{J}^{\pm} = \mathbf{J}(\mathbf{U}^{\pm}, \Psi^{\pm})$ as an 8×8 regular matrix such that

$$\mathbf{X}^{\pm} = (\mathbf{J}^{\pm})^{-1}\mathbf{W}^{\pm} \tag{42}$$

satisfy

$$\begin{cases} X_1^{\pm} = W_1^{\pm} + \sum_{j=5}^{7} U_j^{\pm} W_j^{\pm}, \\ X_2^{\pm} = W_2^{\pm} - (\Psi^{\pm})_{x_2} W_3^{\pm} - (\Psi^{\pm})_{x_3} W_4^{\pm}, \\ X_5^{\pm} = W_5^{\pm} - (\Psi^{\pm})_{x_2} W_6^{\pm} - (\Psi^{\pm})_{x_3} W_7^{\pm}, \\ (X_3^{\pm}, X_4^{\pm}, X_6^{\pm}, X_7^{\pm}, X_8^{\pm}) = (W_3^{\pm}, W_4^{\pm}, W_6^{\pm}, W_7^{\pm}, W_8^{\pm}), \end{cases}$$

which means that X_1, X_2, X_5, and X_8 represent the linearized total pressure, normal velocity, normal magnetic field, and entropy respectively, while (X_3, X_4) and (X_6, X_7) are the associated tangential velocity and magnetic fields.

Under transformation (42), problem (41) for $(\mathbf{W}^{\pm}, \phi)$ is equivalent to the following problem for $(\mathbf{X}^{\pm}, \phi)$:

$$\begin{cases} \tilde{\mathbf{L}}(\mathbf{U}^{\pm}, \Psi^{\pm})\mathbf{X}^{\pm} + \tilde{\mathbf{E}}(\mathbf{U}^{\pm}, \Psi^{\pm})\mathbf{X}^{\pm} = \tilde{\mathbf{F}}^{\pm} & \text{in } \{x_1 > 0\}, \\ \phi_t - X_2^{\pm} + U_3^{\pm}\phi_{x_2} + U_4^{\pm}\phi_{x_3} = h_1^{\pm} & \text{on } \{x_1 = 0\}, \\ X_1^{+} - X_1^{-} = h_2 & \text{on } \{x_1 = 0\}, \\ (\mathbf{X}^{\pm}, \phi)|_{t=0} = 0, \end{cases} \tag{43}$$

where $\tilde{\mathbf{F}}^{\pm} = (\mathbf{J}^{\pm})^{\top}\mathbf{F}^{\pm}$,

$$\tilde{\mathbf{L}}(\mathbf{U}^{\pm}, \Psi^{\pm}) = \tilde{\mathbf{A}}_0(\mathbf{U}^{\pm}, \Psi^{\pm})\partial_t + \sum_{j=1}^{3}\tilde{\mathbf{A}}_j(\mathbf{U}^{\pm}, \Psi^{\pm})\partial_{x_j}$$

with

$$\begin{aligned} &\tilde{\mathbf{A}}_1(\mathbf{U}^{\pm}, \Psi^{\pm}) = (\mathbf{J}^{\pm})^{\top}\bar{\mathbf{A}}_1(\mathbf{U}^{\pm}, \Psi^{\pm})\mathbf{J}^{\pm}, \\ &\tilde{\mathbf{A}}_j(\mathbf{U}^{\pm}, \Psi^{\pm}) = (\mathbf{J}^{\pm})^{\top}\mathbf{A}_j(\mathbf{U}^{\pm})\mathbf{J}^{\pm}, \quad j \neq 1, \\ &\tilde{\mathbf{E}}(\mathbf{U}^{\pm}, \Psi^{\pm})\mathbf{X}^{\pm} = \big((\mathbf{J}^{\pm})^{\top}\mathbf{E}(\mathbf{U}^{\pm}, \Psi^{\pm})\mathbf{J}^{\pm}\big)\mathbf{X}^{\pm} + \big((\mathbf{J}^{\pm})^{\top}\mathbf{L}(\mathbf{U}^{\pm}, \Psi^{\pm})\mathbf{J}^{\pm}\big)\mathbf{X}^{\pm}. \end{aligned}$$

By a direct calculation, we see that the coefficient matrix $\tilde{\mathbf{A}}_1(\mathbf{U}^\pm, \Psi^\pm)$ in $\tilde{\mathbf{L}}(\mathbf{U}^\pm, \Psi^\pm)$ can be decomposed into three parts:

$$\tilde{\mathbf{A}}_1(\mathbf{U}^\pm, \Psi^\pm) = \mathbf{A}_1^{\pm,0} + \mathbf{A}_1^{\pm,1} + \mathbf{A}_1^{\pm,2}$$

with

$$\mathbf{A}_1^{\pm,0} = \frac{1}{(\Psi^\pm)_{x_1}} \begin{bmatrix} 0 & \mathbf{a} \\ \mathbf{a}^\top & \mathbf{O}_{7\times 7} \end{bmatrix}, \qquad \mathbf{A}_1^{\pm,1} = \frac{\Psi_t^\pm - U_{v,N}^\pm}{(\Psi^\pm)_{x_1}} \tilde{\mathbf{A}}_1^{\pm,1},$$
$$\mathbf{A}_1^{\pm,2} = \frac{U_{H,N}^\pm}{(\Psi^\pm)_{x_1}} \tilde{\mathbf{A}}_1^{\pm,2},$$

where

$$\mathbf{a} = (1, 0, 0, -\lambda^\pm, 0, 0, 0),$$
$$U_{v,N}^\pm = U_2^\pm - (\Psi^\pm)_{x_2} U_3^\pm - (\Psi^\pm)_{x_3} U_4^\pm,$$

and

$$U_{H,N}^\pm = U_5^\pm - (\Psi^\pm)_{x_2} U_6^\pm - (\Psi^\pm)_{x_3} U_7^\pm.$$

When the states $(\mathbf{U}^\pm, \Psi^\pm)$ satisfy the boundary conditions given in (36) and constraints (37), i.e.,

$$\psi_t - U_{v,N}^\pm = 0, \qquad U_{H,N}^\pm = 0 \quad \text{on } \{x_1 = 0\},$$

the boundary $\{x_1 = 0\}$ is a characteristic plane of constant multiplicity for the operator $\tilde{\mathbf{L}}(\mathbf{U}^\pm, \Psi^\pm)$. Then, from (43), we obtain that, on $\{x_1 = 0\}$,

$$\left\langle \begin{pmatrix} \tilde{\mathbf{A}}_1(\mathbf{U}^+, \Psi^+) & 0 \\ 0 & \tilde{\mathbf{A}}_1(U^-, \Psi^-) \end{pmatrix} \begin{pmatrix} \mathbf{X}^+ \\ \mathbf{X}^- \end{pmatrix}, \begin{pmatrix} \mathbf{X}^+ \\ \mathbf{X}^- \end{pmatrix} \right\rangle = 2X_1^\pm [X_2 - \lambda X_5], \tag{44}$$

when $[X_1] = 0$ on $\{x_1 = 0\}$.

In order to decouple the front unknown ϕ from the boundary condition $(43)_2$, we use the linearization of the constraints $H_N^\pm|_{x_1=0} = 0$:

$$X_5^\pm - U_6^\pm \phi_{x_2} - U_7^\pm \phi_{x_3} = h_3^\pm \quad \text{on } \{x_1 = 0\} \tag{45}$$

to obtain

$$[X_2 - \lambda X_5] = \phi_{x_2}[U_3 - \lambda U_6] + \phi_{x_3}[U_4 - \lambda U_7] - [h_1 + \lambda h_3]. \tag{46}$$

From the assumption $\mathbf{H}_\tau^+ \nparallel \mathbf{H}_\tau^-$ on $\{x_1 = 0\}$, there exist unique

$$\lambda^\pm = \lambda^\pm(v_2^\pm, v_3^\pm, H_2^\pm, H_3^\pm)$$

such that

$$\begin{pmatrix} v_2^+ \\ v_3^+ \end{pmatrix} - \begin{pmatrix} v_2^- \\ v_3^- \end{pmatrix} = \lambda^+ \begin{pmatrix} H_2^+ \\ H_3^+ \end{pmatrix} - \lambda^- \begin{pmatrix} H_2^- \\ H_3^- \end{pmatrix} \quad \text{on } \{x_1 = 0\}, \tag{47}$$

that is, $[U_3 - \lambda U_6] = [U_4 - \lambda U_7] = 0$. In this case, (46) is simplified as

$$[X_2 - \lambda X_5] = -[h_1 + \lambda h_3].$$

Therefore, with the aid of (45) and the choice of $\lambda^\pm$ in (47), we deduce from (43) that $\mathbf{X}^\pm$ satisfy the following problem:

$$\begin{cases} \tilde{\mathbf{L}}(\mathbf{U}^\pm, \Psi^\pm)\mathbf{X}^\pm + \tilde{\mathbf{E}}(\mathbf{U}^\pm, \Psi^\pm)\mathbf{X}^\pm = \tilde{\mathbf{F}}^\pm & \text{in } \{x_1 > 0\}, \\ [X_2 - \lambda X_5] = -[h_1 + \lambda h_3] & \text{on } \{x_1 = 0\}, \\ X_1^+ - X_1^- = h_2 & \text{on } \{x_1 = 0\}, \\ \mathbf{X}^\pm|_{t\le 0} = 0. \end{cases} \tag{48}$$

From (44), we know that the homogeneous boundary conditions:

$$[X_1] = [X_2 - \lambda X_5] = 0 \quad \text{on } \{x_1 = 0\} \tag{49}$$

are nonnegative for the operator

$$\tilde{\mathbf{L}}(\mathbf{U}^\pm, \Psi^\pm) = \tilde{\mathbf{A}}_0(\mathbf{U}^\pm, \Psi^\pm)\partial_t + \sum_{j=1}^{3} \tilde{\mathbf{A}}_j(\mathbf{U}^\pm, \Psi^\pm)\partial_{x_j}.$$

Moreover, by noting that, on the boundary $\{x_1 = 0\}$,

$$\tilde{A}_1(U^\pm, \Psi^\pm) = \frac{1}{(\Psi^\pm)_{x_1}} \begin{bmatrix} 0 & \mathbf{a} \\ \mathbf{a}^\top & \mathbf{O}_{7\times 7} \end{bmatrix}$$

with $\mathbf{a} = (1, 0, 0, -\lambda^\pm, 0, 0, 0)$, the boundary conditions (49) on $\{x_1 = 0\}$ are also maximally nonnegative for the operator $\tilde{\mathbf{L}}(\mathbf{U}^\pm, \Psi^\pm)$. Then, by employing the Lax–Friedrichs theory (Theorem 1.1 in [27]) for (48), we conclude that there exists a unique solution $\mathbf{X}^\pm$ of problem (48) satisfying the following energy estimates:

Theorem 2 *For any fixed $s_0 > 17/2$, there exist constants C_0 and μ_0 depending only on $\|\dot{coef}\|_{s_0,T}$ for the coefficient functions in (48) such that, for any $s \ge s_0$ and $\mu \ge \mu_0$, the estimate:*

$$\begin{aligned} &\max_{0\le t\le T} \|\mathbf{X}^\pm(t)\|^2_{s,\mu} + \mu\|\mathbf{X}^\pm\|^2_{s,\mu,T} \\ &\quad \le \frac{C_0}{\mu}\Big(\|\mathbf{F}^\pm\|_{s,\mu,T} + \|\mathbf{h}\|^2_{H^{s+1}_\mu(b\Omega_T)} + \|\dot{coef}\|^2_{s,\mu,T}(\|\mathbf{F}^\pm\|^2_{s_0,T} + |\mathbf{h}|^2_{s_0+1,T})\Big) \end{aligned} \tag{50}$$

holds, provided that the eikonal equations:

$$\psi_t = U_2^\pm - \psi_{x_2}U_3^\pm - \psi_{x_3}U_4^\pm$$

and the constraints:

$$U^\pm_{H,N} := U_5^\pm - \psi^\pm_{x_2}U_6^\pm - \psi_{x_3}U_7^\pm = 0$$

are valid for $(\mathbf{U}^\pm, \psi)$ on $\{x_1 = 0\}$, and $\lambda^\pm$ determined in (47) satisfy condition (30), where $\mathbf{h} = (h_1^\pm, h_2, h_3^\pm)^\top$ and the norms in $H^s_\mu(b\Omega_T)$ are defined as that of $B^s_\mu(\Omega_T)$ with functions independent of x_1, and $\dot{coef}(t, \mathbf{x}) = coef(t, \mathbf{x}) - coef(0)$ with $coef(t, x)$ being the coefficient functions in the linear operators $\tilde{L}(U^\pm, \Psi^\pm)$ and $\tilde{E}(U^\pm, \Psi^\pm)$ in the equations in (48).

By fixing $\mu \gg 1$ in (50), we conclude

Corollary 1 *For any fixed $s_0 > 17/2$, there exists a constant $C_0 > 0$ depending only on $\|\dot{coef}\|_{s_0,T}$ and T such that, for any $s \geq s_0$, the following estimate holds*:

$$\|\mathbf{X}^{\pm}\|_{s,T}^2 \leq C_0\Big(\|\mathbf{F}^{\pm}\|_{s,T}^2 + |\mathbf{h}|_{s+1,T}^2 + \|\dot{coef}\|_{s,T}^2(\|\mathbf{F}^{\pm}\|_{s_0,T}^2 + |\mathbf{h}|_{s_0+1,T}^2)\Big). \quad (51)$$

Finally, let us study the determination of the perturbation $\Phi^{\pm}$ of the front functions $\Psi^{\pm}$. From problem (41), the natural idea is to solve the following problems:

$$\begin{cases} \partial_t \Phi^{\pm} - X_2^{\pm} + U_3^{\pm}\partial_{x_2}\Phi^{\pm} + U_4^{\pm}\partial_{x_3}\Phi^{\pm} = h_1^{\pm} \quad \text{in } \{x_1 > 0\}, \\ \Phi^{\pm}|_{t=0} = 0. \end{cases} \quad (52)$$

An important question is whether we have $\Phi^+ = \Phi^-$ on $\{x_1 = 0\}$. This question is answered by the following result.

Proposition 1 *Let $\Phi^+(t, \mathbf{x})$ be given by problem (52) with the plus sign, and $\phi(t, x_2, x_3) = \Phi^+|_{x_1=0}$. If $\lambda^{\pm}$ are given in (47), and the boundary condition*:

$$[X_2 - \lambda X_5] = -[h_1 + \lambda h_3] \quad \text{on } \{x_1 = 0\} \quad (53)$$

holds as in (48), with $h_3^{\pm}$ being given by

$$X_5^{\pm} - U_6^{\pm}\phi_{x_2} - U_7^{\pm}\phi_{x_3} = h_3^{\pm} \quad \text{on } \{x_1 = 0\}, \quad (54)$$

then we have

$$\partial_t \phi - X_2^- + U_3^- \partial_{x_2}\phi + U_4^- \partial_{x_3}\phi = h_1^- \quad \text{on } \{x_1 = 0\}. \quad (55)$$

Proof Notice that (53) and (54) can be rewritten as

$$X_2^+ + h_1^+ - (X_2^- + h_1^-) = \lambda^+(X_5^+ - h_3^+) - \lambda^-(X_5^- - h_3^-), \quad (56)$$

and

$$X_5^{\pm} - h_3^{\pm} = (\partial_{x_2}\phi, \partial_{x_3}\phi)\begin{pmatrix} U_6^{\pm} \\ U_7^{\pm} \end{pmatrix}. \quad (57)$$

Thus, from (52) with the plus sign, we obtain

$$\begin{aligned} &\partial_t \phi + (\partial_{x_2}\phi, \partial_{x_3}\phi)\begin{pmatrix} U_3^+ \\ U_4^+ \end{pmatrix} - (X_2^- + h_1^-) \\ &= (\partial_{x_2}\phi, \partial_{x_3}\phi)\left[\lambda^+\begin{pmatrix} U_6^+ \\ U_7^+ \end{pmatrix} - \lambda^-\begin{pmatrix} U_6^- \\ U_7^- \end{pmatrix}\right] \\ &= (\partial_{x_2}\phi, \partial_{x_3}\phi)\left[\begin{pmatrix} U_3^+ \\ U_4^+ \end{pmatrix} - \begin{pmatrix} U_3^- \\ U_4^- \end{pmatrix}\right] \end{aligned} \quad (58)$$

by using (56)–(57) and (47). From (58), we immediately conclude (55). □

5 Compressible Current-Vortex Sheets in MHD: Nonlinear Stability

In this section, we describe the main steps to prove Theorem 1, the existence of a local solution to the nonlinear problem (36) under constraints (37) for the initial data, by developing an iteration scheme of the Nash–Moser–Hörmander type.

5.1 Construction of the Zero-th Order Approximate Solutions

Suppose that the initial data $(\mathbf{U}_0^{\pm}, \psi_0)$ is a perturbation of a planar current-vortex sheet $(\bar{\mathbf{U}}^{\pm}, \bar{\psi})$ with the constant states $\bar{\mathbf{U}}^{\pm}$ and $\bar{\psi} = 0$ satisfying (25)–(26), $\psi_0 \in H^{s-1}(\mathbb{R}^2)$, and $\dot{\mathbf{U}}_0^{\pm} = \mathbf{U}_0^{\pm} - \bar{\mathbf{U}}^{\pm} \in B^s(\mathbb{R}^3_+)$ for any fixed integer $s > 9/2$. Suppose that $(\mathbf{U}_0^{\pm}, \psi_0)$ satisfy the compatibility conditions of problem (36) up to order $[\frac{s}{2}]$, and the constraints in (37) with $\Psi^{\pm}(0, \mathbf{x})$ being proper extensions of $\psi_0(x_2, x_3)$ in $\{x_1 > 0\}$, satisfying $\pm\partial_{x_1}\Psi^{\pm}|_{t=0} \geq \kappa > 0$. In a classical way, one can construct the zero-th order approximate solutions $(\mathbf{U}_a^{\pm}, \Psi_a^{\pm})$ such that $\dot{\mathbf{U}}_a^{\pm} = \mathbf{U}_a^{\pm} - \bar{\mathbf{U}}^{\pm} \in B^{[s/2]+1}(\mathbb{R}_+ \times \mathbb{R}^3_+)$, $\Psi_a^{\pm} \mp x_1 \in B^{[s/2]+2}(\mathbb{R}_+ \times \mathbb{R}^3_+)$ with $\pm\partial_{x_1}\Psi^{\pm}|_{t=0} \geq \kappa/2 > 0$ satisfying

$$\partial_t^j \big(\mathbf{L}(\mathbf{U}_a^{\pm}, \Psi_a^{\pm})\mathbf{U}_a^{\pm}\big)|_{t=0} = 0 \quad \text{for } 0 \leq j \leq \left[\frac{s}{2}\right] - 1 \tag{59}$$

and

$$\begin{aligned} &\mathbf{B}(\mathbf{U}_a^{+}, \mathbf{U}_a^{-}, \psi_a) = 0, \\ &U_{a,5}^{\pm} - (\psi_a)_{x_2} U_{a,6}^{\pm} - (\psi_a)_{x_3} U_{a,7}^{\pm} = 0 \quad \text{on } \{x_1 = 0\} \end{aligned} \tag{60}$$

with $\Psi_a^{\pm}|_{x_1=0} = \psi_a(t, x_2, x_3)$.

Set

$$\mathbf{V}^{\pm} = \mathbf{U}^{\pm} - \mathbf{U}_a^{\pm}, \qquad \Phi^{\pm} = \Psi^{\pm} - \Psi_a^{\pm}. \tag{61}$$

Then it follows from (59) and (60) that problem (36) is equivalent to the following problem for $(\mathbf{V}^{\pm}, \Phi^{\pm})$:

$$\begin{cases} \mathscr{L}(\mathbf{V}^{\pm}, \Phi^{\pm})\mathbf{V}^{\pm} = \mathbf{f}_a^{\pm} & \text{in } \{t > 0, x_1 > 0\}, \\ \mathscr{B}(\mathbf{V}^{+}, \mathbf{V}^{-}, \phi) = 0 & \text{on } \{x_1 = 0\}, \\ \mathbf{V}^{\pm}|_{t\leq 0} = 0, \quad \phi|_{t \leq 0} = 0, \end{cases} \tag{62}$$

where $\phi(t, x_2, x_3) = \Phi^{\pm}|_{x_1=0}$ and $\mathbf{f}_a^{\pm} = -\mathbf{L}(\mathbf{U}_a^{\pm}, \Psi_a^{\pm})\mathbf{U}_a^{\pm}$,

$$\mathscr{L}(\mathbf{V}^{\pm}, \Phi^{\pm})\mathbf{V}^{\pm} = \mathbf{L}(\mathbf{U}_a^{\pm} + \mathbf{V}^{\pm}, \Psi_a^{\pm} + \Phi^{\pm})(\mathbf{U}_a^{\pm} + \mathbf{V}^{\pm}) - \mathbf{L}(\mathbf{U}_a^{\pm}, \Psi_a^{\pm})\mathbf{U}_a^{\pm},$$

and

$$\mathscr{B}(\mathbf{V}^{+}, \mathbf{V}^{-}, \phi) = \mathbf{B}(\mathbf{U}_a^{+} + \mathbf{V}^{+}, \mathbf{U}_a^{-} + \mathbf{V}^{-}, \psi_a + \phi).$$

5.2 Iteration Scheme

From the linear stability estimate established in Theorem 2, we observe that there exists a loss of regularity for the linearized problem (48). This inspires us to use a suitable iteration scheme of the Nash–Moser–Hörmander type (cf. [31]) to study the nonlinear problem (62).

To do this, we first recall a standard family of smoothing operators (cf. [2, 22]):

$$\{S_\theta\}_{\theta>0} : B^0_\mu(\Omega_T) \longrightarrow \bigcap_{s\geq 0} B^s_\mu(\Omega_T) \tag{63}$$

satisfying

$$\begin{cases} \|S_\theta u\|_{s,T} \leq C\theta^{(s-\alpha)_+}\|u\|_{\alpha,T} & \text{for all } s,\alpha\geq 0,\\ \|S_\theta u - u\|_{s,T} \leq C\theta^{s-\alpha}\|u\|_{\alpha,T} & \text{for all } s\in[0,\alpha],\\ \left\|\dfrac{d}{d\theta}S_\theta u\right\|_{s,T} \leq C\theta^{s-\alpha-1}\|u\|_{\alpha,T} & \text{for all } s,\alpha\geq 0, \end{cases} \tag{64}$$

and

$$\big|(S_\theta u_+ - S_\theta u_-)|_{x_1=0}\big|_{s,T} \leq C\theta^{(s+1-\alpha)_+}\big|(u_+-u_-)|_{x_1=0}\big|_{\alpha,T} \quad \text{for all } s,\alpha\geq 0. \tag{65}$$

Similarly, one has a family of smoothing operators (still denoted by) $\{S_\theta\}_{\theta>0}$ acting on $H^s(b\Omega_T)$, satisfying also (64) for the norms of $H^s(b\Omega_T)$ (cf. [2, 22]).

Now we construct the iteration scheme for solving the nonlinear problem (62) in $\mathbb{R}_+\times\mathbb{R}^3_+$.

Let $\mathbf{V}^{\pm,0}=0$ and $\Phi^{\pm,0}=0$. Assume that $(\mathbf{V}^{\pm,k},\Phi^{\pm,k})$ have been known for $k=0,\ldots,n$, and satisfy $\Phi^{+,k}=\Phi^{-,k}$ on $\{x_1=0\}$,

$$(\mathbf{V}^{\pm,k},\Phi^{\pm,k})=0 \quad \text{in } \{t\leq 0\}. \tag{66}$$

Denote the $(n+1)^{th}$ approximate solutions to (62) in $\mathbb{R}_+\times\mathbb{R}^3_+$ by

$$\mathbf{V}^{\pm,n+1}=\mathbf{V}^{\pm,n}+\delta\mathbf{V}^{\pm,n}, \qquad \Phi^{\pm,n+1}=\Phi^{\pm,n}+\delta\Phi^{\pm,n}. \tag{67}$$

Let $\theta_0\geq 1$ and $\theta_n=\sqrt{\theta_0^2+n}$ for any $n\geq 1$. Let S_{θ_n} be the associated smoothing operator defined as above. Denote by

$$\begin{aligned} &\mathbf{L}'_{e,(\mathbf{U}_a^\pm+\mathbf{V}^{\pm,n+\frac12},\Psi_a^\pm+S_{\theta_n}\Phi^{\pm,n})}\delta\dot{\mathbf{V}}^{\pm,n}\\ &\quad=\mathbf{L}(\mathbf{U}_a^\pm+\mathbf{V}^{\pm,n+\frac12},\Psi_a^\pm+S_{\theta_n}\Phi^{\pm,n})\delta\dot{\mathbf{V}}^{\pm,n}\\ &\qquad+\mathbf{E}(\mathbf{U}_a^\pm+\mathbf{V}^{\pm,n+\frac12},\Psi_a^\pm+S_{\theta_n}\Phi^{\pm,n})\delta\dot{\mathbf{V}}^{\pm,n} \end{aligned} \tag{68}$$

the effective linearized operator, and

$$\delta\dot{\mathbf{V}}^{\pm,n}=\delta\mathbf{V}^{\pm,n}-\delta\Phi^{\pm,n}\frac{(\mathbf{U}_a^\pm+\mathbf{V}^{\pm,n+\frac12})_{x_1}}{(\Psi_a^\pm+S_{\theta_n}\Phi^{\pm,n})_{x_1}} \tag{69}$$

the good unknown.

By a direct computation, we have

$$\mathscr{L}(\mathbf{V}^{\pm,n+1}, \varPhi^{\pm,n+1})\mathbf{V}^{\pm,n+1} = \sum_{j=0}^{n} \big(\mathbf{L}'_{\mathbf{e},(\mathbf{U}_a^{\pm}+\mathbf{V}^{\pm,j+\frac{1}{2}},\varPsi_a^{\pm}+S_{\theta_j}\varPhi^{\pm,j})}\delta\dot{\mathbf{V}}^{\pm,j} + \mathbf{e}_{\pm,j}\big), \tag{70}$$

where the modified states $\mathbf{V}^{\pm,j+\frac{1}{2}}$ will be chosen such that the boundary $\{x_1 = 0\}$ is a uniform characteristic plane of constant multiplicity for the operator $\mathbf{L}'_{\mathbf{e},(\mathbf{U}_a^{\pm}+\mathbf{V}^{\pm,j+\frac{1}{2}},\varPsi_a^{\pm}+S_{\theta_j}\varPhi^{\pm,j})}$ for all $j \geq 0$, and

$$\mathbf{e}_{\pm,j} = \sum_{k=1}^{4} \mathbf{e}_{\pm,j}^{(k)} \tag{71}$$

with

$$\begin{aligned}
\mathbf{e}_{\pm,j}^{(1)} &= \mathbf{L}(\mathbf{U}_a^{\pm} + \mathbf{V}^{\pm,j+1}, \varPsi_a^{\pm} + \varPhi^{\pm,j+1})(\mathbf{U}_a^{\pm} + \mathbf{V}^{\pm,j+1}) \\
&\quad - \mathbf{L}(\mathbf{U}_a^{\pm} + \mathbf{V}^{\pm,j}, \varPsi_a^{\pm} + \varPhi^{\pm,j})(\mathbf{U}_a^{\pm} + \mathbf{V}^{\pm,j}) \\
&\quad - \mathbf{L}'_{(\mathbf{U}_a^{\pm}+\mathbf{V}^{\pm,j},\varPsi_a^{\pm}+\varPhi^{\pm,j})}(\delta\mathbf{V}^{\pm,j}, \delta\varPhi^{\pm,j}), && (72)\\
\mathbf{e}_{\pm,j}^{(2)} &= \mathbf{L}'_{(\mathbf{U}_a^{\pm}+\mathbf{V}^{\pm,j},\varPsi_a^{\pm}+\varPhi^{\pm,j})}(\delta\mathbf{V}^{\pm,j}, \delta\varPhi^{\pm,j}) \\
&\quad - \mathbf{L}'_{(\mathbf{U}_a^{\pm}+S_{\theta_j}\mathbf{V}^{\pm,j},\varPsi_a^{\pm}+S_{\theta_j}\varPhi^{\pm,j})}(\delta\mathbf{V}^{\pm,j}, \delta\varPhi^{\pm,j}), && (73)\\
\mathbf{e}_{\pm,j}^{(3)} &= \mathbf{L}'_{(\mathbf{U}_a^{\pm}+S_{\theta_j}\mathbf{V}^{\pm,j},\varPsi_a^{\pm}+S_{\theta_j}\varPhi^{\pm,j})}(\delta\mathbf{V}^{\pm,j}, \delta\varPhi^{\pm,j}) \\
&\quad - \mathbf{L}'_{(\mathbf{U}_a^{\pm}+\mathbf{V}^{\pm,j+\frac{1}{2}},\varPsi_a^{\pm}+S_{\theta_j}\varPhi^{\pm,j})}(\delta\mathbf{V}^{\pm,j}, \delta\varPhi^{\pm,j}), && (74)
\end{aligned}$$

and

$$\mathbf{e}_{\pm,j}^{(4)} = \frac{\delta\varPhi^{\pm,j}}{(\varPsi_a^{\pm} + S_{\theta_j}\varPhi^{\pm,j})_{x_1}} \big(\mathbf{L}(\mathbf{U}_a^{\pm} + \mathbf{V}^{\pm,j+\frac{1}{2}}, \varPsi_a^{\pm} + S_{\theta_j}\varPhi^{\pm,j})(\mathbf{U}_a^{\pm} + \mathbf{V}^{\pm,j+\frac{1}{2}})\big)_{x_1}. \tag{75}$$

For the boundary condition given in (62), we have

$$\mathscr{B}(\mathbf{V}^{\pm}, \varPhi^{\pm}) = (\mathscr{B}_1^{+}(\mathbf{V}^{+}, \varPhi^{+}), \mathscr{B}_1^{-}(\mathbf{V}^{-}, \varPhi^{-}), \mathscr{B}_2(\mathbf{V}^{+}, \mathbf{V}^{-}))^{\top},$$

with

$$\begin{cases}
\mathscr{B}_1^{\pm}(\mathbf{V}^{\pm}, \varPhi^{\pm}) = (\partial_t + U_{a,3}^{\pm}\partial_{x_2} + U_{a,4}^{\pm}\partial_{x_3})\varPhi^{\pm} - V_2^{\pm} + (\varPsi_a^{\pm} + \varPhi^{\pm})_{x_2} V_3^{\pm} \\
\qquad\qquad + (\varPsi_a^{\pm} + \varPhi^{\pm})_{x_3} V_4^{\pm}, \\
\mathscr{B}_2(\mathbf{V}^{+}, \mathbf{V}^{-}) = V_1^{+} - V_1^{-} + \frac{1}{2}(|\mathbf{V}_H^{+}|^2 - |\mathbf{V}_H^{-}|^2) + \langle \mathbf{U}_{a,\mathbf{H}}^{+}, \mathbf{V}_H^{+}\rangle - \langle \mathbf{U}_{a,\mathbf{H}}^{-}, \mathbf{V}_{\mathbf{H}}^{-}\rangle,
\end{cases} \tag{76}$$

$\mathbf{U}_{a,\mathbf{H}}^{\pm} = (U_{a,5}^{\pm}, U_{a,6}^{\pm}, U_{a,7}^{\pm})^{\top}$, and $\mathbf{V}_{\mathbf{H}}^{\pm} = (V_5^{\pm}, V_6^{\pm}, V_7^{\pm})^{\top}$.

Associated with the constraints $H_N^{\pm}|_{x_1=0} = 0$, denote by

$$\mathscr{B}_3^{\pm}(\mathbf{V}^{\pm},\Phi^{\pm}) = V_5^{\pm} - (\Psi_a^{\pm}+\Phi^{\pm})_{x_2} V_6^{\pm} - (\Psi_a^{\pm}+\Phi^{\pm})_{x_3} V_7^{\pm} - (\Phi^{\pm})_{x_2} U_{a,6}^{\pm} - (\Phi^{\pm})_{x_3} U_{a,7}^{\pm}.$$

By a direct calculation, for $i=1,3$, we have

$$\begin{aligned}&\mathscr{B}_i^{\pm}(\mathbf{V}^{\pm,n+1},\Phi^{\pm,n+1}) - \mathscr{B}_i^{\pm}(\mathbf{V}^{\pm,n},\Phi^{\pm,n})\\ &= \mathscr{B}^{\pm\prime}_{i,(\mathbf{V}^{\pm,n+\frac12},S_{\theta_n}\Phi^{\pm,n})}(\delta\dot{\mathbf{V}}^{\pm,n},\delta\Phi^{\pm,n}) + \widetilde{e_{i,n}^{\pm}},\end{aligned} \tag{77}$$

where

$$\widetilde{e_{i,n}^{\pm}} = \sum_{k=1}^{4} \widetilde{e_{i,n}^{\pm,k}} \tag{78}$$

with the errors:

$$\begin{cases} \widetilde{e_{i,n}^{\pm,1}} = \mathscr{B}_i^{\pm}(\mathbf{V}^{\pm,n+1},\Phi^{\pm,n+1}) - \mathscr{B}_i^{\pm}(\mathbf{V}^{\pm,n},\Phi^{\pm,n}) \\ \qquad - \mathscr{B}^{\pm\prime}_{i,(\mathbf{V}^{\pm,n},\Phi^{\pm,n})}(\delta\mathbf{V}^{\pm,n},\delta\Phi^{\pm,n}) \\ \qquad = \begin{cases} (\delta\Phi^{\pm,n})_{x_2}\delta V_3^{\pm,n} + (\delta\Phi^{\pm,n})_{x_3}\delta V_4^{\pm,n}, & i=1,\\ -(\delta\Phi^{\pm,n})_{x_2}\delta V_6^{\pm,n} - (\delta\Phi^{\pm,n})_{x_3}\delta V_7^{\pm,n}, & i=3,\end{cases} \\ \widetilde{e_{i,n}^{\pm,2}} = \mathscr{B}^{\pm\prime}_{i,(\mathbf{V}^{\pm,n},\Phi^{\pm,n})}(\delta\mathbf{V}^{\pm,n},\delta\Phi^{\pm,n}) - \mathscr{B}^{\pm\prime}_{i,(S_{\theta_n}\mathbf{V}^{\pm,n},S_{\theta_n}\Phi^{\pm,n})}(\delta\mathbf{V}^{\pm,n},\delta\Phi^{\pm,n}) \\ \qquad = \begin{cases} \big((I-S_{\theta_n})\Phi^{\pm,n}\big)_{x_2}\delta V_3^{\pm,n} + \big((I-S_{\theta_n})\Phi^{\pm,n}\big)_{x_3}\delta V_4^{\pm,n} \\ \quad +(\delta\Phi^{\pm,n})_{x_2}(I-S_{\theta_n})V_3^{\pm,n} + (\delta\Phi^{\pm,n})_{x_3}(I-S_{\theta_n})V_4^{\pm,n}, & i=1,\\ -\big((I-S_{\theta_n})\Phi^{\pm,n}\big)_{x_2}\delta V_6^{\pm,n} - \big((I-S_{\theta_n})\Phi^{\pm,n}\big)_{x_3}\delta V_7^{\pm,n} \\ \quad -(\delta\Phi^{\pm,n})_{x_2}(I-S_{\theta_n})V_6^{\pm,n} - (\delta\Phi^{\pm,n})_{x_3}(I-S_{\theta_n})V_7^{\pm,n}, & i=3,\end{cases} \end{cases} \tag{79}$$

and

$$\begin{cases} \widetilde{e_{i,n}^{\pm,3}} = \mathscr{B}^{\pm\prime}_{i,(S_{\theta_n}\mathbf{V}^{\pm,n},S_{\theta_n}\Phi^{\pm,n})}(\delta\mathbf{V}^{\pm,n},\delta\Phi^{\pm,n}) - \mathscr{B}^{\pm\prime}_{i,(\mathbf{V}^{\pm,n+\frac12},S_{\theta_n}\Phi^{\pm,n})}(\delta\mathbf{V}^{\pm,n},\delta\Phi^{\pm,n}) \\ \qquad = \begin{cases} (S_{\theta_n}V_3^{\pm,n} - V_3^{\pm,n+\frac12})(\delta\Phi^{\pm,n})_{x_2} \\ \quad + (S_{\theta_n}V_4^{\pm,n} - V_4^{\pm,n+\frac12})(\delta\Phi^{\pm,n})_{x_3}, & i=1,\\ (V_6^{\pm,n+\frac12} - S_{\theta_n}V_6^{\pm,n})(\delta\Phi^{\pm,n})_{x_2} \\ \quad + (V_7^{\pm,n+\frac12} - S_{\theta_n}V_7^{\pm,n})(\delta\Phi^{\pm,n})_{x_3}, & i=3,\end{cases} \\ \widetilde{e_{i,n}^{\pm,4}} = \mathscr{B}^{\pm\prime}_{i,(\mathbf{V}^{\pm,n+\frac12},S_{\theta_n}\Phi^{\pm,n})}(\delta\mathbf{V}^{\pm,n},\delta\Phi^{\pm,n}) - \mathscr{B}^{\pm\prime}_{i,(\mathbf{V}^{\pm,n+\frac12},S_{\theta_n}\Phi^{\pm,n})}(\delta\dot{\mathbf{V}}^{\pm,n},\delta\Phi^{\pm,n}) \\ \qquad = \begin{cases} \frac{\delta\Phi^{\pm,n}}{(\Psi_a^{\pm}+S_{\theta_n}\Phi^{\pm,n})_{x_1}} E_T\big(\partial_{x_1}\mathscr{B}_1^{\pm}(\mathbf{V}^{\pm,n+\frac12},S_{\theta_n}\phi_n)\big), & i=1,\\ \frac{\delta\Phi^{\pm,n}}{(\Psi_a^{\pm}+S_{\theta_n}\Phi^{\pm,n})_{x_1}} E_T\big(\partial_{x_1}\mathscr{B}_3^{\pm}(\mathbf{V}^{\pm,n+\frac12},S_{\theta_n}\phi_n)\big), & i=3,\end{cases} \end{cases} \tag{80}$$

for $E_T(\cdot)$ being a proper bounded extension from $H^s(b\Omega_T)$ to $B^{s+1}(\Omega_T)$.

Using (77) and noting that $\mathscr{B}_i^{\pm}(\mathbf{V}^{\pm,0}, \Phi^{\pm,0}) = 0$, we obtain

$$\mathscr{B}_i^{\pm}(\mathbf{V}^{\pm,n+1}, \Phi^{\pm,n+1}) = \sum_{j=0}^{n} \left(\mathscr{B}^{\pm\prime}_{i,(\mathbf{V}^{\pm,j+\frac{1}{2}}, S_{\theta_j}\Phi^{\pm,j})}(\delta\dot{\mathbf{V}}^{\pm,j}, \delta\Phi^{\pm,j}) + \widetilde{e_{i,j}^{\pm}} \right) \tag{81}$$

for $i = 1, 3$.

Similarly, one has

$$\mathscr{B}_2(\mathbf{V}^{+,n+1}, \mathbf{V}^{-,n+1}) = \sum_{j=0}^{n} \left(\mathscr{B}'_{2,(\mathbf{V}^{+,j+\frac{1}{2}}, \mathbf{V}^{-,j+\frac{1}{2}})}(\delta\dot{\mathbf{V}}^{+,j}, \delta\dot{\mathbf{V}}^{-,j}) + \widetilde{e_{2,j}} \right), \tag{82}$$

where the errors $\widetilde{e_{2,j}}$ can be defined as that of $\widetilde{e_{i,j}^{\pm}}$ in (78)–(80) with $\mathscr{B}_i^{\pm}$ being replaced by $\mathscr{B}_2$.

Observe that, if the limit of $(\mathbf{V}^{\pm,n}, \Phi^{\pm,n})$ exists which is expected to be a solution of (62), the left-hand sides of Eqs. (70) and (81)–(82) should tend to $\mathbf{f}_a^{\pm}$ and zero respectively when $n \to \infty$. Thus, with respect to the well-posed boundary condition form of the linear problem (48), we define the modified increments $\delta\dot{\mathbf{V}}^{\pm,n}$ of the approximate solutions to be the solutions to the following problem:

$$\begin{cases} \mathbf{L}'_{e,(\mathbf{U}_a^{\pm}+\mathbf{V}^{\pm,n+\frac{1}{2}}, \Psi_a^{\pm}+S_{\theta_n}\Phi^{\pm,n})} \delta\dot{\mathbf{V}}^{\pm,n} = \mathbf{f}_n^{\pm}, & \\ B_1(\delta\dot{\mathbf{V}}^{+,n}, \delta\dot{\mathbf{V}}^{-,n}) = h_{1,n}^{+} - h_{1,n}^{-} + \lambda^{+}(\mathbf{U}_a^{\pm} + \mathbf{V}^{\pm,n+\frac{1}{2}}) h_{3,n}^{+} & \\ \qquad - \lambda^{-}(\mathbf{U}_a^{\pm} + \mathbf{V}^{\pm,n+\frac{1}{2}}) h_{3,n}^{-} & \text{on } b\Omega_T, \\ \mathscr{B}'_{2,(\mathbf{V}^{+,n+\frac{1}{2}}, \mathbf{V}^{-,n+\frac{1}{2}})}(\delta\dot{\mathbf{V}}^{+,n}, \delta\dot{\mathbf{V}}^{-,n}) = \tilde{g}_n & \text{on } b\Omega_T, \end{cases} \tag{83}$$

where

$$\begin{aligned} & B_1(\delta\dot{\mathbf{V}}^{+,n}, \delta\dot{\mathbf{V}}^{-,n}) \\ & \quad = \lambda^{+}(\mathbf{U}_a^{\pm} + \mathbf{V}^{\pm,n+\frac{1}{2}}) \Big(\delta\dot{\mathbf{V}}_5^{+,n} - (\Psi_a^{+} + S_{\theta_n}\Phi^{+,n})_{x_2} \delta\dot{V}_6^{+,n} \\ & \qquad\qquad - (\Psi_a^{+} + S_{\theta_n}\Phi^{+,n})_{x_3} \delta\dot{V}_7^{+,n} \Big) \\ & \qquad - \Big(\delta\dot{V}_2^{+,n} - (\Psi_a^{+} + S_{\theta_n}\Phi^{+,n})_{x_2} \delta\dot{V}_3^{+,n} - (\Psi_a^{+} + S_{\theta_n}\Phi^{+,n})_{x_3} \delta\dot{V}_4^{+,n} \Big) \\ & \qquad - \lambda^{-}(\mathbf{U}_a^{\pm} + \mathbf{V}^{\pm,n+\frac{1}{2}}) \Big(\delta\dot{V}_5^{-,n} - (\Psi_a^{-} + S_{\theta_n}\Phi^{-,n})_{x_2} \delta\dot{V}_6^{-,n} \\ & \qquad\qquad - (\Psi_a^{-} + S_{\theta_n}\Phi^{-,n})_{x_3} \delta\dot{V}_7^{-,n} \Big) \\ & \qquad + \Big(\delta\dot{V}_2^{-,n} - (\Psi_a^{-} + S_{\theta_n}\Phi^{-,n})_{x_2} \delta\dot{V}_3^{-,n} - (\Psi_a^{-} + S_{\theta_n}\Phi^{-,n})_{x_3} \delta\dot{V}_4^{-,n} \Big) \end{aligned}$$

with $\lambda^{\pm}(\cdot)$ being defined in (47), $\mathbf{f}_n^{\pm}$, $\tilde{g}_n$, $h_{1,n}^{\pm}$, and $h_{3,n}^{\pm}$ are defined by

$$\begin{cases} \displaystyle\sum_{j=0}^{n} \mathbf{f}_j^{\pm} + S_{\theta_n}\left(\sum_{j=0}^{n-1} \mathbf{e}_{\pm,j} \right) = S_{\theta_n}\mathbf{f}_a^{\pm}, \qquad \sum_{j=0}^{n} \tilde{g}_j + S_{\theta_n}\left(\sum_{j=0}^{n-1} \widetilde{e_{2,j}} \right) = 0, \\ \displaystyle\sum_{j=0}^{n} h_{1,j}^{\pm} + S_{\theta_n}\left(\sum_{j=0}^{n-1} \widetilde{e_{1,j}^{\pm}} \right) = 0, \qquad \sum_{j=0}^{n} h_{3,j}^{\pm} + S_{\theta_n}\left(\sum_{j=0}^{n-1} \widetilde{e_{3,j}^{\pm}} \right) = 0, \end{cases} \tag{84}$$

by induction on n, with $\mathbf{f}_0^\pm = S_{\theta_0}\mathbf{f}_a^\pm$ and $\tilde{g}_0 = h_{1,0}^\pm = h_{3,0}^\pm = 0$.

To construct $\delta\Phi^{\pm,n}$ from $\mathscr{B}_1^\pm(\mathbf{V}^\pm, \Phi^\pm)$ defined in (76), we clearly have

$$\begin{aligned}\mathscr{B}^{\pm\prime}_{1,(\mathbf{V}^\pm,\Phi^\pm)}(\mathbf{W}^\pm, \Theta^\pm) = \partial_t\Theta^\pm + (U_{a,3}^\pm + V_3^\pm)\partial_{x_2}\Theta^\pm + (U_{a,4}^\pm + V_4^\pm)\partial_{x_3}\Theta^\pm \\ - W_2^\pm + (\Psi_a^\pm + \Phi^\pm)_{x_2} W_3^\pm + (\Psi_a^\pm + \Phi^\pm)_{x_3} W_4^\pm.\end{aligned} \tag{85}$$

From (81) with $i = 1$, we first define $\delta\Phi^{+,n}$ by the following problem:

$$\begin{cases}\mathscr{B}^{+\prime}_{1,(\mathbf{V}^{+,n+\frac12}, S_{\theta_n}\Phi^{+,n})}(\delta\dot{\mathbf{V}}^{+,n}, \delta\Phi^{+,n}) = h_{1,n}^+ \quad \text{in } \Omega_T, \\ \delta\Phi^{+,n}|_{t\le 0} = 0,\end{cases} \tag{86}$$

where $h_{1,n}^+$ is given in (84). Denote by

$$\begin{aligned}\widetilde{h_{3,n}^\pm} = \delta\dot{V}_5^{+,n} - (\Psi_a^+ + S_{\theta_n}\Phi^{+,n})_{x_2}\delta\dot{V}_6^{+,n} - (\Psi_a^+ + S_{\theta_n}\Phi^{+,n})_{x_3}\delta\dot{V}_7^{+,n} \\ - \big((\delta\Phi^{+,n})_{x_2}, (\delta\Phi^{+,n})_{x_3}\big)\begin{pmatrix} U_{a,6}^\pm + V_6^{\pm,n+\frac12} \\ U_{a,7}^\pm + V_7^{\pm,n+\frac12}\end{pmatrix},\end{aligned} \tag{87}$$

and

$$\begin{aligned}\widetilde{h_{1,n}^-} = h_{1,n}^+ + \lambda^+(\mathbf{U}_a^\pm + \mathbf{V}^{\pm,n+\frac12})\widetilde{h_{3,n}^+} \\ - \lambda^-(\mathbf{U}_a^\pm + \mathbf{V}^{\pm,n+\frac12})\widetilde{h_{3,n}^-} - B_1(\delta\dot{\mathbf{V}}^{+,n}, \delta\dot{\mathbf{V}}^{-,n}).\end{aligned}$$

Then we determine $\delta\Phi^{-,n}$ by solving the following problem:

$$\begin{cases}\mathscr{B}^{-\prime}_{1,(\mathbf{V}^{-,n+\frac12}, S_{\theta_n}\Phi^{-,n})}(\delta\dot{\mathbf{V}}^{-,n}, \delta\Phi^{-,n}) = \widetilde{h_{1,n}^-} \quad \text{in } \Omega_T, \\ \delta\Phi^{-,n}|_{t\le 0} = 0.\end{cases} \tag{88}$$

By employing Proposition 1 for problems (86) and (88), we obtain

$$\delta\Phi^{+,n} = \delta\Phi^{-,n} \quad \text{on } \{x_1 = 0\}.$$

In order to keep the boundary $\{x_1 = 0\}$ being a uniform characteristic plane of constant multiplicity at each iteration step (83), we define the modified state $\mathbf{V}^{\pm,n+\frac12}$ by requiring

$$\begin{cases}(\partial_t + U_{a,3}^\pm\partial_{x_2} + U_{a,4}^\pm\partial_{x_3})(S_{\theta_n}\Phi^{\pm,n}) - V_2^{\pm,n+\frac12} \\ \quad + (\Psi_a^\pm + S_{\theta_n}\Phi^{\pm,n})_{x_2} V_3^{\pm,n+\frac12} + (\Psi_a^\pm + S_{\theta_n}\Phi^{\pm,n})_{x_3} V_4^{\pm,n+\frac12} = 0, \\ V_5^{\pm,n+\frac12} - (\Psi_a^\pm + S_{\theta_n}\Phi^{\pm,n})_{x_2} V_6^{\pm,n+\frac12} - (\Psi_a^\pm + S_{\theta_n}\Phi^{\pm,n})_{x_3} V_7^{\pm,n+\frac12} \\ \quad - (S_{\theta_n}\Phi^{\pm,n})_{x_2} U_{a,6}^\pm - (S_{\theta_n}\Phi^{\pm,n})_{x_3} U_{a,7}^\pm = 0\end{cases} \tag{89}$$

on $\{x_1 = 0\}$, which leads to define

$$V_j^{\pm,n+\frac12} = S_{\theta_n} V_j^{\pm,n} \quad \text{for } j \neq 2, 5, \tag{90}$$

and

$$\begin{cases} V_2^{\pm,n+\frac{1}{2}} = (\Psi_a^\pm + S_{\theta_n}\Phi^{\pm,n})_{x_2} S_{\theta_n} V_3^{\pm,n} + (\Psi_a^\pm + S_{\theta_n}\Phi^{\pm,n})_{x_3} S_{\theta_n} V_4^{\pm,n} \\ \qquad\qquad + (\partial_t + U_{a,3}^\pm \partial_{x_2} + U_{a,4}^\pm \partial_{x_3})(S_{\theta_n}\Phi^{\pm,n}), \\ V_5^{\pm,n+\frac{1}{2}} = (\Psi_a^\pm + S_{\theta_n}\Phi^{\pm,n})_{x_2} S_{\theta_n} V_6^{\pm,n} + (\Psi_a^\pm + S_{\theta_n}\Phi^{\pm,n})_{x_3} S_{\theta_n} V_7^{\pm,n} \\ \qquad\qquad + U_{a,6}^\pm (S_{\theta_n}\Phi^{\pm,n})_{x_2} + U_{a,7}^\pm (S_{\theta_n}\Phi^{\pm,n})_{x_3}. \end{cases} \tag{91}$$

The steps for determining $(\delta \mathbf{V}^{\pm,n}, \delta\Phi^{\pm,n})$ are to solve first $\delta\dot{\mathbf{V}}^{\pm,n}$ from (83) and then $\delta\Phi^{\pm,n}$ from (86) and (88), and to obtain $\delta\mathbf{V}^{\pm,n}$ finally from (69).

5.3 Convergence of the Iteration Scheme

Fix any $s_0 \geq 9$, $\alpha \geq s_0 + 6$, and $s_1 \in [\alpha + 7, 2\alpha + 4 - s_0]$. Let the zero-th order approximate solutions for the initial data $(\mathbf{U}_0^\pm, \psi_0)$ constructed in Sect. 5.1 satisfy

$$\|\dot{\mathbf{U}}_a^\pm\|_{s_1+3,T} + \|\dot{\Psi}_a^\pm\|_{s_1+3,T} + \|\mathbf{f}_a^\pm\|_{s_1-4,T} \leq \varepsilon, \qquad \|\mathbf{f}_a^\pm\|_{\alpha+3,T}/\varepsilon \quad \text{is small}, \tag{92}$$

for some small constant $\varepsilon > 0$, with $\dot{\Psi}_a^\pm = \chi(\pm x_1)\psi_a(t, x_2, x_3)$.

The key estimates for proving the convergence of the iteration scheme are as follows:

Proposition 2 *For the solution sequence* $(\delta\mathbf{V}^{\pm n}, \delta\Phi^{\pm,n})$ *given by* (83) *and* (86)–(88), *we have*

$$\begin{cases} \|\delta\mathbf{V}^{\pm,n}\|_{s,T} + \|\delta\Phi^{\pm,n}\|_{s,T} \leq \varepsilon\theta_n^{s-\alpha-2}\Delta_n & \text{for } s \in [s_0, s_1], \\ \|L(\mathbf{V}^{\pm,n}, \Phi^{\pm,n})\mathbf{V}^{\pm,n} - \mathbf{f}_a^\pm\|_{s,T} \leq 2\varepsilon\theta_n^{s-\alpha-3} & \text{for } s \in [s_0, s_1 - 4], \\ \|\mathscr{B}_1^\pm(\mathbf{V}^{\pm,n}, \Phi^{\pm,n})\|_{s,T} \leq 2\varepsilon\theta_n^{s-\alpha-3} & \text{for } s \in [s_0, s_1 - 4], \\ |\mathscr{B}_2(\mathbf{V}^{+,n}, \mathbf{V}^{-,n})|_{s-1,T} \leq \varepsilon\theta_n^{s-\alpha-3} & \text{for } s \in [s_0, s_1 - 2] \end{cases} \tag{93}$$

for any $n \geq 0$, *where* $\Delta_n = \theta_{n+1} - \theta_n$.

This proposition is obtained by induction on $n \geq 0$. Suppose that estimates (93) hold for all $0 \leq n \leq m - 1$. From the definition of $(\widetilde{e_{i,n}^{\pm,1}}, \widetilde{e_{i,n}^{\pm,2}})$ $(i = 1, 3)$ given in (79), we conclude

$$\begin{cases} \|\widetilde{e_{1,n}^{\pm,1}}\|_{s,T} + \|\widetilde{e_{3,n}^{\pm,1}}\|_{s,T} \leq C\varepsilon^2\theta_n^{s+s_0-2\alpha-4}\Delta_n, \\ \|\widetilde{e_{1,n}^{\pm,2}}\|_{s,T} + \|\widetilde{e_{3,n}^{\pm,2}}\|_{s,T} \leq C\varepsilon^2\theta_n^{s+s_0-2\alpha-2}\Delta_n \end{cases} \tag{94}$$

for all $0 \leq n \leq m - 1$ and $s \in [s_0, s_1 - 1]$.

From (91), we have

$$\begin{aligned} V_2^{\pm,n+\frac{1}{2}} - V_2^{\pm,n} &= (\partial_t + U_{a,3}^\pm\partial_{x_2} + U_{a,4}^\pm\partial_{x_3})(S_{\theta_n} - I)\Phi^{\pm,n} \\ &\quad + (\Psi_a^\pm + S_{\theta_n}\Phi^{\pm,n})_{x_2}(S_{\theta_n} - I)V_3^{\pm,n} \end{aligned}$$

$$+ (\Psi_a^{\pm} + S_{\theta_n}\Phi^{\pm,n})_{x_3}(S_{\theta_n} - I)V_4^{\pm,n} + ((S_{\theta_n} - I)\Phi^{\pm,n})_{x_2}V_3^{\pm,n}$$
$$+ ((S_{\theta_n} - I)\Phi^{\pm,n})_{x_3}V_4^{\pm,n} + \mathscr{B}_1^{\pm}(\mathbf{V}^{\pm,n}, \Phi^{\pm,n}),$$

which implies the estimate:

$$\|V_2^{\pm,n+\frac{1}{2}} - V_2^{\pm,n}\|_{s,T} \le C\varepsilon\theta_n^{s-\alpha}$$

for all $s_0 \le s \le s_1 - 4$ and $0 \le n \le m-1$.

Similarly, one has

$$\|V_5^{\pm,n+\frac{1}{2}} - V_5^{\pm,n}\|_{s,T} \le C\varepsilon\theta_n^{s-\alpha}$$

holding for all $s_0 \le s \le s_1 - 4$ and $0 \le n \le m-1$.

Therefore, we obtain

$$\begin{aligned}
&|\mathscr{B}_i^{\pm}(\mathbf{V}^{\pm,n+\frac{1}{2}}, S_{\theta_n}\phi_n)|_{s,T} \\
&\le |\mathscr{B}_i^{\pm}(\mathbf{V}^{\pm,n+\frac{1}{2}}, S_{\theta_n}\phi_n) - \mathscr{B}_i^{\pm}(V^{\pm,n}, \phi_n)|_{s,T} + |\mathscr{B}_i^{\pm}(V^{\pm,n}, \phi_n)|_{s,T} \\
&\le C\varepsilon\theta_n^{s+1-\alpha}
\end{aligned}$$

for $s_0 - 1 \le s \le s_1 - 5$, $0 \le n \le m-1$, and $i = 1, 3$, by using an estimate of $\mathscr{B}_3^{\pm}(\mathbf{V}^{\pm,n}, \Phi^{\pm,n})$ similar to (93) derived from problem (83) (cf. [38]).

Thus, from (80), we deduce

$$\|\widetilde{e_{1,n}^{\pm,4}}\|_{s,T} + \|\widetilde{e_{3,n}^{\pm,4}}\|_{s,T} \le C\varepsilon^2\theta_n^{s+s_0-2\alpha}\Delta_n \tag{95}$$

for all $0 \le n \le m-1$ and $s \in [s_0, s_1 - 6]$.

Combining (94) with (95), it follows that

$$\|\widetilde{e_{1,n}^{\pm}}\|_{s,T} + \|\widetilde{e_{3,n}^{\pm}}\|_{s,T} \le C\varepsilon^2\theta_n^{s+s_0-2\alpha}\Delta_n \tag{96}$$

for all $0 \le n \le m-1$ and $s \in [s_0, s_1 - 6]$.

From (84), one immediately deduces

$$h_{i,m}^{\pm} = (S_{\theta_{m-1}} - S_{\theta_m})\widetilde{E_{i,m-1}^{\pm}} + S_{\theta_m}\widetilde{e_{i,m-1}^{\pm}}$$

with $\widetilde{E_{i,m-1}^{\pm}} = \sum_{n=0}^{m-2}\widetilde{e_{i,n}^{\pm}}$ for $i = 1, 3$, which implies

$$|h_{1,m}^{\pm}|_{s,T} + |h_{3,m}^{\pm}|_{s,T} \le C\varepsilon^2\theta_m^{s+s_0-2\alpha}\Delta_m \tag{97}$$

for all $s \ge s_0$.

On the other hand, for $\mathbf{f}_m^{\pm}$ and $\tilde{g}_m$ given in (83), as in [17], we have

$$\|\mathbf{f}_m^{\pm}\|_{s,T} + |\tilde{g}_m|_{s+1,T} \le C(\varepsilon\delta + \varepsilon^2)\theta_m^{s-\alpha-3}\Delta_m$$

for all $s \ge s_0$, with $\delta = \|\mathbf{f}_a^{\pm}\|_{\alpha+2,T}/\varepsilon$ being small.

Applying Corollary 1 for problem (83) with $n = m$ and using the above estimate and (97), we find

$$\|\delta\dot{\mathbf{V}}^{\pm,m}\|_{s,T} \le C(\varepsilon\delta + \varepsilon^2)\theta_m^{s-\alpha-3}\Delta_m \tag{98}$$

for all $s \ge s_0$, by noting $\alpha \ge s_0 + 5$.

By applying a classical estimate for problem (86) of $\delta\Phi^{+,m}$ and using (98), it follows that

$$\|\delta\Phi^{+,m}\|_{s,T} \leq CT(\varepsilon\delta+\varepsilon^2)\theta_m^{s-\alpha-3}\Delta_m \tag{99}$$

for all $s \geq s_0$.

Thus, from (87), we have

$$\|\widetilde{h_{3,m}^{\pm}}\|_{s,T} \leq C(\varepsilon\delta+\varepsilon^2)\theta_m^{s-\alpha-2}\Delta_m \tag{100}$$

for all $s \geq s_0$.

For the function $\widetilde{h_{1,m}^{-}}$ given in (88), it is easy to have

$$\begin{aligned}\widetilde{h_{1,m}^{-}} = h_{1,m}^{-} &+ \lambda^{+}(\mathbf{U}_a^{\pm}+\mathbf{V}^{\pm,m+\frac{1}{2}})(\widetilde{h_{3,m}^{+}}-h_{3,m}^{+})\\ &-\lambda^{-}(\mathbf{U}_a^{\pm}+\mathbf{V}^{\pm,m+\frac{1}{2}})(\widetilde{h_{3,m}^{-}}-h_{3,m}^{-}),\end{aligned}$$

which implies

$$\|\widetilde{h_{1,m}^{-}}\|_{s,T} \leq C(\varepsilon\delta+\varepsilon^2)\theta_m^{s-\alpha-2}\Delta_m \tag{101}$$

for all $s \geq s_0$.

Applying the classical estimate again for problem (88) and using (101), we have

$$\|\delta\Phi^{-,m}\|_{s,T} \leq CT(\varepsilon\delta+\varepsilon^2)\theta_m^{s-\alpha-2}\Delta_m \tag{102}$$

for all $s \geq s_0$.

Thus, from (98), (99), and (102), we have

$$\|\delta\mathbf{V}^{\pm,m}\|_{s,T} \leq C(\varepsilon\delta+\varepsilon^2)\theta_m^{s-\alpha-2}\Delta_m \tag{103}$$

for all $s \geq s_0$.

From (99) and (102)–(103), we immediately obtain the estimates of $\|\delta\mathbf{V}^{\pm,m}\|_{s,T}$ and $\|\delta\Phi^{\pm,m}\|_{s,T}$ given in (93) by choosing $\delta = \|\mathbf{f}_a^{\pm}\|_{\alpha+2,T}/\varepsilon$ small. The remaining estimates of (93) can be verified directly, and the details can be found in [17].

Convergence of the Iteration Scheme From the first result of (93), we have

$$\sum_{n\geq 0}\|(\delta\mathbf{V}^{\pm,n},\delta\Phi^{\pm,n})\|_{\alpha,T} < \infty, \tag{104}$$

which implies that there exist $(\mathbf{V}^{\pm},\Phi^{\pm}) \in B^{\alpha}(\Omega_T)$ such that

$$(\mathbf{V}^{\pm,n},\Phi^{\pm,n}) \longrightarrow (\mathbf{V}^{\pm},\Phi^{\pm}) \quad \text{in } B^{\alpha}(\Omega_T). \tag{105}$$

From the other results given in (93), we obtain that the limit functions $(\mathbf{V}^{\pm},\Phi^{\pm})$ satisfy

$$\begin{cases}\vec{\mathscr{L}}(\mathbf{V}^{\pm},\Phi^{\pm})\mathbf{V}^{\pm} = \mathbf{f}_a^{\pm} & \text{in } \Omega_T,\\ \mathscr{B}_1^{\pm}(\mathbf{V}^{\pm},\Phi^{\pm}) = 0 & \text{in } \Omega_T,\\ \mathscr{B}_2(\mathbf{V}^{+},\mathbf{V}^{-}) = 0 & \text{on } b\Omega_T.\end{cases} \tag{106}$$

On the other hand, from the second result given in (106), we obtain that the constraint

$$\mathscr{B}_3^{\pm}(\mathbf{V}^{\pm}, \Phi^{\pm}) = 0$$

also holds on $b\Omega_T$ if it is true at $\{t = 0\}$, by using Lemma 1 in Sect. 3.

Note that $\delta\Phi^{+,n} = \delta\Phi^{-,n}$ for all n immediately imply $\Phi^+ = \Phi^-$ on $\{x_1 = 0\}$ as well. Thus the second result given in (106) leads to one of the Rankine–Hugoniot conditions:

$$v_N^+ = v_N^- \quad \text{on } \Gamma$$

given in (26).

Therefore, we conclude

Theorem 3 *Let* $\alpha \geq 15$ *and* $s_1 \in [\alpha + 7, 2\alpha - 5]$. *Let* $\psi_0 \in H^{2s_1+3}(\mathbb{R}^2)$ *and* $\mathbf{U}_0^{\pm} - \bar{\mathbf{U}}^{\pm} \in B^{2(s_1+2)}(\mathbb{R}_+^3)$ *satisfy the compatibility conditions of problem* (36) *up to order* $s_1 + 2$, *and let conditions* (25)–(26) *and* (92) *be satisfied. Then there exists a solution* $\mathbf{V}^{\pm} \in B^{\alpha}(\Omega_T)$, $\phi \in H^{\alpha-1}(b\Omega_T)$ *to problem* (62).

Then Theorem 1 in Sect. 3 directly follows from Theorem 3.

6 Concluding Remarks and Open Problems

Characteristic discontinuities (compressible vortex sheets and entropy waves), along with shock and rarefaction waves, occur ubiquitously in nature and are fundamental waves in the entropy solutions to hyperbolic systems of conservation laws in several space variables. The stability problems for characteristic discontinuities are fundamental, especially in shock reflection-diffraction and various wave interactions. Their mathematical rigorous treatments are truly challenging. What we have known is still very limited. Most of problems involving characteristic discontinuities are longstanding and still open. In particular, the following problems have not well understood, which deserve our attention:

1 As discussed in Sect. 2.2, another kind of characteristic discontinuities for the two-dimensional full Euler equations in gas dynamics is entropy waves. Similarly, they occur in the higher dimensional situations. It would be interesting to analyze entropy waves to explore new phenomena and features of these waves in two-dimensions and even higher dimensions.

2 In Sects. 3–5, we have shown the stability of current-vortex sheets when the jump of the tangential velocity is dominated by the jump of the non-paralleled tangential magnetic fields in the sense that $\lambda^{\pm}$ determined in (39) satisfy condition (38); also see Remark 1. The next concern is the stability/instability issue of current-vortex sheets in three-dimensional MHD when the jump of the tangential velocity is not dominated by the jump of the tangential magnetic fields, especially when the magnetic fields are parallel to each other on both sides of the front. As a special example, Wang–Yu [42] recently obtained a stability criterion on the current-vortex sheets in two-dimensional MHD, in which the tangential magnetic fields are parallel to each other always on both sides of the front: This stability criterion shows that there is certain stabilization effect of the magnetic fields in this case.

3 From the Rankine–Hugoniot conditions in (24) with $m_N = 0$ on Γ, besides the case (25)–(26) for the current-vortex sheets, there is another kind of characteristic discontinuities on which

$$\psi_t = v_N^+ = v_N^-, \qquad H_N^+ = H_N^- \neq 0 \quad \text{on } \Gamma,$$

which implies

$$[\mathbf{v}_\tau] = [\mathbf{H}_\tau] = 0,$$

that is,

$$[\mathbf{H}] = [\mathbf{v}] = 0, \qquad [p] = 0,$$

but

$$[S] \neq 0 \quad \text{equivalently} \quad [\rho] \neq 0.$$

Such a wave is called a current-entropy wave (an Alfvén wave). It is important to understand the stability/instability of current-entropy waves in three-dimensional MHD.

4 There are other different characteristic/noncharacteristic discontinuities in MHD; see Blokhin–Trakhinin [10], Trakhinin [39], and the references cited therein. It would be interesting to study these discontinuities and related problems in MHD and explore their new phenomena/features.

5 For the Euler equations in gas dynamics, it has been shown in Chen–Zhang–Zhu [16] and Chen–Kukreja [15] that two-dimensional steady-state vortex sheets are always stable under the two-dimensional steady perturbations of the incoming supersonic fluid flow. For shock reflection-diffraction problems, the solutions are self-similar, and most of Mach reflection-diffraction configurations involve a vorticity wave formed by a vortex sheet. It is important to understand the compressible vortex sheets for the Euler equations in the self-similar coordinates. In particular, when a vortex sheet forms a vorticity wave, it is useful to understand to which spaces of functions the solutions of the vorticity waves belong.

6 Another important direction is to analyze various interaction between shock fronts and vortex sheets/entropy waves in multidimensional compressible fluid flows.

It would be interesting to explore possible nonlinear approaches to see whether the corresponding estimates of solutions have no derivative loss with respect to initial data for the problems addressed; also see Coutand–Shkoller [19, 20]. It is clear that the solution to these problems involving characteristic discontinuities requires further new mathematical ideas, techniques, and approaches, which will be also useful for solving other longstanding problems in nonlinear partial differential equations, especially various boundary value problems, free boundary problems, among others, in hyperbolic conservation laws.

Acknowledgements The research of Gui-Qiang Chen was supported in part by the National Science Foundation under Grants DMS-0935967, DMS-0807551, the UK EPSRC Science and Innovation award to the Oxford Centre for Nonlinear PDE (EP/E035027/1), the NSFC under a joint project Grant 10728101, and the Royal Society—Wolfson Research Merit Award (UK). The research of Ya-Guang Wang was supported in part by the National Science Foundation of China under Grants 10971134 and 11031001. The second author would like to express his gratitude to the Department of Mathematics of Northwestern University (USA) for the hospitality, where this work was initiated when he visited there during the Spring Quarter 2005. Both authors would like to thank Y. Trakhinin for the helpful discussion on this problem.

References

1. Alì, G., Hunter, J.K.: Nonlinear surface waves on a tangential discontinuity in magnetohydrodynamics. Q. Appl. Math. **61**, 451–474 (2003)
2. Alinhac, S.: Existence d'ondes de raréfaction pour des systèmes quasi-linéaires hyperboliques multidimensionnels. Commun. Partial Differ. Equ. **14**, 173–230 (1989)
3. Anile, A.M.: Relativistic Fluids and Magneto-Fluids. With Applications in Astrophysics and Plasma Physics. Cambridge University Press, Cambridge (1989)
4. Artola, M., Majda, A.: Nonlinear development of instability in supersonic vortex sheets, I: The basic kink modes. Physica D **28**, 253–281 (1987)
5. Artola, M., Majda, A.: Nonlinear development of instability in supersonic vortex sheets, II. SIAM J. Appl. Math. **49**, 1310–1349 (1989)
6. Artola, M., Majda, A.: Nonlinear kind modes for supersonic vortex sheets. Phys. Fluids A **1**, 583–596 (1989)
7. Bae, M., Chen, G.-Q., Feldman, M.: Regularity of solutions to regular shock reflection for potential flow. Invent. Math. **175**, 505–543 (2009)
8. Begelman, M.C., Blandford, R.D., Rees, M.J.: Theory of extragalactic radio sources. Rev. Mod. Phys. **56**, 255–351 (1984)
9. Ben-Dor, G.: Shock Wave Reflection Phenomena. Springer, New York (1991)
10. Blokhin, A., Trakhinin, Y.: Stability of strong discontinuities in fluids and MHD. In: Handbook of Mathematical Fluid Dynamics, I, pp. 545–652. North-Holland, Amsterdam (2002)
11. Chen, G.-Q., Feldman, M.: Multidimensional transonic shocks and free boundary problems for nonlinear equations of mixed type. J. Am. Math. Soc. **3**, 461–494 (2003)
12. Chen, G.-Q., Feldman, M.: Global solutions to shock reflection by large-angle wedges for potential flow. Ann. Math. **171**, 1019–1134 (2010)
13. Chen, G.-Q., Feldman, M.: Shock reflection-diffraction phenomena and multidimensional conservation laws. In: Hyperbolic Problems: Theory, Numerics and Applications, Part 1. Proc. Sympos. Appl. Math., vol. 67, pp. 25–51. AMS, Providence (2009)
14. Chen, G.-Q., Feldman, M.: Mathematics of shock reflection-diffraction and von Neumann's conjectures. Monograph. Preprint (July 2011)
15. Chen, G.-Q., Kukreja, V.: Existence and stability of compressible vortex sheets in steady supersonic Euler flows between two Lipschitz walls. Preprint (2011)
16. Chen, G.-Q., Zhang, Y., Zhu, D.: Stability of compressible vortex sheets in steady supersonic Euler flows over Lipschitz walls. SIAM J. Math. Anal. **38**, 1660–1693 (2006/2007)

17. Chen, G.-Q., Wang, Y.-G.: Existence and stability of compressible current-vortex sheets in three-dimensional magnetohydrodynamics. Arch. Ration. Mech. Anal. **187**, 369–408 (2008)
18. Chen, S.: On the initial-boundary value problems for quasilinear symmetric hyperbolic systems with characteristic boundary. Chin. Ann. Math. **3**, 222–232 (1982) (in Chinese)
19. Coutand, D., Shkoller, S.: Well-posedness in smooth function spaces for the moving-boundary 3-D compressible Euler equations in physical vacuum. arXiv:1003.4721 (2010)
20. Coutand, D., Shkoller, S.: Well-posedness of the free-surface incompressible Euler equations with or without surface tension. J. Am. Math. Soc. **20**(3), 829–930 (2007)
21. Coulombel, J.F., Secchi, P.: Stability of compressible vortex sheet in two space dimensions. Indiana Univ. Math. J. **53**, 941–1012 (2004)
22. Coulombel, J.F., Secchi, P.: Nonlinear compressible vortex sheets in two space dimensions. Ann. Sci. Éc. Norm. Supér. (4) **41**, 85–139 (2008)
23. Coulombel, J.F., Secchi, P.: Uniqueness of 2-D compressible vortex sheets. Commun. Pure Appl. Anal. **8**, 1439–1450 (2009)
24. Courant, R., Friedrichs, K.: Supersonic Flow and Shock Waves. Springer, New York (1948)
25. Fejer, J.A., Miles, J.W.: On the stability of a plane vortex sheet with respect to three-dimensional disturbances. J. Fluid Mech. **15**, 335–336 (1963)
26. Francheteau, J., Métivier, G.: Existence de chocs faibles pour des systèmes quasi-linéaires hyperboliques multidimensionnels. Astérisque **268**, 1–198 (2000)
27. Friedrichs, K.O., Lax, P.D.: Boundary value problems for first order operators. Commun. Pure Appl. Math. **18**, 355–388 (1965)
28. Glimm, J., Majda, A.: Multidimensional Hyperbolic Problems and Computations. The IMA Volumes in Mathematics and Its Applications, vol. 29. Springer, New York (1991)
29. Guès, O.: Problème mixte hyperbolique quasi-linéaire caractéristique. Commun. Partial Differ. Equ. **15**, 595–645 (1990)
30. Guès, O., Métivier, G., Williams, M., Zumbrun, K.: Existence and stability of multidimensional shock fronts in the vanishing viscosity limit. Arch. Ration. Mech. Anal. **175**, 151–244 (2005)
31. Hörmander, L.: The boundary problems of physical geodesy. Arch. Ration. Mech. Anal. **62**, 1–52 (1982)
32. Majda, A.: Compressible Fluid Flow and Systems of Conservation Laws in Several Space Variables. Springer, New York (1984)
33. Métivier, G.: Stability of multidimensional shocks. In: Advances in the Theory of Shock Waves. PNDEA, vol. 47, pp. 25–103. Birkhäuser, Boston (2001)
34. Miles, J.W.: On the reflection of sound at an interface of relative motion. J. Acoust. Soc. Am. **29**, 226–228 (1957)
35. Miles, J.W.: On the disturbed motion of a plane vortex sheet. J. Fluid Mech. **4**, 538–552 (1958)
36. Smart, L.L., Norman, M.L., Winkler, K.A.: Shocks, interfaces, and patterns in supersonic jets. Physica D **12**, 83–106 (1984)
37. Trakhinin, Y.: Existence of compressible current-vortex sheets: variable coefficients linear analysis. Arch. Ration. Mech. Anal. **177**, 331–366 (2005)
38. Trakhinin, Y.: The existence of current-vortex sheets in ideal compressible magnetohydrodynamics. Arch. Ration. Mech. Anal. **191**, 245–310 (2009)
39. Trakhinin, Y.: On the well-posedness of a linearized plasma-vacuum interface problem in ideal compressible MHD. J. Differ. Equ. **249**, 2577–2599 (2010)
40. Van Dyke, M.: An Album of Fluid Motion. The Parabolic Press, Stanford (1982)
41. von Neumann, J.: Collect Works, vol. 5. Pergamon, New York (1963)
42. Wang, Y.-G., Yu, F.: Stabilization effect of magnetic fields on two-dimensional compressible current-vortex sheets. Preprint (2011)

h-Principle and Rigidity for $C^{1,\alpha}$ Isometric Embeddings

Sergio Conti, Camillo De Lellis, and László Székelyhidi Jr.

Abstract In this paper we study the embedding of Riemannian manifolds in low codimension. The well-known result of Nash and Kuiper (Nash in Ann. Math. 60:383–396, 1954; Kuiper in Proc. Kon. Acad. Wet. Amsterdam A 58:545–556, 1955; Kuiper in Proc. Kon. Acad. Wet. Amsterdam A 58:683–689, 1955) says that any short embedding in codimension one can be uniformly approximated by C^1 isometric embeddings. This statement clearly cannot be true for C^2 embeddings in general, due to the classical rigidity in the Weyl problem. In fact Borisov extended the latter to embeddings of class $C^{1,\alpha}$ with $\alpha > 2/3$ in (Borisov in Vestn. Leningr. Univ. 14(13):20–26, 1959; Borisov in Vestn. Leningr. Univ. 15(19):127–129, 1960). On the other hand he announced in (Borisov in Doklady 163:869–871, 1965) that the Nash–Kuiper statement can be extended to local $C^{1,\alpha}$ embeddings with $\alpha < (1 + n + n^2)^{-1}$, where n is the dimension of the manifold, provided the metric is analytic. Subsequently a proof of the 2-dimensional case appeared in (Borisov in Sib. Mat. Zh. 45(1):25–61, 2004). In this paper we provide analytic proofs of all these statements, for general dimension and general metric.

1 Introduction

Let M^n be a smooth compact manifold of dimension $n \geq 2$, equipped with a Riemannian metric g. An isometric immersion of (M^n, g) into $\mathbb{R}^m$ is a map $u \in C^1(M^n; \mathbb{R}^m)$ such that the induced metric agrees with g. In local coordinates

S. Conti
Institut für Angewandte Mathematik, Universität Bonn, 53115 Bonn, Germany
e-mail: sergio.conti@uni-bonn.de

C. De Lellis (✉)
Institut für Mathematik, Universität Zürich, 8057 Zürich, Switzerland
e-mail: camillo.delellis@math.unizh.ch

L. Székelyhidi Jr.
Hausdorff Center for Mathematics, Universität Bonn, 53115 Bonn, Germany
e-mail: laszlo.szekelyhidi@hcm.uni-bonn.de

H. Holden, K.H. Karlsen (eds.), *Nonlinear Partial Differential Equations*,
Abel Symposia 7, DOI 10.1007/978-3-642-25361-4_5,

this amounts to the system

$$\partial_i u \cdot \partial_j u = g_{ij} \tag{1}$$

consisting of $n(n+1)/2$ equations in m unknowns. If in addition u is injective, it is an isometric embedding. Assume for the moment that $g \in C^\infty$. The two classical theorems concerning the solvability of this system are:

(A) if $m \geq (n+2)(n+3)/2$, then any short embedding can be uniformly approximated by isometric embeddings of class C^∞ (Nash [23], Gromov [16]);
(B) if $m \geq n+1$, then any short embedding can be uniformly approximated by isometric embeddings of class C^1 (Nash [22], Kuiper [20, 21]).

Recall that a short embedding is an injective map $u : M^n \to \mathbb{R}^m$ such that the metric induced on M by u is shorter than g. In coordinates this means that $(\partial_i u \cdot \partial_j u) \leq (g_{ij})$ in the sense of quadratic forms. Thus, (A) and (B) are not merely existence theorems, they show that there exists a huge (essentially C^0-dense) set of solutions. This type of abundance of solutions is a central aspect of Gromov's h-principle, for which the isometric embedding problem is a primary example (see [12, 16]).

Naively, this type of flexibility could be expected for high codimension as in (A), since then there are many more unknowns than equations in (1). The h-principle for C^1 isometric embeddings is on the other hand rather striking, especially when compared to the classical rigidity result concerning the Weyl problem: if (S^2, g) is a compact Riemannian surface with positive Gauss curvature and $u \in C^2$ is an isometric immersion into $\mathbb{R}^3$, then u is uniquely determined up to a rigid motion ([8, 17], see also [31] for a thorough discussion). Thus it is clear that isometric immersions have a completely different qualitative behavior at low and high regularity (i.e. below and above C^2).

This qualitative difference is further highlighted by the following optimal mapping properties in the case when m is allowed to be sufficiently high:

(C) if $g \in C^{l,\beta}$ with $l+\beta > 2$ and m is sufficiently large, then there exists a solution $u \in C^{l,\beta}$ (Nash [23], Jacobowitz [18]);
(D) if $g \in C^{l,\beta}$ with $0 < l+\beta < 2$ and m is sufficiently large, then there exists a solution $u \in C^{1,\alpha}$ with $\alpha < (l+\beta)/2$ (Källen [19]).

These results are optimal in the sense that in both cases there exists $g \in C^{l,\beta}$ to which no solution u has better regularity than stated.

The techniques are also different: whereas the proofs of (A) and (C) rely on the Nash–Moser implicit function theorem, the proofs of (B) and (D) involve an iteration technique called convex integration. This technique was developed by Gromov [15, 16] into a very powerful tool to prove the h-principle in a wide variety of geometric problems (see also [12, 33]). In general the regularity of solutions obtained using convex integration agrees with the highest derivatives appearing in the equations (see [32]). Thus, an interesting question raised in [16], p. 219 is how one could extend the methods to produce more regular solutions. Essentially the same question, in the case of isometric embeddings, is also mentioned in [34] (see Problem 27). For high codimension this is resolved in (D).

Our primary aim in this paper is to consider the low codimension case, i.e. when $m = n+1$. This range was first considered by Borisov. In [6] it was announced that if g is analytic, then the h-principle holds for local isometric embeddings $u \in C^{1,\alpha}$ for $\alpha < \frac{1}{1+n+n^2}$. A proof for the case $n = 2$ appeared in [7]. Our main result is to provide a proof of the h-principle in this range for g which is not necessarily analytic and general $n \geq 2$ (see Sect. 1.1 for precise statements). Moreover, at least for $l = 0$ and sufficiently small $\beta > 0$, we recover the optimal mapping range corresponding to (D). Thus, there seems to be a direct trade-off between codimension and regularity.

The novelty of our approach, compared to Borisov's, is that only a finite number of derivatives need to be controlled. This is achieved by introducing a smoothing operator in the iteration step, analogous to the device of Nash used to overcome the loss of derivative problem in [23]. A similar method was used by Källen in [19]. See Sect. 3 for an overview of the iteration procedure. In addition, the errors coming from the smoothing operator are controlled by using certain commutator estimates on convolutions. These estimates are in Sect. 2.

Concerning rigidity in the Weyl problem, it is known from the work of Pogorelov and Sabitov that

1. closed C^1 surfaces with positive Gauss curvature and bounded extrinsic curvature are convex (see [26]);
2. closed convex surfaces are rigid in the sense that isometric immersions are unique up to rigid motion [25];
3. a convex surface with metric $g \in C^{l,\beta}$ with $l \geq 2, 0 < \beta < 1$ and positive curvature is of class $C^{l,\beta}$ (see [26, 27]).

Thus, extending the rigidity in the Weyl problem to $C^{1,\alpha}$ isometric immersions can be reduced to showing that the image of the surface has bounded extrinsic curvature (for definitions see Sect. 7). Using geometric arguments, in a series of papers [1–5] Borisov proved that for $\alpha > 2/3$ the image of surfaces with positive Gauss curvature has indeed bounded extrinsic curvature. Consequently, rigidity holds in this range and in particular $2/3$ is an upper bound on the range of Hölder exponents that can be reached using convex integration.

Using the commutator estimates from Sect. 2, at the end of this paper (in Sect. 7) we provide a short and self-consistent analytic proof of this result.

1.1 The h-Principle for Small Exponents

In this subsection we state our main existence results for $C^{1,\alpha}$ isometric immersions. One is of local nature, whereas the second is global. Note that for the local result the exponent matches the one announced in [6]. In what follows, we denote by sym_n^+ the cone of positive definite symmetric $n \times n$ matrices. Moreover, given an immersion $u : M^n \to \mathbb{R}^m$, we denote by $u^\sharp e$ the pullback of the standard Euclidean metric through u, so that in local coordinates

$$(u^\sharp e)_{ij} = \partial_i u \cdot \partial_j u.$$

Finally, let

$$n_* = \frac{n(n+1)}{2}.$$

Theorem 1 (Local existence) *Let $n \in \mathbb{N}$ and $g_0 \in \mathrm{sym}_n^+$. There exists $r > 0$ such that the following holds for any smooth bounded open set $\Omega \subset \mathbb{R}^n$ and any Riemannian metric $g \in C^\beta(\overline{\Omega})$ with $\beta > 0$ and $\|g - g_0\|_{C^0} \leq r$. There exists a constant $\delta_0 > 0$ such that, if $u \in C^2(\overline{\Omega}; \mathbb{R}^{n+1})$ and α satisfy*

$$\|u^\sharp e - g\|_0 \leq \delta_0^2 \quad and \quad 0 < \alpha < \min\left\{\frac{1}{1+2n_*}, \frac{\beta}{2}\right\},$$

then there exists a map $v \in C^{1,\alpha}(\overline{\Omega}; \mathbb{R}^{n+1})$ with

$$v^\sharp e = g \quad and \quad \|v - u\|_{C^1} \leq C\|u^\sharp e - g\|_{C^0}^{1/2}.$$

Corollary 1 (Local h-principle) *Let $n, g_0, \Omega, g, \alpha$ be as in Theorem 1. Given any short map $u \in C^1(\overline{\Omega}; \mathbb{R}^{n+1})$ and any $\varepsilon > 0$ there exists an isometric immersion $v \in C^{1,\alpha}(\overline{\Omega}; \mathbb{R}^{n+1})$ with $\|u - v\|_{C^0} \leq \varepsilon$.*

Theorem 2 (Global existence) *Let M^n be a smooth, compact manifold with a Riemannian metric $g \in C^\beta(M)$ and let $m \geq n+1$. There is a constant $\delta_0 > 0$ such that, if $u \in C^2(M; \mathbb{R}^m)$ and α satisfy*

$$\|u^\sharp e - g\|_{C^0} \leq \delta_0^2 \quad and \quad 0 < \alpha < \min\left\{\frac{1}{1+2(n+1)n_*}, \frac{\beta}{2}\right\},$$

then there exists a map $v \in C^{1,\alpha}(M; \mathbb{R}^m)$ with

$$v^\sharp e = g \quad and \quad \|v - u\|_{C^1} \leq C\|u^\sharp e - g\|_{C^0}^{1/2}.$$

Corollary 2 (Global h-principle) *Let (M^n, g) and α be as in Theorem 2. Given any short map $u \in C^1(M; \mathbb{R}^m)$ with $m \geq n+1$ and any $\varepsilon > 0$ there exists an isometric immersion $v \in C^{1,\alpha}(M; \mathbb{R}^m)$ with $\|u - v\|_{C^0} \leq \varepsilon$.*

Remark 1 In both corollaries, if u is an embedding, then there exists a corresponding v which in addition is an embedding.

1.2 Rigidity for Large Exponents

The following is a crucial estimate on the metric pulled back by standard regularizations of a given map.

Proposition 1 (Quadratic estimate) *Let $\Omega \subset \mathbb{R}^n$ be an open set, $v \in C^{1,\alpha}(\Omega, \mathbb{R}^m)$ with $v^\sharp e \in C^2$ and $\varphi \in C_c^\infty(\mathbb{R}^n)$ a standard symmetric convolution kernel. Then, for every compact set $K \subset \Omega$,*

$$\|(v * \varphi_\ell)^\sharp e - v^\sharp e\|_{C^1(K)} = O(\ell^{2\alpha-1}). \tag{2}$$

In particular, fix a map u and a kernel φ satisfying the assumptions of the Proposition with $\alpha > 1/2$. Then the Christoffel symbols of $(v * \varphi_\ell)^\sharp e$ converge to those of $v^\sharp e$. This corresponds to the results of Borisov in [1, 2], and hints at the absence of h-principle for $C^{1,\frac{1}{2}+\varepsilon}$ immersions. Relying mainly on this estimate we can give a fairly short proof of Borisov's theorem:

Theorem 3 *Let (M^2, g) be a surface with C^2 metric and positive Gauss curvature, and let $u \in C^{1,\alpha}(M^2; \mathbb{R}^3)$ be an isometric immersion with $\alpha > 2/3$. Then $u(M)$ is a surface of bounded extrinsic curvature.*

This leads to the following corollaries, which follow from the work of Pogorelov and Sabitov.

Corollary 3 *Let (S^2, g) be a closed surface with $g \in C^2$ and positive Gauss curvature, and let $u \in C^{1,\alpha}(S^2; \mathbb{R}^3)$ be an isometric immersion with $\alpha > 2/3$. Then, $u(S^2)$ is the boundary of a bounded convex set and any two such images are congruent. In particular if the Gauss curvature is constant, then $u(S^2)$ is the boundary of a ball $B_r(x)$.*

Corollary 4 *Let $\Omega \subset \mathbb{R}^2$ be open and $g \in C^{2,\beta}$ a metric on Ω with positive Gauss curvature. Let $u \in C^{1,\alpha}(\Omega; \mathbb{R}^3)$ be an isometric immersion with $\alpha > 2/3$. Then $u(\Omega)$ is $C^{2,\beta}$ and locally uniformly convex (that is, for every $x \in \Omega$ there exists a neighborhood V such that $u(\Omega) \cap V$ is the graph of a $C^{2,\beta}$ function with positive definite second derivative).*

1.3 Connections to the Euler Equations

There is an interesting analogy between isometric immersions in low codimension (in particular the Weyl problem) and the incompressible Euler equations. In [10] a method, which is very closely related to convex integration, was introduced to construct highly irregular energy-dissipating solutions of the Euler equations. Being in conservation form, the "expected" regularity space for convex integration for the Euler equations should be C^0. This is still beyond reach, and in [10] a weak version of convex integration was applied instead, to produce solutions in L^∞ (see also [11] for a slightly better space) and, moreover, to show that a weak version of the h-principle holds.

Nevertheless, just like for isometric immersions, for the Euler equations there is particular interest to go beyond C^0: in [24] L. Onsager, motivated by the phenomenon of anomalous dissipation in turbulent flows, conjectured that there exist weak solutions of the Euler equations of class C^α with $\alpha < 1/3$ which dissipate energy, whereas for $\alpha > 1/3$ the energy is conserved. The latter was proved in [9, 13], but on the construction of energy-dissipating weak solutions nothing is known beyond L^∞ (for previous work see [28–30]). It should be mentioned that the critical exponent $1/3$ is very natural—it agrees with the scaling of the energy cascade predicted by Kolmogorov's theory of turbulence (see for instance [14]).

For the analogous problem for isometric immersions there does not seem to be a universally accepted critical exponent (cf. Problem 27 of [34]), even though $1/2$ seems likely (cf. Sect. 1.2 and the discussion in [7]). In fact, the regularization and the commutator estimates used in our proof of Proposition 1 and Theorem 3 have been inspired by (and are closely related to) the arguments of [9].

2 Estimates on Convolutions: Proof of Proposition 1

As usual, we denote the norm on the Hölder space $C^{k,\alpha}(\overline{\Omega})$ by

$$\|f\|_{k,\alpha} := \sup_{x\in\Omega}\sum_{|a|\le k}|\partial^a f(x)| + \sup_{x,y\in\Omega, x\neq y}\sum_{|a|=k}\frac{|\partial^a f(x)-\partial^a f(y)|}{|x-y|^\alpha}.$$

Here $k = 0, 1, 2, \dots$, $a = (a_1, \dots, a_n)$ is a multi-index with $|a| = a_1 + \dots + a_n$ and $\alpha \in [0, 1[$. For simplicity we will also use the abbreviation $\|f\|_k = \|f\|_{k,0}$ and $\|f\|_\alpha = \|f\|_{0,\alpha}$.

Recall the following interpolation inequalities for these norms:

$$\|f\|_{k,\alpha} \le C\|f\|_{k_1,\alpha_1}^{\lambda}\|f\|_{k_2,\alpha_2}^{1-\lambda},$$

where C depends on the various parameters, $0 < \lambda < 1$ and

$$k + \alpha = \lambda(k_1 + \alpha_1) + (1-\lambda)(k_2 + \alpha_2).$$

The following estimates are well known and play a fundamental role in both the constructions and the proof of rigidity.

Lemma 1 *Let $\varphi \in C_c^\infty(\mathbb{R}^n)$ be symmetric and such that $\int\varphi = 1$. Then for any $r, s \ge 0$ and $\alpha \in]0, 1]$ we have*

$$\|f * \varphi_\ell\|_{r+s} \le C\ell^{-s}\|f\|_r, \tag{3}$$

$$\|f - f*\varphi_\ell\|_r \le C\ell^2\|f\|_{r+2}, \tag{4}$$

$$\|(fg)*\varphi_\ell - (f*\varphi_\ell)(g*\varphi_\ell)\|_r \le C\ell^{2\alpha - r}\|f\|_\alpha\|g\|_\alpha. \tag{5}$$

Proof For any multi-indices a, b with $|a| = r$, $|b| = s$ we have $\partial^{a+b}(f*\varphi_\ell) = \partial^a f * \partial^b\varphi_\ell$, hence

$$|\partial^{a+b}(f*\varphi_\ell)| \le C_s\ell^{-s}\|f\|_r.$$

This proves (3).

Next, by considering the Taylor expansion of f at x we see that

$$f(x-y) - f(x) = f'(x)y + r_x(y),$$

where $\sup_x |r_x(y)| \le C|y|^2 \|f\|_2$. Moreover, since φ is symmetric,

$$\int \varphi_\ell(y) y \, dy = 0.$$

Thus,

$$|f - f * \varphi_\ell| = \left| \int \varphi_\ell(y)(f(x-y) - f(x))\, dy \right| \tag{6}$$

$$\le C\|f\|_2 \int \ell^{-n} \left|\varphi\left(\frac{y}{\ell}\right)\right| |y|^2 \, dy = C\ell^2 \|f\|_2. \tag{7}$$

This proves (4) for the case $r = 0$. To obtain the estimate for general r, repeat the same argument for the partial derivatives $\partial^a f$ with $|a| = r$.

For the proof of estimate (5) let a be any multi-index with $|a| = r$. By the product rule

$$\partial^a \big[\varphi_\ell * (fg) - (\varphi_\ell * f)(\varphi_\ell * g)\big] \tag{8}$$

$$= \partial^a \varphi_\ell * (fg) - \sum_{b \le a} \binom{a}{b} (\partial^b \varphi_\ell * f)(\partial^{a-b} \varphi_\ell * g) \tag{9}$$

$$= \partial^a \varphi_\ell * (fg) - (\partial^a \varphi_\ell * f)(\varphi_\ell * g) + (\varphi_\ell * f)(\partial^a \varphi_\ell * g) \tag{10}$$

$$- \sum_{0 < b < a} \binom{a}{b} [\partial^b \varphi_\ell * (f - f(x))][\partial^{a-b} \varphi_\ell * (g - g(x))] \tag{11}$$

$$= \partial^a \varphi_\ell * [(f - f(x))(g - g(x))] \tag{12}$$

$$- \sum_{b \le a} \binom{a}{b} \partial^b \varphi_\ell * (f - f(x)) \cdot \partial^{a-b} \varphi_\ell * (g - g(x)), \tag{13}$$

where we have used the fact that

$$\partial^a \varphi_\ell * f(x) = \begin{cases} f(x) & \text{if } a = 0, \\ 0 & \text{if } a \ne 0. \end{cases}$$

Now observe that

$$|\partial^a \varphi_\ell * [(f - f(x))(g - g(x))]| \tag{14}$$

$$= \left| \int \partial^a \varphi_\ell(y)(f(x-y) - f(x))(g(x-y) - g(x))\, dy \right| \tag{15}$$

$$\leq \int |\partial^a \varphi_\ell(y)||y|^{2\alpha}\, dy \|f\|_\alpha \|g\|_\alpha = C_r \ell^{2\alpha-r} \|f\|_\alpha \|g\|_\alpha. \tag{16}$$

Similarly, all the terms in the sum over b obey the same estimate. This concludes the proof of (5). □

Proof of Proposition 1 Set $g := v^\sharp e$ and $g^\ell := (v * \varphi_\ell)^\sharp e$. We have

$$\|g^\ell_{ij} - g_{ij}\|_1 \leq \|g^\ell_{ij} - g_{ij} * \varphi_\ell\|_1 + \|g_{ij} * \varphi_\ell - g_{ij}\|_1.$$

The first term can be written as

$$\|g^\ell_{ij} - g_{ij} * \varphi_\ell\|_1 = \left\| \partial_j v * \varphi_\ell \cdot \partial_i v * \varphi_\ell - (\partial_j v \cdot \partial_i v) * \varphi_\ell \right\|_1, \tag{17}$$

so that (5) applies, to yield the bound $\ell^{2\alpha-1}\|v\|^2_{1,\alpha}$. For the second term (4) gives the bound $\ell\|g\|_2$. Combining these two we obtain

$$\|g^\ell_{ij} - g_{ij}\|_k \leq C(\ell^{2\alpha-1}\|v\|^2_{1,\alpha} + \ell\|g\|_2),$$

from which (2) readily follows. □

3 *h*-Principle: The General Scheme

The general scheme of our construction follows the method of Nash and Kuiper [20–22]. For convenience of the reader we sketch this scheme in this section. Assume for simplicity that g is smooth.

The existence theorems are based on an iteration of *stages*, and each *stage* consists of several *steps*. The purpose of a *stage* is to correct the error $g - u^\sharp e$. In order to achieve this correction, the error is decomposed into a sum of primitive metrics as

$$g - u^\sharp e = \sum_{k=1}^{n_*} a_k^2 \nu_k \otimes \nu_k \quad \text{(locally)},$$

$$g - u^\sharp e = \sum_j \sum_{k=1}^{n_*} (\psi_j a_{j,k})^2 \nu_{j,k} \otimes \nu_{j,k} \quad \text{(globally)}.$$

The natural estimates associated with this decomposition are

$$\|a_k\|_0 \sim \|g - u^\sharp e\|_0^{1/2}, \tag{18}$$

$$\|a_k\|_{N+1} \sim \|u\|_{N+2} \quad \text{for } N = 0, 1, 2, \ldots. \tag{19}$$

A *step* then involves adding one primitive metric. In other words the goal of a *step* is the metric change

$$u^\sharp e \mapsto u^\sharp e + a^2 \nu \otimes \nu.$$

Nash used spiraling perturbations (also known as the Nash twist) to achieve this; for the codimension one case Kuiper replaced the spirals by corrugations. Using the same ansatz (see formula (36)) one easily checks that addition of a primitive metric is possible with the following estimates (see Proposition 2):

$$\begin{aligned} C^0\text{-error in the metric} &\sim \|g - u^\sharp e\|_0 \frac{1}{K}, \\ \text{increase of } C^1\text{-norm of } u &\sim \|g - u^\sharp e\|_0^{1/2}, \\ \text{increase of } C^2\text{-norm of } u &\sim \|u\|_2 K \end{aligned}$$

for any $K \geq 1$. Observe that the first two of these estimates is essentially the same as in [20–22]. Furthermore, the third estimate is only valid modulo a loss of derivative (see Remark 2).

The low codimension forces the steps to be performed serially. This is in contrast with the method of Källen in [19], where the whole *stage* can be performed in one step due to the high codimension. Thus the number of *steps* in a *stage* equals the number of primitive metrics in the above decomposition which interact. This equals n_* for the local construction and $(n+1)n_*$ for the global construction. To deal with the *loss of derivative* problem we mollify the map u at the start of every stage, in a similar manner as is done in a Nash–Moser iteration. Because of the quadratic estimate (5) in Lemma 1 there will be no additional error coming from the mollification. Therefore, iterating the estimates for one step over a single stage (that is, over N_* steps) leads to

$$\begin{aligned} C^0\text{-error in the metric} &\sim \|g - u^\sharp e\|_0 \frac{1}{K}, \\ \text{increase of } C^1\text{-norm of } u &\sim \|g - u^\sharp e\|_0^{1/2}, \\ \text{increase of } C^2\text{-norm of } u &\sim \|u\|_2 K^{N_*}. \end{aligned}$$

With these estimates, iterating over the *stages* leads to exponential convergence of the metric error, leading to a controlled growth of the C^1 norm and an exponential growth of the C^2 norm of the map. In particular, interpolating between these two norms leads to convergence in $C^{1,\alpha}$ for $\alpha < \frac{1}{1+2N_*}$.

4 *h*-Principle: Construction Step

The main step of our construction is given by the following proposition.

Proposition 2 (Construction step) *Let $\Omega \subset \mathbb{R}^n$, $\nu \in S^{n-1}$ and $N \in \mathbb{N}$. Let $u \in C^{N+2}(\overline{\Omega}; \mathbb{R}^{n+1})$ and $a \in C^{N+1}(\overline{\Omega})$. Assume that $\gamma \geq 1$ and $\ell, \delta \leq 1$ are constants such that*

$$\frac{1}{\gamma} I \leq u^\sharp e \leq \gamma I \quad \text{in } \Omega, \tag{20}$$

$$\|a\|_0 \leq \delta, \tag{21}$$

$$\|u\|_{k+2} + \|a\|_{k+1} \leq \delta \ell^{-(k+1)} \quad \text{for } k = 0, 1, \ldots, N. \tag{22}$$

Then, for any

$$\lambda \geq \ell^{-1} \tag{23}$$

there exists $v \in C^{N+1}(\overline{\Omega}; \mathbb{R}^{n+1})$ *such that*

$$\|v^\sharp e - (u^\sharp e + a^2 \nu \otimes \nu)\|_0 \leq C \frac{\delta^2}{\lambda \ell} \tag{24}$$

and

$$\|u - v\|_j \leq C \delta \lambda^{j-1} \quad \text{for } j = 0, 1, \ldots, N+1, \tag{25}$$

where C *is a constant depending only on* n, N *and* γ.

Remark 2 Observe that if (25) would hold for $j = N+2$, then the conclusion of the proposition would say essentially (with $N = 0$) that the equation

$$v^\sharp e = u^\sharp e + a^2 \nu \otimes \nu$$

admits approximate solutions in C^2 with estimates

$$\|v^\sharp e - (u^\sharp e + a^2 \nu \otimes \nu)\|_0 \leq C \delta^2 \frac{1}{K},$$
$$\|u - v\|_2 \leq C \|u\|_2 K.$$

Here $K = \lambda \ell \geq 1$. The fact that (25) holds only for $j \leq N+1$ amounts to a loss of derivative in the estimate.

In the higher codimension case we need an additional technical assumption in order to carry on the same result. As usual the oscillation $\operatorname{osc} u$ of a vector-valued map u is defined as $\sup_{x,y} |u(x) - u(y)|$.

Proposition 3 (Step in higher codim) *Let* $m, n, N \in \mathbb{N}$ *with* $n, N \geq 1$ *and* $m \geq n+1$. *Then there exist a constant* $\eta_0 > 0$ *with the following property. Let* Ω, g, a, ν *and* $u \in C^{2+N}(\overline{\Omega}, \mathbb{R}^m)$ *satisfy the assumptions of Proposition 2 and assume in addition* $\operatorname{osc} \nabla u \leq \eta_0$. *Then there exists a map* $v \in C^{1+N}(\overline{\Omega}, \mathbb{R}^m)$ *satisfying the same conclusion as in Proposition 2.*

4.1 Basic Building Block

In order to prove the Proposition we need the following lemma. The function Γ will be our corrugation.

Lemma 2 *There exists* $\delta_* > 0$ *and a function* $\Gamma \in C^\infty([0, \delta_*] \times \mathbb{R}; \mathbb{R}^2)$ *with* $\Gamma(\delta, t + 2\pi) = \Gamma(\delta, t)$ *and having the following properties*:

$$|\partial_t \Gamma(s,t) + e_1|^2 = 1 + s^2, \tag{26}$$

$$|\partial_s \partial_t^k \Gamma_1(s,t)| + |\partial_t^k \Gamma(s,t)| \le C_k s \quad \text{for } k \ge 0. \tag{27}$$

Proof Define $H : \mathbb{R}^2 \to \mathbb{R}^2$ as $H(\tau, t) = (\cos(\tau \sin t), \sin(\tau \sin t))$. Then

$$\int_0^{2\pi} H_2(\tau, t)\, dt = \int_0^{2\pi} \sin(\tau \sin t)\, dt = \int_{-\pi}^{\pi} \sin(\tau \sin t)\, dt = 0 \tag{28}$$

by the symmetry of the sine function. Set

$$J_0(\tau) := \frac{1}{2\pi} \int_0^{2\pi} H_1(\tau, t)\, dt = \frac{1}{2\pi} \int_0^{2\pi} \cos(\tau \sin t)\, dt. \tag{29}$$

Note that $J_0 \in C^\infty(\mathbb{R})$ with $J_0(0) = 1$, $J_0'(0) = 0$ and $J''(0) < 0$. We claim that there exists $\delta > 0$ and a function $f \in C^\infty(-\delta, \delta)$ such that $f(0) = 0$ and

$$J_0(f(s)) = \frac{1}{\sqrt{1 + s^2}}. \tag{30}$$

This is a consequence of the implicit function theorem. To see this, set

$$F(s, r) = J_0(r^{1/2}) - (1 + s^2)^{-1/2}.$$

Then $F \in C^\infty(\mathbb{R}^2)$. Indeed, since the Taylor expansion of $\cos x$ contains only even powers of x, $J_0(r^{1/2})$ is obviously analytic. Moreover,

$$J_0(r^{1/2}) = \frac{1}{2\pi} \int_0^{2\pi} \left(1 - \frac{r}{2} \sin^2 t\right) dt + O(r^2).$$

In particular $\partial_r F(0,0) = -1/4$. Since also $F(0,0) = 0$, the implicit function theorem yields $\delta > 0$ and $g \in C^\infty(-\delta, \delta)$ such that $g(0) = 0$ and

$$F(s, g(s)) = 0.$$

Next, observe that $\partial_s F(0,0) = 0$ and $\partial_s^2 F(0,0) = 1$. Therefore

$$g'(0) = 0 \quad \text{and} \quad g''(0) = 4.$$

This implies that $f(s) := g(s)^{1/2}$ is also a smooth function, with

$$f(0) = 0 \quad \text{and} \quad f'(0) = \sqrt{2},$$

thus proving our claim.

Having found $f \in C^\infty(-\delta,\delta)$ with $f(0)=0$ and (30), we finally set

$$\Gamma(s,t) := \int_0^t \left[\sqrt{1+s^2}H(f(s),t') - e_1\right] dt'. \tag{31}$$

By construction $|\partial_t \Gamma(s,t) + e_1|^2 = 1+s^2$. Moreover

$$\begin{aligned}\Gamma(s,t+2\pi) - \Gamma(s,t) &= \int_t^{t+2\pi} \left[\sqrt{1+s^2}H(f(s),t') - e_1\right] dt' \\ &= \sqrt{1+s^2}\int_0^{2\pi} H(f(s),t')\,dt' - 2\pi e_1 \\ &\overset{(28)(29)}{=} 2\pi e_1 \left[\sqrt{1+s^2}J_0(f(s)) - 1\right] \overset{(30)}{=} 0.\end{aligned}$$

Thus the function Γ is 2π-periodic in the second argument.

We now come to the estimates. Fix $\delta_* < \delta$. Then $\Gamma \in C([0,\delta_*]\times\mathbb{R};\mathbb{R}^2)$, and since it is periodic in the second variable, Γ and all its partial derivatives are uniformly bounded. Straightforward computations show that for any $k = 0, 1, \dots$

$$\partial_t^k \Gamma(0,t) = 0 \quad \text{and} \quad \partial_s \partial_t^k \Gamma_1(0,t) = 0 \quad \text{for all } t.$$

Hence, integrating in s, we conclude that

$$\begin{aligned}|\partial_t^k \Gamma(s,t)| &\le s\|\partial_s\partial_t^k\Gamma\|_0, \\ |\partial_s\partial_t^k \Gamma_1(s,t)| &\le s\|\partial_s^2\partial_t^k\Gamma_1\|_0,\end{aligned}$$

which give the desired estimates. □

4.2 *Proof of Proposition 2*

Throughout the proof the letter C will denote a constant, whose value might change from line to line, but otherwise depends only on n, N and γ. Fix a choice of orthonormal coordinates in $\mathbb{R}^n$. In these coordinates the pullback metric can be written as $(u^\sharp e)_{ij} = \partial_i u \cdot \partial_j u$ or, denoting the matrix differential of u by $\nabla u = (\partial_j u^i)_{ij}$, as

$$u^\sharp e = \nabla u^T \nabla u.$$

From now on we will work with this notation.

Let

$$\xi = \nabla u \cdot (\nabla u^T \nabla u)^{-1} \cdot \nu, \qquad \zeta = \partial_1 u \wedge \partial_2 u \wedge \dots \wedge \partial_n u. \tag{32}$$

Because of (20) the vectorfields ξ, ζ are well-defined and satisfy

$$\frac{1}{C} \le |\xi(x)|, |\zeta(x)| \le C \quad \text{for } x \in \Omega \tag{33}$$

with some $C \geq 1$. Now let

$$\xi_1 = \frac{\xi}{|\xi|^2}, \qquad \xi_2 = \frac{\zeta}{|\xi||\zeta|}, \qquad \Psi(x) = \xi_1(x) \otimes e_1 + \xi_2(x) \otimes e_2,$$

and

$$\tilde{a} = |\xi| a.$$

Then

$$\nabla u^T \Psi = \frac{1}{|\xi|^2} \nu \otimes e_1, \qquad \Psi^T \Psi = \frac{1}{|\xi|^2} I, \tag{34}$$

and

$$\begin{aligned} \|\Psi\|_j &\leq C\|u\|_{j+1}, \\ \|\tilde{a}\|_j &\leq C(\|a\|_j + \|a\|_0 \|u\|_{j+1}), \end{aligned} \tag{35}$$

for $j = 0, 1, \ldots, N+1$. Finally, let

$$v(x) := u(x) + \frac{1}{\lambda} \Psi(x) \Gamma\big(\tilde{a}(x), \lambda x \cdot \nu\big), \tag{36}$$

where $\Gamma = \Gamma(s,t)$ is the function constructed in Lemma 2.

Proof of (24) First we compute $\nabla v^T \nabla v$. We have

$$\nabla v = \underbrace{\nabla u + \Psi \cdot \partial_t \Gamma \otimes \nu}_{A} + \underbrace{\lambda^{-1} \Psi \cdot \partial_s \Gamma \otimes \nabla \tilde{a}}_{E_1} + \underbrace{\lambda^{-1} \nabla \Psi \cdot \Gamma}_{E_2}. \tag{37}$$

Using the notation $\mathrm{sym}(A) = (A + A^T)/2$ one has

$$\nabla v^T \nabla v = A^T A + 2\,\mathrm{sym}(A^T E_1 + A^T E_2) + (E_1 + E_2)^T (E_1 + E_2). \tag{38}$$

Using (34) and (26):

$$\begin{aligned} A^T A &= \nabla u^T \nabla u + \frac{1}{|\xi|^2} (2\partial_t \Gamma_1 + |\partial_t \Gamma|^2) \nu \otimes \nu \\ &= \nabla u^T \nabla u + \frac{1}{|\xi|^2} \tilde{a}^2 \nu \otimes \nu = \nabla u^T \nabla u + a^2 \nu \otimes \nu. \end{aligned} \tag{39}$$

Next we estimate the error terms. First of all

$$\begin{aligned} A^T E_1 &= \frac{1}{\lambda} (\nabla u^T \Psi)(\partial_s \Gamma \otimes \nabla \tilde{a}) + \frac{1}{\lambda} (\nu \otimes \partial_t \Gamma)(\Psi^T \Psi)(\partial_s \Gamma \otimes \nabla \tilde{a}) \\ &= \frac{1}{\lambda |\xi|^2} (\partial_s \Gamma_1 + \partial_t \Gamma \cdot \partial_s \Gamma)(\nu \otimes \nabla \tilde{a}). \end{aligned} \tag{40}$$

Note that (27) together with (35) implies:

$$\|\Gamma\|_0, \|\partial_t \Gamma\|_0, \|\partial_s \Gamma_1\|_0 \leq C\|a\|_0.$$

Therefore

$$\|\operatorname{sym}(A^T E_1)\|_0 \le \frac{C}{\lambda}\|a\|_0\|\tilde{a}\|_1 \le C\frac{\delta^2}{\lambda\ell},$$

and similarly

$$\|\operatorname{sym}(A^T E_2)\|_0 \le \frac{C}{\lambda}\|a\|_0\|u\|_2 \le C\frac{\delta^2}{\lambda\ell}. \tag{41}$$

Finally,

$$\|E_1 + E_2\|_0 \le \frac{C}{\lambda}(\|\tilde{a}\|_1 + \|a\|_0\|u\|_2) \le \frac{C}{\lambda}(\|a\|_1 + \delta\|u\|_2) \le C\frac{\delta}{\lambda\ell}. \tag{42}$$

In particular $\|E_1 + E_2\|_0 \le C\delta$ and hence

$$\|(E_1 + E_2)^T(E_1 + E_2)\|_0 \le C\frac{\delta^2}{\lambda\ell}. \tag{43}$$

Putting these estimates together we obtain (24) as required. □

Proof of (25) In fact

$$\|u - v\|_0 \le C\delta\frac{1}{\lambda}$$

is obvious, whereas the estimates for $j = 1, \ldots, N$ will follow by interpolation, provided the case $j = N + 1$ holds. Therefore, we now prove this case. A simple application of the product rule and interpolation yields

$$\begin{aligned}\|v - u\|_{N+1} &\le \frac{C}{\lambda}(\|\Psi\|_{N+1}\|\Gamma\|_0 + \|\Psi\|_0\|\Gamma\|_{N+1})\\ &\le \frac{C}{\lambda}(\|u\|_{N+2}\|\tilde{a}\|_0 + \|\Gamma\|_{N+1}).\end{aligned} \tag{44}$$

Denoting by D_x^j any partial derivative in the variables $x_1, \ldots, x_n$ of order j, the chain rule can be written symbolically as

$$D_x^{N+1}\Gamma = \sum_{i+j\le N+1}(\partial_s^i\partial_t^j\Gamma)\lambda^j\sum_\sigma C_{i,j,\sigma}(D_x\tilde{a})^{\sigma_1}(D_x^2\tilde{a})^{\sigma_2}\cdots(D_x^{N+1}\tilde{a})^{\sigma_{N+1}},$$

where the inner sum is over all σ with

$$\sigma_1 + \cdots + \sigma_{N+1} = i,$$
$$\sigma_1 + 2\sigma_2 + \cdots + (N+1)\sigma_{N+1} + j = N + 1.$$

These relations can be checked by counting the order of differentiation. Therefore, by using (21), (22) and (23)

$$\|D_x^{N+1}\Gamma\|_0 \le C \sum_{i+j\le N+1} \|\partial_s^i\partial_t^j\Gamma\|_0 \lambda^j \delta^i \ell^{-(N+1-j)}$$
$$\le C \sum_{i+j\le N+1} \|\partial_s^i\partial_t^j\Gamma\|_0 \delta^i \lambda^{N+1} \le C\delta\lambda^{N+1}. \tag{45}$$

In particular, since $\|\Gamma\|_0 \le \delta$, we deduce that $\|\Gamma\|_{N+1} \le C\delta\lambda^{N+1}$. Therefore

$$\|v-u\|_{N+1} \le \frac{C}{\lambda}(\delta\|u\|_{N+2} + \delta\lambda^{N+1}) \le C\delta\lambda^N. \tag{46}$$

This concludes the proof of the proposition. □

4.3 *Proof of Proposition 3*

The proof of Proposition 2 would carry over to this case if we can choose an appropriate normal vector field ζ as at the beginning of the proof of Proposition 2, enjoying the estimate (33) with a fixed constant.

To obtain $\zeta(x)$ let $T(x)$ be the tangent plane to $u(\mathbb{R}^n)$ at the point $u(x)$, i.e. the plane generated by $\{\partial_1 u, \dots, \partial_n u\}$. Denote by π_x the orthogonal projection of $\mathbb{R}^m$ onto $T(x)$. Assuming that ∇u has oscillation smaller than η_0, there exists a vector $w \in S^{n-1}$ such that $|\pi_x w| \le 1/2$ for every $x \in \overline{\Omega}$. Hence, we can define

$$\zeta(x) := w - \pi_x w.$$

It is straightforward to see that this choice of ζ gives a map enjoying the same estimates as the ζ used in the proof of Proposition 2.

5 *h*-Principle: Stage

Proposition 4 (Stage, local) *For all* $g_0 \in \mathrm{sym}_n^+$ *there exists* $0 < r < 1$ *such that the following holds for any* $\Omega \subset \mathbb{R}^n$ *and* $g \in C^\beta(\overline{\Omega})$ *with* $\|g - g_0\|_0 \le r$. *There exists a* $\delta_0 > 0$ *such that, if* $K \ge 1$ *and* $u \in C^2(\overline{\Omega}, \mathbb{R}^{n+1})$ *satisfies*

$$\|u^\sharp e - g\|_0 \le \delta^2 \le \delta_0^2 \quad \textit{and} \quad \|u\|_2 \le \mu,$$

then there exists $v \in C^2(\overline{\Omega}, \mathbb{R}^{n+1})$ *with*

$$\|v^\sharp e - g\|_0 \le C\delta^2\left(\frac{1}{K} + \delta^{\beta-2}\mu^{-\beta}\right), \tag{47}$$

$$\|v\|_2 \le C\mu K^{n_*}, \tag{48}$$

$$\|u - v\|_1 \leq C\delta. \tag{49}$$

Here C is a constant depending only on n, g_0, g and Ω.

The Proposition above is the basic stage of the iteration scheme which will prove Theorem 1. A similar proposition, to be used in the proof of Theorem 2 will be stated later.

5.1 Decomposing a Metric into Primitive Metrics

Lemma 3 *Let $g_0 \in \mathrm{sym}_n^+$. Then there exists $r > 0$, vectors $\nu_1, \ldots, \nu_{n_*} \in \mathbb{S}^{n-1}$ and linear maps $L_k : \mathrm{sym}_n \to \mathbb{R}$ such that*

$$g = \sum_{k=1}^{n_*} L_k(g)\nu_k \otimes \nu_k \quad \textit{for every } g \in \mathrm{sym}_n$$

and, moreover, $L_k(g) \geq r$ for every k and every $g \in \mathrm{sym}_n^+$ with $|g - g_0| \leq r$.

Proof Consider the set $S := \{(e_i + e_j) \otimes (e_i + e_j), i \leq j\}$, where $\{e_i\}$ is the standard basis of $\mathbb{R}^n$. Since the span of S contains all matrices of the form $e_i \otimes e_j + e_j \otimes e_i$, clearly S generates sym_n. On the other hand S consists of n_* matrices with $n_* = \dim(\mathrm{sym}_n)$. So S is a basis for sym_n. Let us relabel the vectors $e_i + e_j$ $(i \leq j)$ as $f_1, \ldots, f_{n_*}$, and let

$$h = \sum_{k=1}^{n_*} f_k \otimes f_k.$$

Then $h \in \mathrm{sym}_n^+$ and hence there exists an invertible linear transformation L such that $LhL^T = g_0$. In particular, writing $\nu_k = Lf_k/|Lf_k| \in \mathbb{S}^{n-1}$, we have

$$g_0 = \sum_{k=1}^{n_*} Lf_k \otimes Lf_k = \sum_{k=1}^{n_*} |Lf_k|^2 \nu_k \otimes \nu_k.$$

Note that the set $\{\nu_k \otimes \nu_k\}$ is also a basis for sym_n and therefore there exist linear maps $L_k : \mathrm{sym}_n \to \mathbb{R}$ such that $\sum L_k(A)\nu_k \otimes \nu_k$ is the unique representation of $A \in \mathrm{sym}_n$ as linear combination of $\nu_k \otimes \nu_k$. In particular, $L_i(g_0) = |Lf_k|^2 > 0$. The existence of $r > 0$ satisfying the claim of the lemma follows easily. □

5.2 Proof of Proposition 4

Choose $r > 0$ and $\gamma > 1$ so that the statement of Lemma 3 holds with g_0 and $2r$, and so that

$$\frac{1}{\gamma} I \leq h \leq \gamma \quad \text{for any } h \in \text{sym}_n^+ \text{ with } |h - g_0| < 2r.$$

Moreover, extend u and g to $\mathbb{R}^n$ so that

$$\|u\|_{C^2(\mathbb{R}^n)} \leq C\|u\|_{C^2(\overline{\Omega})}, \qquad \|g\|_{C^\beta(\mathbb{R}^n)} \leq C\|g\|_{C^\beta(\overline{\Omega})}.$$

The procedure of such an extension is well known, with the constant C depending on n, β and Ω. In what follows, the various constants will be allowed to depend in addition on r and γ.

Step 1. Mollification We set

$$\ell = \frac{\delta}{\mu},$$

and let

$$\tilde{u} = u * \varphi_\ell, \qquad \tilde{g} = g * \varphi_\ell, \tag{50}$$

where $\varphi \in C_c^\infty(B_1(0))$ is a symmetric nonnegative convolution kernel with $\int \varphi = 1$. Lemma 1 implies

$$\|\tilde{u} - u\|_1 \leq C\|u\|_2 \ell \leq C\delta, \tag{51}$$

$$\|\tilde{g} - g\|_0 \leq C\|g\|_\beta \ell^\beta, \tag{52}$$

$$\|\tilde{u}\|_{k+2} \leq C\|u\|_2 \ell^{-k} \leq C\delta\ell^{-(k+1)}, \tag{53}$$

and

$$\begin{aligned} \|\tilde{u}^\sharp e - \tilde{g}\|_k &\leq \|\tilde{u}^\sharp e - (u^\sharp e) * \varphi_\ell\|_k + \|(u^\sharp e) * \varphi_\ell - g * \varphi_\ell\|_k \\ &\leq C\ell^{2-k}\|u\|_2^2 + C\ell^{-k}\|u^\sharp e - g\|_0 \leq C\delta^2\ell^{-k}, \end{aligned} \tag{54}$$

where $k = 0, 1, \ldots, n_*$. Moreover, since the set $\{h \in \text{sym}_n^+ : |h - g_0| \leq r\}$ is convex, $\tilde{g}$ also satisfies $\|\tilde{g} - g_0\|_0 \leq r$.

Step 2. Rescaling First of all, observe that

$$\tilde{h} := \tilde{g} + \frac{r}{C\delta^2}(\tilde{g} - \tilde{u}^\sharp e)$$

satisfies the condition $|\tilde{h}(x) - g_0| \leq \frac{r}{C\delta^2}\|\tilde{g} - \tilde{u}^\sharp e\|_0 + r \leq 2r$. Therefore, using Lemma 3 we have

$$(1 + Cr^{-1}\delta^2)\tilde{g} - \tilde{u}^\sharp e = \frac{C\delta^2}{r}\tilde{h} = \sum_{i=1}^{n_*} \tilde{a}_i^2 \nu_i \otimes \nu_i,$$

where $\tilde{a}_i(x) = (C\frac{\delta^2}{r}L_i(\tilde{h}(x)))^{1/2}$. In particular $\tilde{a}_i$ is smooth and

$$\begin{aligned}\|\tilde{a}_i\|_k &\le C\delta\frac{\|L_i(\tilde{h})\|_k}{\|L_i(\tilde{h})\|_0^{1/2}} \le C\delta\|\tilde{h}\|_k\\ &\le C\delta\left(\|\tilde{g}\|_k + \frac{1}{\delta^2}\|\tilde{g}-\tilde{u}^\sharp e\|_k\right) \le C\delta\ell^{-k}\end{aligned}$$

for $k = 0, 1, 2, \dots, n_*$ (note that the first inequality is achieved through interpolation). Let

$$u_0 = \frac{1}{(1+Cr^{-1}\delta^2)^{1/2}}\tilde{u}, \qquad a_i = \frac{1}{(1+Cr^{-1}\delta^2)^{1/2}}\tilde{a}_i.$$

Then we have

$$\tilde{g} - u_0^\sharp e = \sum_{i=1}^{n_*} a_i^2 \nu_i \otimes \nu_i,$$

with

$$\|\tilde{u} - u_0\|_1 \le C\delta, \tag{55}$$

$$\|a_i\|_0 \le C\delta, \tag{56}$$

$$\|u_0\|_{k+2} + \|a_i\|_{k+1} \le C\delta\ell^{-(k+1)}, \tag{57}$$

for $k = 0, 1, \dots, n^*$. Notice that the constants above depend also on k, but since we will only use these estimates for $k \le n_*$, this dependence can be suppressed.

Finally, using (54) we have $\|u_0^\sharp e - g_0\|_0 \le r + C\delta^2$, so that $\gamma^{-1}I \le u_0^\sharp e \le \gamma I$, provided δ_0 is sufficiently small.

Step 3. Iterating One-Dimensional Oscillations We now apply n_* times successively Proposition 2, with

$$\ell_j = \ell K^{-j}, \qquad \lambda_j = K^{j+1}\ell^{-1}, \qquad N_j = n_* - j$$

for $j = 0, 1, \dots, n_*$. In other words we construct a sequence of immersions u_j such that $\frac{1}{\gamma}I \le u_j^\sharp e \le \gamma I$ and

$$\|u_j\|_{k+2} \le C\delta\ell_j^{-(k+1)} \quad \text{for } k = 0, 1, \dots, N_j. \tag{58}$$

To see that Proposition 2 is applicable, observe that $\lambda_j = K\ell_j^{-1}$. Therefore it suffices to check inductively the validity of (58). This follows easily from (25). The constants will depend on j, but this can again be suppressed because $j \le n_*$.

In this way we obtain the functions $u_1, u_2, \dots, u_{n_*}$ with estimates

$$\|u_j\|_2 \leq C\delta\ell^{-1}K^j,$$

$$\|u_{j+1}^\sharp e - (u_j^\sharp e + a_{j+1}^2 \nu_{j+1} \otimes \nu_{j+1})\|_0 \leq C\frac{\delta^2}{\lambda_j \ell_j} = C\delta^2 \frac{1}{K},$$

and moreover

$$\|u_{j+1} - u_j\|_1 \leq C\delta. \tag{59}$$

Observe also that $\|u_j^\sharp e - g_0\|_0 \leq r + C\delta^2$, so that, provided δ_0 is sufficiently small, $\gamma^{-1} I \leq u_j^\sharp e \leq \gamma I$ for all j.

Thus $v := u_{n_*}$ satisfies the estimates

$$\|v^\sharp e - \tilde{g}\|_0 \leq C\delta^2 \frac{1}{K},$$

$$\|v\|_2 \leq C\mu K^{n_*},$$

$$\|v - u_0\|_1 \leq C\delta.$$

The estimates (47), (48) and (49) follow from the above combined with (51), (52) and (55).

5.3 *Stage for General Manifolds*

Given M as in Theorem 2 we fix a finite atlas of M with charts Ω_i and a corresponding partition of unity $\{\phi_i\}$, so that $\sum \phi_i = 1$ and $\phi_i \in C_c^\infty(\Omega_i)$. Furthermore, on each Ω_i we fix a choice of coordinates.

Using the partition of unity we define the space $C^k(M)$. In particular, let

$$\|u\|_k := \sum_i \|\phi_i u\|_k.$$

Similarly, we define mollification on M via the partition of unity. In other words we fix $\varphi \in C_c^\infty(B_1(0))$, and for a function u on M we define

$$u * \varphi_\ell := \sum_i (\phi_i u) * \varphi_\ell. \tag{60}$$

It is not difficult to check that the estimates in Lemma 1 continue to hold on M with these definitions.

Next, let g be a metric on M as in Theorem 2. Since M is compact and g is continuous, there exists $\gamma > 0$ such that

$$\frac{1}{\gamma} I \leq g \leq \gamma I \quad \text{in } M. \tag{61}$$

Moreover, also by compactness, there exists $r_0 > 0$ such that Lemma 3 holds with $r = 2r_0$ for any g_0 satisfying $\frac{1}{\gamma} I \le g_0 \le \gamma I$. Therefore there exists $\rho_0 > 0$ so that

$$\begin{aligned} U \subset \Omega_i \quad &\text{for some } i \text{ and } \operatorname{osc}_U g < r_0 \\ &\text{whenever } U \subset M \text{ with } \operatorname{diam} U < \rho_0. \end{aligned} \tag{62}$$

Here $\operatorname{osc}_U g$ is to be evaluated in the coordinates of the chart Ω_i.

In the following we will need coverings of M with the following property:

Definition 1 (Minimal cover of M) For $\rho > 0$ a finite open covering $\mathcal{C}$ of M is a minimal cover of diameter ρ if:

1. the diameter of each $U \in \mathcal{C}$ is less than ρ;
2. $\mathcal{C}$ can be subdivided into $n + 1$ subfamilies $\mathcal{F}_i$, each consisting of pairwise disjoint sets.

The existence of such coverings is a well-known fact. For the convenience of the reader we give a short proof at the end of this section.

We are now ready to state the iteration stage needed for the proof of Theorem 2. Recall that $\eta_0 > 0$ is the constant from Proposition 3.

Proposition 5 (Stage, global) *Let (M^n, g) be a smooth, compact Riemannian manifold with $g \in C^\beta(M)$, and let $\mathcal{C}$ be a minimal cover of M of diameter $\rho < \rho_0$, where ρ_0 is as in (62). There exists $\delta_0 > 0$ such that, if $K \ge 1$ and $u \in C^2(M, \mathbb{R}^m)$ satisfies*

$$\|u^\sharp e - g\|_0 \le \delta^2 < \delta_0^2, \tag{63}$$

$$\|u\|_2 \le \mu, \tag{64}$$

$$\operatorname{osc}_U \nabla u \le \eta_0/2 \quad \text{for all } U \in \mathcal{C}, \tag{65}$$

then there exists $v \in C^2(M, \mathbb{R}^m)$ with

$$\|v^\sharp e - g\|_0 \le C\delta^2 \left(\frac{1}{K} + \delta^{\beta-2} \mu^{-\beta} \right), \tag{66}$$

$$\|v\|_2 \le C\mu K^{(n+1)n_*}, \tag{67}$$

$$\|u - v\|_1 \le C\delta. \tag{68}$$

The constants C depend only (M^n, g) and $\mathcal{C}$.

5.4 *Proof of Proposition 5*

We proceed as in the proof of Proposition 4. Enumerate the covering as $\mathcal{C} = \{U_j\}_{j \in J}$, and for each j choose a matrix $g_j \in \operatorname{sym}_n^+$ such that

$$|g(x) - g_j| \le r_0 \quad \text{for } x \in U_j.$$

Furthermore, fix a partition of unity $\{\psi_j\}$ for $\mathcal{C}$ in the sense that $\psi_j \in C_c^\infty(U_j)$ and $\sum_j \psi_j^2 = 1$ on M.

Step 1. Mollification The mollification step is precisely as in Proposition 4. We set

$$\ell = \frac{\delta}{\mu},$$

and let

$$\tilde{u} = u * \varphi_\ell, \qquad \tilde{g} = g * \varphi_\ell, \tag{69}$$

where now the convolution is defined in (60) above. Then, as before,

$$\|\tilde{u} - u\|_1 \le C\delta, \tag{70}$$

$$\|\tilde{g} - g\|_0 \le C\|g\|_\beta \ell^\beta, \tag{71}$$

$$\|\tilde{u}\|_{k+2} \le C\delta\ell^{-(k+1)}, \tag{72}$$

$$\|\tilde{u}^\sharp e - \tilde{g}\|_k \le C\delta^2\ell^{-k}, \tag{73}$$

for $k = 0, 1, \dots, (n+1)n_*$. In particular, for any $j \in J$ and any $x \in U_j$

$$|\tilde{g}(x) - g_j| \le r_0 + C\ell^\beta \le r_0 + C\delta_0^\beta \le \frac{3}{2}r_0$$

provided $\delta_0 > 0$ is sufficiently small.

Step 2. Rescaling We rescale the map analogously to Step 2 in Proposition 4. Accordingly,

$$\tilde{h} := \tilde{g} + \frac{r_0}{2C\delta^2}(\tilde{g} - \tilde{u}^\sharp e)$$

satisfies

$$|\tilde{h}(x) - g_j| \le \frac{r_0}{2C\delta^2}\|\tilde{g} - \tilde{u}^\sharp e\|_0 + \frac{3}{2}r_0 \le 2r_0 \quad \text{in } U_j.$$

Therefore, using Lemma 3 for each g_j and introducing

$$u_0 = \frac{1}{(1 + Cr_0^{-1}\delta^2)^{1/2}}\tilde{u}$$

we obtain (as in Proposition 4)

$$\tilde{g} - u_0^\sharp e = \sum_{i=1}^{n_*} a_{i,j}^2 \nu_{i,j} \otimes \nu_{i,j} \quad \text{in } U_j$$

for some functions $a_{i,j} \in C^\infty(U_j)$ satisfying the estimates

$$\|a_{i,j}\|_{C^{k+1}(U_j)} \le C\delta\ell^{-(k+1)} \quad \text{for } j \in J \text{ and } k = 0, 1, \dots, (n+1)n_*.$$

In particular, using the partition of unity $\{\psi_j\}$ we obtain

$$\tilde{g} - u_0^\sharp e = \sum_{j\in J}\sum_{i=1}^{n_*}(\psi_j a_{i,j})^2 \nu_{i,j}\otimes \nu_{i,j}, \tag{74}$$

with

$$\|u-u_0\|_1 \le C\delta, \tag{75}$$

$$\|\psi_j a_{i,j}\|_0 \le C\delta, \tag{76}$$

$$\|u_0\|_{k+2} + \|\psi_j a_{i,j}\|_{k+1} \le C\delta\ell^{-(k+1)} \tag{77}$$

for $k = 0, 1, \dots, (n+1)n_*$.

Step 3. Iterating One-Dimensional Oscillations We now argue as in the Step 3 of the proof of Proposition 4. However, there are two differences. First of all we apply Proposition 3 in place of Proposition 2. This requires an additional control of the oscillation of ∇u in each U_j. Second, the number of steps is $(n+1)n_*$. Indeed, observe that (74) can be written as

$$\tilde{g} - u_0^\sharp e = \sum_{\sigma=1}^{n+1}\sum_{i=1}^{n_*}\sum_{j\in J_\sigma}(\psi_j a_{i,j})^2 \nu_{i,j}\otimes \nu_{i,j}, \tag{78}$$

where the index set J is decomposed as $J = J_1 \cup \dots \cup J_{n+1}$ so that $U_j \in \mathcal{F}_\sigma$ if and only if $j \in J_\sigma$. The point is that the sum in j consists of functions with disjoint supports, and hence for this sum Proposition 3 can be performed in parallel, in one step. Thus, the number of steps to be performed serially is the number of summands in σ and i, which is precisely $(n+1)n_*$.

To deal with the restriction on the oscillation of u_k in each step, observe that $\operatorname{osc}_{U_j}\nabla u \le \eta_0/2$ by assumption, and clearly the same holds for u_0. Also, at each step we have the estimate $\|u_{k+1}-u_k\|_1 \le C\delta \le C\delta_0$. Therefore, choosing $\delta_0 > 0$ sufficiently small (only depending on the constants and on η_0), we ensure that the condition remains satisfied inductively $(n+1)n_*$ times.

Thus, proceeding as in the proof of Proposition 4 we apply Proposition 3 successively with $\ell_k = \ell K^{-k}$, $\lambda_k = K^{k+1}\ell^{-1}$, and $N_k = (n+1)n_* - k$. In this way we obtain a final map $v := u_{(n+1)n_*}$ such that

$$\|v^\sharp e - \tilde{g}\|_0 \le C\delta^2\frac{1}{K},$$

$$\|v\|_2 \le C\mu K^{(n+1)n_*},$$

$$\|v-u_0\|_1 \le C\delta.$$

The above inequalities combined with (70), (71) and (75) imply the estimates (66), (67) and (68). This concludes the proof.

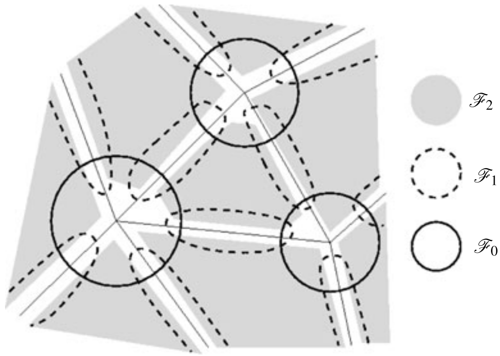


Fig. 1 The triangulation T and the covering for a 2-dimensional manifold

5.5 Existence of Minimal Covers

We fix a triangulation T of M with simplices having diameter smaller than $\rho/3$. We let S_0 be the vertices of the triangulation, S_1 be the edges, S_k be the k-faces. $\mathcal{F}_0$ is made by pairwise disjoint balls centered on the elements of S_0, with radius smaller than $\rho/2$. We let M_0 be the union of these balls. Next, for any element $\sigma \in S_1$, we consider $\sigma' = \sigma \setminus M_0$. The σ' are therefore pairwise disjoint compact sets and we let $\mathcal{F}_1$ be a collection of pairwise disjoint neighborhoods of σ', each with diameter less than ρ. We define M_1 to be the union of the elements of $\mathcal{F}_1$ and $\mathcal{F}_0$. We proceed inductively. At the step k, for every k-dim. face $F \in S_k$ we define $F' = F \setminus A_{k-1}$. Clearly, the F' are pairwise disjoint compact sets and hence we can find pairwise disjoint neighborhoods of the F' with diameter smaller than ρ. Figure 1 shows the elements of $\mathcal{F}_i$ for a 2-d triangulation.

Clearly, the collection $\mathcal{F}_0 \cup \cdots \cup \mathcal{F}_n$ covers any simplex of T, and hence is a covering of M.

6 *h*-Principle: Iteration

6.1 Proof of Theorem 1

Let $\mu_0, \delta_0 > 0$ be such that

$$\|u^\sharp e - g\|_0 \le \delta_0^2,$$
$$\|u\|_2 \le \mu_0.$$

Let also $K \ge 1$. Later on we are going to adjust the parameters μ_0 and K in order to achieve the required convergence in $C^{1,\alpha}$. Applying Proposition 4 successively, we obtain a sequence of maps $u_k \in C^2(\overline{\Omega}, \mathbb{R}^{n+1})$ such that

$$\|u_k^\sharp e - g\|_0 \le \delta_k^2,$$
$$\|u_k\|_2 \le \mu_k,$$
$$\|u_{k+1} - u_k\|_1 \le C\delta_k,$$

where

$$\delta_{k+1}^2 = C\delta_k^2\left(\frac{1}{K} + \delta_k^{\beta-2}\mu_k^{-\beta}\right), \tag{79}$$
$$\mu_{k+1} = C\mu_k K^{n_*}. \tag{80}$$

Substituting K with $\max\{C^{1/n_*}K, K\}$ we can absorb the constant in (80) to achieve $\mu_{k+1} = \mu_k K^{n_*}$, at the price of getting a possibly worse constant in (79). In particular $\mu_k = \mu_0 K^{kn_*}$. Next, we show by induction that for any

$$a < \min\left\{\frac{1}{2}, \frac{\beta n_*}{2-\beta}\right\} \tag{81}$$

there exists a suitable initial choice of K and μ_0 so that

$$\delta_k \le \delta_0 K^{-ak}.$$

The case $k = 0$ is obvious. Assuming the inequality to hold for k, we have

$$\delta_{k+1}^2 \le C\delta_0^2 K^{-2ak-1} + C\delta_0^\beta \mu_0^{-\beta} K^{-\beta k(a+n_*)}.$$

Therefore $\delta_{k+1} \le \delta_0 K^{-a(k+1)}$ provided

$$2C \le K^{1-2a} \quad \text{and} \quad 2C \le \mu_0^\beta \delta_0^{2-\beta} K^{k[\beta(a+n_*)-2a]-2a}.$$

By choosing first K and then $\mu_0 \ge \|u\|_2$ sufficiently large, these two inequalities can be satisfied for any given a in the range prescribed in (81). This proves our claim.

Next we show that for any

$$\alpha < \min\left\{\frac{1}{1+2n_*}, \frac{\beta}{2}\right\} \tag{82}$$

the parameters μ_0 and K can be chosen so that the sequence u_k converges in $C^{1,\alpha}(\Omega;\mathbb{R}^{n+1})$. To this end observe that to any α satisfying (82) there exists an a satisfying (81) such that

$$\alpha < \frac{a}{a+n_*}.$$

Then, choosing μ_0 and K sufficiently large as above, we obtain a sequence u_k such that

$$\|u_{k+1} - u_k\|_1 \le C\delta_0 K^{-ak},$$
$$\|u_{k+1} - u_k\|_2 \le \mu_{k+1} + \mu_k \le 2\mu_0 K^{(k+1)n_*}.$$

Therefore, by interpolation

$$\|u_{k+1} - u_k\|_{1,\alpha} \le \|u_{k+1} - u_k\|_1^{1-\alpha} \|u_{k+1} - u_k\|_2^{\alpha} \le \tilde{C} K^{-[(1-\alpha)a - \alpha n_*]k}. \tag{83}$$

Thus the sequence converges in $C^{1,\alpha}$ to some limit map $v \in C^{1,\alpha}(\overline{\Omega}; \mathbb{R}^{n+1})$. Since $\delta_k \to 0$, the limit satisfies $v^\sharp e = g$ in Ω.

Finally, choosing K so large that $K^{-a} \le 1/2$, we have

$$\|v - u\|_1 \le C\delta_0 \sum_k K^{-ak} \le 2C\delta_0.$$

6.2 Proof of Theorem 2

Recall from Sect. 5.3 that for the whole construction we work with a fixed atlas $\{\Omega_i\}$ of the manifold M, and that to the given metric $g \in C^\beta(M)$ there exist constants $\gamma > 1$ and $\rho_0 > 0$ such that (61) and (62) hold.

Since $u \in C^2(M; \mathbb{R}^m)$ and there are a finite number of charts Ω_i, there exists $\rho < \rho_0$ such that

$$\operatorname{osc}_U \nabla u < \eta_0/4 \quad \text{whenever } U \subset M \text{ with } \operatorname{diam} U < \rho.$$

Fix a minimal cover $\mathcal{C}$ of M with diameter ρ and let $\mu_0, \delta_0 > 0$ be such that

$$\|u^\sharp e - g\|_0 \le \delta_0^2,$$
$$\|u\|_2 \le \mu_0.$$

The iteration now proceeds with respect to this fixed cover, parallel to the proof of Theorem 1. More precisely, arguing as in Theorem 1, Proposition 5 yields a sequence $u_k \in C^2(M; \mathbb{R}^m)$ with

$$\|u_k^\sharp e - g\|_0 \le \delta_k^2,$$
$$\|u_k\|_2 \le \mu_0 K^{k(n+1)n_*},$$
$$\|u_{k+1} - u_k\|_1 \le C\delta_k,$$

where

$$\delta_{k+1}^2 = C\delta_k^2 \left(\frac{1}{K} + \delta_k^{\beta-2} K^{-\beta k(n+1)n_*} \right). \tag{84}$$

The proof that μ_0 and K can be chosen so that u_k converges in $C^{1,\alpha}$ for

$$\alpha < \min\left\{\frac{1}{1+2(n+1)n_*}, \frac{\beta}{2}\right\} \tag{85}$$

follows entirely analogously. Recall that this argument yields in particular

$$\delta_k \le \delta_0 K^{-ak}.$$

The only difference is that the estimates (63) and (65) need to be fulfilled at each stage. To this end note that $\delta_k \le \delta_0$, so that (63) will hold at stage k if it holds at the initial stage. Moreover,

$$\mathrm{osc}_U \nabla u_k \le \mathrm{osc}_U \nabla u + \sum_{j=0}^{k-1} 2\|u_{j+1} - u_j\|_1 \le \frac{\eta_0}{4} + 2C\delta_0 \sum_j K^{-aj} \le \frac{\eta_0}{4} + 4C\delta_0,$$

so that (65) is fulfilled by u_k provided δ_0 is sufficiently small (depending only on the various constants).

6.3 Proof of Corollaries 1 and 2

The corollaries are a direct consequence of the Nash–Kuiper theorem combined with Theorems 1 and 2 respectively. For simplicity, we allow M to be either $\overline{\Omega}$ for a smooth bounded open set $\Omega \subset \mathbb{R}^n$ or a compact Riemannian manifold of dimension n, and assume that $g \in C^\beta(M)$ is satisfying either the assumptions of Theorem 1 or those of Theorem 2. We then set $\alpha_0 = \min\{(2n_* + 1)^{-1}, \beta/2\}$ in the first case, and $\alpha_0 = \min\{(2(n+1)n_* + 1)^{-1}, \beta/2\}$ in the second.

Let $u \in C^1(M;\mathbb{R}^m)$ be a short map and $\varepsilon > 0$. We may assume without loss of generality that $\varepsilon < \delta_0$. Using the Nash–Kuiper theorem together with a standard regularization, there exists $u_0 \in C^2(M;\mathbb{R}^m)$ such that

$$\|u - u_0\|_1 \le \varepsilon/2,$$
$$\|u_0^\sharp e - g\|_0 \le \left(\frac{\varepsilon}{2C}\right)^2,$$

where C is the constant in Theorems 1 and 2 respectively. Then the theorem, applied to u_0, yields an isometric immersion $v \in C^{1,\alpha}(M;\mathbb{R}^m)$ for any $\alpha < \alpha_0$, such that $\|v - u_0\|_1 \le \varepsilon/2$, so that $\|v - u\|_1 \le \varepsilon$. This proves the corollaries.

We now come to Remark 1. This follows immediately from the fact that the Nash–Kuiper theorem also works for embeddings, and that the set of embeddings of a compact manifold is an open set in $C^1(M;\mathbb{R}^m)$. Indeed, if u is an embedding, the Nash–Kuiper theorem gives the existence of an embedding u_0 with the estimates above. Ensuring in addition that ε is so small that any map $v \in C^1(M;\mathbb{R}^m)$ with $\|v - u\|_1 \le \varepsilon$ is an embedding, we reach the required conclusion.

7 Rigidity: Proof of Theorem 3

7.1 Curvature and Brouwer Degree

Let (M, g) be as in Theorem 3. As usual, we denote by dA the area element in M and by κ the Gauss curvature of (M, g). Consider next a C^2 isometric embedding $v : M \to \mathbb{R}^3$. The unit normal $N(p)$ to $v(M)$ is the unique vector of $\mathbb{R}^3$ such that, given a positively oriented basis e_1, e_2 for $T_p(M)$, the triple $(dv_p(e_1), dv_p(e_2), N(p))$ is an orthonormal positively oriented frame of $\mathbb{R}^3$.

As it is well known, if $d\sigma$ denotes the area element in $\mathbb{S}^2$, then $N^\sharp\, d\sigma = \kappa\, dA$. Therefore, for every open set $V \Subset M$ and for every $f \in C^1(\mathbb{S}^2)$, the usual change of variable formula yields

$$\int_V f(N(x))\kappa(x)\, dA(x) = \int_{\mathbb{S}^2} f(y)\deg(y, V, N)\, d\sigma(y), \tag{86}$$

where $\deg(y, V, N)$ denotes the Brouwer degree of the map N. Though the differential definition of deg makes sense only for regular values of N, it is a classical observation that deg is constant on connected components of $\mathbb{S}^2 \setminus N(\partial V)$. Thus it has a unique continuous extension to $\mathbb{S}^2 \setminus N(\partial V)$, which will be denoted as well by deg.

Consider next an isometric embedding $v \in C^1$. In this case $N \in C^0$. The Brouwer degree $\deg(y, V, N)$ can still be defined and we recall the following well-known theorem.

Theorem 4 *Let $N \in C(V, \mathbb{S}^2)$ and $\{N_k\} \subset C^\infty(V, \mathbb{S}^2)$ be a sequence converging uniformly to N. Let $K \subset \mathbb{S}^2 \setminus N(\partial V)$ be a closed set. For any k sufficiently large,* $\deg(\cdot, V, N_k) \equiv \deg(\cdot, V, N)$ *on K.*

Thus $\deg(\cdot, V, N) \in L^1_{\mathrm{loc}}(\mathbb{S}^2 \setminus N(\partial V))$. A key step to the proof of Theorem 3 is to show that formula (86) holds for $v \in C^{1,\alpha}$ with $\alpha > 2/3$.

Proposition 6 *Let $v \in C^{1,\alpha}(M, \mathbb{R}^3)$ be an isometric embedding with $\alpha > 2/3$. Then (86) holds for every open set $V \Subset M$ diffeomorphic to a subset of $\mathbb{R}^2$ and every $f \in L^\infty$ with* $\mathrm{supp}(f) \subset \mathbb{S}^2 \setminus N(\partial V)$.

In order to deal with $N(\partial V)$ we recall the following elementary fact.

Lemma 4 *Let M and $\tilde{M}$ be 2-dimensional Riemannian manifolds, $\beta > \frac{1}{2}$ and $N \in C^{0,\beta}(M, \tilde{M})$. If $E \subset M$ has Hausdorff dimension 1, then the area of $N(E)$ is 0.*

The following is then a corollary of Proposition 6 and Lemma 4.

Corollary 5 *Let (M, g) and v be as in Proposition 6, with $\kappa \geq 0$. For any open $V \Subset M$, $\deg(\cdot, V, N)$ is a nonnegative L^1 function and (86) holds for every $f \in L^\infty(\mathbb{S}^2 \setminus N(\partial V))$.*

7.2 Proof of Proposition 6

By a standard approximation argument, it suffices to prove the statement when f is smooth. Under this additional assumption the proof is a direct consequence of Theorem 4 and of the convergence result below, which is a consequence of Proposition 1. Since V is diffeomorphic to an open set of the euclidean plane, we can consider global coordinates x_1, x_2 on it. Fix a symmetric kernel $\varphi \in C_c^\infty(\mathbb{R}^2)$, set $\varphi_\varepsilon(x) = \varepsilon^{-2}\varphi(x/\varepsilon)$ and let $v^\varepsilon := (v\mathbf{1}_V) * \varphi_\varepsilon$ (we consider here the convolution of the two functions in $\mathbb{R}^2$ using the coordinates x_1, x_2 and the corresponding Lebesgue measure).

Proposition 7 *Let v and v^ε be defined as above and denote by N^ε, g^ε, A^ε and κ^ε respectively, the normal to $v^\varepsilon(M)$, the pull-back of the metric on $v^\varepsilon(M)$, and the corresponding area element and Gauss curvature. Then,*

$$\lim_{\varepsilon\downarrow 0}\int_V f(N^\varepsilon)\kappa^\varepsilon\, dA^\varepsilon = \int_V f(N)\kappa\, dA \quad \forall f \in C_c^\infty(\mathbb{S}^2 \setminus N(\partial V)). \tag{87}$$

Proof In coordinates, our aim is to show that

$$\lim_{\varepsilon\downarrow 0}\int_V f(N^\varepsilon(x))\kappa^\varepsilon(x)(\det g^\varepsilon(x))^{\frac{1}{2}}\, dx = \int_V f(N(x))\kappa(x)(\det g(x))^{\frac{1}{2}}\, dx. \tag{88}$$

We recall the formulas for the Christoffel symbols, the Riemann tensor and the Gauss curvature in V, in the system of coordinates already fixed:

$$\Gamma^i_{jk} = \frac{1}{2} g^{im}\big(\partial_k g_{jm} + \partial_j g_{mk} - \partial_m g_{kj}\big), \tag{89}$$

$$R_{iljk} = g_{lm}\big(\partial_k \Gamma^m_{ij} - \partial_j \Gamma^m_{ik} + \Gamma^l_{ij}\Gamma^m_{kl} - \Gamma^l_{ik}\Gamma^m_{jl}\big), \tag{90}$$

$$\kappa = \frac{R_{1212}}{\det(g_{ij})}. \tag{91}$$

After obvious computations we conclude that

$$\kappa = (\det g)^{-1}(c_{ijkl}\partial_{kl} g_{ij} + d_{ijklmn}(g)\partial_k g_{ij}\partial_l g_{mn}) \tag{92}$$

where c_{ijkl} are *constant* coefficients and the functions d_{ijklmn} are smooth.

Proposition 1 implies that $\partial_k g^\varepsilon_{ij}$ and g^ε_{ij} converge locally uniformly to $\partial_k g_{ij}$ and g_{ij} respectively. Moreover, N^ε converges locally uniformly to N. Since there is a compact set containing $f(N^\varepsilon)$ and $f(N)$, we only need to show that

$$\lim_{\varepsilon\downarrow 0}\int_V f(N^\varepsilon(x))(\det g^\varepsilon(x))^{-\frac{1}{2}}\partial_{kl} g^\varepsilon_{ij}(x)\, dx = \int_V f(N(x))(\det g(x))^{-\frac{1}{2}}\partial_{kl} g_{ij}(x)\, dx. \tag{93}$$

Denote by ψ^ε the function $f(N^\varepsilon(x))(\det g^\varepsilon(x))^{-\frac{1}{2}}$. Since $f(N^\varepsilon)$ is smooth and compactly supported in V we can integrate by parts to get

$$\int_V \psi^\varepsilon \partial_{kl} g^\varepsilon_{ij} = \int_V \partial_k \psi^\varepsilon \partial_l g^\varepsilon_{ij}. \tag{94}$$

Note that $\|\partial_k \psi^\varepsilon\| \le C\varepsilon^{\alpha-1}$ by obvious estimates on convolutions. Hence, (2) gives

$$\int_V \partial_k \psi^\varepsilon \big(\partial_l g^\varepsilon_{ij} - \partial_l g_{ij}\big) = O(\varepsilon^{3\alpha-2}) \tag{95}$$

which converges to 0 because $\alpha > 3/2$. Integrating again by parts, we get

$$\begin{aligned}\lim_{\varepsilon\downarrow 0} \int_V & f(N^\varepsilon(x))(\det g^\varepsilon(x))^{-\frac{1}{2}} \partial_{kl} g^\varepsilon_{ij}(x)\,dx \\ &= \lim_{\varepsilon\downarrow 0} \int_V f(N^\varepsilon(x))(\det g^\varepsilon(x))^{-\frac{1}{2}} \partial_{kl} g_{ij}(x)\,dx.\end{aligned}$$

Using the uniform convergence of N^ε to N and of g^ε to g we then conclude (93) and hence the proof of the Proposition. □

7.3 *Proof of Lemma 4 and Corollary 5*

Proof of Lemma 4 By the definition of Hausdorff dimension, for every $\varepsilon > 0$ and $\eta > 1$ there exists a covering of E with closed sets E_i such that

$$\sum_i (\mathrm{diam}(E_i))^\eta \le \varepsilon. \tag{96}$$

On the other hand, $\mathrm{diam}(g(E_i)) \le C(\mathrm{diam}(E_i))^\beta$ and hence the area $|g(E_i)|$ can be estimated with $C(\mathrm{diam}(E_i))^{2\beta}$. Since $\beta > 1/2$, we can pick $\eta = 2\beta$ to conclude that

$$|g(E)| \le C\sum_i (\mathrm{diam}(E_i))^\eta \le C\varepsilon.$$

The arbitrariness of ε implies $|g(E)| = 0$. □

Proof of Corollary 5 First of all, we know from Proposition 6 that the formula (86) is valid for any open set V which is diffeomorphic to an open set of $\mathbb{R}^2$, and any $f \in L^\infty$ compactly supported in $\mathbb{S}^2 \setminus N(\partial V)$. Since κ is nonnegative, we conclude that $\deg(\cdot, N, V) \ge 0$. Testing (86) with a sequence of compactly supported functions $f_k \uparrow \mathbf{1}_{\mathbb{S}^2\setminus N(\partial V)}$ we derive that

$$\int \deg(y, N, V)\,d\sigma(y) = \int_V \kappa\,dA < \infty,$$

which implies $\deg(\cdot, N, V) \in L^1$.

Next, consider a V with smooth boundary. We decompose it into the union of finitely many nonoverlapping Lipschitz open sets V_i diffeomorphic to open sets of the euclidean plane. Then

$$\deg(y, N, V) = \sum_i \deg(y, N, V_i) \quad \text{for every } y \notin \bigcup N(\partial V_i).$$

On the other hand, by Lemma 4, $\bigcup_i N(\partial V_i)$ is a negligible set, and hence we conclude the formula for V from the previous step.

Finally, fix a generic V and an $f \in L^\infty$ with $\mathrm{supp}(f) \subset \mathbb{S}^2 \setminus N(\partial V)$. Choose an open set V' with smooth boundary $\partial V'$ sufficiently close to ∂V. Then $\deg(\cdot, V, N)$ and $\deg(\cdot, V', N)$ coincide on the support of f, whereas the support of $f(N(\cdot))$ is contained in V'. From the formula for V' and f we conclude then the validity of the formula for V and f. Arguing again as above, we conclude that $\deg(\cdot, N, V)$ is summable and nonnegative and that the formula (86) holds for any V and any $f \in L^\infty(\mathbb{S}^2 \setminus N(\partial V))$. □

7.4 Bounded Extrinsic Curvature. The Proof of Theorem 3

We recall the notion of bounded extrinsic curvature for a C^1 immersed surface (see p. 590 of [26]).

Definition 2 Let $\Omega \subset \mathbb{R}^2$ be open and $u \in C^1(\Omega, \mathbb{R}^3)$ an immersion. The surface $u(\Omega)$ has *bounded extrinsic curvature* if there is a C such that

$$\sum_{i=1}^N |N(E_i)| \le C \tag{97}$$

for any finite collection $\{E_i\}$ of pairwise disjoint closed subsets of Ω.

The proof of Theorem 3 follows now from Corollary 5.

Proof of Theorem 3 The theorem follows easily from the claim:

$$\deg(\cdot, V, N) \ge \mathbf{1}_{N(V)\setminus N(\partial V)} \quad \text{for every open } V \subset \Omega. \tag{98}$$

In fact, given disjoint closed sets $E_1, \ldots, E_N$, we can cover them with disjoint open sets $V_1, \ldots V_N$ with smooth boundaries. By (98) and Corollary 5,

$$\sum_i |N(E_i) \setminus N(\partial V_i)| \le \sum_i |N(V_i) \setminus N(\partial V_i)| \le \sum_i \int_{V_i} \kappa \le \int_\Omega \kappa. \tag{99}$$

On the other hand, by Lemma 4, $|N(\partial V_i)| = 0$. Thus, (99) shows (97).

We now come to the proof of (98). Obviously $\deg(y, V, N) = 0$ if $y \notin N(V)$. Moreover, by Corollary 5, $\deg(\cdot, V, N) \geq 0$. Therefore, fix $y_0 \in N(V) \setminus N(\partial V)$ and assume, by contradiction, that $\deg(y_0, V, N) = 0$. Consider a small open disk D centered at y_0 such that $N^{-1}(D) \cap \partial V = \emptyset$ and let $W := N^{-1}(D) \cap V$. Then $N(\partial W) \subset \partial D$ and $N(W) \subset D$. So, $\deg(\cdot, W, N)$ vanishes on $\mathbb{S}^2 \setminus \overline{D}$ and is a constant integer k on D. On the other hand $k = \deg(y_0, W, N) = \deg(y_0, V, N) - \deg(y_0, V \setminus \overline{W}, N) = -\deg(y_0, V \setminus \overline{W}, N)$. Since $y_0 \notin N(V \setminus \overline{W})$, we conclude $k = 0$ and hence

$$0 = \int \deg(y, W, N)\, dy = \int_W \kappa \, dA$$

which is a contradiction because $W \neq \emptyset$ and $\kappa > 0$. □

Corollary 3 follows from Theorem 3 and the results of Pogorelov cited in the introduction. More precisely, by Theorem 9 on p. 650 [26], $u(S^2)$ is a closed convex surface, which by [25] is rigid.

Corollary 4 also follows from the results in [26] and [27]. However, we were unable to find an exact reference for open surfaces, and therefore, for the reader's convenience, we have included a proof in the Appendix.

Acknowledgements Camillo De Lellis has been supported by the SFB grant TR 71.

Appendix

Proof of Corollary 4 First of all, since the theorem is local, without loss of generality we can assume that:

1. $\Omega = B_r(0)$, $u \in C^{1,\alpha}(\overline{B}_r(x))$, $g \in C^{2,\beta}(\overline{B}_r(x))$ and u is an embedding;
2. $u(\Omega)$ has bounded extrinsic curvature.

Step 1. Density of Regular Points For any point $z \in \mathbb{S}^2$ we let $n(z)$ be the cardinality of $N^{-1}(z)$. It is easy to see that, for a surface of bounded extrinsic curvature, $\int_{\mathbb{S}^2} n < \infty$ (cf. with Theorem 3 of p. 590 in [26]). Therefore, the set $E := \{n = \infty\}$ has measure zero. Let $\Omega_r := N^{-1}(\mathbb{S}^2 \setminus E)$. Observe that

$$\Omega_r \quad \text{is dense in } \Omega. \tag{100}$$

Otherwise there is a nontrivial smooth open set V such that $N(V) \subset E$. But then, $\deg(\cdot, V, N) = 0$ for every $y \notin N(\overline{V})$, and since $|N(V)| = |N(\partial V)| = 0$, it follows that $\deg(\cdot, V, N) = 0$ a.e. By Corollary 5, $\int_V \kappa = 0$, which contradicts $\kappa > 0$.

Step 2. Convexity Around Regular Points Note next that, for every $x \in \Omega_r$ there is a neighborhood U of x such that $N(y) \neq N(x)$ for all $y \in U \setminus \{x\}$, i.e. x is regular in the sense of [26] p. 582. Recalling (98), $\deg(\cdot, V, N) \geq \mathbf{1}_{V \setminus \partial V}$ for every V: therefore the index of the map N at every point $x \in \Omega_r$ is at least 1. So, by the

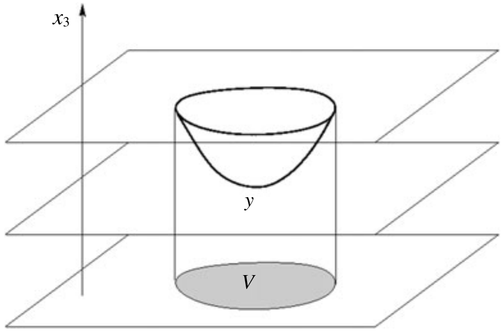


Fig. 2 The convex sets of type $V \times]-a, a[$ among which we choose the maximal one U_m

Lemma of page 594 in [26], any point $x \in \Omega_r$ is an elliptic point relative to the mapping N (that is, there is a neighborhood U of x such that the tangent plane π to $u(\Omega)$ in x intersects $U \cap u(\Omega)$ only in $u(x)$; cf. with page 593 of [26]).

By the discussion of page 650 in [26], $u(\Omega)$ has nonnegative extrinsic curvature as defined in IX.5 of [26]. Then, Lemma 2 of page 612 shows that, for every elliptic point $y \in u(\Omega)$ there is a neighborhood where $u(\Omega)$ is convex. This conclusion applies, therefore, to any $y \in \Omega_r$. We next claim the existence of a constant C with the following property. Set $\rho(y) := C^{-1} \min\{1, \mathrm{dist}(u(y), u(\partial\Omega))\}$. Then

$$u(\Omega) \cap B_{\rho(y)}(y) \quad \text{is convex for all } y \in \Omega_r. \tag{101}$$

Recall that u is an embedding and hence $\mathrm{dist}(u(y), u(\partial\Omega)) > 0$ for every $y \in \Omega$. By (100), (101) gives for any $y \in \Omega$ there is a neighborhood where $u(\Omega)$ is convex. This would complete the proof.

Step 3. Proof of (101) First of all, since u is an embedding and $\|u\|_{C^{1,\alpha}}$ is finite, there is a constant c_0 such that, for any point x, $B_{c_0}(x) \cap u(\Omega)$ is the graph of a $C^{1,\alpha}$ function with $\|\cdot\|_{C^{1,\alpha}}$ norm smaller than 1. In order to prove (101) we assume, without loss of generality, that $y = 0$ and that the tangent plane to $u(\Omega)$ at y is $\{x_3 = 0\}$. Denote by π the projection on $\{x_3 = 0\}$. By [27] there is a constant $\lambda > 0$ (depending only on $\|g\|_{C^{2,\beta}}$, $\|\kappa\|_{C^0}$ and $\|\kappa^{-1}\|_{C^0}$) with the following property.

(Est) Let U be an open convex set such that $U \cap u(\partial\Omega) = \emptyset$, $\mathrm{diam}(U) \le c_0$ and $U \cap u(\Omega)$ is locally convex. Then $U \cap u(\Omega)$ is the graph of a function $f : \pi(u(\Omega) \cap U) \to \mathbb{R}$ with $\|f\|_{C^{2,1/2}} \le \lambda^{-1}$ and $D^2 f \ge \lambda \mathrm{Id}$.

We now look for sets U as in (Est) with the additional property that $U = V \times]-a, a[$ and $f|_{\partial V} = a$ (see Fig. 2). Let U_m be the maximal set of this form for which the assumptions of (Est) hold. We claim that, either $\partial U_m \cap u(\partial\Omega) \neq \emptyset$, or $\mathrm{diam}(U_m) = c_0$. By (Est), this claim easily implies (101). To prove the claim, assume by contradiction that it is wrong and let $U_m = W_m \times]-a_m, a_m[$ be the maximal set. Let $\gamma = \partial U_m \cap u(\Omega)$. By the choice of c_0, γ is necessarily the curve $\partial W_m \times \{a\}$. On the other hand, by the estimates of (Est), it follows that every tangent plane to $u(\Omega)$ at a point of γ is transversal to $\{x_3 = 0\}$. So, for a sufficiently small $\varepsilon > 0$, the intersection $\{x_3 = a_m + \varepsilon\} \cap u(\Omega)$ contains a curve γ' bounding a connected region $D \subset u(\Omega)$ which contains $u(\Omega) \cap U_m$. By Theorem 8 of page 650 in [26], D is a convex set. This easily shows that U_m was not maximal. □

References

1. Borisov, J.F.: The parallel translation on a smooth surface. I. Vestn. Leningr. Univ. **13**(7), 160–171 (1958)
2. Borisov, J.F.: The parallel translation on a smooth surface. II. Vestn. Leningr. Univ. **13**(19), 45–54 (1958)
3. Borisov, J.F.: On the connection between the spatial form of smooth surfaces and their intrinsic geometry. Vestn. Leningr. Univ. **14**(13), 20–26 (1959)
4. Borisov, J.F.: The parallel translation on a smooth surface. III. Vestn. Leningr. Univ. **14**(1), 34–50 (1959)
5. Borisov, J.F.: On the question of parallel displacement on a smooth surface and the connection of space forms of smooth surfaces with their intrinsic geometries. Vestn. Leningr. Univ. **15**(19), 127–129 (1960)
6. Borisov, J.F.: $C^{1,\alpha}$-isometric immersions of Riemannian spaces. Doklady **163**, 869–871 (1965)
7. Borisov, Y.F.: Irregular $C^{1,\beta}$-surfaces with analytic metric. Sib. Mat. Zh. **45**(1), 25–61 (2004)
8. Cohn-Vossen, S.: Zwei Sätze über die Starrheit der Eiflächen. Nachr. Gött. **1927**, 125–137 (1927)
9. Constantin, P., E, W., Titi, E.S.: Onsager's conjecture on the energy conservation for solutions of Euler's equation. Commun. Math. Phys. **165**(1), 207–209 (1994)
10. De Lellis, C., Székelyhidi, L.J.: The Euler equations as a differential inclusion. Ann. of Math. (2) **170**(3), 1417–1436 (2009) (English summary)
11. De Lellis, C., Székelyhidi, L.J.: On admissibility criteria for weak solutions of the Euler equations. Arch. Ration. Mech. Anal. **195**(1), 225–260 (2010) (English summary)
12. Eliashberg, Y., Mishachev, N.: Introduction to the h-Principle. Graduate Studies in Mathematics, vol. 48. Am. Math. Soc., Providence (2002)
13. Eyink, G.L.: Energy dissipation without viscosity in ideal hydrodynamics. I. Fourier analysis and local energy transfer. Physica D **78**(3–4), 222–240 (1994)
14. Frisch, U.: Turbulence. The Legacy of A.N. Kolmogorov, p. 296. Cambridge University Press, Cambridge (1995). ISBN:0-521-45103-5
15. Gromov, M.: Convex integration of differential relations. Izv. Akad. Nauk USSR **37**, 329–343 (1973)
16. Gromov, M.: Partial Differential Relations. Springer, Berlin (1986)
17. Herglotz, G.: Über die Starrheit der Eiflächen. Abh. Math. Semin. Hansische Univ. **15**, 127–129 (1943)
18. Jacobowitz, H.: Implicit function theorems and isometric embeddings. Ann. of Math. (2) **95**, 191–225 (1972)
19. Källén, A.: Isometric embedding of a smooth compact manifold with a metric of low regularity. Ark. Mat. **16**(1), 29–50 (1978)
20. Kuiper, N.: On C^1 isometric imbeddings I. Proc. Kon. Acad. Wet. Amsterdam A **58**, 545–556 (1955)
21. Kuiper, N.: On C^1 isometric imbeddings II. Proc. Kon. Acad. Wet. Amsterdam A **58**, 683–689 (1955)
22. Nash, J.: C^1 isometric imbeddings. Ann. Math. **60**, 383–396 (1954)
23. Nash, J.: The imbedding problem for Riemannian manifolds. Ann. Math. **63**, 20–63 (1956)
24. Onsager, L.: Statistical hydrodynamics. Nuovo Cimento (9) **6**(Supplemento, 2(Convegno Internazionale di Meccanica Statistica)), 279–287 (1949)
25. Pogorelov, A.V.: The rigidity of general convex surfaces. Dokl. Akad. Nauk SSSR **79**, 739–742 (1951)

26. Pogorelov, A.V.: Extrinsic Geometry of Convex Surfaces. Translations of Mathematical Monographs, vol. 35. Am. Math. Soc., Providence (1973)
27. Sabitov, I.H.: Regularity of convex domains with a metric that is regular on Hölder classes. Sib. Mat. Zh. **17**(4), 907–915 (1976)
28. Scheffer, V.: An inviscid flow with compact support in space-time. J. Geom. Anal. **3**(4), 343–401 (1993)
29. Shnirelman, A.: On the nonuniqueness of weak solution of the Euler equation. Commun. Pure Appl. Math. **50**(12), 1261–1286 (1997)
30. Shnirelman, A.: Weak solutions with decreasing energy of incompressible Euler equations. Commun. Math. Phys. **210**(3), 541–603 (2000)
31. Spivak, M.: A Comprehensive Introduction to Differential Geometry, vol. V, 2nd edn. Publish or Perish, Berkeley (1979)
32. Spring, D.: On the regularity of solutions in convex integration theory. Invent. Math. **104**(1), 165–178 (1991). doi:10.1007/BF01245070
33. Spring, D.: Convex Integration Theory. Birkhäuser, Basel (1998)
34. Yau, S.-T.: Open problems in geometry. In: Differential Geometry: Partial Differential Equations on Manifolds, Los Angeles, CA, 1990. Proc. Sympos. Pure Math., vol. 54, pp. 1–28. Am. Math. Soc., Providence (1993)

About Existence, Symmetry and Symmetry Breaking for Extremal Functions of Some Interpolation Functional Inequalities

Jean Dolbeault and Maria J. Esteban

Abstract This chapter is devoted to a review of some recent results on existence, symmetry and symmetry breaking of optimal functions for Caffarelli–Kohn–Nirenberg (CKN) and weighted logarithmic Hardy (WLH) inequalities. These results have been obtained in a series of papers (Dolbeault et al. in Ann. Sc. Norm. Super. Pisa, Cl. Sci. (5) 7(2):313–341, 2008; Dolbeault et al. in Adv. Nonlinear Stud. 9(4):713–726, 2009; Dolbeault, Esteban in Extremal functions for Caffarelli–Kohn–Nirenberg and logarithmic Hardy inequalities Calc. Var. Partial Differ. Equ., 2010; del Pino et al. in J. Funct. Anal. 259(8):2045–2072, 2010; Dolbeault et al. in Radial symmetry and symmetry breaking for some interpolation inequalities, preprint, 2010) in collaboration with M. del Pino, S. Filippas, M. Loss, G. Tarantello and A. Tertikas. Here we put the highlights on a symmetry breaking result: extremals of some inequalities are not radially symmetric in regions where the symmetric extremals are linearly stable. Special attention is paid to the study of the critical cases for (CKN) and (WLH).

1 Two Families of Interpolation Inequalities

For any dimension $d \in \mathbb{N}^*$ and any $\theta \in [0,1]$, let us consider the set $\mathscr{D}$ of all smooth functions which are compactly supported in $\mathbb{R}^d \setminus \{0\}$. Define the numbers

$$\vartheta(p,d) := \frac{d\,(p-2)}{2\,p}, \qquad a_c := \frac{d-2}{2},$$

$$\Lambda(a) := (a-a_c)^2 \quad \text{and} \quad p(a,b) := \frac{2d}{d-2+2(b-a)}.$$

J. Dolbeault · M.J. Esteban (✉)
Ceremade, UMR CNRS No. 7534, Université Paris-Dauphine, Place de Lattre de Tassigny, 75775 Paris Cédex 16, France
e-mail: esteban@ceremade.dauphine.fr

J. Dolbeault
e-mail: dolbeaul@ceremade.dauphine.fr

H. Holden, K.H. Karlsen (eds.), *Nonlinear Partial Differential Equations*, Abel Symposia 7, DOI 10.1007/978-3-642-25361-4_6,

We shall also set $2^* := \frac{2d}{d-2}$ if $d \geq 3$ and $2^* := \infty$ if $d = 1$ or 2. For any $a < a_c$, we consider the following two families of interpolation inequalities, which have been introduced in [1, 4]:

(CKN) *Caffarelli–Kohn–Nirenberg inequalities*—Let $b \in (a + 1/2, a + 1]$ and $\theta \in (1/2, 1]$ if $d = 1$, $b \in (a, a + 1]$ if $d = 2$ and $b \in [a, a + 1]$ if $d \geq 3$. Assume that $p = p(a, b)$, and $\theta \in [\vartheta(p, d), 1]$ if $d \geq 2$. Then, there exists a finite positive constant $\mathsf{C}_{\mathrm{CKN}}(\theta, p, a)$ such that, for any $u \in \mathscr{D}$,

$$\| |x|^{-b} u\|^2_{\mathrm{L}^p(\mathbb{R}^d)} \leq \mathsf{C}_{\mathrm{CKN}}(\theta, p, a) \| |x|^{-a} \nabla u\|^{2\theta}_{\mathrm{L}^2(\mathbb{R}^d)} \| |x|^{-(a+1)} u\|^{2(1-\theta)}_{\mathrm{L}^2(\mathbb{R}^d)}.$$

(WLH) *Weighted logarithmic Hardy inequalities*—Let $\gamma \geq d/4$ and $\gamma > 1/2$ if $d = 2$. There exists a positive constant $\mathsf{C}_{\mathrm{WLH}}(\gamma, a)$ such that, for any $u \in \mathscr{D}$, normalized by $\| |x|^{-(a+1)} u\|_{\mathrm{L}^2(\mathbb{R}^d)} = 1$,

$$\int_{\mathbb{R}^d} \frac{|u|^2 \log(|x|^{2(a_c-a)}|u|^2)}{|x|^{2(a+1)}}\, dx \leq 2\gamma \log\big[\mathsf{C}_{\mathrm{WLH}}(\gamma, a) \| |x|^{-a} \nabla u\|^2_{\mathrm{L}^2(\mathbb{R}^d)}\big].$$

(WLH) appears as a limiting case of (CKN) in the limit $\theta = \gamma(p-2)$, $p \to 2_+$. See [4, 5] for details. By a standard completion argument, these inequalities can be extended to the set

$$\mathscr{D}^{1,2}_a(\mathbb{R}^d) := \{u \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^d) : |x|^{-a} \nabla u \in \mathrm{L}^2(\mathbb{R}^d) \text{ and } |x|^{-(a+1)} u \in \mathrm{L}^2(\mathbb{R}^d)\}.$$

In the sequel, we shall assume that all constants in the inequalities are taken with their optimal values. For brevity, we shall call *extremals* the functions which realize equality in (CKN) or in (WLH).

Let $\mathsf{C}^*_{\mathrm{CKN}}(\theta, p, a)$ and $\mathsf{C}^*_{\mathrm{WLH}}(\gamma, a)$ denote the optimal constants when admissible functions are restricted to the set of radially symmetric functions. *Radial extremals*, that is, the extremals of the above inequalities when restricted to the set of radially symmetric functions, are explicit and the values of the constants, $\mathsf{C}^*_{\mathrm{CKN}}(\theta, p, a)$ and $\mathsf{C}^*_{\mathrm{WLH}}(\gamma, a)$, are known. According to [4], we have:

(CKN*) *Radial Caffarelli–Kohn–Nirenberg inequalities*:

$$\frac{\mathsf{C}^*_{\mathrm{CKN}}(\theta, p, a)}{|\mathbb{S}^{d-1}|^{-\frac{p-2}{p}}} = \left[\frac{(a-a_c)^2(p-2)^2}{2+(2\theta-1)p}\right]^{\frac{p-2}{2p}} \left[\frac{2+(2\theta-1)p}{2p\theta(a-a_c)^2}\right]^{\theta} \times \left[\frac{4}{p+2}\right]^{\frac{6-p}{2p}} \left[\frac{\Gamma(\frac{2}{p-2}+\frac{1}{2})}{\sqrt{\pi}\,\Gamma(\frac{2}{p-2})}\right]^{\frac{p-2}{p}},$$

and if $\theta > \vartheta(p, 1)$, the best constant is achieved by an optimal radial function u such that $u(r) = r^{a-a_c}\,\overline{w}(-\log r)$, where $\overline{w}$ is unique up to multiplication by constants and translations in s, and given by

$$\overline{w}(s) = \big(\cosh(\lambda s)\big)^{-\frac{2}{p-2}}, \quad \text{with } \lambda = \frac{1}{2}(p-2)\left[\frac{(a-a_c)^2(p+2)}{2+(2\theta-1)p}\right]^{\frac{1}{2}}.$$

(WLH*) *Radial weighted logarithmic Hardy inequalities*:

$$\mathsf{C}^*_{\rm WLH}(\gamma,a)=\begin{cases}\frac{1}{4\gamma}\,\frac{[\Gamma(\frac{d}{2})]^{\frac{1}{2\gamma}}}{(2\pi^{d+1}e)^{\frac{1}{4\gamma}}}\,(\frac{4\gamma-1}{(a-a_c)^2})^{\frac{4\gamma-1}{4\gamma}} & \text{if } \gamma>\frac{1}{4},\\ \mathsf{C}^*_{\rm WLH}=\frac{[\Gamma(\frac{d}{2})]^2}{2\pi^{d+1}e} & \text{if } \gamma=\frac{1}{4},\end{cases}$$

and if $\gamma>\frac{1}{4}$, equality in the weighted logarithmic Hardy inequality is achieved by an optimal radial function u such that $u(r)=r^{a-a_c}\,w(-\log r)$, where

$$w(s)=\frac{\tilde{w}(s)}{\int_{\mathscr{C}}\tilde{w}^2\,dy}\quad\text{and}\quad\tilde{w}(s)=\exp\left(-\frac{(a-a_c)^2s^2}{(4\gamma-1)}\right).$$

Moreover we have

$$\begin{aligned}&\mathsf{C}_{\rm CKN}(\theta,p,a)\ge\mathsf{C}^*_{\rm CKN}(\theta,p,a)=\mathsf{C}^*_{\rm CKN}(\theta,p,a_c-1)\,\Lambda(a)^{\frac{p-2}{2p}-\theta},\\&\mathsf{C}_{\rm WLH}(\gamma,a)\ge\mathsf{C}^*_{\rm WLH}(\gamma,a)=\mathsf{C}^*_{\rm WLH}(\gamma,a_c-1)\,\Lambda(a)^{-1+\frac{1}{4\gamma}},\end{aligned}\tag{1}$$

where in both cases, the inequalities follow from the definitions. Radial symmetry for the extremals of (CKN) and (WLH) implies that $\mathsf{C}_{\rm CKN}(\theta,p,a)=\mathsf{C}^*_{\rm CKN}(\theta,p,a)$ and $\mathsf{C}_{\rm WLH}(\gamma,a)=\mathsf{C}^*_{\rm WLH}(\gamma,a)$, while *symmetry breaking* means that inequalities in (1) are strict. As we shall see later, there are cases where $\mathsf{C}_{\rm CKN}(\theta,p,a)=\mathsf{C}^*_{\rm CKN}(\theta,p,a)$ and for which radial and non-radial extremal functions coexist. This may happen only for the limiting value of a beyond which the equality does not hold anymore. On the contrary, when $\mathsf{C}_{\rm CKN}(\theta,p,a)>\mathsf{C}^*_{\rm CKN}(\theta,p,a)$, none of the extremals of (CKN) is radially symmetric.

Section 2 is devoted to the attainability of the best constants in the above inequalities. In Sect. 3 we describe the best available symmetry breaking results. In Sect. 4 we give some plots and also prove some new asymptotic results in the limit $p\to 2_+$.

2 Existence of Extremals

In this section, we describe the set of parameters for which the inequalities are achieved. The following result is taken from [5].

Theorem 1 (Existence based on *a priori* estimates) *Equality in* (*CKN*) *is attained for any* $p\in(2,2^*)$ *and* $\theta\in(\vartheta(p,d),1)$ *or for* $\theta=\vartheta(p,d)$, $d\ge 2$, *and* $a\in(a_\star,a_c)$, *for some* $a_\star<a_c$. *It is not attained if* $p=2$, *or* $a<0$, $p=2^*$, $\theta=1$ *and* $d\ge 3$, *or* $d=1$ *and* $\theta=\vartheta(p,1)$.

Equality in (*WLH*) *is attained if* $\gamma\ge 1/4$ *and* $d=1$, *or* $\gamma>1/2$ *if* $d=2$, *or for* $d\ge 3$ *and either* $\gamma>d/4$ *or* $\gamma=d/4$ *and* $a\in(a_{\star\star},a_c)$, *for some* $a_{\star\star}<a_c$.

A complete proof of these results is given in [5]. In the sequel, we shall only give some indications on how they are established.

First of all, it is very convenient to reformulate (CKN) and (WLH) inequalities in cylindrical variables. By means of the Emden–Fowler transformation

$$s=\log|x|\in\mathbb{R},\qquad \omega=x/|x|\in\mathbb{S}^{d-1},\qquad y=(s,\omega),\qquad v(y)=|x|^{a_c-a}u(x),$$

inequality (CKN) for u is equivalent to a Gagliardo–Nirenberg–Sobolev inequality on the cylinder $\mathcal{C}:=\mathbb{R}\times\mathbb{S}^{d-1}$ for v, namely

$$\|v\|^2_{\mathrm{L}^p(\mathcal{C})}\le \mathsf{C}_{\mathrm{CKN}}(\theta,p,a)\big(\|\nabla v\|^2_{\mathrm{L}^2(\mathcal{C})}+\Lambda\|v\|^2_{\mathrm{L}^2(\mathcal{C})}\big)^\theta\|v\|^{2(1-\theta)}_{\mathrm{L}^2(\mathcal{C})}\quad\forall v\in\mathrm{H}^1(\mathcal{C})$$

with $\Lambda=\Lambda(a)$. Similarly, with $w(y)=|x|^{a_c-a}u(x)$, inequality (WLH) is equivalent to

$$\int_{\mathcal{C}}|w|^2\log|w|^2\,dy\le 2\gamma\log\big[\mathsf{C}_{\mathrm{WLH}}(\gamma,a)(\|\nabla w\|^2_{\mathrm{L}^2(\mathcal{C})}+\Lambda)\big],$$

for any $w\in\mathrm{H}^1(\mathcal{C})$ such that $\|w\|_{\mathrm{L}^2(\mathcal{C})}=1$. *Notice that radial symmetry for u means that v and w depend only on s*. For brevity, we shall call them *s-symmetric functions*.

On $\mathrm{H}^1(\mathcal{C})$, consider the functional

$$\mathcal{E}^p_{\theta,\Lambda}[v]:=\big(\|\nabla v\|^2_{\mathrm{L}^2(\mathcal{C})}+\Lambda\|v\|^2_{\mathrm{L}^2(\mathcal{C})}\big)^\theta\|v\|^{2(1-\theta)}_{\mathrm{L}^2(\mathcal{C})}.$$

Assume that $d\ge3$, let $t:=\|\nabla v\|^2_{\mathrm{L}^2(\mathcal{C})}/\|v\|^2_{\mathrm{L}^2(\mathcal{C})}$ and $\Lambda=\Lambda(a)$. If v is a minimizer of $\mathcal{E}^p_{\theta,\Lambda}[v]$ such that $\|v\|_{\mathrm{L}^p(\mathcal{C})}=1$, then, as in [5], we have

$$\begin{aligned}(t+\Lambda)^\theta=\mathcal{E}^p_{\theta,\Lambda}[v]\frac{\|v\|^2_{\mathrm{L}^p(\mathcal{C})}}{\|v\|^2_{\mathrm{L}^2(\mathcal{C})}}&=\frac{\|v\|^2_{\mathrm{L}^p(\mathcal{C})}}{\mathsf{C}_{\mathrm{CKN}}(\theta,p,a)\|v\|^2_{\mathrm{L}^2(\mathcal{C})}}\\&\le\frac{\mathsf{S}_d^{\vartheta(p,d)}}{\mathsf{C}_{\mathrm{CKN}}(\theta,p,a)}(t+a_c^2)^{\vartheta(p,d)}\end{aligned}\tag{2}$$

where $\mathsf{S}_d=\mathsf{C}_{\mathrm{CKN}}(1,2^*,0)$ is the optimal Sobolev constant, while we know from (1), that $\lim_{a\to a_c}\mathsf{C}_{\mathrm{CKN}}(\theta,p,a)=\infty$ if $d\ge2$. This provides a bound on t if $\theta>\vartheta(p,d)$.

Consider now a sequence $(v_n)_n$ of functions in $\mathrm{H}^1(\mathcal{C})$, which minimizes $\mathcal{E}^p_{\theta,\Lambda}[v]$ under the constraint $\|v\|_{\mathrm{L}^p(\mathcal{C})}=1$. Assume therefore that $\|v_n\|_{\mathrm{L}^p(\mathcal{C})}=1$ for any $n\in\mathbb{N}$. If $\theta>\vartheta(p,d)$, an estimate similar to (2) asymptotically holds for $(v_n)_n$, thus providing bounds on $t_n:=\|\nabla v_n\|_{\mathrm{L}^2(\mathcal{C})}/\|v_n\|_{\mathrm{L}^2(\mathcal{C})}$ and $\|v_n\|_{\mathrm{H}^1(\mathcal{C})}$, for n large enough.

Then, standard tools of the concentration-compactness method allow to conclude that $(v_n)_n$ is relatively compact and converges up to translations and the extraction of a subsequence towards a minimizer of $\mathcal{E}^p_{\theta,\Lambda}$. The only specific idea concerning the use of concentration-compactness in this context relies on the use of the following inequality: for any $x,y>0$ and any $\eta\in(0,1)$,

$$(1+x)^\eta(1+y)^{1-\eta}\ge1+x^\eta y^{1-\eta},\quad\text{with strict inequality unless } x=y.$$

A similar approach holds for (CKN) if $d = 2$.

In the case of (WLH), for $\gamma > d/4$, the method of proof is similar to that of (CKN). The energy functional to be considered is now

$$\mathscr{F}_\gamma[w] := (\|\nabla w\|^2_{\mathrm{L}^2(\mathscr{C})} + \Lambda)\exp\left[-\frac{1}{2\gamma}\int_{\mathscr{C}} |w|^2 \log|w|^2\,dy\right].$$

If w is a minimizer of $\mathscr{F}_\gamma[w]$ under the constraint $\|w\|_{\mathrm{L}^2(\mathscr{C})} = 1$, then we have

$$\frac{t+\Lambda}{[\mathsf{C}_{\mathrm{CKN}}(1,p,\alpha)(t+\Lambda(\alpha))]^{\frac{1}{2\gamma}\frac{p}{p-2}}} \le \mathscr{F}_\gamma[w] = \frac{1}{\mathsf{C}_{\mathrm{WLH}}(\gamma,a)} \le \frac{\Lambda(a)^{1-\frac{1}{4\gamma}}}{\mathsf{C}^*_{\mathrm{WLH}}(\gamma,a_c-1)} \tag{3}$$

for an arbitrary $\alpha < a_c$. The concentration-compactness method applies using the following inequality: for any $x, y > 0$ and $\eta \in (0,1)$,

$$\eta x^{1/\eta} + (1-\eta)y^{1/(1-\eta)} \ge xy,$$

with strict inequality unless $x = y$ and $\eta = 1/2$.

Let us now consider the *critical case* $\theta = \vartheta(p,d)$ for (CKN). Estimate (3) still provides *a priori* bounds for minimizing sequences whenever $a \in (a_1, a_c)$ where a_1 can be obtained as follows. When $\theta = \vartheta(p,d)$, we can rewrite (2) as

$$(t+\Lambda) \le K(t+a_c^2) \quad \text{where } \mathsf{K} = \frac{\mathsf{S}_d^\theta}{\mathsf{C}_{\mathrm{CKN}}(\theta,p,a)}.$$

Hence we can deduce that $0 \le t \le \frac{\mathsf{K}a_c^2-\Lambda}{1-\mathsf{K}}$, if $\mathsf{K} < 1$ and $\Lambda \le \mathsf{K}a_c^2$. These two inequalities define the constant

$$\Lambda_1 := \min\left\{\left(\frac{\mathsf{C}^*_{\mathrm{CKN}}(\theta,p,a_c-1)^{1/\theta}}{\mathsf{S}_d}\right)^{\frac{d}{d-1}}, \left(\frac{a_c^2\mathsf{S}_d}{\mathsf{C}^*_{\mathrm{CKN}}(\theta,p,a_c-1)^{1/\theta}}\right)^d\right\}, \tag{4}$$

so that t is bounded if $a \in (a_1, a_c)$ with $a_1 := a_c - \sqrt{\Lambda_1}$. See [5] for more details.

Such an estimate is not anymore available in the *critical case* for (WLH), that is, if $\gamma = d/4$, $d \ge 3$. We may indeed notice that $p \le 2^*$ and $\gamma = d/4$ mean $1 - \frac{1}{2\gamma}\frac{p}{p-2} \le 0$. A more detailed analysis of the possible losses of compactness is therefore necessary. This can actually be done in the two critical cases, $\theta = \vartheta(p,d)$ for (CKN) and $\gamma = d/4$, $d \ge 3$, for (WLH).

Let $\mathsf{C}_{\mathrm{GN}}(p)$ be the optimal constant in the Gagliardo–Nirenberg–Sobolev interpolation inequalities

$$\|u\|^2_{\mathrm{L}^p(\mathbb{R}^d)} \le \mathsf{C}_{\mathrm{GN}}(p)\|\nabla u\|^{2\vartheta(p,d)}_{\mathrm{L}^2(\mathbb{R}^d)}\|u\|^{2(1-\vartheta(p,d))}_{\mathrm{L}^2(\mathbb{R}^d)} \quad \forall u \in \mathrm{H}^1(\mathbb{R}^d)$$

with $p \in (2, 2^*)$ if $d = 2$ or $p \in (2, 2^*]$ if $d \ge 3$. Also consider Gross' logarithmic

Sobolev inequality in Weissler's form (see [11])

$$\int_{\mathbb{R}^d} |u|^2 \log |u|^2 \, dx \le \frac{d}{2} \log\left(\mathsf{C}_{\rm LS} \|\nabla u\|^2_{\mathrm{L}^2(\mathbb{R}^d)}\right) \tag{5}$$

for any $u \in \mathrm{H}^1(\mathbb{R}^d)$ such that $\|u\|_{\mathrm{L}^2(\mathbb{R}^d)} = 1$, with optimal constant $\mathsf{C}_{\rm LS} := \frac{2}{\pi d e}$.

The Gaussian function $u(x) = \frac{1}{(2\pi)^{d/4}} e^{-|x|^2/4}$ is an extremal for (5). By taking $u_n(x) := u(x + n\mathsf{e})$ for some $\mathsf{e} \in \mathbb{S}^{d-1}$ and any $n \in \mathbb{N}$ as test functions for (WLH), and letting $n \to +\infty$, we find that

$$\mathsf{C}_{\rm LS} \le \mathsf{C}_{\rm WLH}(d/4, a).$$

If equality holds, this is a mechanism of loss of compactness for minimizing sequences. On the opposite, if $\mathsf{C}_{\rm LS} < \mathsf{C}_{\rm WLH}(d/4, a)$, we can establish a compactness result (see Theorem 2 below) which proves that, for some $a_{\star\star} < a_c$, equality is attained in (WLH) in the critical case $\gamma = d/4$ for any $a \in (a_{\star\star}, a_c)$. Indeed, we know that $\lim_{a \to a_c} \mathsf{C}_{\rm WLH}(d/4, a) = \lim_{a \to a_c} \mathsf{C}^*_{\rm WLH}(d/4, a) = \infty$.

A similar analysis for (CKN) shows that

$$\mathsf{C}_{\rm GN}(p) \le \mathsf{C}_{\rm CKN}(\theta, p, a)$$

in the critical case $\theta = \vartheta(p, d)$. Exactly as for (WLH), we also have an existence result, which has been established in [5], if $\mathsf{C}_{\rm GN}(p) < \mathsf{C}_{\rm CKN}(\theta, p, a)$.

Theorem 2 (Existence in the critical cases) *With the above notations,*

(i) *if* $\theta = \vartheta(p, d)$ *and* $\mathsf{C}_{\rm GN}(p) < \mathsf{C}_{\rm CKN}(\theta, p, a)$, *then* (*CKN*) *admits an extremal function in* $\mathcal{D}^{1,2}_a(\mathbb{R}^d)$,

(ii) *if* $\gamma = d/4$, $d \ge 3$, *and* $\mathsf{C}_{\rm LS} < \mathsf{C}_{\rm WLH}(\gamma, a)$, *then* (*WLH*) *admits an extremal function in* $\mathcal{D}^{1,2}_a(\mathbb{R}^d)$. *Additionally, if* $a \in (a^{\rm WLH}_{\star\star}, a_c)$, *then* $\mathsf{C}_{\rm LS} < \mathsf{C}_{\rm WLH}(d/4, a)$ *where* $a^{\rm WLH}_{\star\star}$ *is defined by*

$$a^{\rm WLH}_{\star\star} := a_c - \sqrt{\Lambda^{\rm WLH}_{\star\star}} \quad \text{and} \quad \Lambda^{\rm WLH}_{\star\star} := (d-1) e \left[\frac{\Gamma(\frac{d}{2})^2}{2^{d+1}\pi}\right]^{\frac{1}{d-1}}.$$

The values of $\mathsf{C}_{\rm GN}(p)$ and $\mathsf{C}_{\rm CKN}(\vartheta(p,d), p, a)$ are not explicitly known if $d \ge 2$, so we cannot get an explicit interval of existence in terms of a for (CKN). The strict inequality of Theorem 2(i) holds if $\mathsf{C}_{\rm GN}(p) < \mathsf{C}^*_{\rm CKN}(\vartheta(p,d), p, a)$ since we know that $\mathsf{C}^*_{\rm CKN}(\vartheta(p,d), p, a) \le \mathsf{C}_{\rm CKN}(\vartheta(p,d), p, a)$. The condition $\mathsf{C}_{\rm GN}(p) = \mathsf{C}^*_{\rm CKN}(\vartheta(p,d), p, a)$ defines a number $a^{\rm CKN}_\star$ for which existence is granted if $a \in (a^{\rm CKN}_\star, a_c)$, hence proving that $a_\star \le a^{\rm CKN}_\star$ (if we consider the lowest possible value of $a_\star$ in Theorem 1). Still we do not know the explicit value of $\mathsf{C}_{\rm GN}(p)$, but, since the computation of a_1 only involves the optimal constants among radial functions, at least we know that $a^{\rm CKN}_\star \le a_1 < a_c$.

On the opposite, we know the explicit values of $\mathsf{C}_{\rm LS}$ and $\mathsf{C}^*_{\rm WLH}(d/4)$, so that the computation of the value of $a^{\rm WLH}_\star$, which is determined by the condition $\mathsf{C}_{\rm LS} = \mathsf{C}^*_{\rm WLH}(d/4)$, is tedious but explicit.

We may observe from the expression of (CKN) and (WLH) when they are written on the cylinder (after the Emden–Fowler transformation) that $\mathsf{C}_{\mathrm{CKN}}$ and $\mathsf{C}_{\mathrm{WLH}}$ are monotone non-decreasing functions of a in $(-\infty, a_c)$, and actually increasing if there is an extremal. As long as it is finite, the optimal function $a_\star$ in Theorem 1 is continuous as a function of p, as a consequence of Theorem 2 and of the compactness of minimizing sequences. So, finally, $a_\star$ and $a_{\star\star}$ can be chosen such that $a_\star \le a_\star^{\mathrm{CKN}}$ and $a_{\star\star} \le a_{\star\star}^{\mathrm{WLH}}$. It is not difficult to observe that $a_\star^{\mathrm{CKN}}$ can be seen as a continuous, but not explicit, function of p and we shall see later (in Corollary 1, below) that $\lim_{p\to 2_+} a_\star^{\mathrm{CKN}}(p) = a_{\star\star}^{\mathrm{WLH}}$.

Next, note that if $\mathsf{C}_{\mathrm{CKN}} = \mathsf{C}^*_{\mathrm{CKN}}$ is known, then there are radially symmetric extremals, whose existence has been established in [4]. Anticipating on the results of the next section, we can state the following result which arises as a consequence of the Schwarz symmetrization method (see Theorem 4, below, and [8]).

Proposition 1 (Existence of radial extremals) *Let $d \ge 3$. Then (CKN) with $\theta = \vartheta(p,d)$ admits a radial extremal if $a \in [a_0, a_c)$ where $a_0 := a_c - \sqrt{\Lambda_0}$ and $\Lambda = \Lambda_0$ is defined by the condition*

$$\Lambda^{(d-1)/d} = \vartheta(p,d)\mathsf{C}^*_{\mathrm{CKN}}(\theta, p, a_c - 1)^{1/\vartheta(p,d)}/\mathsf{S}_d .$$

A similar estimate also holds if $\theta > \vartheta(p,d)$, with less explicit computations. See [8] for details.

The proof of this symmetry result follows from a not straightforward use of the Schwarz symmetrization. If $u(x) = |x|^a v(x)$, (CKN) is equivalent to

$$\| |x|^{a-b} v\|^2_{\mathrm{L}^p(\mathbb{R}^d)} \le \mathsf{C}_{\mathrm{CKN}}(\theta, p, \Lambda)(\mathscr{A} - \lambda\mathscr{B})^\theta \mathscr{B}^{1-\theta}$$

with $\mathscr{A} := \|\nabla v\|^2_{\mathrm{L}^2(\mathbb{R}^d)}$, $\mathscr{B} := \||x|^{-1} v\|^2_{\mathrm{L}^2(\mathbb{R}^d)}$ and $\lambda := a(2a_c - a)$.

We observe that the function $B \mapsto h(\mathscr{B}) := (\mathscr{A} - \lambda\mathscr{B})^\theta \mathscr{B}^{1-\theta}$ satisfies

$$\frac{h'(\mathscr{B})}{h(\mathscr{B})} = \frac{1-\theta}{\mathscr{B}} - \frac{\lambda\theta}{\mathscr{A} - \lambda\mathscr{B}} .$$

By Hardy's inequality ($d \ge 3$), we know that

$$\mathscr{A} - \lambda\mathscr{B} \ge \inf_{a>0} \big(\mathscr{A} - a(2a_c - a)\mathscr{B}\big) = \mathscr{A} - a_c^2 \mathscr{B} > 0,$$

and so $h'(\mathscr{B}) \le 0$ if $(1-\theta)\mathscr{A} < \lambda\mathscr{B}$, which is equivalent to $\mathscr{A}/\mathscr{B} < \lambda/(1-\theta)$. By interpolation $\mathscr{A}/\mathscr{B}$ is small if $a_c - a > 0$ is small enough, for $\theta > \vartheta(p,d)$ and $d \ge 3$. The precise estimate of when $\mathscr{A}/\mathscr{B}$ is smaller than $\lambda/(1-\theta)$ provides us with the definition of a_0.

3 Symmetry and Symmetry Breaking

Define

$$\underline{a}(\theta,p):=a_c-\frac{2\sqrt{d-1}}{p+2}\sqrt{\frac{2p\theta}{p-2}-1},\qquad \tilde a(\gamma):=a_c-\frac{1}{2}\sqrt{(d-1)(4\gamma-1)},$$

$$a_{\mathrm{SB}}:=a_c-\sqrt{\Lambda_{\mathrm{SB}}(\gamma)},$$
$$\Lambda_{\mathrm{SB}}(\gamma):=\frac{4\gamma-1}{8}e\left(\frac{\pi^{4\gamma-d-1}}{16}\right)^{\frac{1}{4\gamma-1}}\left(\frac{d}{\gamma}\right)^{\frac{4\gamma}{4\gamma-1}}\Gamma\left(\frac{d}{2}\right)^{\frac{2}{4\gamma-1}}$$

and take into account the definitions of $a_\star^{\mathrm{CKN}}$ and $a_{\star\star}^{\mathrm{WLH}}$ previously given. Thus we have the following result, which has been established in [4, 8].

Theorem 3 *Let $d\ge 2$ and $p\in(2,2^*)$. Symmetry breaking holds in (CKN) if either $a<\underline{a}(\theta,p)$ and $\theta\in[\vartheta(p,d),1]$, or $a<a_\star^{\mathrm{CKN}}$ and $\theta=\vartheta(p,d)$.*

Assume that $\gamma>1/2$ if $d=2$ and $\gamma\ge d/4$ if $d\ge 3$. Symmetry breaking holds in (WLH) if $a<\max\{\tilde a(\gamma),a_{\mathrm{SB}}\}$.

When $\gamma=d/4$, $d\ge 3$, we observe that $\Lambda_{\star\star}^{\mathrm{WLH}}=\Lambda_{\mathrm{SB}}(d/4)<\Lambda(\tilde a(d/4))$ with the notations of Theorem 1 and there is symmetry breaking if $a\in(-\infty,a_{\star\star}^{\mathrm{WLH}})$, in the sense that $\mathsf{C}_{\mathrm{WLH}}(d/4,a)>\mathsf{C}^*_{\mathrm{WLH}}(d/4,a)$ in that interval, although we do not know if extremals for (WLH) exist when $\gamma=d/4$ and $a<a_{\star\star}^{\mathrm{WLH}}$.

Concerning (CKN) with $\theta\ge\vartheta(p,d)$, results of symmetry breaking for $a<\underline{a}(\theta,p)$ have been established first in [3, 6, 9] when $\theta=1$ and later extended in [4] to $\theta<1$. The main idea in case of (CKN) is to consider the quadratic form associated to the second variation of $\mathscr{E}^p_{\theta,\Lambda}$, restricted to $\{v\in\mathrm{H}^1(\mathscr{C}):\|v\|_{\mathrm{L}^p(\mathscr{C})}=1\}$, around a minimizer among functions depending on s only and observe that the linear operator $\mathscr{L}^p_{\theta,\Lambda}$ associated to the quadratic form has a negative eigenvalue if $a<\underline{a}$.

Because of the homogeneity in (CKN), if v is an s-symmetric extremal, then λv is also an s-symmetric extremal for any $\lambda\in\mathbb{R}$ and v is therefore in the kernel of $\mathscr{L}^p_{\theta,\Lambda}$. When v generates $\mathrm{Ker}(\mathscr{L}^p_{\theta,\Lambda})$ and all non-zero eigenvalues are positive, that is for $a\in(\underline{a}(\theta,p),a_c)$, we shall say that v is *linearly* stable, without further precision. In such a case, the operator $\mathscr{L}^p_{\theta,\Lambda}$ has the property of *spectral gap*.

Results in [4] for (WLH), $a<\tilde a(\gamma)$, are based on the same method.

For any $a<a_\star^{\mathrm{CKN}}$, we have

$$\mathsf{C}^*_{\mathrm{CKN}}(\vartheta(p,d),p,a)<\mathsf{C}_{\mathrm{GN}}(p)\le\mathsf{C}_{\mathrm{CKN}}(\vartheta(p,d),p,a),$$

which proves symmetry breaking. Using well-chosen test functions, it has been proved [8] that $\underline{a}(\vartheta(p,d),p)<a_\star^{\mathrm{CKN}}$ for $p-2>0$, small enough, thus also proving symmetry breaking for $a-\underline{a}(\vartheta(p,d),p)>0$, small, and $\theta-\vartheta(p,d)>0$, small. *This shows that in some cases, symmetry can be broken even in regions where the* *radial extremals are linearly stable. In Sect.* 4, *we give a more quantitative result about this* (*see Corollary* 1).

Next we will describe how the set of parameters involved in our inequalities is cut into two subsets, both of them simply connected. They are separated by a continuous surface which isolates the symmetry region from the region of symmetry breaking. See [7, 8] for detailed statements and proofs.

Theorem 4 *For all* $d \ge 2$, *there exists a continuous function* a^* *defined on the set* $\{(\theta, p) \in (0,1] \times (2, 2^*) : \theta > \vartheta(p,d)\}$ *such that* $\lim_{p \to 2_+} a^*(\theta, p) = -\infty$ *with the property that* (*CKN*) *has only radially symmetric extremals if* $(a, p) \in (a^*(\theta, p), a_c) \times (2, 2^*)$, *and none of the extremals is radially symmetric if* $(a, p) \in (-\infty, a^*(\theta, p)) \times (2, 2^*)$.

Similarly, for all $d \ge 2$, *there exists a continuous function* $a^{**} : (d/4, \infty) \to (-\infty, a_c)$ *such that, for any* $\gamma > d/4$ *and* $a \in [a^{**}(\gamma), a_c)$, *there is a radially symmetric extremal for* (*WLH*), *while for* $a < a^{**}(\gamma)$ *no extremal is radially symmetric.*

We sketch below the main steps of the proof. First note that as previously explained (see [8] for details), the Schwarz symmetrization allows to characterize a nonempty subdomain of $(0, a_c) \times (0, 1) \ni (a, \theta)$ in which symmetry holds for extremals of (CKN), when $d \ge 3$. If $\theta = \vartheta(p, d)$ and $p > 2$, there are radially symmetric extremals if $a \in [a_0, a_c)$ where a_0 is given in Proposition 1.

Symmetry also holds if $a_c - a$ is small enough, for (CKN) as well as for (WLH), or when $p \to 2_+$ in (CKN), for any $d \ge 2$, as a consequence of the existence of the spectral gap of $\mathscr{L}^p_{\theta,\Lambda}$ when $a > \underline{a}(\theta, p)$.

According to [7, 8], for given θ and p, there is a unique $a^* \in (-\infty, a_c)$ for which there is symmetry breaking in $(-\infty, a^*)$ and for which all extremals are radially symmetric when $a \in (a^*, a_c)$. This follows from the observation that, if $v_\sigma(s, \omega) := v(\sigma s, \omega)$ for $\sigma > 0$, then the quantity

$$(\mathscr{E}^p_{\theta,\sigma^2\Lambda}[v_\sigma])^{1/\theta} - \sigma^{(2\theta-1+2/p)/\theta^2} (\mathscr{E}^p_{\theta,\Lambda}[v])^{1/\theta}$$

is equal to 0 if v depends only on s, while it has the sign of $\sigma - 1$ otherwise. The method also applies to (WLH) and gives a similar result for a^{**}.

From Theorem 3, we can infer that radial and non-radial extremals for (CKN) with $\theta > \vartheta(p, d)$ coexist on the threshold, in some cases.

4 Numerical Computations and Asymptotic Results for (CKN)

In the critical case for (CKN), that is for $\theta = \vartheta(p, d)$, numerical results illustrating our results on existence and on symmetry *versus* symmetry breaking have been collected in Figs. 1 and 2 below.

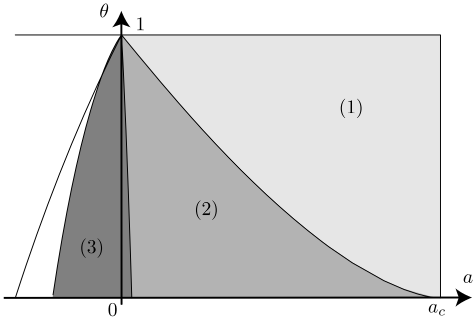


Fig. 1 Existence in the critical case for (CKN). Here we assume that $d = 5$

4.1 Existence for (CKN)

In Fig. 1, the zones in which existence is known are:

(1) $a \geq a_0$: extremals are achieved among radial functions, by the Schwarz symmetrization method (Proposition 1),

(1) + (2) $a > a_1$: this follows from the explicit *a priori* estimates (Theorem 1); see (4) for the definition of $\Lambda_1 = (a_c - a_1)^2$,

(1) + (2) + (3) $a > a_\star^{\mathrm{CKN}}$: this follows by comparison of the optimal constant for (CKN) with the optimal constant in the corresponding Gagliardo–Nirenberg–Sobolev inequality (Theorem 2).

4.2 Symmetry and Symmetry Breaking for (CKN)

In Fig. 2, the zone of symmetry breaking contains:

(1) $a < \underline{a}(\theta, p)$: by linearization around radial extremals (Theorem 3),

(1) + (2) $a < a_\star^{\mathrm{CKN}}$: by comparison with the Gagliardo–Nirenberg–Sobolev inequality (Theorem 3).

In (3) it is not known whether symmetry holds or if there is symmetry breaking, while in (4), that is, for $a_0 \leq a < a_c$, according to Proposition 1, symmetry holds by the Schwarz symmetrization.

4.3 When (CKN) Approaches (WLH)

In the critical case $\theta = \vartheta(p, d) = d(p-2)/(2p)$, when p approaches 2_+, it is possible to obtain detailed results for (CKN) and to compare (CKN) and (WLH), or at least get explicit results for the various curves of Figs. 1 and 2.

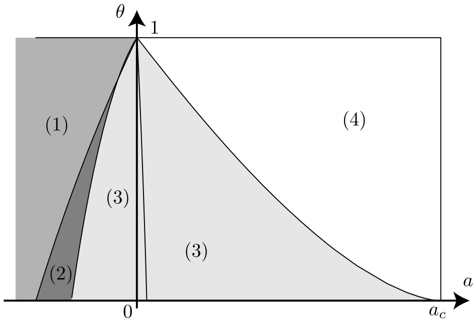


Fig. 2 Symmetry and symmetry breaking results in the critical case for (CKN). Here we assume that $d = 5$

(1) *Cases covered by the Schwarz symmetrization method.* With Λ_0 defined by $\Lambda_0^{(d-1)/d} = \vartheta(p,d)\mathsf{C}^*_{\mathrm{CKN}}(\theta, p, a_c - 1)^{1/\vartheta(p,d)}/\mathsf{S}_d$, since

$$\lim_{p\to 2_+} \mathsf{C}^*_{\mathrm{CKN}}(\theta, p, a_c - 1)^{1/\vartheta(p,d)} = \frac{(d-1)^{\frac{d-1}{d}}}{d(2e)^{1/d}\pi^{\frac{d+1}{d}}}\Gamma\left(\frac{d}{2}\right)^{2/d}$$

it follows that a_0 defined in Proposition 1 by $a_0 = a_c - \sqrt{\Lambda_0}$ converges to a_c as $p \to 2_+$.

(2) *Existence range obtained by a priori estimates.* The expression of $a_1 = a_c - \sqrt{\Lambda_1}$ is explicit for any p and $p \mapsto \Lambda_1(p)$ has a limit $\Lambda_1(2)$ as $p \to 2_+$, which is given by

$$\min\left\{\frac{1}{4}\left[\frac{2}{e}(d-2)^d(d-1)^{d-3}\left(\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d-1}{2})}\right)^2\right]^{\frac{1}{d-1}}, \frac{e}{8}\,\frac{(d-2)^d}{(d-1)^{d-3}}\left(\frac{\Gamma(\frac{d-1}{2})}{\Gamma(\frac{d}{2})}\right)^2\right\}.$$

A careful investigation shows that $\Lambda_1(2)$ is given by the first term in the above min. As a function of d, $a_c - \sqrt{\Lambda_1(2)}$ is monotone decreasing in $(3, \infty)$ and converges to 0_+ as $d \to \infty$. Moreover, for all $d \geq 2$, $\Lambda_1(2) \leq \Lambda_{\star\star}^{\mathrm{WLH}}$, since both estimates are done among radial functions and the latter is optimal among those.

(3) *Symmetry breaking range obtained by linearization around radial extremals.* Computations are explicit and it has already been observed in [4] that $\underline{a}(\theta, p)$ (see Theorem 3) is such that $\lim_{p\to 2_+} \underline{a}(\vartheta(p,d), p) = -1/2$.

(4) *Existence range obtained by comparison with Gagliardo–Nirenberg–Sobolev inequalities.* Although the value of $\mathsf{C}_{\mathrm{GN}}(p)$ is not known explicitly, we can get an estimate by using a Gaussian as a test function. This estimate turns out to be sharp as p approaches 2_+. More precisely, we get a lower bound for $\mathsf{C}_{\mathrm{GN}}(p)$ by computing

$$Q(p) := \frac{\|u_2\|^2_{\mathrm{L}^p(\mathbb{R}^d)}}{\|\nabla u_2\|^{2\vartheta(p,d)}_{\mathrm{L}^2(\mathbb{R}^d)}\|u_2\|^{2(1-\vartheta(p,d))}_{\mathrm{L}^2(\mathbb{R}^d)}}$$

with $u_2(x) := \pi^{-d/4} e^{-|x|^2/2}$, which is such that

$$\lim_{p\to 2_+} \frac{Q(p)-1}{p-2} = \frac{d}{4}\log \mathsf{C}_{\mathrm{LS}} = \frac{d}{4}\log\left(\frac{2}{\pi d e}\right) \le \lim_{p\to 2_+} \frac{\mathsf{C}_{\mathrm{GN}}(p)-1}{p-2}.$$

This estimate is not only a lower bound for the limit, but gives its exact value, as shown by the following new result.

Proposition 2 *With the above notations, we have*

$$\lim_{p\to 2_+} \frac{\mathsf{C}_{\mathrm{GN}}(p)-1}{p-2} = \frac{d}{4}\log \mathsf{C}_{\mathrm{LS}}.$$

Hence, in the regime $p \to 2_+$, the condition which defines $a = a_\star^{\mathrm{CKN}}$, namely the equality $\mathsf{C}_{\mathrm{GN}}(p) = \mathsf{C}^*_{\mathrm{CKN}}(\vartheta(p,d),p,a)$ leads to

$$\begin{aligned} &1 + \frac{d}{4}\log \mathsf{C}_{\mathrm{LS}}(p-2) + o(p-2) \\ &= \mathsf{C}_{\mathrm{GN}}(p) = \mathsf{C}^*_{\mathrm{CKN}}(\vartheta(p,d),p,a) \\ &= 1 + \frac{d}{4}\log \mathsf{C}^*_{\mathrm{WLH}}(d/4,a)(p-2) + o(p-2) \end{aligned}$$

(for the second line in the inequality, see [4, Lemma 4]), which asymptotically amounts to solve

$$\mathsf{C}^*_{\mathrm{WLH}}(d/4,a) = \mathsf{C}_{\mathrm{LS}}.$$

In other words, we have

$$\lim_{p\to 2_+} a_\star^{\mathrm{CKN}}(p) = a_{\star\star}^{\mathrm{WLH}}.$$

As a consequence, we have the following symmetry breaking result, which allows to refine an earlier result of [8] in the subcritical case and is new in the critical case.

Corollary 1 *Let $d \ge 2$ and $p \in (2,2^*)$. For p sufficiently close to 2_+, $\underline{a}(\vartheta(p,d), p) < a_\star^{\mathrm{CKN}}$, and so, there is symmetry breaking in a region where the radial extremals are linearly stable.*

Notice that the case $d = 2$ is not covered, for instance in Theorem 2(ii), but the computations can be justified after noticing that among radial functions, (WLH) also makes sense with $\gamma = d/2$ if $d = 2$ (see [4]). By symmetry breaking, we mean $\mathsf{C}^*_{\mathrm{CKN}}(\vartheta(p,d),p,a) < \mathsf{C}_{\mathrm{CKN}}(\vartheta(p,d),p,a)$, since existence of extremals is not known for $a < a_\star^{\mathrm{CKN}}$.

Proof of Proposition 2 Optimal functions for Gagliardo–Nirenberg–Sobolev inequalities are, up to translations, radial solutions of the Euler–Lagrange equations

$$-\Delta u = \mathsf{a} u^{p-1} - \mathsf{b} u \tag{6}$$

where a and b are two positive coefficients which can be chosen arbitrarily because of the invariance of the inequality under a multiplication by a positive constant and the invariance under scalings. As a special choice, we can impose

$$\mathsf{a} = \frac{2}{p-2} \quad \text{and} \quad \mathsf{b} = \frac{2}{p-2} - \frac{d}{2}(2 + \log \pi)$$

and denote by u_p the (unique) corresponding solution so that, by passing to the limit as $p \to 2_+$, we get the equation

$$-\Delta u = 2u \log u + \frac{d}{2}(2 + \log \pi) u .$$

Note that the function $u_2(x)$ is a positive radial solution of this equation in $\mathrm{H}^1(\mathbb{R}^d)$, which is normalized in $\mathrm{L}^2(\mathbb{R}^d)$: $\|u\|_{\mathrm{L}^2(\mathbb{R}^d)} = 1$. According to [2], it is an extremal function for the logarithmic Sobolev inequality: for any $u \in \mathrm{H}^1(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} |u|^2 \log\left(\frac{|u|^2}{\|u\|^2_{\mathrm{L}^2(\mathbb{R}^d)}}\right) dx + \frac{d}{2}(2 + \log \pi)\|u\|^2_{\mathrm{L}^2(\mathbb{R}^d)} \le \int_{\mathbb{R}^d} |\nabla u|^2 \, dx ,$$

which, after optimization under scalings, is equivalent to (5). Moreover, it is unique as can be shown by considering for instance the remainder integral term arising from the Bakry–Emery method (see for instance [10], and [2] for an earlier proof by a different method). A standard analysis shows that the solution u_p converges to u_2 and $\lim_{p \to 2_+} \|u_p\|_{\mathrm{L}^p(\mathbb{R}^d)} = 1$. Multiplying (6) by u and by $x \cdot \nabla u$, one gets after a few integrations by parts that

$$\|\nabla u_p\|^2_{\mathrm{L}^2(\mathbb{R}^d)} = \mathsf{a}\|u_p\|^p_{\mathrm{L}^p(\mathbb{R}^d)} - \mathsf{b}\|u_p\|^2_{\mathrm{L}^2(\mathbb{R}^d)} ,$$
$$\frac{d-2}{2d}\|\nabla u_p\|^2_{\mathrm{L}^2(\mathbb{R}^d)} = \frac{\mathsf{b}}{2}\|u_p\|^2_{\mathrm{L}^2(\mathbb{R}^d)} - \frac{\mathsf{a}}{p}\|u_p\|^p_{\mathrm{L}^p(\mathbb{R}^d)}$$

so that

$$\mathsf{C}_{\mathrm{GN}}(p) = \frac{\|u_p\|^2_{\mathrm{L}^2(\mathbb{R}^d)}}{\|\nabla u_p\|^{2\vartheta(p,d)}_{\mathrm{L}^2(\mathbb{R}^d)} \|u_p\|^{2(1-\vartheta(p,d))}_{\mathrm{L}^2(\mathbb{R}^d)}} = g(p)\|u_p\|^{2-p}_{\mathrm{L}^p(\mathbb{R}^d)}$$

with $g(p) := \frac{1}{2}(\frac{2p}{d})^{\vartheta(p,d)}[p\frac{4-d(p-2)(2+\log\pi)}{2p-d(p-2)}]^{1-\vartheta(p,d)}$ and the conclusion holds since $g'(2) = \frac{d}{4}\log \mathsf{C}_{\mathrm{LS}}$. □

Notice that it is possible to rephrase the Gagliardo–Nirenberg–Sobolev inequalities in a non-scale invariant form as

$$\mathsf{a}\|u_p\|^{p-2}_{\mathrm{L}^p(\mathbb{R}^d)}\|u\|^2_{\mathrm{L}^p(\mathbb{R}^d)} \le \|\nabla u\|^2_{\mathrm{L}^2(\mathbb{R}^d)} + \mathsf{b}\|u\|^2_{\mathrm{L}^2(\mathbb{R}^d)} \quad \forall u \in \mathrm{H}^1(\mathbb{R}^d),$$

which can itself be recast into

$$2\,\frac{\|u\|^2_{\mathrm{L}^p(\mathbb{R}^d)} - \|u\|^2_{\mathrm{L}^2(\mathbb{R}^d)}}{p-2} \le \frac{\|\nabla u\|^2_{\mathrm{L}^2(\mathbb{R}^d)}}{\|u_p\|^{p-2}_{\mathrm{L}^p(\mathbb{R}^d)}} + \left(\|u_p\|^{2-p}_{\mathrm{L}^p(\mathbb{R}^d)}\mathsf{b} - \mathsf{a}\right)\|u\|^2_{\mathrm{L}^2(\mathbb{R}^d)}.$$

It is then straightforward to understand why the limit as $p \to 2_+$ in the above inequality gives the logarithmic Sobolev inequality (with optimal constant).

Acknowledgements The authors have been supported by the ANR projects CBDif-Fr and EVOL.

References

1. Caffarelli, L., Kohn, R., Nirenberg, L.: First order interpolation inequalities with weights. Compos. Math. **53**(3), 259–275 (1984)
2. Carlen, E.A., Loss, M.: Extremals of functionals with competing symmetries. J. Funct. Anal. **88**(2), 437–456 (1990). doi:10.1016/0022-1236(90)90114-Z
3. Catrina, F., Wang, Z.-Q.: On the Caffarelli–Kohn–Nirenberg inequalities: sharp constants, existence (and nonexistence), and symmetry of extremal functions. Commun. Pure Appl. Math. **54**(2), 229–258 (2001)
4. del Pino, M., Dolbeault, J., Filippas, S., Tertikas, A.: A logarithmic Hardy inequality. J. Funct. Anal. **259**(8), 2045–2072 (2010). doi:10.1016/j.jfa.2010.06.005
5. Dolbeault, J., Esteban, M.J.: Extremal functions for Caffarelli–Kohn–Nirenberg and logarithmic Hardy inequalities. Proc. R. Soc. Edinb., Sect. A (2011, to appear) (Ref: 2010110)
6. Dolbeault, J., Esteban, M.J., Tarantello, G.: The role of Onofri type inequalities in the symmetry properties of extremals for Caffarelli–Kohn–Nirenberg inequalities, in two space dimensions. Ann. Sc. Norm. Super. Pisa, Cl. Sci. (5) **7**(2), 313–341 (2008)
7. Dolbeault, J., Esteban, M.J., Loss, M., Tarantello, G.: On the symmetry of extremals for the Caffarelli–Kohn–Nirenberg inequalities. Adv. Nonlinear Stud. **9**(4), 713–726 (2009)
8. Dolbeault, J., Esteban, M.J., Tarantello, G., Tertikas, A.: Radial symmetry and symmetry breaking for some interpolation inequalities. Calc. Var. Partial Differ. Eqn. **42**, 461–485 (2011)
9. Felli, V., Schneider, M.: Perturbation results of critical elliptic equations of Caffarelli–Kohn–Nirenberg type. J. Differ. Equ. **191**(1), 121–142 (2003)
10. Toscani, G.: Sur l'inégalité logarithmique de Sobolev. C. R. Acad. Sci. Paris, Sér. I Math. **324**(6), 689–694 (1997). doi:10.1016/S0764-4442(97)86991-1
11. Weissler, F.B.: Logarithmic Sobolev inequalities for the heat-diffusion semigroup. Trans. Am. Math. Soc. **237**, 255–269 (1978)

On the Oberbeck–Boussinesq Approximation on Unbounded Domains

Eduard Feireisl and Maria E. Schonbek

Abstract We study the Oberbeck–Boussinesq approximation describing the motion of an incompressible, heat-conducting fluid occupying a general unbounded domain in R^3. We provide a rigorous justification of the model by means of scale analysis of the full Navier–Stokes–Fourier system in the low Mach and Froude number regime on large domains, the diameter of which is proportional to the speed of sound. Finally, we show that the total energy of any solution of the resulting Oberbeck–Boussinesq system tends to zero with growing time.

Keywords Oberbeck–Boussinesq system · Singular limit · Unbounded domain

1 Introduction

Stratified flows occur frequently in the atmosphere or oceans. The Oberbeck–Boussinesq approximation is a mathematical model of a stratified fluid flow, where the fluid is assumed to be incompressible and yet convecting a diffusive quantity creating positive or negative buoyancy force. The diffusive quantity is identified with the deviation of temperature from its equilibrium value. The resulting system of equations reads:

$$\mathrm{div}_x \mathbf{U} = 0, \tag{1}$$

$$\overline{\rho}\big(\partial_t \mathbf{U} + \mathrm{div}_x \mathbf{U} \otimes \mathbf{U}\big) + \nabla_x \Pi = \mu \Delta \mathbf{U} + r \nabla_x G, \tag{2}$$

E. Feireisl (✉)
Institute of Mathematics of the Academy of Sciences of the Czech Republic, Žitná 25, 115 67 Praha 1, Czech Republic
e-mail: feireisl@math.cas.cz

M.E. Schonbek
Department of Mathematics, University of California at Santa Cruz, Santa Cruz, CA 95064, USA
e-mail: maria.schonbek@googlemail.com

H. Holden, K.H. Karlsen (eds.), *Nonlinear Partial Differential Equations*, Abel Symposia 7, DOI 10.1007/978-3-642-25361-4_7,

$$\overline{\rho} c_p \big(\partial_t \Theta + \mathrm{div}_x (\Theta \mathbf{U}) \big) - \kappa \Delta \Theta - \overline{\rho} \overline{\vartheta} \alpha \, \mathrm{div}_x (G \mathbf{U}) = 0, \tag{3}$$

$$r + \overline{\rho} \alpha \Theta = 0, \tag{4}$$

where the unknowns are the fluid velocity $\mathbf{U} = \mathbf{U}(t,x)$ and the temperature deviation $\Theta = \Theta(t,x)$. The symbol Π denotes the pressure, $\mu > 0$ is the viscosity coefficient, $\kappa > 0$ the heat conductivity coefficient, $\overline{\rho} > 0$ stands for the fluid density, and $\overline{\vartheta} > 0$ is the reference temperature. Here, $c_p > 0$ is the specific heat at constant pressure and $\alpha > 0$ denotes the coefficient of thermal expansion of the fluid, both evaluated at the reference density $\overline{\rho}$ and temperature $\overline{\vartheta}$. The function $G = G(x)$ is a given gravitational potential acting on the fluid. Thus the fluid density is constant in the Oberbeck–Boussinesq approximation except in the buoyancy force, where it is interrelated to the temperature deviation through *Boussinesq relation* (4), cf. Zeytounian [30, 31].

In real world applications, it is customary to take the x_3-coordinate to be vertical parallel to the gravitational force $\nabla_x G = g[0,0,-1]$. This is indeed a reasonable approximation provided the fluid occupies a bounded domain $\Omega \subset R^3$, where the gravitational field can be taken constant. Recently, several authors studied system (1–4) on the whole space $\Omega = R^3$, with $\nabla_x G = g[0,0,-1]$, see [4], Danchin and Paicu [7]. Such an "extrapolation" of the model is quite natural from the mathematical viewpoint, however, a bit awkward physically. Indeed, if the self-gravitation of the fluid is neglected, the origin of the gravitational force must be an object placed *outside* the fluid domain Ω therefore

$$G(x) = \int_{R^3} \frac{1}{|x-y|} m(y) \, \mathrm{d}y, \quad \text{with } m \geq 0, \qquad \mathrm{supp}[m] \subset R^3 \setminus \Omega, \tag{5}$$

where m denotes the mass density of the object acting on the fluid by means of gravitation. In other words, G is a harmonic function in Ω, $G(x) \approx 1/|x|$ as $|x| \to \infty$.

Motivated by the previous observations, we consider the Oberbeck–Boussinesq system on a domain $\Omega = R^3 \setminus K$ exterior to a compact set K. Accordingly, we take G such that

$$-\Delta G = m \quad \text{in } R^3, \quad \nabla_x G \in L^2(R^3; R^3), \qquad \mathrm{supp}[m] \subset K. \tag{6}$$

In particular, introducing a new variable $\theta = \Theta - \overline{\vartheta} \alpha G / c_p$ we can rewrite the system (1–4) in the more frequently used form

$$\mathrm{div}_x \, \mathbf{U} = 0, \tag{7}$$

$$\overline{\rho} \Big(\partial_t \mathbf{U} + \mathrm{div}_x \, \mathbf{U} \otimes \mathbf{U} \Big) + \nabla_x P = \mu \Delta \mathbf{U} - \overline{\rho} \alpha \theta \nabla_x G, \tag{8}$$

$$\overline{\rho} c_p \Big(\partial_t \theta + \mathrm{div}_x (\theta \mathbf{U}) \Big) - \kappa \Delta \theta = 0, \tag{9}$$

where we have set $P = \Pi - G^2 \overline{\rho} \overline{\vartheta} \alpha^2 / 2 c_p$.

We will show in Sect. 2 that the Oberbeck–Boussinesq approximation (1–4), supplemented with suitable boundary conditions, may be viewed as a singular limit of the full Navier–Stokes–Fourier system considered on a family of "large domains", where the Mach and Froude numbers tend simultaneously to zero. This part of the paper can be viewed as an application of the abstract method developed in [13] in order to control the propagation and the final filtering of acoustic waves in the limit system. Furthermore, we discuss the basic properties of the limit system (1–4), in particular, validity of the energy inequality, see Sect. 3. Finally, in Sect. 4, we show that the total energy of any weak solution to the Oberbeck–Boussinesq approximation (7–9), supplemented with the homogeneous Dirichlet boundary conditions, tends to zero with growing time. To this end, we first establish the result for the temperature deviations represented by θ, and then use the standard estimates for the incompressible Navier–Stokes in the spirit of Miyakawa and Sohr [24].

1.1 Notation and Preliminaries

We use the symbol $\langle \cdot, \cdot \rangle$ to denote duality product, in particular,

$$\langle f, g \rangle = \int_O fg,$$

provided f, g are square integrable on a set O.

The symbol $L^p(O)$ denotes the space of measurable functions v, with $|v|^p$ integrable in O. $W^{k,p}$ denotes the Sobolev space of functions having derivatives up to order k in L^p. Finally, we introduce the homogeneous Sobolev spaces:

$$\widehat{W}^{m,p} = \left\{ v \in L^1_{loc}(\Omega),\ D^\alpha u \in L^p(\Omega), |\alpha| = m \right\}, \quad m \geq 0,\ p \geq 1.$$

By the symbol c we denote a generic constant that may change line by line.

Most of the results of the paper concern problems on an exterior domain $\Omega \subset R^3$. In order to avoid technicalities, we assume that the boundary $\partial\Omega$ is smooth, say of class $C^{2+\nu}$, in particular, Ω satisfies the *cone property*:

The domain Ω is said to satisfy the *cone property* if there exists a finite cone $\mathscr{C}$ such that each point $x \in \Omega$ is the vertex of a finite cone $\mathscr{C}_x$ contained in Ω and congruent to $\mathscr{C}$.

To conclude the preliminary part, we record a variant of the Gagliardo–Nirenberg inequalities for exterior domains proved by Crispo and Maremonti [6].

Proposition 1.1 *Let $\Omega \subset R^N$ be an exterior domain with cone property. Let $w \in \widehat{W}^{m,p}(\Omega) \cap L^q(\Omega)$, with $1 \leq p \leq \infty$, $1 \leq q < \infty$.*

Then

$$\|D^k w\|_{L^r(\Omega)} \leq c \|D^m w\|^a_{L^p(\Omega)} \|w\|^{1-a}_{L^q(\Omega)} \tag{10}$$

for any integer $k \in [0, m-1]$, where

$$\frac{1}{r} = \frac{k}{N} + a\left(\frac{1}{p} - \frac{m}{N}\right) + (1-a)\frac{1}{q},$$

with $a \in [\frac{k}{m}, 1]$, *either if* $p = 1$ *or* $p > 1$ *and* $m - k - \frac{N}{p} \notin \mathcal{N} \cup \{0\}$, *while* $a \in [\frac{k}{m}, 1)$ *if* $p > 1$ *and* $m - k - \frac{N}{p} \in \mathcal{N} \cup \{0\}$.

2 The Oberbeck–Boussinesq Approximation as a Singular Limit of the Full Navier–Stokes–Fourier System

Motivated by the mathematical theory developed in [14], we introduce a scaled *Navier–Stokes–Fourier system* in the form:

MASS CONSERVATION

$$\partial_t \rho + \mathrm{div}_x(\rho \mathbf{u}) = 0, \tag{11}$$

MOMENTUM BALANCE

$$\partial_t(\rho\mathbf{u}) + \mathrm{div}_x(\rho\mathbf{u} \otimes \mathbf{u}) + \frac{1}{\varepsilon^2}\nabla_x p(\rho, \vartheta) = \mathrm{div}_x\, \mathbb{S}(\vartheta, \nabla_x \mathbf{u}) + \frac{1}{\varepsilon}\rho\nabla_x G, \tag{12}$$

ENTROPY BALANCE

$$\partial_t(\rho s(\rho, \vartheta)) + \mathrm{div}_x(\rho s(\rho, \vartheta)\mathbf{u}) + \mathrm{div}_x\left(\frac{\mathbf{q}(\vartheta, \nabla_x \vartheta)}{\vartheta}\right) = \sigma, \tag{13}$$

TOTAL ENERGY CONSERVATION

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \left(\frac{1}{2}\rho|\mathbf{u}|^2 + \frac{1}{\varepsilon^2}\rho e(\rho, \vartheta) - \frac{1}{\varepsilon}\rho G\right) \mathrm{d}x = 0, \tag{14}$$

where S is the viscous stress given by *Newton's rheological law*

$$\mathbb{S}(\vartheta, \nabla_x \mathbf{u}) = \mu(\vartheta)\left(\nabla_x \mathbf{u} + \nabla_x^t \mathbf{u} - \frac{2}{3}\mathrm{div}_x\, \mathbf{u}\mathbb{I}\right) + \eta(\vartheta)\,\mathrm{div}_x\, \mathbf{u}\mathbb{I}, \tag{15}$$

$\mathbf{q}$ is the heat flux determined by *Fourier's law*

$$\mathbf{q}(\vartheta, \nabla_x \vartheta) = -\kappa(\vartheta)\nabla_x \vartheta, \tag{16}$$

whereas the *entropy production rate* σ satisfies

$$\sigma \geq \frac{1}{\vartheta}\left(\varepsilon^2 \mathbb{S} : \nabla_x \mathbf{u} - \frac{\mathbf{q} \cdot \nabla_x \vartheta}{\vartheta}\right). \tag{17}$$

The unknowns in (11)–(14) are the fluid mass density $\rho = \rho(t, x)$, the velocity field $\mathbf{u} = \mathbf{u}(t, x)$, and the absolute temperature $\vartheta = \vartheta(t, x)$. The pressure p, the specific internal energy e, and the specific entropy s are given numerical functions of ρ and ϑ interrelated through *Gibbs' equation*

$$\vartheta Ds = De + pD\left(\frac{1}{\rho}\right). \tag{18}$$

The system (11)–(14) is supplemented with the *conservative* boundary conditions, specifically,

$$\mathbf{u} \cdot \mathbf{n}|_{\partial\Omega} = 0, \qquad \beta[\mathbf{u}]_{\tan} + [\mathbb{S}\mathbf{n}]_{\tan}|_{\partial\Omega} = 0, \qquad \mathbf{q} \cdot \mathbf{n}|_{\partial\Omega} = -\beta|\mathbf{u}|^2|_{\partial\Omega}, \quad \beta > 0, \tag{19}$$

where $\mathbf{n}$ denotes the outer normal vector to $\partial\Omega$. The first two conditions in (19) are usually termed *Navier's slip boundary condition* with a friction coefficient $\beta > 0$, see Málek and Rajagopal [22]. In accordance with (19), the total energy of the fluid is a conserved quantity as stated in (14).

The small parameter ε appearing in (12), (14), and (17) results from the scaling analysis of the Navier–Stokes–Fourier system, where the *Mach number* and the *Froude number* are proportional to ε, see [14, Chaps. 4, 5], Klein et al. [16], Zeytounian [32]. Physically this means that the characteristic speed of the fluid is largely dominated by the speed of sound and the fluid is stratified. Note that a similar system of equations may be obtained by *constitutive* scaling, where the rheological properties of the fluid are changing rather than the characteristic geometrical parameters of the flow, see Novotný, Růžička, Thaeter [25], Rajagopal, Růžička, and Srinivasa [26].

2.1 Weak Solutions

In the framework of *weak solutions*, the equation of continuity (11) is replaced by a family of integral identities

$$\int_\Omega \left[\rho(\tau, \cdot)\varphi(\tau, \cdot) - \rho_0\varphi(0, \cdot)\right] \mathrm{d}x$$
$$= \int_0^\tau \int_\Omega \left(\rho\partial_t\varphi + \rho\mathbf{u} \cdot \nabla_x\varphi\right) \mathrm{d}x\, \mathrm{d}t \quad \text{for any } \tau \in [0, T], \tag{20}$$

for any test function $\varphi \in C^1([0, T] \times \overline{\Omega})$. In particular, the mapping $\tau \mapsto \rho(\tau, \cdot)$ is weakly continuous, and ρ satisfies the initial condition

$$\rho(0, \cdot) = \rho_0.$$

Similarly, the momentum equation (12), together with Navier's slip boundary conditions (19), read

$$\int_{\Omega} \big[\rho\mathbf{u}(\tau,\cdot)\cdot\varphi(\tau,\cdot) - \rho_0\mathbf{u}_0\cdot\varphi(0,\cdot)\big]\,\mathrm{d}x$$
$$= \int_0^{\tau}\int_{\Omega}\left(\rho\mathbf{u}\cdot\partial_t\varphi + \rho\mathbf{u}\otimes\mathbf{u}:\nabla_x\varphi + \frac{p}{\varepsilon^2}\operatorname{div}_x\varphi - \mathbb{S}:\nabla_x\varphi + \frac{\rho}{\varepsilon}\nabla_x G\cdot\varphi\right)\mathrm{d}x\,\mathrm{d}t$$
$$+ \int_0^{\tau}\int_{\partial\Omega}\beta\mathbf{u}\cdot\varphi\,\mathrm{dS}_x\ \mathrm{d}t, \tag{21}$$

for any $\tau\in[0,T]$, and any $\varphi\in C^1([0,T]\times\overline{\Omega};R^3)$, $\varphi\cdot\mathbf{n}|_{\partial\Omega}=0$. Thus the momentum $\tau\mapsto(\rho\mathbf{u})(\tau,\cdot)$ is weakly continuous and

$$(\rho\mathbf{u})(0,\cdot) = \rho_0\mathbf{u}_0.$$

Finally, we may write the entropy balance (13) in the form

$$\int_{\Omega}\big[\rho s(\rho,\vartheta)(\tau,\cdot)\varphi(\tau,\cdot) - \rho_0 s(\rho_0,\vartheta_0)\varphi(0,\cdot)\big]\,\mathrm{d}x$$
$$= \langle\sigma, 1_{[0,\tau]}\varphi\rangle + \int_0^{\tau}\int_{\Omega}\left(\rho s\partial_t\varphi + \rho s\mathbf{u}\cdot\nabla_x\varphi + \frac{\mathbf{q}}{\vartheta}\cdot\nabla_x\varphi\right)\mathrm{d}x\,\mathrm{d}t$$
$$+ \varepsilon^2\int_0^{\tau}\int_{\partial\Omega}\frac{\beta}{\vartheta}|\mathbf{u}|^2\varphi\,\mathrm{dS}_x\ \mathrm{d}t, \tag{22}$$

for any test function $\varphi\in C^1([0,T]\times\overline{\Omega})$, where the entropy production rate σ is interpreted as a non-negative measure on $[0,T]\times\overline{\Omega}$ satisfying

$$\sigma \geq \frac{1}{\vartheta}\left(\varepsilon^2\mathbb{S}:\nabla_x\mathbf{u} - \frac{\mathbf{q}\cdot\nabla_x\vartheta}{\vartheta}\right). \tag{23}$$

The total energy balance (14) reads

$$\int_{\Omega}\left(\frac{1}{2}\rho|\mathbf{u}|^2 + \frac{1}{\varepsilon^2}\rho e - \frac{1}{\varepsilon}\rho G\right)(\tau,\cdot)\,\mathrm{d}x$$
$$= \int_{\Omega}\left(\frac{1}{2}\rho_0|\mathbf{u}_0|^2 + \frac{1}{\varepsilon^2}\rho_0 e(\rho_0,\vartheta_0) - \frac{1}{\varepsilon}\rho_0 G\right)\mathrm{d}x. \tag{24}$$

The interested reader may consult [14, Chap. 2] for a formal interpretation of the weak solutions to the Navier–Stokes–Fourier system. We only note that the entropy production rate σ associated to a weak solution that is sufficiently smooth necessarily satisfies

$$\sigma = \frac{1}{\vartheta}\left(\varepsilon^2\mathbb{S}:\nabla_x\mathbf{u} - \frac{\mathbf{q}\cdot\nabla_x\vartheta}{\vartheta}\right),$$

in agreement with the classical theory.

Unlike (20), (21), relations (22), (24) are satisfied only for a.a. $\tau \in [0,T]$. In particular, the total entropy $\rho s(\rho,\vartheta)$ may not be a weakly continuous function of time due to hypothetical jumps in σ. Introducing a *time lifting* Σ of the measure σ in the form

$$\langle \Sigma, \varphi \rangle \equiv \langle \sigma, I[\varphi] \rangle,$$

where

$$I[\varphi](t,x) = \int_0^t \varphi(z,x)\,\mathrm{d}z \quad \text{for any } \varphi \in L^1(0,T;C(\overline{\Omega})),$$

we check easily that Σ can be identified with a mapping $\Sigma \in L^\infty_{\text{weak}}(0,T;\mathcal{M}^+(\overline{\Omega}))$, where

$$\langle \Sigma(\tau), \varphi \rangle = \lim_{\delta \to 0+} \langle \sigma, \psi_\delta \varphi \rangle,$$

with

$$\psi_\delta(t) = \begin{cases} 0, & \text{for } t \in [0,\tau), \\ \frac{1}{\delta}(t-\tau), & \text{for } t \in (\tau, \tau+\delta), \\ 1, & \text{for } t \geq \tau + \delta. \end{cases}$$

In particular, the measure Σ is well-defined for *any* $\tau \in [0,T]$, and the mapping $\tau \mapsto \Sigma_\varepsilon$ is non-increasing in the sense of measures. Here the subscript in L^∞_{weak} means "weakly measurable".

The entropy balance (22) can be therefore rewritten as

$$\begin{aligned} &\int_\Omega \big[\rho s(\rho,\vartheta)(\tau,\cdot)\varphi(\tau,\cdot) - \rho_0 s(\rho_0,\vartheta_0)\varphi(0,\cdot)\big]\,\mathrm{d}x \\ &\qquad + \langle \Sigma(\tau), \varphi(\tau,\cdot)\rangle - \langle \Sigma(0), \varphi(0,\cdot)\rangle \\ &= \int_0^\tau \langle \Sigma, \partial_t \varphi \rangle\,\mathrm{d}t + \int_0^\tau \int_\Omega \Big(\rho s(\rho,\vartheta)\partial_t \varphi + \rho s(\rho,\vartheta)\mathbf{u}\cdot\nabla_x \varphi + \frac{\mathbf{q}}{\vartheta}\cdot\nabla_x\varphi \Big)\,\mathrm{d}x\,\mathrm{d}t \\ &\qquad + \varepsilon^2 \int_0^\tau \int_{\partial\Omega} \frac{\beta}{\vartheta}|\mathbf{u}|^2 \varphi\,\mathrm{dS}_x\,\mathrm{d}t, \end{aligned} \tag{25}$$

for any $\varphi \in C^1([0,T]\times\overline{\Omega})$, where the mapping

$$\tau \mapsto \rho s(\rho,\vartheta)(\tau,\cdot) + \Sigma(\tau) \quad \text{is continuous with values in } \mathcal{M}(\overline{\Omega})$$

provided the space of measures $\mathcal{M}$ is endowed with the weak-(*) topology.

2.2 Existence Theory for the Navier–Stokes–Fourier System

The framework of weak solutions introduced in Sect. 2.1 is broad enough to develop an existence theory without any essential restrictions imposed on the initial data as well as the length of the time interval $(0,T)$. We start with a list of technical hypotheses imposed on the constitutive equations and the transport coefficients. The reader may consult [14, Chap. 3] for the physical background and further discussion.

The pressure p will be given by a general formula

$$p(\rho,\vartheta) = \vartheta^{5/2} P\left(\frac{\rho}{\vartheta^{3/2}}\right) + \frac{a}{3}\vartheta^4, \quad a > 0, \tag{26}$$

where

$$P \in C^1[0,\infty), \qquad P(0) = 0, \qquad P'(Z) > 0 \quad \text{for all } Z \geq 0, \tag{27}$$

in particular, the compressibility $\partial_\rho p(\rho,\vartheta)$ is always positive. The former component in (26) represents the standard molecular pressure of a general monoatomic gas while the latter is a contribution due to thermal radiation.

In accordance with Gibbs' relation (18), the specific internal energy can be taken in the form

$$e(\rho,\vartheta) = \frac{3}{2}\vartheta\left(\frac{\vartheta^{3/2}}{\rho}\right) P\left(\frac{\rho}{\vartheta^{3/2}}\right) + a\frac{\vartheta^4}{\rho}, \tag{28}$$

where, in addition to (27), we assume that

$$0 < \frac{\frac{5}{3}P(Z) - P'(Z)Z}{Z} < c \quad \text{for all } Z > 0. \tag{29}$$

The awkwardly looking condition (29) has a clear physical meaning, namely the specific heat at constant volume—$\partial_\vartheta e(\rho,\vartheta)$—is positive and bounded. In particular, (29) implies that the function $Z \mapsto P(Z)/Z^{5/3}$ is decreasing, and we assume

$$\lim_{Z\to\infty} \frac{P(Z)}{Z^{5/3}} = P_\infty > 0. \tag{30}$$

We remark that the molecular pressure $\vartheta^{5/2}P(\rho/\vartheta^{3/2})$ coincides with the standard perfect gas law $\mathrm{R}\vartheta\rho$ as long as $P(Z) \approx \mathrm{R}Z$, see Eliezer, Ghatak, and Hora [11] and [14, Chap. 1].

In addition to the previous hypotheses, we suppose that the transport coefficients $\mu = \mu(\vartheta)$, $\eta = \eta(\vartheta)$, and $\kappa = \kappa(\vartheta)$ are continuously differentiable functions of $\vartheta \in [0,\infty)$ such that

$$0 < \underline{\mu}(1+\vartheta) \leq \mu(\vartheta), \qquad |\mu'(\vartheta)| \leq \mu_1 \quad \text{for all } \vartheta \geq 0, \tag{31}$$

$$0 \leq \eta(\vartheta) \leq \overline{\eta}(1+\vartheta) \quad \text{for all } \vartheta \geq 0, \tag{32}$$

and

$$0 < \underline{\kappa}(1+\vartheta^3) \leq \kappa(\vartheta) \leq \overline{\kappa}(1+\vartheta^3) \quad \text{for all } \vartheta \geq 0. \tag{33}$$

We report the following result (see [14, Chap. 3, Theorem 3.1]):

Theorem 2.1 *Assume that $\Omega \subset R^3$ is a bounded domain of class $C^{2+\nu}$. Let $\varepsilon > 0$ $\beta > 0$ be given, let the initial data satisfy*

$$\rho_0 \in L^\infty(\Omega), \qquad \rho_0 > 0,$$
$$\vartheta_0 \in L^\infty(\Omega), \qquad \vartheta_0 > 0, \qquad \mathbf{u}_0 \in L^\infty(\Omega; R^3),$$

and let $G \in W^{1,\infty}(\Omega)$. Suppose that the thermodynamic functions p, e, and s satisfy Gibbs' equation (18), *together with the structural hypotheses* (26)–(30), *and the transport coefficients comply with* (31)–(33).
Then the Navier–Stokes–Fourier system possesses a weak solution ρ, ϑ, $\mathbf{u}$ on the set $(0, T) \times \Omega$ in the sense specified in Sect. 2.1.

Remark 2.1 As a matter of fact, the existence theorem [14, Chap. 3, Theorem 3.1] is proved for $\beta = 0$, however, the case $\beta > 0$ requires only straightforward modifications.

Remark 2.2 The weak solution, the existence of which is claimed in Theorem 2.1, satisfies $\rho \geq 0$, $\vartheta > 0$ a.a. in $(0, T) \times \Omega$. In addition, the weak solutions can be constructed to satisfy the equation of continuity (11) in the sense of renormalized solutions introduced by DiPerna and Lions [9]. Other regularity properties of the weak solutions are discussed in [14, Chap. 3, Sect. 3.8].

Remark 2.3 The hypotheses imposed on the initial data in Theorem 2.1 are not optimal. As a matter of fact, it is enough to assume that the initial energy and entropy of the system is finite, see [14, Chap. 3]. Similarly, the hypotheses imposed on the structural properties of thermodynamic functions as well as the transport coefficients may be considerably relaxed, see [14, Chap. 3].

2.3 Uniform Bounds and Stability with Respect to the Singular Parameter

Our goal is to identify the Oberbeck–Boussinesq approximation (1)–(4) with the asymptotic limit for $\varepsilon \to 0$ of the scaled Navier–Stokes–Fourier system (11)–(14). Moreover, we want the limit system to be defined on an exterior (unbounded) domain $\Omega \subset R^3$. To this end, we consider the scaled Navier–Stokes–Fourier system on a family of (bounded) domains

$$\Omega_\varepsilon = \Omega \cap \left\{ x \in R \;\middle|\; |x| < \frac{1}{\varepsilon^r} \right\}, \quad r > 1, \tag{34}$$

supplemented, for simplicity, with the complete slip boundary condition (Navier's slip with $\beta = 0$),

$$\mathbf{u} \cdot \mathbf{n}|_{\partial \Omega_\varepsilon} = 0, \qquad [\mathbb{S}\mathbf{n}] \times \mathbf{n}|_{\partial \Omega_\varepsilon} = 0, \qquad \mathbf{q} \cdot \mathbf{n}|_{\partial \Omega_\varepsilon} = 0, \tag{35}$$

cf. (19).

Thus, at least formally, $\Omega_\varepsilon \to \Omega$ as $\varepsilon \to 0$. As we shall see, the major problem in the limit passage is filtering the acoustic waves represented by the gradient component of the velocity field. Since the speed of sound in the fluid is proportional to $1/\varepsilon$, hypothesis (34) ensures that the outer boundary of Ω_ε becomes irrelevant, at least for what concerns the behavior of acoustic waves on compact subsets of the physical space, and, accordingly, we may use the dispersive phenomena to eliminate the presence of acoustic waves in the asymptotic limit.

Uniform Bounds Based on Energy Dissipation

Let $\{\rho_\varepsilon, \vartheta_\varepsilon, \mathbf{u}_\varepsilon\}$ be a weak solution of the scaled Navier–Stokes–Fourier system on the set $(0, T) \times \Omega_\varepsilon$ in the sense of Sect. 2.1. We start by deriving *uniform* bounds independent of $\varepsilon \to 0$. The key quantity is the *ballistic free energy* introduced by Ericksen [12, Chap. 1.3]:

$$H(\rho, \vartheta) = \rho e(\rho, \vartheta) - \overline{\vartheta} \rho s(\rho, \vartheta),$$

where $\overline{\vartheta}$ is a positive constant. It is easy to check that

$$\frac{\partial^2 H(\rho, \overline{\vartheta})}{\partial \rho^2} = \frac{1}{\rho} \frac{\partial p(\rho, \overline{\vartheta})}{\partial \rho}, \qquad \frac{\partial H(\rho, \vartheta)}{\partial \vartheta} = \frac{\rho}{\vartheta} (\vartheta - \overline{\vartheta}) \frac{\partial e(\rho, \vartheta)}{\partial \vartheta},$$

in particular, hypotheses (27), (29) imply that

$$\left[\begin{array}{c} \rho \mapsto H(\rho, \overline{\vartheta}) \text{ is strictly convex,} \\ \vartheta \mapsto H(\rho, \vartheta) \text{ is strictly decreasing for } \vartheta < \overline{\vartheta} \\ \text{and strictly increasing for } \vartheta > \overline{\vartheta}. \end{array} \right]$$

Conditions (27), (29) guarantee *thermodynamic stability* of the system, see Bechtel, Rooney, and Forest [3]. As we will see, they are crucial to control the norm of solutions to the scaled system.

In the so-called *static* density and temperature distribution for the scaled Navier–Stokes–Fourier system, the temperature equals a positive constant $\overline{\vartheta}$ while the density $\tilde{\rho}_\varepsilon$ satisfies

$$\nabla_x p(\tilde{\rho}_\varepsilon, \overline{\vartheta}) = \varepsilon \tilde{\rho}_\varepsilon \nabla_x G.$$

It is easy to check that

$$\frac{\partial H(\tilde{\rho}_\varepsilon, \overline{\vartheta})}{\partial \rho} = \varepsilon G + \text{const} \quad \text{in } \Omega_\varepsilon \tag{36}$$

provided $\tilde{\rho}_\varepsilon$ is strictly positive in Ω_ε.

Taking advantage of (36), we may combine total energy balance (24) with the entropy equation (22) to obtain

$$\int_{\Omega_\varepsilon} \left(\frac{1}{2} \rho_\varepsilon |\mathbf{u}_\varepsilon|^2 + \frac{1}{\varepsilon^2} \left(H(\rho_\varepsilon, \vartheta_\varepsilon) - \frac{\partial H(\tilde{\rho}_\varepsilon, \overline{\vartheta})}{\partial \rho} (\rho_\varepsilon - \tilde{\rho}_\varepsilon) - H(\tilde{\rho}_\varepsilon, \overline{\vartheta}) \right) \right) (\tau, \cdot) \, \mathrm{d}x$$
$$+ \frac{\overline{\vartheta}}{\varepsilon^2} \sigma_\varepsilon [[0, \tau] \times \overline{\Omega}]$$
$$= \int_{\Omega_\varepsilon} \left(\frac{1}{2} \rho_{0,\varepsilon} |\mathbf{u}_{0,\varepsilon}|^2 + \frac{1}{\varepsilon^2} \left(H(\rho_{0,\varepsilon}, \vartheta_{0,\varepsilon}) \right.\right.$$
$$\left.\left. - \frac{\partial H(\tilde{\rho}_\varepsilon, \overline{\vartheta})}{\partial \rho} (\rho_{0,\varepsilon} - \tilde{\rho}_\varepsilon) - H(\tilde{\rho}_\varepsilon, \overline{\vartheta}) \right) \right) \mathrm{d}x \tag{37}$$

for a.a. $\tau \in (0, T)$ provided we fix the static density so that

$$\int_{\Omega_\varepsilon} \rho_\varepsilon(\tau, \cdot) \, \mathrm{d}x = \int_{\Omega_\varepsilon} \rho_{0,\varepsilon} \, \mathrm{d}x = \int_{\Omega_\varepsilon} \tilde{\rho}_\varepsilon \, \mathrm{d}x,$$

meaning the total mass of the fluid contained in Ω_ε coincides with the total mass of the static distribution $\tilde{\rho}_\varepsilon$.

As a matter of fact, it is more convenient to consider a static solution $\tilde{\rho}_\varepsilon$ defined on the whole space R^3, specifically,

$$\nabla_x p(\tilde{\rho}_\varepsilon, \overline{\vartheta}) = \varepsilon \tilde{\rho}_\varepsilon \nabla_x G \quad \text{in } R^3,$$

satisfying

$$\lim_{|x| \to \infty} \tilde{\rho}_\varepsilon(x) = \overline{\rho}.$$

Consequently, we have

$$\tilde{\rho}_\varepsilon - \overline{\rho} = \frac{\varepsilon}{\mathscr{P}'(\overline{\rho})} G + \varepsilon^2 h_\varepsilon G, \qquad \mathscr{P}'(\rho) = \frac{1}{\rho} \partial_\rho p(\rho, \overline{\vartheta}), \tag{38}$$

with

$$\|h_\varepsilon\|_{L^\infty(R^3)} \le c, \qquad |\nabla_x \tilde{\rho}_\varepsilon(x)| \le \varepsilon c |\nabla_x G(x)| \quad \text{for } x \in R^3. \tag{39}$$

In order to exploit (37), the initial data must be chosen in such a way that the right-hand side of (37) remains bounded uniformly for $\varepsilon \to 0$. To this end, we take

$$\rho_{0,\varepsilon} = \tilde{\rho}_\varepsilon + \varepsilon \rho_{0,\varepsilon}^{(1)}, \qquad \vartheta_{0,\varepsilon} = \overline{\vartheta} + \varepsilon \vartheta_{0,\varepsilon}^{(1)}, \tag{40}$$

where

$$\|\rho_{0,\varepsilon}^{(1)}\|_{L^2 \cap L^\infty(\Omega_\varepsilon)} \le c, \qquad \|\vartheta_{0,\varepsilon}^{(1)}\|_{L^2 \cap L^\infty(\Omega_\varepsilon)} \le c, \tag{41}$$

$$\int_{\Omega_\varepsilon} \rho_{0,\varepsilon}^{(1)} \, \mathrm{d}x = \int_{\Omega_\varepsilon} \vartheta_{0,\varepsilon}^{(1)} \, \mathrm{d}x = 0; \tag{42}$$

and

$$\|\mathbf{u}_{0,\varepsilon}\|_{L^2 \cap L^\infty(\Omega_\varepsilon; R^3)} \le c, \tag{43}$$

where all constants are independent of ε.

By virtue of (27), (29), the ballistic free energy possesses remarkable coercivity properties, specifically,

$$H(\rho,\vartheta)-\frac{\partial H(\tilde\rho,\overline{\vartheta})}{\partial\rho}(\rho-\tilde\rho)-H(\tilde\rho,\overline{\vartheta})$$
$$\geq c(K)(|\rho-\tilde\rho|^2+|\vartheta-\overline{\vartheta}|^2)\quad\text{for all }(\rho,\vartheta)\in K,\tag{44}$$

and

$$H(\rho,\vartheta)-\frac{\partial H(\tilde\rho,\overline{\vartheta})}{\partial\rho}(\rho-\tilde\rho)-H(\tilde\rho,\overline{\vartheta})$$
$$\geq c(K)\big(1+\rho|e(\rho,\vartheta)|+\rho|s(\rho,\vartheta)|\big)\quad\text{for all }(\rho,\vartheta)\in(0,\infty)^2\setminus K,\tag{45}$$

for any compact $K\subset(0,\infty)^2$ containing $(\tilde\rho,\overline{\vartheta})$, see [14, Chap. 5, Lemma 5.1]. Consequently, introducing the decomposition

$$h=h_{\text{ess}}+h_{\text{res}},\qquad h_{\text{ess}}=\chi(\rho_\varepsilon,\vartheta_\varepsilon)h,\qquad h_{\text{res}}=(1-\chi(\rho_\varepsilon,\vartheta_\varepsilon))h,\tag{46}$$

for any measurable function h, where $\chi\in C_c^\infty((0,\infty)^2)$ such that

$$0\leq\chi\leq 1,\qquad \chi\equiv 1\quad\text{on the rectangle }[\overline{\vartheta}/2,2\overline{\vartheta}]\times[\overline{\rho}/2,2\overline{\rho}],$$

we deduce from (37) the following list of uniform bounds:

$$\operatorname*{ess\,sup}_{t\in(0,T)}\int_{\Omega_\varepsilon}\rho_\varepsilon|\mathbf{u}_\varepsilon|^2(t,\cdot)\,\mathrm{d}x\leq c,\tag{47}$$

and, by virtue of (44),

$$\operatorname*{ess\,sup}_{t\in(0,T)}\left\|\left[\frac{\rho_\varepsilon-\tilde\rho_\varepsilon}{\varepsilon}\right]_{\text{ess}}(t,\cdot)\right\|_{L^2(\Omega_\varepsilon)}\leq c,\tag{48}$$

$$\operatorname*{ess\,sup}_{t\in(0,T)}\left\|\left[\frac{\vartheta_\varepsilon-\overline{\vartheta}}{\varepsilon}\right]_{\text{ess}}(t,\cdot)\right\|_{L^2(\Omega_\varepsilon)}\leq c,\tag{49}$$

where we have used (38), (39) and the fact that the static density $\tilde\rho_\varepsilon$ remains uniformly close to the constant $\overline{\rho}$ as soon as ε is small enough.

Furthermore, by virtue of (45) and the hypotheses (26)–(30), it follows that

$$\operatorname*{ess\,sup}_{t\in[0,T]}\int_{\Omega_\varepsilon}[\rho_\varepsilon]_{\text{res}}^{5/3}(t,\cdot)\,\mathrm{d}x\leq\varepsilon^2c,\tag{50}$$

$$\operatorname*{ess\,sup}_{t\in[0,T]}\int_{\Omega_\varepsilon}[\vartheta_\varepsilon]_{\text{res}}^{4}(t,\cdot)\,\mathrm{d}x\leq\varepsilon^2c,\tag{51}$$

and

$$\operatorname*{ess\,sup}_{t\in[0,T]} \int_{\Omega_\varepsilon} \big(|[\rho_\varepsilon e(\rho_\varepsilon, \vartheta_\varepsilon)]_{\rm res}| + |[p(\rho_\varepsilon, \vartheta_\varepsilon)]_{\rm res}| + |[\rho_\varepsilon s(\rho_\varepsilon, \vartheta_\varepsilon)]_{\rm res}| \big) \, {\rm d}x \le c. \tag{52}$$

Finally, by the same token, the measure of the "residual" set is also small, specifically,

$$\operatorname*{ess\,sup}_{t\in[0,T]} \int_{\Omega_\varepsilon} 1_{\rm res}(t,\cdot) \, {\rm d}x \le \varepsilon^2 c, \tag{53}$$

where all the constants "c" are independent of ε. It is remarkable that the measure of the "residual" set remains small although the measure of Ω_ε tends to infinity as $\varepsilon \to 0$.

Going back to (37) we get

$$\|\sigma_\varepsilon\|_{\mathcal{M}^+([0,T]\times\overline{\Omega}_\varepsilon)} \le \varepsilon^2 c; \tag{54}$$

whence, in view of (23) and hypotheses (31)–(33),

$$\int_0^T \left\| \nabla_x \mathbf{u}_\varepsilon + \nabla_x^t \mathbf{u}_\varepsilon - \frac{2}{3} {\rm div}_x \, \mathbf{u}_\varepsilon \mathbb{I} \right\|^2_{L^2(\Omega_\varepsilon; R^{3\times 3})} {\rm d}t \le c, \tag{55}$$

and

$$\int_0^T \left\| \nabla_x \frac{\vartheta_\varepsilon - \overline{\vartheta}}{\varepsilon} \right\|^2_{L^2(\Omega_\varepsilon; R^3)} {\rm d}t + \int_0^T \left\| \nabla_x \frac{\log(\vartheta_\varepsilon) - \log(\overline{\vartheta})}{\varepsilon} \right\|^2_{L^2(\Omega_\varepsilon; R^3)} {\rm d}t \le c. \tag{56}$$

Moreover, since the measure of the residual set is small (see (53)), we can apply Poincaré's inequality to conclude that

$$\int_0^T \left\| \frac{\vartheta_\varepsilon - \overline{\vartheta}}{\varepsilon} \right\|^2_{W^{1,2}(\Omega_\varepsilon)} {\rm d}t + \int_0^T \left\| \frac{\log(\vartheta_\varepsilon) - \log(\overline{\vartheta})}{\varepsilon} \right\|^2_{W^{1,2}(\Omega_\varepsilon)} {\rm d}t \le c. \tag{57}$$

A similar argument, based on a generalized version of Korn's inequality due to Reshetnyak [28] (see also [14, Chap. 10, Theorem 10.16]), can be applied to (47), (48) to conclude that

$$\int_0^T \|\mathbf{u}_\varepsilon\|^2_{W^{1,2}(\Omega_\varepsilon; R^3)} \, {\rm d}t \le c. \tag{58}$$

Here we have also used the fact that $[\rho]_{\rm ess}$ is bounded below away from zero on a set, the complement of which is of small measure (see (53)).

2.4 Convergence to the Limit System—Part I

Our goal now is to exploit the uniform bounds obtained in the previous part to pass to the limit in the sequence $\{\rho_\varepsilon, \mathbf{u}_\varepsilon, \vartheta_\varepsilon\}_{\varepsilon>0}$ for $\varepsilon \to 0$. To begin, we observe that

(48), (50) yield

$$\operatorname*{ess\,sup}_{t\in(0,T)} \|\rho_\varepsilon(t,\cdot) - \tilde{\rho}_\varepsilon\|_{(L^2\oplus L^{5/3})(\Omega_\varepsilon)} \to 0 \quad \text{as } \varepsilon \to 0. \tag{59}$$

In particular, by virtue of (38),

$$\operatorname*{ess\,sup}_{t\in(0,T)} \|\rho_\varepsilon(t,\cdot) - \overline{\rho}\|_{L^{5/3}(K)} \to 0 \quad \text{as } \varepsilon \to 0 \text{ for any compact } K \subset \Omega. \tag{60}$$

Thus the fluid density becomes constant provided the Mach number tends to zero.

Similarly, relations (49), (51), and (53) yield

$$\operatorname*{ess\,sup}_{t\in(0,T)} \|\vartheta_\varepsilon(t,\cdot) - \overline{\vartheta}\|_{L^2(\Omega_\varepsilon)} \to 0 \quad \text{as } \varepsilon \to 0. \tag{61}$$

Next, in order to control the temperature deviations from the equilibrium state $\overline{\vartheta}$, we use (57), (58) to deduce that

$$\Theta_\varepsilon \equiv \frac{\vartheta_\varepsilon - \overline{\vartheta}}{\varepsilon} \to \Theta \quad \text{weakly in } L^2(0,T; W^{1,2}(\Omega)). \tag{62}$$

Moreover, by the same token,

$$\mathbf{u}_\varepsilon \to \mathbf{U} \quad \text{weakly in } L^2(0,T; W^{1,2}(\Omega; R^3)), \tag{63}$$

passing to subsequences if necessary. Here, we have assumed that ϑ_ε, $\mathbf{u}_\varepsilon$ were extended to the whole domain Ω.

A short inspection of the scaled Navier–Stokes–Fourier system (11–13) reveals the most difficult step, namely we need to show strong (pointwise) convergence of the velocity in order to control the convective term. More specifically, we need to show that

$$\mathbf{u}_\varepsilon \to \mathbf{U} \quad \text{(strongly) in } L^2((0,T)\times K; R^3) \text{ for any compact } K \subset \Omega. \tag{64}$$

As a matter of fact, it is enough to prove that

$$\rho_\varepsilon \mathbf{u}_\varepsilon \to \overline{\rho}\mathbf{U} \quad \text{in } L^2(0,T; W^{-1,2}(K)). \tag{65}$$

Indeed, for any $\varphi \in C_c^\infty(\Omega)$, we have

$$\overline{\rho}\int_0^T \int_\Omega \varphi |\mathbf{u}_\varepsilon|^2 \,\mathrm{d}x\,\mathrm{d}t = \int_0^T \int_\Omega \varphi(\overline{\rho} - \rho_\varepsilon)|\mathbf{u}_\varepsilon|^2 \,\mathrm{d}x\,\mathrm{d}t + \int_0^T \int_\Omega \varphi\rho_\varepsilon \mathbf{u}_\varepsilon \cdot \mathbf{u}_\varepsilon \,\mathrm{d}x\,\mathrm{d}t,$$

where, by virtue of the previous estimates and the embedding relation $W^{1,2}(\Omega) \hookrightarrow L^6(\Omega)$,

$$\int_0^T \int_\Omega \varphi(\overline{\rho} - \rho_\varepsilon)|\mathbf{u}_\varepsilon|^2 \,\mathrm{d}x\,\mathrm{d}t \to 0,$$

while, as a consequence of (63), (65),

$$\int_0^T \int_\Omega \varphi \rho_\varepsilon \mathbf{u}_\varepsilon \cdot \mathbf{u}_\varepsilon \, dx \, dt \to \overline{\rho} \int_0^T \int_\Omega \varphi |\mathbf{U}|^2 \, dx \, dt.$$

The final observation is that for (65) to hold it is enough to show that

$$\left\{ t \mapsto \int_\Omega (\rho_\varepsilon \mathbf{u}_\varepsilon)(t, \cdot) \cdot \varphi \, dx \right\} \quad \text{is precompact in } L^2(0,T) \tag{66}$$

for any fixed $\varphi \in C_c^\infty(\Omega)$ since, as a consequence of (47), (48), and (50),

$$\operatorname*{ess\,sup}_{t \in (0,T)} \| \rho_\varepsilon \mathbf{u}_\varepsilon \|_{L^{5/4}(K;R^3)} \leq c(K) \quad \text{for any compact } K \subset \Omega$$

and the embedding $L^{5/4}(K) \hookrightarrow W^{-1,2}(K)$ is compact. Accordingly, we fix $\varphi \in C_c^\infty(\Omega)$ for the remaining part of this section and focus on proving (66).

2.5 Acoustic Equation

As already pointed out, our main goal is to show (66) for any fixed $\varphi \in C_c^\infty(\Omega)$. To this end, we rewrite the Navier–Stokes–Fourier system in the form

$$\varepsilon \partial_t R_\varepsilon + \omega \operatorname{div}_x \mathbf{V}_\varepsilon = \varepsilon f_\varepsilon^1, \tag{67}$$

$$\varepsilon \partial_t \mathbf{V}_\varepsilon + \nabla_x R_\varepsilon = \varepsilon \mathbf{f}_\varepsilon^2, \tag{68}$$

where we have set

$$R_\varepsilon = A \left(\frac{\rho_\varepsilon - \overline{\rho}}{\varepsilon} \right) + B \left(\frac{\rho_\varepsilon s(\rho_\varepsilon, \vartheta_\varepsilon) - \overline{\rho} s(\overline{\rho}, \overline{\vartheta})}{\varepsilon} \right) - \overline{\rho} G, \qquad \mathbf{V}_\varepsilon = \rho_\varepsilon \mathbf{u}_\varepsilon,$$

$$f_\varepsilon^1 = B \left[\operatorname{div}_x \left(\rho_\varepsilon \frac{s(\overline{\rho}, \overline{\vartheta}) - s(\rho_\varepsilon, \vartheta_\varepsilon)}{\varepsilon} \mathbf{u}_\varepsilon \right) + \operatorname{div}_x \left(\frac{\kappa(\vartheta_\varepsilon)}{\vartheta_\varepsilon} \frac{\nabla_x \vartheta_\varepsilon}{\varepsilon} \right) + \frac{1}{\varepsilon} \sigma_\varepsilon \right],$$

and

$$\mathbf{f}_\varepsilon^2 = \frac{1}{\varepsilon} \nabla_x \left[A \left(\frac{\rho_\varepsilon - \overline{\rho}}{\varepsilon} \right) + B \left(\frac{\rho_\varepsilon s(\rho_\varepsilon, \vartheta_\varepsilon) - \overline{\rho} s(\overline{\rho}, \overline{\vartheta})}{\varepsilon} \right) \right.$$
$$\left. - \left(\frac{p(\rho_\varepsilon, \vartheta_\varepsilon) - p(\overline{\rho}, \overline{\vartheta})}{\varepsilon} \right) \right] - \operatorname{div}_x (\rho_\varepsilon \mathbf{u}_\varepsilon \otimes \mathbf{u}_\varepsilon) + \operatorname{div}_x \mathbb{S}_\varepsilon + \frac{\rho_\varepsilon - \overline{\rho}}{\varepsilon} \nabla_x G,$$

and where the constants A, B, ω are chosen so that

$$B \overline{\rho} \partial_\vartheta s(\overline{\rho}, \overline{\vartheta}) = \partial_\vartheta p(\overline{\rho}, \overline{\vartheta}), \qquad A + B \partial_\rho (\rho s)(\overline{\rho}, \overline{\vartheta}) = p_\rho(\overline{\rho}, \overline{\vartheta}),$$

and

$$\omega = p_\rho(\overline{\rho}, \overline{\vartheta}) + \frac{|p_\vartheta(\overline{\rho}, \overline{\vartheta})|^2}{\overline{\rho}^2 s_\vartheta(\overline{\rho}, \overline{\vartheta})} > 0.$$

System (67), (68) is usually termed *acoustic equation*, or, *Lighthill's acoustic analogy*, see Lighthill [19, 20].

The inevitable presence of the measure σ_ε in the forcing term f_ε^1 may cause discontinuities (in time) in solutions of the system (67), (68); therefore it seems more convenient to use the time-lifting Σ_ε of the measure σ_ε introduced in Sect. 2.1. With the new variables

$$S_\varepsilon = A\left(\frac{\rho_\varepsilon - \overline{\rho}}{\varepsilon}\right) + B\left(\frac{\rho_\varepsilon s(\rho_\varepsilon, \vartheta_\varepsilon) - \overline{\rho} s(\overline{\rho}, \overline{\vartheta})}{\varepsilon}\right) - \overline{\rho} G + \frac{B}{\varepsilon}\Sigma_\varepsilon, \qquad \mathbf{V}_\varepsilon = \rho_\varepsilon \mathbf{u}_\varepsilon, \tag{69}$$

we may write the acoustic equation in the form

$$\varepsilon \partial_t S_\varepsilon + \omega \operatorname{div}_x \mathbf{V}_\varepsilon = \varepsilon F_\varepsilon^1, \tag{70}$$

$$\varepsilon \partial_t \mathbf{V}_\varepsilon + \nabla_x S_\varepsilon = \varepsilon \mathbf{F}_\varepsilon^2, \tag{71}$$

with

$$F_\varepsilon^1 = B\left[\operatorname{div}_x\left(\rho_\varepsilon \frac{s(\overline{\rho}, \overline{\vartheta}) - s(\rho_\varepsilon, \vartheta_\varepsilon)}{\varepsilon}\mathbf{u}_\varepsilon\right) + \operatorname{div}_x\left(\frac{\kappa(\vartheta_\varepsilon)}{\vartheta_\varepsilon}\frac{\nabla_x \vartheta_\varepsilon}{\varepsilon}\right)\right], \tag{72}$$

and

$$\begin{aligned}\mathbf{F}_\varepsilon^2 = \frac{1}{\varepsilon}\nabla_x\Bigg[&A\left(\frac{\rho_\varepsilon - \overline{\rho}}{\varepsilon}\right) + B\left(\frac{\rho_\varepsilon s(\rho_\varepsilon, \vartheta_\varepsilon) - \overline{\rho} s(\overline{\rho}, \overline{\vartheta})}{\varepsilon}\right) \\ &- \left(\frac{p(\rho_\varepsilon, \vartheta_\varepsilon) - p(\overline{\rho}, \overline{\vartheta})}{\varepsilon}\right)\Bigg] - \operatorname{div}_x(\rho_\varepsilon \mathbf{u}_\varepsilon \otimes \mathbf{u}_\varepsilon) + \operatorname{div}_x \mathbb{S}_\varepsilon + \frac{\rho_\varepsilon - \overline{\rho}}{\varepsilon}\nabla_x G \\ &+ \frac{B}{\varepsilon^2}\nabla_x \Sigma_\varepsilon,\end{aligned} \tag{73}$$

supplemented with the homogeneous Neumann boundary condition

$$\mathbf{V}_\varepsilon \cdot \mathbf{n}|_{\partial\Omega_\varepsilon} = 0. \tag{74}$$

Of course, system (69–74) should be understood in the weak sense as specified in the following section.

Boundedness of the Data in the Acoustic Equation

As suggested by the previous discussion, the system (70), (71) will describe the behavior of the velocity field or rather the momentum $\mathbf{V}_\varepsilon$, while the remaining quantities appearing F_ε^1, $\mathbf{F}_\varepsilon^2$ are given. Using the uniform bounds established in Sect. 2.3, we estimate the forcing terms as well as the initial data in the acoustic equation. To begin write, using the decomposition introduced by (46)

$$\frac{\varrho_\varepsilon - \overline{\varrho}}{\varepsilon} = \frac{\varrho_\varepsilon - \tilde{\varrho}_\varepsilon}{\varepsilon} + \frac{\tilde{\varrho}_\varepsilon - \overline{\varrho}}{\varepsilon} = \left[\frac{\varrho_\varepsilon - \tilde{\varrho}_\varepsilon}{\varepsilon}\right]_{\rm ess} + \left[\frac{\varrho_\varepsilon - \tilde{\varrho}_\varepsilon}{\varepsilon}\right]_{\rm res} + \frac{\tilde{\varrho}_\varepsilon - \overline{\varrho}}{\varepsilon},$$

where, in accordance with (48), (50), and (53), we have

$$\begin{aligned} \operatorname*{ess\,sup}_{t\in(0,T)} \left\| \left[\frac{\varrho_\varepsilon - \tilde{\varrho}_\varepsilon}{\varepsilon}\right]_{\rm ess} \right\|_{L^2(\Omega_\varepsilon)} &\leq c, \\ \operatorname*{ess\,sup}_{t\in(0,T)} \left\| \left[\frac{\varrho_\varepsilon - \tilde{\varrho}_\varepsilon}{\varepsilon}\right]_{\rm res} \right\|_{L^1(\Omega_\varepsilon)} &\leq \varepsilon c, \end{aligned} \tag{75}$$

and, moreover, using (38), (39) it follows that

$$\begin{aligned} \left\| \frac{\tilde{\varrho}_\varepsilon - \overline{\varrho}}{\varepsilon} \right\|_{(L^\infty \cap L^q)(R^3)} &\leq c \quad \text{for any } q > 3, \\ \left\| \nabla_x \left(\frac{\tilde{\varrho}_\varepsilon - \overline{\varrho}}{\varepsilon} \right) \right\|_{L^2(R^3;R^3)} &\leq c. \end{aligned} \tag{76}$$

The next step is to write

$$\begin{aligned} &\frac{\varrho_\varepsilon s(\varrho_\varepsilon, \vartheta_\varepsilon) - \overline{\varrho} s(\overline{\varrho}, \overline{\vartheta})}{\varepsilon} \\ &\quad = \frac{\varrho_\varepsilon s(\varrho_\varepsilon, \vartheta_\varepsilon) - \tilde{\varrho}_\varepsilon s(\tilde{\varrho}_\varepsilon, \overline{\vartheta})}{\varepsilon} + \frac{\tilde{\varrho}_\varepsilon s(\tilde{\varrho}_\varepsilon, \overline{\vartheta}) - \overline{\varrho} s(\overline{\varrho}, \overline{\vartheta})}{\varepsilon} \\ &\quad = \left[\frac{\varrho_\varepsilon s(\varrho_\varepsilon, \vartheta_\varepsilon) - \tilde{\varrho}_\varepsilon s(\tilde{\varrho}_\varepsilon, \overline{\vartheta})}{\varepsilon}\right]_{\rm ess} + \left[\frac{\varrho_\varepsilon s(\varrho_\varepsilon, \vartheta_\varepsilon) - \tilde{\varrho}_\varepsilon s(\tilde{\varrho}_\varepsilon, \overline{\vartheta})}{\varepsilon}\right]_{\rm res} \\ &\qquad + \frac{\tilde{\varrho}_\varepsilon s(\tilde{\varrho}_\varepsilon, \overline{\vartheta}) - \overline{\varrho} s(\overline{\varrho}, \overline{\vartheta})}{\varepsilon}, \end{aligned}$$

where, in accordance with the uniform bounds established in Sect. 2.3,

$$\begin{aligned} \operatorname*{ess\,sup}_{t\in(0,T)} \left\| \left[\frac{\varrho_\varepsilon s(\varrho_\varepsilon, \vartheta_\varepsilon) - \tilde{\varrho}_\varepsilon s(\tilde{\varrho}_\varepsilon, \overline{\vartheta})}{\varepsilon}\right]_{\rm ess} \right\|_{L^2(\Omega_\varepsilon)} &\leq c, \\ \operatorname*{ess\,sup}_{t\in(0,T)} \left\| \left[\frac{\varrho_\varepsilon s(\varrho_\varepsilon, \vartheta_\varepsilon) - \tilde{\varrho}_\varepsilon s(\tilde{\varrho}_\varepsilon, \overline{\vartheta})}{\varepsilon}\right]_{\rm res} \right\|_{L^1(\Omega_\varepsilon)} &\leq \varepsilon c, \end{aligned}$$

and

$$\left\| \frac{\tilde{\varrho}_\varepsilon s(\tilde{\varrho}_\varepsilon, \overline{\vartheta}) - \overline{\varrho} s(\overline{\varrho}, \overline{\vartheta})}{\varepsilon} \right\|_{(L^\infty \cap L^q)(R^3)} \leq c \quad \text{for all } q > 3,$$

$$\left\| \nabla_x \left(\frac{\tilde{\rho}_\varepsilon s(\tilde{\rho}_\varepsilon, \overline{\vartheta}) - \overline{\rho} s(\overline{\rho}, \overline{\vartheta})}{\varepsilon} \right) \right\|_{L^2(R^3;R^3)} \leq c.$$

Furthermore, by virtue of (54),

$$\operatorname*{ess\,sup}_{t\in(0,T)} \left\| \frac{\Sigma_\varepsilon(t,\cdot)}{\varepsilon} \right\|_{\mathscr{M}^+(\overline{\Omega}_\varepsilon)} \leq \varepsilon c,$$

therefore we may write

$$S_\varepsilon(t) = S_\varepsilon^1(t) + S_\varepsilon^2(t) + S_\varepsilon^3,$$

with

$$\operatorname*{ess\,sup}_{t\in(0,T)} \|S_\varepsilon^1\|_{\mathscr{M}^+(\overline{\Omega}_\varepsilon)} \leq \varepsilon c, \qquad \operatorname*{ess\,sup}_{t\in(0,T)} \|S_\varepsilon^2\|_{L^2(\Omega_\varepsilon)} \leq c,$$
$$\|S_\varepsilon^3\|_{D^{1,2}(R^3)} \leq c,$$

where the symbol $D^{1,2}$ denotes the homogeneous Sobolev space—a completion of compactly supported smooth functions with respect to the L^2-norm of their gradients.

Next, writing

$$\mathbf{V}_\varepsilon = [\rho_\varepsilon \mathbf{u}_\varepsilon]_{\text{ess}} + [\rho_\varepsilon \mathbf{u}_\varepsilon]_{\text{res}},$$

we have, in agreement with (47), (50), (53),

$$\begin{aligned} &\operatorname*{ess\,sup}_{t\in(0,T)} \|[\rho_\varepsilon \mathbf{u}_\varepsilon]_{\text{ess}}\|_{L^2(\Omega_\varepsilon;R^3)} \leq c, \\ &\operatorname*{ess\,sup}_{t\in(0,T)} \|[\rho_\varepsilon \mathbf{u}_\varepsilon]_{\text{res}}\|_{L^1(\Omega_\varepsilon;R^3)} \leq \varepsilon c. \end{aligned} \tag{77}$$

Other terms appearing in F_ε^1, $\mathbf{F}_\varepsilon^2$ can be treated in a similar manner. We focus only on the most complicated expression:

$$\begin{aligned} &A\left(\frac{\rho_\varepsilon - \overline{\rho}}{\varepsilon^2}\right) + B\left(\frac{\rho_\varepsilon s(\rho_\varepsilon, \vartheta_\varepsilon) - \overline{\rho} s(\overline{\rho}, \overline{\vartheta})}{\varepsilon^2}\right) - \left(\frac{p(\rho_\varepsilon, \vartheta_\varepsilon) - p(\overline{\rho}, \overline{\vartheta})}{\varepsilon^2}\right) \\ &\quad = A\left(\frac{\rho_\varepsilon - \tilde{\rho}_\varepsilon}{\varepsilon^2}\right) + B\left(\frac{\rho_\varepsilon s(\rho_\varepsilon, \vartheta_\varepsilon) - \tilde{\rho}_\varepsilon s(\tilde{\rho}_\varepsilon, \overline{\vartheta})}{\varepsilon^2}\right) - \left(\frac{p(\rho_\varepsilon, \vartheta_\varepsilon) - p(\tilde{\rho}_\varepsilon, \overline{\vartheta})}{\varepsilon^2}\right) \\ &\qquad + A\left(\frac{\tilde{\rho}_\varepsilon - \overline{\rho}}{\varepsilon^2}\right) + B\left(\frac{\tilde{\rho}_\varepsilon s(\tilde{\rho}_\varepsilon, \overline{\vartheta}) - \overline{\rho} s(\overline{\rho}, \overline{\vartheta})}{\varepsilon^2}\right) - \left(\frac{p(\tilde{\rho}_\varepsilon, \overline{\vartheta}) - p(\overline{\rho}, \overline{\vartheta})}{\varepsilon^2}\right). \end{aligned}$$

Seeing that

$$A + B\partial_\rho(\rho s)(\overline{\rho}, \overline{\vartheta}) - \partial_\rho p(\overline{\rho}, \overline{\vartheta}) = 0,$$

the quantity

$$A\left(\frac{\rho_\varepsilon - \tilde{\rho}_\varepsilon}{\varepsilon^2}\right) + B\left(\frac{\rho_\varepsilon s(\rho_\varepsilon, \vartheta_\varepsilon) - \tilde{\rho}_\varepsilon s(\tilde{\rho}_\varepsilon, \overline{\vartheta})}{\varepsilon^2}\right) - \left(\frac{p(\rho_\varepsilon, \vartheta_\varepsilon) - p(\tilde{\rho}_\varepsilon, \overline{\vartheta})}{\varepsilon^2}\right)$$

contains only quadratic terms proportional to $\rho_\varepsilon - \tilde{\rho}_\varepsilon$, $\vartheta - \overline{\vartheta}$ and as such may be estimated in terms of (48)–(53). Similarly,

$$\begin{aligned}&\left\| A\left(\frac{\tilde{\rho}_\varepsilon - \overline{\rho}}{\varepsilon^2}\right) + B\left(\frac{\tilde{\rho}_\varepsilon s(\tilde{\rho}_\varepsilon, \overline{\vartheta}) - \overline{\rho} s(\overline{\rho}, \overline{\vartheta})}{\varepsilon^2}\right)\right. \\ &\quad \left. - \left(\frac{p(\tilde{\rho}_\varepsilon, \overline{\vartheta}) - p(\overline{\rho}, \overline{\vartheta})}{\varepsilon^2}\right)\right\|_{(L^\infty \cap L^q)(R^3)} \leq c \quad \text{for all } q > 3/2.\end{aligned}$$

Summing up the previous estimates we may write down a weak formulation of the acoustic equation in the form:

$$\begin{aligned}&\varepsilon \int_0^T \langle S_\varepsilon(t,\cdot), \partial_t \varphi \rangle \, \mathrm{d}t + \omega \int_0^T \int_{\Omega_\varepsilon} \mathbf{V}_\varepsilon \cdot \nabla_x \varphi \, \mathrm{d}x \, \mathrm{d}t \\ &\quad = -\varepsilon \langle S_{0,\varepsilon}, \varphi(0,\cdot) \rangle + \varepsilon \int_0^T \int_{\Omega_\varepsilon} \left(\mathbf{H}^1_\varepsilon \cdot \nabla_x \varphi + \mathbf{H}^2_\varepsilon \cdot \nabla_x \varphi\right) \mathrm{d}x \, \mathrm{d}t, \end{aligned} \tag{78}$$

for any $\varphi \in C^1_c([0,T) \times \overline{\Omega}_\varepsilon)$,

$$\begin{aligned}&\varepsilon \int_0^T \int_{\Omega_\varepsilon} \mathbf{V}_\varepsilon \cdot \partial_t \varphi \, \mathrm{d}x \, \mathrm{d}t + \int_0^T \langle S_\varepsilon(t,\cdot), \mathrm{div}_x \, \varphi \rangle \, \mathrm{d}t \\ &\quad = -\varepsilon \int_{\Omega_\varepsilon} \mathbf{V}_{0,\varepsilon} \cdot \varphi(0,\cdot) \, \mathrm{d}x + \varepsilon \int_0^T \langle \mathbb{G}^1_\varepsilon(t,\cdot), \nabla_x \varphi \rangle \, \mathrm{d}t \\ &\quad\quad + \varepsilon \int_0^T \int_\Omega \mathbb{G}^2_\varepsilon : \nabla_x \varphi \, \mathrm{d}x \, \mathrm{d}t + \varepsilon \int_0^T \int_\Omega \mathbf{G}^3_\varepsilon \cdot \varphi \, \mathrm{d}x \, \mathrm{d}t \end{aligned} \tag{79}$$

for any $\varphi \in C^1_c([0,T) \times \overline{\Omega}_\varepsilon; R^3)$, $\varphi \cdot \mathbf{n}|_{\partial\Omega} = 0$, where

$$\begin{aligned}&S_\varepsilon = S^1_\varepsilon + S^2_\varepsilon + S^{1,2}_\varepsilon, \\ &\operatorname*{ess\,sup}_{t \in (0,T)} \|S^1_\varepsilon(t,\cdot)\|_{\mathscr{M}^1(\overline{\Omega}_\varepsilon)} \leq \varepsilon c, \\ &\operatorname*{ess\,sup}_{t \in (0,T)} \|S^2_\varepsilon(t,\cdot)\|_{L^2(\overline{\Omega}_\varepsilon)} + \|S^{1,2}_\varepsilon\|_{D^{1,2}(R^3)} \leq c, \\ &S_{0,\varepsilon} = S^1_{0,\varepsilon} + S^2_{0,\varepsilon} + S^{1,2}_\varepsilon,\end{aligned} \tag{80}$$

$$\|S^1_{0,\varepsilon}\|_{\mathscr{M}^1(\overline{\Omega}_\varepsilon)} \leq \varepsilon c, \qquad \|S^2_{0,\varepsilon}\|_{L^2(\Omega_\varepsilon)} \leq c, \tag{81}$$

and, moreover,

$$S_\varepsilon \in C_{\text{weak-}(*)}([0,T]; \mathscr{M}^+(\overline{\Omega}_\varepsilon)).$$

Furthermore,

$$\mathbf{V}_\varepsilon = \mathbf{V}^1_\varepsilon + \mathbf{V}^2_\varepsilon,$$
$$\operatorname*{ess\,sup}_{t\in(0,T)} \|\mathbf{V}^1_\varepsilon\|_{L^1(\Omega_\varepsilon;R^3)} \le \varepsilon c, \qquad \operatorname*{ess\,sup}_{t\in(0,T)} \|\mathbf{V}^2_\varepsilon\|_{L^2(\Omega_\varepsilon;R^3)} \le c, \tag{82}$$

$$\|\mathbf{V}_{0,\varepsilon}\|_{(L^\infty\cap L^2)(\Omega_\varepsilon;R^3)} \le c, \tag{83}$$

and

$$\mathbf{V}_\varepsilon \in C_{\text{weak}}([0,T]; L^1(\Omega_\varepsilon)).$$

Finally,

$$\int_0^T \left(\|\mathbf{H}^1_\varepsilon\|^2_{L^1(\Omega_\varepsilon;R^3)} + \|\mathbf{H}^2_\varepsilon\|^2_{L^2(\Omega_\varepsilon;R^3)} \right) \mathrm{d}t \le c, \tag{84}$$

$$\int_0^T \left(\|\mathbb{G}^1_\varepsilon\|^2_{\mathscr{M}^+(\Omega_\varepsilon;R^{3\times3})} + \|\mathbb{G}^2_\varepsilon\|^2_{L^2(\Omega_\varepsilon;R^{3\times3})} \right) \mathrm{d}t \le c, \tag{85}$$

and

$$\operatorname*{ess\,sup}_{t\in(0,T)} \|\mathbf{G}^3_\varepsilon(t,\cdot)\|_{L^{5/3}(R^3;R^3)} \le c, \tag{86}$$

where all constants are independent of ε.

Reduction to Smooth Solutions

With the notation introduced in the previous section, the desired relation (66) reads:

$$\left\{ t \mapsto \int_\Omega \mathbf{V}_\varepsilon(t,\cdot)\cdot\varphi \,\mathrm{d}x \right\} \quad \text{is precompact in } L^2(0,T) \text{ for any } \varphi \in C^\infty_c(\Omega; R^3). \tag{87}$$

In order to see (87), it is more convenient to deal with the classical (smooth) solutions of acoustic equation (78), (79). Since $\varphi \in C^\infty_c(\Omega; R^3)$ is fixed, the idea is to replace the data in (78), (79) by smooth ones in such a way that the resulting smooth solution of (78), (79) is close to $\mathbf{V}_\varepsilon$ at least on the support of φ. To this end, fixing $\varepsilon > 0$ for a moment, we consider

$$S^i_{0,\varepsilon,\delta} \in C^\infty_c(\Omega_\varepsilon), \quad i = 1,2,3,$$
$$\|S^1_{0,\varepsilon,\delta}\|_{L^1(\Omega)} + \|S^2_{0,\varepsilon,\delta}\|_{L^2(\Omega)} + \|S^3_{0,\varepsilon,\delta}\|_{D^{1,2}(R^3)} \le c, \tag{88}$$

such that

$$S^1_{0,\varepsilon,\delta} \to S^1_{0,\varepsilon} \quad \text{weakly-}(*) \text{ in } \mathscr{M}^+(\overline{\Omega}_\varepsilon),$$
$$S^j_{0,\varepsilon,\delta} \to S^j_{0,\varepsilon} \quad \text{in } L^2(\Omega_\varepsilon),\ j = 2,3, \text{ for } \delta \to 0.$$

Similarly, take

$$\mathbf{V}^i_{0,\varepsilon,\delta} \in C^\infty_c(\Omega_\varepsilon; R^3), \quad i = 1, 2, \qquad \|\mathbf{V}^1_{0,\varepsilon,\delta}\|_{L^1(\Omega;R^3)} + \|\mathbf{V}^2_{0,\varepsilon,\delta}\|_{L^2(\Omega;R^3)} \le c,$$
$$\mathbf{V}^1_{0,\varepsilon,\delta} \to \mathbf{V}_{0,\varepsilon} \quad \text{in } L^1(\Omega_\varepsilon; R^3), \qquad \mathbf{V}^2_{0,\varepsilon,\delta} \to \mathbf{V}_{0,\varepsilon} \quad \text{in } L^2(\Omega_\varepsilon; R^3) \text{ as } \delta \to 0, \tag{89}$$

and, finally,

$$\begin{gathered}
\mathbf{H}^i_{\varepsilon,\delta} \in C^\infty_c((0,T) \times \Omega_\varepsilon; R^3), \quad i = 1, 2, \\
\|\mathbf{H}^1_{\varepsilon,\delta}\|_{L^2(0,T;L^1(\Omega;R^3))} + \|\mathbf{H}^2_{\varepsilon,\delta}\|_{L^2(0,T;L^2(\Omega;R^3))} \le c, \\
\mathbf{H}^1_{\varepsilon,\delta} \to \mathbf{H}^1_\varepsilon \quad \text{in } L^2(0,T; L^1(\Omega_\varepsilon; R^3)), \\
\mathbf{H}^2_{\varepsilon,\delta} \to \mathbf{H}^2_\varepsilon \quad \text{in } L^2(0,T; L^2(\Omega_\varepsilon; R^3)) \text{ as } \delta \to 0
\end{gathered} \tag{90}$$

with

$$\mathbb{G}^i_{\varepsilon,\delta} \in C^\infty_c((0,T) \times \Omega_\varepsilon; R^{3\times 3}), \quad i = 1, 2,$$

$$\|\mathbb{G}^1_{\varepsilon,\delta}\|_{L^2(0,T;L^1(\Omega;R^{3\times3}))} + \|\mathbb{G}^2_{\varepsilon,\delta}\|_{L^2(0,T;L^2(\Omega;R^{3\times3}))} \le c,$$

$$\mathbb{G}^1_{\varepsilon,\delta} \to \mathbb{G}^1_\varepsilon \quad \text{weakly-}(*) \text{ in } \ L^2(0,T; \mathscr{M}^+(\Omega_\varepsilon; R^{3\times3})), \tag{91}$$

$$\mathbb{G}^2_{\varepsilon,\delta} \to \mathbb{G}^2_\varepsilon \quad \text{in } L^2(0,T; L^2(\Omega_\varepsilon; R^{3\times3})) \text{ as } \delta \to 0, \tag{92}$$

$$\mathbf{G}^3_{\varepsilon,\delta} \in C^\infty_c(\Omega_\varepsilon; R^3), \qquad \|\mathbf{G}^3_{\varepsilon,\delta}\|_{L^{5/3}(\Omega;R^3)} \le c,$$

$$\mathbf{G}^3_{\varepsilon,\delta} \to \mathbf{G}^3_\varepsilon \quad \text{in } L^{5/3}((0,T) \times \Omega_\varepsilon; R^3) \text{ as } \delta \to 0. \tag{93}$$

Assume that $S_{\varepsilon,\delta}$, $\mathbf{V}_{\varepsilon,\delta}$ is the (unique) classical solution of the acoustic equation (78), (79), with the initial data and the forcing terms replaced by their δ-approximations specified in (88–93). Keeping (87) in mind we will show that

$$\operatorname*{ess\,sup}_{t\in(0,T)} \left| \int_{\Omega_\varepsilon} \big(\mathbf{V}_{\varepsilon,\delta}(t,\cdot) - \mathbf{V}_\varepsilon(t,\cdot)\big) \cdot \varphi \, \mathrm{d}x \right| \le \varepsilon \quad \text{whenever } \delta \text{ is small enough}, \tag{94}$$

for any fixed $\varepsilon > 0$. Consequently, it follows from (94) that it is enough to show (87) for $\mathbf{V}_{\varepsilon,\delta(\varepsilon)}$. In other words, we may assume that all the quantities appearing in the acoustic equation are smooth and all the data is compactly supported in Ω_ε.

To see (94), we fix ε and write the function φ in terms of its *Helmholtz decomposition*,

$$\boldsymbol{\varphi} = \mathbf{H}[\varphi] + \mathbf{H}^\perp[\varphi],$$

where

$$\mathbf{H}^\perp[\varphi] = \nabla_x \psi, \qquad \Delta\psi = \operatorname{div}_x \varphi \quad \text{in } \Omega_\varepsilon, \qquad \nabla_x \psi \cdot \mathbf{n}|_{\partial\Omega_\varepsilon} = 0.$$

Taking $\mathbf{H}[\varphi]$ as a test function in (79) we easily deduce that

$$\sup_{t\in(0,T)} \left| \int_{\Omega_\varepsilon} \big(\mathbf{V}_{\varepsilon,\delta}(t,\cdot) - \mathbf{V}_\varepsilon(t,\cdot)\big) \cdot \mathbf{H}[\varphi]\,\mathrm{d}x \right| \le \varepsilon \tag{95}$$

whenever $\delta = \delta(\varepsilon)$ is small enough.

Now, let $\{\psi_n\}_{n=0}^\infty$ be an orthonormal system of eigenfunctions of the Laplace operator in Ω_ε endowed with the homogeneous Neumann boundary conditions, specifically,

$$-\Delta\psi_n = \lambda_n\psi_n \quad \text{in } \Omega_\varepsilon, \qquad \nabla_x\psi_n \cdot \mathbf{n}|_{\partial\Omega_\varepsilon} = 0, \quad n = 0, 1, \dots.$$

Taking the quantities $\phi(t)\psi_n(x)$, $\phi(t)\nabla_x\psi_n$, $\phi \in C_c^\infty(0,T)$ as test functions in (78), (79), respectively, we obtain a system of two ordinary differential equations:

$$\varepsilon\partial_t \int_{\Omega_\varepsilon} \mathbf{V}(t,\cdot)\cdot\nabla_x\psi_n\,\mathrm{d}x - \lambda_n\langle S(t,\cdot),\psi_n\rangle = \varepsilon f_n^1,$$

$$\varepsilon\partial_t\langle S(t,\cdot),\psi_n\rangle + \int_{\Omega_\varepsilon} \mathbf{V}(t,\cdot)\cdot\nabla_x\psi_n\,\mathrm{d}x = \varepsilon f_n^2$$

for the unknown functions of time:

$$\left\{t \mapsto \int_{\Omega_\varepsilon} \mathbf{V}(t,\cdot)\cdot\nabla_x\psi_n\,\mathrm{d}x\right\}, \qquad \{t\mapsto\langle S(t,\cdot),\psi_n\rangle\},$$

where the initial data as well as the forcing terms f_n^1, f_ε^2 can be evaluated in terms of the (ε,δ)-quantities. Consequently, we infer that for given $\varepsilon > 0$, $N > 0$, there exists $\delta = \delta(N,\varepsilon) > 0$ such that

$$\sup_{t\in(0,T)} \left| \int_{\Omega_\varepsilon} \big(\mathbf{V}_{\varepsilon,\delta}(t,\cdot) - \mathbf{V}_\varepsilon(t,\cdot)\big)\cdot\nabla_x\psi_n\,\mathrm{d}x \right| \le \varepsilon \quad \text{whenever } \delta \le \delta(N,\varepsilon) \tag{96}$$

for any $n \le N$.

Finally, since $\mathbf{V}_\varepsilon$ admits the bound (82), we have

$$\sup_{t\in(0,T)} \left| \int_{\Omega_\varepsilon} \mathbf{V}_\varepsilon\cdot\nabla_x\left(\Delta_N^{-1}\operatorname{div}_x\varphi - P_M[\Delta_N^{-1}\operatorname{div}_x\varphi]\right)\mathrm{d}x \right| \le \varepsilon \quad \text{for all } M > M(\varphi), \tag{97}$$

where P_M denotes the orthogonal projection onto $\operatorname{span}\{\psi_1,\dots,\psi_M\}$. Moreover,

$$\begin{aligned}
&\int_{\Omega_\varepsilon} \mathbf{V}_{\varepsilon,\delta}\cdot\nabla_x\big(\Delta_N^{-1}\operatorname{div}_x\varphi - P_M[\Delta_N^{-1}\operatorname{div}_x\varphi]\big)\,\mathrm{d}x \\
&\quad = \int_{\Omega_\varepsilon} \nabla_x\Psi_{\varepsilon,\delta}\cdot\nabla_x\big(\Delta_N^{-1}\operatorname{div}_x\varphi - P_M[\Delta_N^{-1}\operatorname{div}_x\varphi]\big)\,\mathrm{d}x \\
&\quad = -\int_{\Omega_\varepsilon} \Psi_{\varepsilon,\delta}\big(\operatorname{div}_x\varphi - P_M[\operatorname{div}_x\varphi]\big)\,\mathrm{d}x,
\end{aligned}$$

where $\Psi_{\varepsilon,\delta}$ solves a wave equation

$$\varepsilon \partial_t S_{\varepsilon,\delta} + \omega \Delta \Psi_{\varepsilon,\delta} = \varepsilon \operatorname{div}_x \big(\mathbf{H}^1_{\varepsilon,\delta} + \mathbf{H}^2_{\varepsilon,\delta} \big), \tag{98}$$

$$\varepsilon \partial_t \Psi_{\varepsilon,\delta} + S_{\varepsilon,\delta} = \varepsilon \Delta_N^{-1} \big[\operatorname{div}_x \operatorname{div}_x \big(\mathbb{G}^1_{\varepsilon,\delta} + \mathbb{G}^2_{\varepsilon,\delta} \big) \big] + \varepsilon \Delta_N^{-1} [\operatorname{div}_x \mathbf{G}^3_{\varepsilon,\delta}], \tag{99}$$

supplemented with the boundary conditions

$$\nabla_x \Psi_{\varepsilon,\delta} \cdot \mathbf{n}|_{\partial \Omega_\varepsilon} = 0. \tag{100}$$

Thus in view of the uniform bounds (88–93), we can find $M = M(\varepsilon) > 0$ such that

$$\sup_{t \in (0,T)} \left| \int_{\Omega_\varepsilon} \mathbf{V}_{\varepsilon,\delta} \cdot \nabla_x \big(\Delta_N^{-1} \operatorname{div}_x \varphi - P_M [\Delta_N^{-1} \operatorname{div}_x \varphi] \big) \, \mathrm{d}x \right| < \varepsilon$$
$$\text{for all } M > M(\varepsilon), \quad \delta > 0. \tag{101}$$

Combining the estimates (95–101) we obtain the desired conclusion (94). Consequently, we may assume that all quantities appearing in the acoustic equation are smooth, with the data compactly supported in $(0, T) \times \Omega_\varepsilon$. Accordingly, the acoustic equation reads:

$$\varepsilon \partial_t S_\varepsilon + \omega \operatorname{div}_x \mathbf{V}_\varepsilon = \varepsilon \operatorname{div}_x \big(\mathbf{H}^1_\varepsilon + \mathbf{H}^2_\varepsilon \big), \tag{102}$$

$$\varepsilon \partial_t \mathbf{V}_\varepsilon + \nabla_x S_\varepsilon = \varepsilon \operatorname{div}_x \big(\mathbb{G}^1_\varepsilon + \mathbb{G}^2_\varepsilon \big) + \varepsilon \mathbf{G}^3_\varepsilon, \tag{103}$$

supplemented with the boundary conditions

$$\mathbf{V}_\varepsilon \cdot \mathbf{n}|_{\partial \Omega_\varepsilon} = 0, \tag{104}$$

and the initial conditions

$$S_\varepsilon(0, \cdot) = S^1_{0,\varepsilon} + S^2_{0,\varepsilon} + S^3_{0,\varepsilon}, \qquad \mathbf{V}_\varepsilon(0, \cdot) = \mathbf{V}^1_{0,\varepsilon} + \mathbf{V}^2_\varepsilon, \tag{105}$$

where

$$\|S^1_{0,\varepsilon}\|_{L^1(\Omega)} + \|S^2_{0,\varepsilon}\|_{L^2(\Omega)} + \|S^3_{0,\varepsilon}\|_{D^{1,2}(R^3)} \le c, \tag{106}$$

$$\|\mathbf{V}^1_{0,\varepsilon}\|_{L^1(\Omega; R^3)} + \|\mathbf{V}^2_{0,\varepsilon}\|_{L^2(\Omega; R^3)} \le c, \tag{107}$$

and

$$\|\mathbf{H}^1_\varepsilon\|_{L^2(0,T; L^1(\Omega; R^3))} + \|\mathbf{H}^2_\varepsilon\|_{L^2(0,T; L^2(\Omega; R^3))} \le c, \tag{108}$$

$$\|\mathbb{G}^1_\varepsilon\|_{L^2(0,T; L^1(\Omega; R^{3\times 3}))} + \|\mathbb{G}^2_\varepsilon\|_{L^2(0,T; L^2(\Omega; R^{3\times 3}))} \le c, \tag{109}$$

$$\|\mathbf{G}^3_\varepsilon\|_{L^\infty(0,T; L^{5/3}(\Omega; R^3))} \le c. \tag{110}$$

Finite Speed of Propagation

System (102), (103) admits a finite speed of propagation proportional to $\sqrt{\omega}/\varepsilon$, specifically, if the initial data for two solutions coincide on the set

$$B_{T\sqrt{\omega}/\varepsilon} = \{x \in \Omega \mid |x| < R + T\sqrt{\omega}/\varepsilon\} \subset \Omega_\varepsilon,$$

and the forcing terms are the same on the space-time cylinder $(0,T) \times B_{T\sqrt{\omega}/\varepsilon}$, then the two solutions are the same on the cone

$$\left\{(t,x) \mid t \in (0,T), x \in B_{T\sqrt{\omega}/\varepsilon}, \operatorname{dist}[x, \partial B_{T\sqrt{\omega}/\varepsilon}] > t\sqrt{\omega}/\varepsilon\right\}.$$

Since we are interested only in the local behavior of solutions, specifically we want to show

$$\left\{t \mapsto \int_\Omega \mathbf{V}_\varepsilon(t,\cdot) \cdot \varphi \,\mathrm{d}x\right\} \quad \text{is precompact in } L^2(0,T) \text{ for any } \varphi \in C_c^\infty(\Omega; R^3), \tag{111}$$

we may assume that the acoustic system (102), (103) is satisfied on the whole set $(0,T) \times \Omega$ and that its solutions have compact support in $[0,T] \times \overline{\Omega}$.

Compactness of the Solenoidal Component

A short inspection of (103) implies that the family

$$\left\{t \mapsto \int_\Omega \mathbf{V}_\varepsilon \cdot \mathbf{H}[\varphi] \,\mathrm{d}x\right\} \quad \text{is precompact in } C[0,T]$$

for any $\varphi \in C_c^\infty(\Omega; R^3)$. Consequently, writing the field $\mathbf{V}_\varepsilon$ in the form of its Helmholtz decomposition in Ω:

$$\mathbf{V}_\varepsilon = \mathbf{H}[\mathbf{V}_\varepsilon] + \nabla_x \Psi_\varepsilon,$$

we can see that (87) follows as soon as we show

$$\left\{t \mapsto \int_\Omega \nabla_x \Psi_\varepsilon \cdot \varphi \,\mathrm{d}x\right\} \to 0 \quad \text{in} \quad L^2(0,T) \tag{112}$$

for any $\varphi \in C_c^\infty(\Omega; R^3)$, where Ψ_ε is the acoustic potential.

2.6 Acoustic Equation—Abstract Formulation

In order to show (112), we introduce an abstract formulation of the acoustic equation in terms of the *Neumann Laplacean* Δ_N,

$$\Delta_N v = \Delta v \quad \text{in } \Omega, \qquad \nabla_x v \cdot \mathbf{n}|_{\partial\Omega} = 0, \quad v \in C_c^\infty(\overline{\Omega}).$$

It is standard that Δ_N can be extended as a self-adjoint operator on the Hilbert space $L^2(\Omega)$. As a consequence of Rellich's theorem, the point spectrum of Δ_N is empty. Moreover, the spectrum of $-\Delta_N$ is absolutely continuous and coincides with $[0, \infty)$, see Leis [18].

Since all quantities in the acoustic equation (102), (103) are smooth, $\mathbf{V}_\varepsilon \cdot \mathbf{n}|_{\partial\Omega} = 0$, and the data $\mathbb{G}^i_\varepsilon$, $i = 1, 2$, $\mathbf{G}^3_\varepsilon$ are compactly supported, we deduce that $\nabla_x S_\varepsilon \cdot \mathbf{n}|_{\partial\Omega} = 0$. In particular, system (102), (103) converts to a *wave equation*:

$$\varepsilon \partial_t S_\varepsilon + \omega \Delta_N \Psi_\varepsilon = \varepsilon \operatorname{div}_x \left(\mathbf{H}^1_\varepsilon + \mathbf{H}^2_\varepsilon \right), \tag{113}$$

$$\varepsilon \partial_t \Psi_\varepsilon + S_\varepsilon = \varepsilon \Delta_N^{-1} \operatorname{div}_x \operatorname{div}_x \left(\mathbb{G}^1_\varepsilon + \mathbb{G}^2_\varepsilon \right) + \varepsilon \Delta_N^{-1} \operatorname{div}_x \mathbf{G}^3_\varepsilon, \tag{114}$$

supplemented with the homogeneous Neumann boundary conditions

$$\nabla_x \Psi_\varepsilon \cdot \mathbf{n}|_{\partial\Omega} = 0, \tag{115}$$

and the initial conditions

$$S_\varepsilon(0, \cdot) = S_{0,\varepsilon}, \qquad \Psi_\varepsilon(0, \cdot) = \Delta_N^{-1} \operatorname{div}_x \mathbf{V}_{0,\varepsilon}, \tag{116}$$

where $\nabla_x \Psi_\varepsilon = \mathbf{H}^\perp[\mathbf{V}_\varepsilon]$ is the gradient component of the Helmholtz decomposition of $\mathbf{V}_\varepsilon$.

Our goal is to rewrite system (113), (114) solely in terms of the operator Δ_N and functions ranging in the Hilbert space $L^2(\Omega)$. To this end, observe first that the expression $\operatorname{div}_x \operatorname{div}_x (\mathbb{G}^1_\varepsilon + \mathbb{G}^2_\varepsilon)(t, \cdot)$ may be viewed as a continuous linear form on $\mathscr{D}((-\Delta_N)^2) \cap \mathscr{D}((-\Delta_N)^{1/2})$ for any fixed t. Indeed it is enough to show that if

$$h \in \mathscr{D}((-\Delta_N)^2) \cap \mathscr{D}((-\Delta_N)^{1/2}),$$

then h possesses second derivatives bounded and continuous in $\overline{\Omega}$, and, in addition,

$$\nabla_x h \in L^2(\Omega; R^3), \qquad \nabla_x^2 h \in L^2(\Omega, R^{3\times 3}).$$

Since $\mathscr{D}((-\Delta_N)^{1/2}) = D^{1,2}(\Omega)$, we immediately get $\nabla_x h \in L^2(\Omega; R^3)$, $h \in L^6(\Omega)$. Next, taking $\psi \in C^\infty(\Omega)$, $\operatorname{supp}[\psi] \subset \Omega$, $\psi \equiv 1$ outside some ball, we get

$$\Delta(\psi h) = \psi \Delta h + 2\nabla_x \psi \cdot \nabla_x h + \Delta \psi h \quad \text{in } R^3,$$

where the right-hand side is bounded in $L^2(R^3)$. We conclude, by means of the well-known regularity properties of Δ on R^3, that $\nabla_x^2 h \in L^2(\Omega; R^{3\times 3})$, in particular, h is Hölder continuous and bounded in Ω. Finally, since $\Delta h \in L^2(\Omega)$, and $\Delta^2[h] \in L^2(\Omega)$, we have Δh Hölder continuous, and the standard elliptic theory provides the desired conclusion.

Estimating the remaining terms in a similar fashion, we arrive at the following system:

$$\varepsilon \partial_t S_\varepsilon + \omega \Delta_N \Psi_\varepsilon = \varepsilon \left((-\Delta_N)^2 [h^1_\varepsilon] + h^2_\varepsilon \right), \tag{117}$$

$$\varepsilon \partial_t \Psi_\varepsilon + S_\varepsilon = \varepsilon\big((-\Delta_N)[g_\varepsilon^1 + g_\varepsilon^3] + (-\Delta_N)^{-1/2}[g_\varepsilon^2 + g_\varepsilon^4]\big), \tag{118}$$

supplemented with the initial data

$$S_\varepsilon(0) = (-\Delta_N)^2[s_{0,\varepsilon}^1] + (-\Delta_N)^{-1/2}[s_{0,\varepsilon}^2], \tag{119}$$

$$\Psi_\varepsilon(0) = \Delta_N[v_{0,\varepsilon}^1] + \Delta_N^{-1}[v_{0,\varepsilon}^2], \tag{120}$$

with

$$\{h_\varepsilon^i\}_{\varepsilon>0}, \quad i = 1, 2, \qquad \{g_\varepsilon^j\}_{\varepsilon>0}, \quad j = 1, \ldots, 4, \text{ bounded in } L^2(0, T; L^2(\Omega)), \tag{121}$$

$$\{s_{0,\varepsilon}^i\}_{\varepsilon>0}, \quad i = 1, 2, \qquad \{v_{0,\varepsilon}^j\}_{\varepsilon>0}, \quad j = 1, 2, \text{ bounded in } L^2(\Omega). \tag{122}$$

Variation-of-Constants Formula

In accordance with (117)–(122), the acoustic potential Ψ_ε is determined through *variation-of-constants formula*, specifically,

$$\begin{aligned}
\Psi_\varepsilon(t) = {} & \frac{1}{2}\exp\left(\mathrm{i}\,\frac{t}{\varepsilon}\sqrt{-\omega\Delta_N}\right)\left[(-\Delta_N)[v_{0,\varepsilon}^1 + \mathrm{i}\,s_{0,\varepsilon}^1] + \frac{1}{(-\Delta_N)}[v_{0,\varepsilon}^2 + \mathrm{i}\,s_{0,\varepsilon}^2]\right]\\
& + \frac{1}{2}\exp\left(-\mathrm{i}\,\frac{t}{\varepsilon}\sqrt{-\omega\Delta_N}\right)\left[(-\Delta_N)[v_{0,\varepsilon}^1 - \mathrm{i}\,s_{0,\varepsilon}^1] + \frac{1}{(-\Delta_N)}[v_{0,\varepsilon}^2 - \mathrm{i}\,s_{0,\varepsilon}^2]\right]\\
& + \frac{1}{2}\int_0^t \exp\left(\mathrm{i}\,\frac{t-s}{\varepsilon}\sqrt{-\omega\Delta_N}\right)\left[(-\Delta_N)[g_\varepsilon^1 + g_\varepsilon^3] + \frac{1}{\sqrt{(-\Delta_N)}}[g_\varepsilon^2 + g_\varepsilon^4]\right.\\
& \left. + \mathrm{i}(-\Delta_N)^{3/2}[h_\varepsilon^1] + \mathrm{i}\,\frac{1}{\sqrt{-\Delta_N}}[h_\varepsilon^2]\right]\mathrm{d}s\\
& + \frac{1}{2}\int_0^t \exp\left(-\mathrm{i}\,\frac{t-s}{\varepsilon}\sqrt{-\omega\Delta_N}\right)\left[(-\Delta_N)[g_\varepsilon^1 + g_\varepsilon^3]\right.\\
& + \frac{1}{\sqrt{(-\Delta_N)}}[g_\varepsilon^2 + g_\varepsilon^4]\\
& \left. - \mathrm{i}(-\Delta_N)^{3/2}[h_\varepsilon^1] - \mathrm{i}\,\frac{1}{\sqrt{-\Delta_N}}[h_\varepsilon^2]\right]\mathrm{d}s.
\end{aligned} \tag{123}$$

Strong Convergence of Velocities

We are ready to show (112), specifically,

$$\left\{t \mapsto \int_\Omega \Psi_\varepsilon(t,\cdot)\,\mathrm{div}_x\,\varphi\,\mathrm{d}x\right\} \to 0 \quad \text{in } L^2(0,T) \text{ as } \varepsilon \to 0 \tag{124}$$

for any fixed $\varphi \in C_c^\infty(\Omega; R^3)$.

First of all, observe that it is enough to show

$$\left\{ t \mapsto \int_{\Omega} \chi H(-\Delta_N)[\Psi_\varepsilon(t,\cdot)]\,\mathrm{d}x \right\} \to 0 \quad \text{in } L^2(0,T) \tag{125}$$

for any fixed $\chi \in C_c^\infty(\Omega)$, $H \in C_c^\infty(0,\infty)$. Indeed, taking $\chi \in C_c^\infty(\Omega)$ such that $\chi|_{\mathrm{supp}[\varphi]} = 1$, we have

$$\begin{aligned}\int_{\Omega} \Psi_\varepsilon \,\mathrm{div}_x \varphi \,\mathrm{d}x &= \int_{\Omega} \chi \Psi_\varepsilon \,\mathrm{div}_x \varphi \,\mathrm{d}x \\ &= \int_{\Omega} \chi \Big(\mathrm{Id} - H(-\Delta_N) \Big)[\Psi_\varepsilon]\,\mathrm{div}_x \varphi \,\mathrm{d}x \\ &\quad + \int_{\Omega} \chi \,\mathrm{div}_x \varphi H(-\Delta_N)[\Psi_\varepsilon]\,\mathrm{d}x,\end{aligned}$$

where, as stated in (125),

$$\left\{ t \mapsto \int_{\Omega} \chi \,\mathrm{div}_x \varphi H(-\Delta_N)[\Psi_\varepsilon(t,\cdot)]\,\mathrm{d}x \right\} \to 0 \quad \text{in } L^2(0,T) \text{ as } \varepsilon \to 0.$$

On the other hand,

$$\begin{aligned}\int_{\Omega} \chi \Big(\mathrm{Id} - H(-\Delta_N) \Big)[\Psi_\varepsilon]\,\mathrm{div}_x \varphi \,\mathrm{d}x &= \int_{\Omega} \Big(\mathrm{Id} - H(-\Delta_N) \Big)[\Psi_\varepsilon]\,\mathrm{div}_x \varphi \,\mathrm{d}x \\ &= \int_{\Omega} \Big(\mathrm{Id} - H(-\Delta_N) \Big)[\mathrm{div}_x \varphi]\Psi_\varepsilon \,\mathrm{d}x. \end{aligned} \tag{126}$$

Taking a family of functions $H(\lambda) \nearrow 1$, in particular,

$$(H(-\Delta_N) - \mathrm{Id})[h] \to 0 \quad \text{for any fixed } h \in L^2(\Omega),$$

we observe that the integral (126) is small, uniformly with respect to $t \in (0,T)$ for a suitable choice of H, as soon as we can show that

$$(-\Delta_N)^{3/2}[\mathrm{div}_x \varphi], \qquad \frac{1}{(-\Delta_N)}[\mathrm{div}_x \varphi] \in L^2(\Omega) \tag{127}$$

since Ψ_ε is given by (123). To see (127), it is enough to observe that

$$\Delta_N[h] = \mathrm{div}_x \varphi \quad \text{implies} \quad \nabla_x h \in L^q(\Omega; R^3) \quad \text{for any } q > 1;$$

whence, by virtue of Sobolev's theorem $h \in L^2(\Omega)$.

In view of the previous discussion, the proof of strong (a.a. pointwise) convergence of velocities reduces to showing (125). This will be done in the following section.

Spectral Measures

Our goal in this section is to show (125). Since Ψ_ε is given by (123), it is sufficient to check that

$$\left(\int_0^T \left|\left\langle \exp\left(\mathrm{i}\sqrt{-\Delta_N}\frac{t}{\varepsilon}\right)H(-\Delta_N)[h],\varphi\right\rangle\right|^2 \mathrm{d}t\right)^{1/2} \leq \omega(\varepsilon,H,\varphi)\|h\|_{L^2(\Omega)} \quad (128)$$

for any $h \in L^2(\Omega)$, with

$$\omega(\varepsilon,H,\varphi)\to 0 \quad \text{as } \varepsilon\to 0 \text{ for any fixed } \varphi, H,$$

where $\langle,\rangle$ denotes the standard (complex) scalar product in $L^2(\Omega)$. Uniformity with respect to h is needed when handling the time integrals in (123).

The integrand in (128) may be written by *spectral theorem* (see Reed and Simon [27, Chap. VIII]) as follows

$$\left\langle \exp\left(\mathrm{i}\sqrt{-\Delta_N}\frac{t}{\varepsilon}H(-\Delta_N)\right)[h],\varphi\right\rangle = \int_0^\infty \exp\left(\mathrm{i}\sqrt{\lambda}\frac{t}{\varepsilon}\right)H(\lambda)\tilde{h}(\lambda)\,\mathrm{d}\mu_\varphi(\lambda), \quad (129)$$

where μ_φ is the spectral measure associated to the function φ, and

$$\tilde{h}\in L^2(\Omega;\mathrm{d}\mu_\varphi), \qquad \|\tilde{h}\|_{L^2_{\mu_\varphi}(\Omega)} \leq \|h\|_{L^2(\Omega)}.$$

Following Last [17], we obtain

$$\begin{aligned}
&\int_0^T \left|\left\langle \exp\left(\mathrm{i}\sqrt{-\Delta_N}\frac{t}{\varepsilon}\right)H(-\Delta_N)[h],\varphi\right\rangle\right|^2 \mathrm{d}t \\
&= \int_0^T\int_0^\infty\int_0^\infty \exp\left(\mathrm{i}(\sqrt{x}-\sqrt{y})\frac{t}{\varepsilon}\right)H(x)\tilde{h}(x)\;H(y)\bar{\tilde{h}}(y)\,\mathrm{d}\mu_\varphi(x)\,\mathrm{d}\mu_\varphi(y)\;\mathrm{d}t \\
&\leq c(H)\int_0^\infty\int_0^\infty\left(\int_{-\infty}^\infty \exp(-(t/T)^2)\exp\left(\mathrm{i}(\sqrt{x}-\sqrt{y})\frac{t}{\varepsilon}\right)\,\mathrm{d}t\right) \\
&\quad\times H(x)\tilde{h}(x)H(y)\bar{\tilde{h}}(y)\,\mathrm{d}\mu_\varphi(x)\,\mathrm{d}\mu_\varphi(y) \\
&\leq c(T,H)\sqrt{\pi}\int_0^\infty\int_0^\infty |\tilde{h}(x)||\tilde{h}(y)|\exp\left(-\frac{T^2|\sqrt{x}-\sqrt{y}|^2}{4\varepsilon^2}\right)\mathrm{d}\mu_\varphi(x)\,\mathrm{d}\mu_\varphi(y).
\end{aligned} \quad (130)$$

Consequently, by virtue of the Cauchy–Schwartz inequality,

$$\int_0^T \left|\left\langle \exp\left(\mathrm{i}\sqrt{-\Delta_N}\frac{t}{\varepsilon}\right)H(-\Delta)[h],\varphi\right\rangle\right|^2 \mathrm{d}t \leq c(H)\omega^2(\varepsilon,\varphi)\|h\|^2_{L^2(\Omega)}, \quad (131)$$

where

$$\omega(\varepsilon,\varphi) = \sqrt{2}\left(\int_0^\infty\int_0^\infty \exp\left(-\frac{T^2|\sqrt{x}-\sqrt{y}|^2}{2\varepsilon^2}\right)\mathrm{d}\mu_\varphi(x)\,\mathrm{d}\mu_\varphi(y)\right)^{1/4}.$$

Now, it is easy to check that $\omega(\varepsilon, H, \varphi) \to 0$ as $\varepsilon \to 0$ provided the spectral measure μ_φ does not charge points in $[0, \infty)$, in other words, as long as the point spectrum of the operator Δ_N is empty. As a matter of fact, the rate of convergence is independent of the specific choice of H. Thus we have proved (125) yielding the desired conclusion

$$\mathbf{u}_\varepsilon \to \mathbf{U} \quad \text{in } L^2((0,T) \times K; R^3) \text{ for any compact set } K \subset \Omega. \tag{132}$$

2.7 Convergence to the Limit System—Part II

Since we have shown strong pointwise (a.a.) convergence of the family of the velocity fields $\{\mathbf{u}_\varepsilon\}_{\varepsilon>0}$ it is a routine matter to let $\varepsilon \to 0$ in the weak formulation of the primitive system to deduce that

$$\frac{\rho_\varepsilon - \overline{\rho}}{\varepsilon} \to r \quad \text{weakly-}(*) \text{ in } L^\infty(0,T; L^{5/3}(K)) \text{ for any compact } K \subset \Omega,$$

$$\frac{\vartheta_\varepsilon - \overline{\vartheta}}{\varepsilon} \to \Theta \quad \text{weakly in } L^2(0,T; W^{1,2}(\Omega)),$$

$$\mathbf{u}_\varepsilon \to \mathbf{U} \quad \text{weakly in } L^2(0,T; W^{1,2}(\Omega; R^3)),$$

and

$$\mathbf{u}_\varepsilon \to \mathbf{U} \quad \text{in } L^2((0,T) \times K) \text{ for any compact } K \subset \Omega,$$

where $\mathbf{U}$, Θ, r is a weak solution of the Oberbeck–Boussinesq approximation (1)–(4), together with the boundary conditions

$$\mathbf{U} \cdot \mathbf{n}|_{\partial\Omega} = 0, \qquad [\mathbb{S}(\nabla_x \mathbf{U})\mathbf{n}] \times \mathbf{n}|_{\partial\Omega} = 0, \qquad \nabla_x \Theta \cdot \mathbf{n}|_{\partial\Omega} = 0.$$

More specifically, we have

$$\operatorname{div}_x \mathbf{U} = 0 \quad \text{a.a. on } (0,T) \times \Omega,$$

$$\int_0^T \int_\Omega (\overline{\rho}(\mathbf{U} \cdot \partial_t \varphi + (\mathbf{U} \otimes \mathbf{U}) : \nabla_x \varphi))\, \mathrm{d}x\, \mathrm{d}t,$$
$$= -\int_\Omega \overline{\rho}\mathbf{U}_0 \cdot \varphi\, \mathrm{d}x + \int_0^T \int_\Omega \mathbb{S} : \nabla_x \varphi - r \nabla_x G\, \mathrm{d}x\, \mathrm{d}t \tag{133}$$

for any test function $\varphi \in C_c^\infty([0,T) \times \overline{\Omega}; R^3)$, $\operatorname{div}_x \varphi = 0$, $\varphi \cdot \mathbf{n}|_{\partial\Omega} = 0$, where we have set

$$\mathbb{S} = \mu(\overline{\vartheta})(\nabla_x \mathbf{U} + \nabla_x^t \mathbf{U}).$$

Furthermore,

$$\overline{\rho} c_p(\overline{\rho}, \overline{\vartheta})\big[\partial_t \Theta + \operatorname{div}_x(\Theta \mathbf{U})\big] - \kappa \Delta \Theta - \overline{\rho}\overline{\vartheta}\alpha(\overline{\rho}, \overline{\vartheta}) \operatorname{div}_x(G\mathbf{U}) = 0$$
$$\text{a.a. in } (0,T) \times \Omega, \tag{134}$$
$$\nabla_x \Theta \cdot \mathbf{n}|_{\partial\Omega} = 0, \qquad \Theta(0,\cdot) = \Theta_0,$$

and

$$r + \overline{\rho}\alpha(\overline{\rho}, \overline{\vartheta})\Theta = 0 \quad \text{a.a. in } (0,T) \times \Omega.$$

We remark that the uniform bounds established above yield

$$\Theta \in L^\infty\big(0,T; L^2(\Omega)\big),$$

while

$$\mathbf{U} \in L^\infty\big(0,T; L^2(\Omega; R^3)\big),$$

in particular, the standard maximal regularity theory of the heat equation justifies validity of (134) a.a. in $(0,T) \times \Omega$.

It is interesting to note that the initial conditions for the velocity are determined through

$$\mathbf{u}_{0,\varepsilon} \to \mathbf{U}_0 \quad \text{weakly in } L^2(\Omega; R^3),$$

while the initial value Θ_0 reads

$$\Theta_0 = \frac{\overline{\vartheta}}{c_p(\overline{\rho}, \overline{\vartheta})}\left(\frac{\partial s(\overline{\rho}, \overline{\vartheta})}{\partial \rho}\rho_0^{(1)} + \frac{\partial s(\overline{\rho}, \overline{\vartheta})}{\partial \vartheta}\vartheta_0^{(1)} + \alpha(\overline{\rho}, \overline{\vartheta})G\right), \tag{135}$$

where

$$\rho_{0,\varepsilon}^{(1)} \to \rho_0^{(1)}, \quad \vartheta_{0,\varepsilon}^{(1)} \to \vartheta_0^{(1)} \quad \text{weakly in } L^2(\Omega).$$

Moreover, we can check if $\rho_0^{(1)}, \vartheta_0^{(1)}$ satisfy a compatibility condition

$$\frac{\partial p(\overline{\rho}, \overline{\vartheta})}{\partial \rho}\rho_0^{(1)} + \frac{\partial p(\overline{\rho}, \overline{\vartheta})}{\partial \vartheta}\vartheta_0^{(1)} = \overline{\rho}G,$$

where the expression on the left-hand side is nothing other than the linearization of the pressure at the constant state $(\overline{\rho}, \overline{\vartheta})$ applied to the vector $[\rho_0^{(1)}, \vartheta_0^{(1)}]$, relation (135) reduces to

$$\Theta_0 = \vartheta_0^{(1)}.$$

The reader may consult [14, Chap. 5, Sect. 5.5] for other aspects of the "data adjustment" problem related to incompressible limits.

Summarizing the arguments of this section we have proved the following result:

Theorem 2.2 *Under the hypotheses of Theorem 2.1, let* $\{\rho_\varepsilon, \vartheta_\varepsilon, \mathbf{u}_\varepsilon\}_{\varepsilon>0}$ *be a family of weak solutions to the Navier–Stokes–Fourier system on the set* $(0,T)\times\Omega_\varepsilon$*, where* Ω_ε *are given by* (34)*, and the initial* $\{\rho_{0,\varepsilon}, \vartheta_{0,\varepsilon}, \mathbf{u}_{0,\varepsilon}\}_{\varepsilon>0}$ *data satisfy* (40–43)*, with*

$$\rho_{0,\varepsilon}^{(1)} \to \rho_0^{(1)}, \qquad \vartheta_{0,\varepsilon}^{(1)} \to \vartheta_0^{(1)} \quad \text{weakly in } L^2(\Omega),$$

$$\mathbf{u}_{0,\varepsilon} \to \mathbf{U}_0 \quad \text{weakly in } L^2(\Omega; R^3).$$

Then, extracting a suitable subsequence, yields

$$\rho_\varepsilon \to \overline{\rho} \quad \text{in } L^\infty\big(0,T; L^{5/3}(K)\big),$$

$$\frac{\vartheta_\varepsilon - \overline{\vartheta}}{\varepsilon} \to \Theta \quad \text{weakly in } L^2\big(0,T; W^{1,2}(K)\big),$$

and

$$\mathbf{u}_\varepsilon \to \mathbf{U} \quad \text{weakly in } L^2\big(0,T; W^{1,2}(K;R^3)\big) \quad \text{and,}$$
$$\text{strongly in } L^2\big((0,T)\times K; R^3\big)$$

for any compact $K \subset \Omega$*, where* $\mathbf{U}$, Θ *is a weak solution of the Oberbeck–Boussinesq approximation in* $(0,T)\times\Omega$ *in the sense specified in* (133), (134)*, and the initial data* (135) *and*

$$\mathbf{U}(0,\cdot) = \mathbf{H}[\mathbf{U}_0].$$

Note that dispersive (Strichartz' estimates) for the wave equation considered in the whole space R^3 were used by Desjardins and Grenier [8] in order to eliminate the acoustic waves in the low Mach number limit for the compressible Navier–Stokes system. Similar technique was used by Alazard [1] and Isozaki [15] in the context of Euler equations. *Weak* convergence of the convective term could be also established by a "local" method developed by Lions and Masmoudi [21] (see also a nice survey by Masmoudi [23]).

3 Oberbeck–Boussinesq Approximation

In the remaining part of the paper, we examine the Oberbeck–Boussinesq approximation written in the form introduced in (7–9), specifically,

$$\mathrm{div}_x\, \mathbf{U} = 0, \tag{136}$$

$$\partial_t \mathbf{U} + \mathrm{div}_x \mathbf{U} \otimes \mathbf{U} + \nabla_x P = \Delta \mathbf{U} - \theta \nabla_x G, \tag{137}$$

$$\partial_t \theta + \mathrm{div}_x(\theta \mathbf{U}) - \Delta\theta = 0, \tag{138}$$

where, for the sake of simplicity, all physical constants have been set to one. As we have seen in Sect. 2, the system (136)–(137), modulo an obvious change of variables specified in Sect. 1, can be identified as a singular limit of the full Navier–Stokes–Fourier system, where the Mach and Froude numbers tend to zero. In contrast with Sect. 2, where the boundary $\partial\Omega$ was supposed to be *acoustically hard* (cf. (19)), we consider the more common no-slip boundary condition

$$\mathbf{U}|_{\partial\Omega} = 0, \tag{139}$$

supplemented with a similar homogeneous Dirichlet boundary condition for the temperature deviation

$$\theta|_{\partial\Omega} = 0. \tag{140}$$

Let us remark that (139), (140) could be justified by similar arguments as in Sect. 2, provided (19) was replaced by more general "penalized" boundary conditions in the spirit of [10].

In addition to (136)–(140), we suppose that the (weak) solutions satisfy the *energy inequality*

$$\|\mathbf{U}(\tau)\|^2_{L^2(\Omega;R^3)} + 2\int_s^\tau \|\nabla_x \mathbf{U}\|^2_{L^2(\Omega;R^{3\times3})}\,\mathrm{d}t$$
$$\leq \|\mathbf{U}(s)\|^2_{L^2(\Omega;R^3)} - \int_s^\tau \int_\Omega \theta\nabla_x G\cdot \mathbf{u}\,\mathrm{d}x\,\mathrm{d}t \tag{141}$$

for any $\tau > 0$ and a.a. $s \leq \tau$ including $s = 0$. If the velocity field $\mathbf{U}$ is smooth, formula (141) follows easily by multiplying (137) by $\mathbf{U}$ and integrating by parts.

Similarly, a formal manipulation of (138) yields

$$\int_\Omega H(\theta(\tau))\,\mathrm{d}x + \int_s^\tau \int_\Omega H''(\theta)|\nabla_x\theta|^2\,\mathrm{d}x\,\mathrm{d}t \leq \int_\Omega H(\theta(s))\,\mathrm{d}x \tag{142}$$

for any $\tau > 0$ and a.a. $s \leq \tau$ including $s = 0$ for any smooth convex H.

The interested reader may consult [14, Sect. 5.5.4, Chap. 5] for a rigorous derivation of the energy inequalities (141), (142) via a singular limit process.

3.1 Suitable Weak Solutions

We consider the initial data for system (136)–(138) in the form

$$\mathbf{U}(0,\cdot) = \mathbf{U}_0 \in L^2(\Omega), \qquad \theta(0,\cdot) = \theta_0 \in L^1\cap L^\infty(\Omega). \tag{143}$$

Motivated by the previous discussion, we shall say that $\mathbf{U}$, θ is a *suitable weak solution* to problem (136)–(140), supplemented with the initial data (143) if

$$\mathbf{U} \in C_{\text{weak}}\big([0,T]; L^2(\Omega; R^3)\big) \cap L^2\big(0,T; W_0^{1,2}(\Omega; R^3)\big);$$

$$\theta \in C_{\text{weak}}\big([0,T]; L^2(\Omega)\big) \cap L^\infty\big(0,T; L^p(\Omega)\big) \quad \text{for any } 1 \le p \le \infty,$$

$$\theta \in L^2\big(0,T; W_0^{1,2}(\Omega)\big);$$

$$\mathrm{div}_x\, \mathbf{U} = 0 \quad \text{a.a. in } (0,T) \times \Omega;$$

the integral identity

$$\int_0^T \int_\Omega \big(\mathbf{U} \cdot \partial_t \varphi + \mathbf{U} \otimes \mathbf{U} : \nabla_x \varphi\big)\, \mathrm{d}x\, \mathrm{d}t$$
$$= \int_0^T \int_\Omega \big(\nabla_x \mathbf{U} : \nabla_x \varphi + \theta \nabla_x G \cdot \varphi\big)\, \mathrm{d}x - \int_\Omega \mathbf{U}_0 \cdot \varphi(0,\cdot)\, \mathrm{d}x$$

holds for any $\varphi \in C_c^\infty([0,T) \times \Omega; R^3)$, $\mathrm{div}_x\, \varphi = 0$;

$$\partial_t \theta + \mathrm{div}_x(\theta \mathbf{U}) - \Delta\theta = 0 \quad \text{a.a. in } (0,T) \times \Omega;$$

the energy inequalities (141), (142) are satisfies for a.a. $\tau \in [0,T]$.

Remark 3.1 Given the anticipated regularity of $\mathbf{U}$, θ enforced by (141), (142), we may use the maximal regularity theory for the heat equation (138) in order to conclude that $\partial_t \theta, \Delta\theta \in L^q(0,T; L^q(\Omega))$ for a certain $q > 1$.

Given the *a priori* bounds induced by (141), (142), the *existence* of suitable weak solutions can be proved, besides a rather complicated undirect proof in the spirit of Sect. 2, by means of nowadays standard methods, see the monograph by Sohr [29]. In the last part of this study, we examine the asymptotic behavior of suitable weak solutions to the Oberbeck–Boussinesq approximation for $t \to \infty$.

4 Long-Time Behavior of Solutions to the Oberbeck–Boussinesq Approximation

We conclude the present study of the Oberbeck–Boussinesq system by investigating the asymptotic behavior of solutions for large times. In contrast with [4], we show that the physically relevant choice of the forcing term $\nabla_x G$ yields strong convergence to zero of the total energy associated to system (136)–(140). Our approach is based on the available results by Miyakawa and Sohr [24] for the forced Navier–Stokes system. More precisely, we derive suitable decay estimates for the temperature deviation θ resulting from the "entropy" inequality (142) and then use the fact that, in accordance with (5), (6),

$$\nabla_x G \in L^p \cap L^\infty(\Omega; R^3) \quad \text{for } p > 3/2, \tag{144}$$

in particular, the forcing term in the Navier–Stokes system decays to zero sufficiently fast for $|x| \to \infty$.

4.1 Decay Estimates for the Temperature Deviations

In this subsection we show that the solutions to the Oberbeck–Boussinesq system decay in L^p, $1 < p \leq \infty$ at the same rate as the solutions of the underlying linear counterpart, namely the solutions to the heat equations.

Theorem 4.1 *Let* $\mathbf{U}(0,\cdot) = \mathbf{U}_0 \in L^2(\Omega)$, $\theta(0,\cdot) = \theta_0 \in L^1 \cap L^\infty(\Omega)$. *Suppose* $\mathbf{U}, \theta$ *is a* suitable weak solution *to problem* (136)–(140), *with the initial data* $(\mathbf{U}_0, \theta_0)$, *then*

$$\|\theta(t,\cdot)\|_{L^p(\Omega)} \leq c(\|\theta_0\|_{L^1\cap L^\infty(\Omega)}) t^{-\frac{3}{2}(1-1/p)}, \quad 1 \leq p \leq \infty,\ t > 0, \tag{145}$$

where the constant c *is independent of* p.

Proof We note first that by appropriate choices of H, the estimate (142) yields

$$\|\theta(t,\cdot)\|_{L^p(\Omega)} \leq \|\theta_0\|_{L^p(\Omega)} \quad \text{for any } t \geq 0,\ 1 \leq p \leq \infty. \tag{146}$$

Following the well-known argument of Alikakos [2] (cf. also Cordoba, Cordoba [5]), we multiply (138) by $2j|\theta|^{2j-2}\theta$ and integrate the resulting expression over Ω, obtaining

$$\partial_t \int_\Omega |\theta^j|^2 \,\mathrm{d}x + \frac{2j(2j-1)}{j^2} \int_\Omega |\nabla_x \theta^j|^2 \,\mathrm{d}x \leq 0,$$

in particular choosing $j = 2^{k-1}$ yields

$$\partial_t \left(\frac{1}{2} \int_\Omega \left| |\theta|^{2^{k-1}} \right|^2 \mathrm{d}x \right) \leq -\frac{2^k - 1}{2^{k-2}} \int_\Omega \left| \nabla_x |\theta|^{2^{k-1}} \right|^2 \mathrm{d}x, \quad k = 1, 2, \ldots. \tag{147}$$

By means of the standard Gagliardo–Nirenberg inequality (see Proposition 1.1), interpolating L^2 between L^1 and $\dot{H}^1$ we have

$$\left\| |\theta|^{2^{k-1}} \right\|^2_{L^2(\Omega)} \leq c \left\| \nabla_x |\theta|^{2^{k-1}} \right\|^{6/5}_{L^2(\Omega)} \left\| |\theta|^{2^{k-1}} \right\|^{4/5}_{L^1(\Omega)}. \tag{148}$$

Let $m_s = 2^s$. We proceed by induction on s. For m_0 the conclusion of the Theorem follows by (146). For $s = k - 1$ we assume by induction that

$$\int_\Omega |\theta|^{2^{k-1}} \, dx \le b_{k-1} t^{-\frac{3}{2}(2^{k-1}-1)}. \tag{149}$$

Let $s = k$, define

$$\Phi_k = \int_\Omega \left| |\theta|^{2^{k-1}} \right|^2 dx.$$

Combining (147), (148) and (149) yields

$$\partial_t \Phi_k \le -\Phi_k^{5/3} c^{-5/3} b_{k-1}^{-4/3} t^{2^k-2},$$

therefore, integrating in time over $[0, t]$ and a simple reordering of the terms gives

$$\Phi_k(t) \le \left[\Phi_k(0) + \frac{2}{3} c^{-5/3} b_{k-1}^{-4/3} \frac{1}{2^k - 1} t^{2^k-1} \right]^{-3/2}. \tag{150}$$

By virtue of (146), we can take

$$b_0 = \|\theta_0\|_{L^1(\Omega)}, \tag{151}$$

and, consequently, formula (150) yields, by induction and interpolation,

$$\|\theta(t, \cdot)\|_{L^p(\Omega)} \le c(\|\theta_0\|_{L^1 \cap L^\infty(\Omega)}) t^{-\frac{3}{2}(1-1/p)} \quad \text{for any } 1 \le p < \infty, \; t > 0. \tag{152}$$

The constant in (152) may, in principle, depend on p, however, a close inspection of (150) reveals, similarly to Alikakos [2, Theorem 3.1] that

$$b_k \le c b_{k-1}^2, \quad \text{meaning,} \qquad b_k \le C^k M^{2^k} \quad \text{for certain } C, M > 0,$$

It follows from (151) that

$$\Phi_k^{1/2k} \le c(\|\theta_0\|_{L^1 \cap L^\infty(\Omega)}) t^{-\frac{3}{2}(1-1/2^k)}.$$

Taking the limit as k tends to infinity extends the decay rate to $p = \infty$, specifically,

$$\|\theta(t, \cdot)\|_{L^\infty(\Omega)} \le c(\|\theta_0\|_{L^1 \cap L^\infty(\Omega)}) t^{-\frac{3}{2}}, \quad t > 0.$$

This concludes the proof of the theorem. □

4.2 Decay Estimates for the Velocity

In view of the specific choice of the potential G (cf. (144)), and the uniform decay estimates of the temperature deviation θ established in (145), (152), the decay of the velocity field $\mathbf{U}$ follows from the results of Miyakawa and Sohr [24]. Indeed it is enough to check that (144), (152) imply that

$$\theta \nabla_x G \in L^1 \cap L^\infty\big(0,\infty; L^2(\Omega; R^3)\big),$$

therefore, by virtue of [24, Theorem 1],

$$\lim_{t\to\infty} \|\mathbf{U}(t)\|_{L^2(\Omega;R^3)} = 0. \tag{153}$$

Moreover, the velocity becomes ultimately more regular, specifically, there exists $T_0 > 0$ such that

$$\begin{aligned} &\mathbf{U} \in L^2(T_0, T_0 + T; W^{2,2}(\Omega; R^3)),\\ &\partial_t \mathbf{U} \in L^2(T_0, T_0 + T, L^2(\Omega; R^3)), \quad \text{for any } T > 0. \end{aligned} \tag{154}$$

Let us summarize the results obtained in this section:

Theorem 4.2 *Let $\Omega \subset R^3$ be an unbounded (exterior) domain with compact boundary of class $C^{2+\nu}$. Let $\mathbf{U}, \theta$ be a suitable weak solution to the Oberbeck–Boussinesq approximation in $(0,\infty) \times \Omega$ specified in Sect. 3.1, emanating from the initial data*

$$\mathbf{U}_0 \in L^2(\Omega; R^3), \qquad \theta_0 \in L^1 \cap L^\infty(\Omega).$$

Then

$$\begin{aligned} &\mathbf{U}(t,\cdot) \to 0 \quad \text{in } L^2(\Omega; R^3),\\ &\theta(t,\cdot) \to 0 \quad \text{in } L^p(\Omega) \text{ for any } 1 < p \le \infty \text{ as } t \to \infty. \end{aligned}$$

Remark 4.1 Since G is a harmonic (regular) function in Ω, the nowadays standard ultimate regularity results for the Navier–Stokes system (see e.g. the monograph by Sohr [29, Chap. V, Theorem 4.2.2]), together with a simple bootstrap argument applied to the heat equation (138), could be used to deduce that the solution $\mathbf{U}$, θ becomes regular if time is large enough. Similarly, decay in stronger Sobolev norms can be shown.

Acknowledgements The work of E. Feireisl was supported by Grant 201/09/0917 of GA ČR as a part of the general research programme of the Academy of Sciences of the Czech Republic, Institutional Research Plan AV0Z10190503. The work of M. Schonbek was partially supported by NSF Grant DMS-0900909.

References

1. Alazard, T.: Incompressible limit of the nonisentropic Euler equations with the solid wall boundary conditions. Adv. Differ. Equ. **10**(1), 19–44 (2005)
2. Alikakos, N.D.: L^p-bounds of solutions of reaction diffusion equations. Commun. Partial Differ. Equ. **4**, 827–868 (1979)
3. Bechtel, S.E., Rooney, F.J., Forest, M.G.: Connection between stability, convexity of internal energy, and the second law for compressible Newtonian fluids. J. Appl. Mech. **72**, 299–300 (2005)
4. Brandolese, L., Schonbek, M.: Large time decay and growth for solutions of a viscous Boussinesq system. Trans. Am. Math. Soc. (to appear)
5. Cordoba, A., Cordoba, D.: A maximum principle applied to quasi-geostrophic equations. Commun. Math. Phys. **249**, 511–528 (2004)
6. Crispo, F., Maremonti, P.: An interpolation inequality in exterior domains. Rend. Semin. Mat. Univ. Padova **112** (2004)
7. Danchin, R., Paicu, M.: Existence and uniqueness results for the Boussinesq system with data in Lorentz spaces. Physica D **237**, 1444–1460 (2008)
8. Desjardins, B., Grenier, E.: Low Mach number limit of viscous compressible flows in the whole space. R. Soc. Lond. Proc., Ser. A, Math. Phys. Eng. Sci. **455**(1986), 2271–2279 (1999)
9. DiPerna, R.J., Lions, P.-L.: Ordinary differential equations, transport theory and Sobolev spaces. Invent. Math. **98**, 511–547 (1989)
10. Donatelli, D., Feireisl, E., Novotný, A.: On incompressible limits for the Navier–Stokes system on unbounded domains under slip boundary conditions. Discrete Contin. Dyn. Syst., Ser. B **13**(4), 783–798 (2010)
11. Eliezer, S., Ghatak, A., Hora, H.: An Introduction to Equations of States, Theory and Applications. Cambridge University Press, Cambridge (1986)
12. Ericksen, J.L.: Introduction to the Thermodynamics of Solids. Applied Mathematical Sciences, vol. 131, revised edn. Springer, New York (1998)
13. Feireisl, E.: Incompressible limits and propagation of acoustic waves in large domains with boundaries. Commun. Math. Phys. **294**, 73–95 (2010)
14. Feireisl, E., Novotný, A.: Singular Limits in Thermodynamics of Viscous Fluids. Birkhäuser, Basel (2009)
15. Isozaki, H.: Singular limits for the compressible Euler equation in an exterior domain. J. Reine Angew. Math. **381**, 1–36 (1987)
16. Klein, R., Botta, N., Schneider, T., Munz, C.D., Roller, S., Meister, A., Hoffmann, L., Sonar, T.: Asymptotic adaptive methods for multi-scale problems in fluid mechanics. J. Eng. Math. **39**, 261–343 (2001)
17. Last, Y.: Quantum dynamics and decompositions of singular continuous spectra. J. Funct. Anal. **142**, 406–445 (1996)
18. Leis, R.: Initial-Boundary Value Problems in Mathematical Physics. Teubner, Stuttgart (1986)
19. Lighthill, J.: On sound generated aerodynamically I. General theory. Proc. R. Soc. Lond. A **211**, 564–587 (1952)
20. Lighthill, J.: On sound generated aerodynamically II. General theory. Proc. R. Soc. Lond. A **222**, 1–32 (1954)
21. Lions, P.-L., Masmoudi, N.: Une approche locale de la limite incompressible. C. R. Acad. Sci. Paris, Sér. I Math. **329**(5), 387–392 (1999)
22. Málek, J., Rajagopal, K.R.: Mathematical issues concerning the Navier–Stokes equations and some of its generalizations. In: Evolutionary Equations, vol. II. Handbook of Differential Equations, pp. 371–459. Elsevier/North-Holland, Amsterdam (2005)
23. Masmoudi, N.: Examples of singular limits in hydrodynamics. In: Dafermos, C., Feireisl, E. (eds.) Handbook of Differential Equations, vol. III. Elsevier, Amsterdam (2006)
24. Miakawa, T., Sohr, H.: On energy inequality, smoothness, and large time behavior in L^2 for weak solutions of the Navier–Stokes system in an exterior domain. Math. Z. **199**, 455–478 (1988)

25. Novotný, A., Růžička, M., Thaeter, G.: Rigorous derivation of the anelastic approximation to the Oberbeck–Boussinesq equations. Preprint IMATH 2010-12 Toulon
26. Rajagopal, K.R., Růžička, M., Shrinivasa, A.R.: On the Oberbeck–Boussinesq approximation. Math. Models Methods Appl. Sci. **6**, 1157–1167 (1996)
27. Reed, M., Simon, B.: Methods of Modern Mathematical Physics. I. Functional Analysis. Academic Press, New York (1978)
28. Reshetnyak, Yu.G.: Stability Theorems in Geometry and Analysis. Mathematics and Its Applications, vol. 304. Kluwer Academic, Dordrecht (1994). Translated from the 1982 Russian original by N.S. Dairbekov and V.N. Dyatlov, and revised by the author. Translation edited and with a foreword by S.S. Kutateladze
29. Sohr, H.: The Navier–Stokes Equations: An Elementary Functional Analytic Approach. Birkhäuser, Basel (2001)
30. Zeytounian, R.Kh.: Asymptotic Modeling of Atmospheric Flows. Springer, Berlin, (1990)
31. Zeytounian, R.Kh.: Joseph Boussinesq and his approximation: a contemporary view. C. R., Méc. **331**, 575–586 (2003)
32. Zeytounian, R.Kh.: Theory and Applications of Viscous Fluid Flows. Springer, Berlin, (2004)

Universal Profiles and Rigidity Theorems for the Energy Critical Wave Equation

Carlos Kenig

Abstract In this note we review recent joint works with F. Merle and with T. Duyckaerts and F. Merle on global existence, scattering and finite time blow-up for the focusing energy critical non-linear wave equation in three space dimensions.

In this note we consider the energy critical non-linear wave equation in $\mathbb{R}^{3+1}$,

$$\begin{cases} \partial_t^2 u - \Delta u = \pm u^5, \\ u|_{t=0} = u_0 \in \dot{H}^1(\mathbb{R}^3), \qquad (x,t) \in \mathbb{R}^3 \times \mathbb{R}. \\ \partial_t u|_{t=0} = u_1 \in L^2(\mathbb{R}^3), \end{cases} \tag{NLW}$$

Here the $-$ sign corresponds to the defocussing case, while the $+$ sign corresponds to the focusing case. The problem is "critical" in the space $\dot{H}^1 \times L^2$, because if u is a solution, so is $\frac{1}{\lambda^{1/2}} u(\frac{x}{\lambda}, \frac{t}{\lambda})$ and the norm of the corresponding initial data in $\dot{H}^1 \times L^2$ is independent of λ. There is also an energy $E_\pm(u_0, u_1) = \frac{1}{2}\int |\nabla u_0|^2 + \frac{1}{2}\int u_1^2 \pm \frac{1}{6}\int u_0^6$, which for solutions is constant in time. Here the $+$ sign corresponds to the defocussing case, while the $-$ sign corresponds to the focusing case.

The so-called "local theory of the Cauchy problem" (see [5] for instance, for an explanation of this terminology) has been well-understood since the 1980's, thanks to work of Pecher, Ginibre–Velo, Lindblad–Sogge and others (again see [5] for precise references). As a consequence, there exists $\delta > 0$, small, so that if $\|(u_0, u_1)\|_{\dot{H}^1 \times L^2} < \delta$, there exists a unique solution $u \in C(\mathbb{R}; \dot{H}^1 \times L^2) \cap L^8(dx\,dt)$ which depends continuously on (u_0, u_1) and which scatters, i.e., there exist $w_{0,\pm}, w_{1,\pm}$ with $\lim_{t\to\pm\infty} \|(u(t), \partial_t u(t)) - (w_\pm(t), \partial_t w_\pm(t))\|_{\dot{H}^1 \times L^2} = 0$, where $w_\pm$ solves the linear Cauchy problem

$$\begin{cases} \partial_t^2 w_\pm - \Delta w_\pm = 0, \\ w_\pm|_{t=0} = w_{0,\pm}, \\ \partial_t w_\pm|_{t=0} = w_{1,\pm}. \end{cases} \tag{LW}$$

C. Kenig (✉)
Department of Mathematics, The University of Chicago, 5734 S. University Ave., Chicago, IL 60637, USA
e-mail: cek@math.uchicago.edu

H. Holden, K.H. Karlsen (eds.), *Nonlinear Partial Differential Equations*, Abel Symposia 7, DOI 10.1007/978-3-642-25361-4_8,

Moreover, given any data (u_0, u_1) (without any size restriction) in $\dot{H}^1 \times L^2$, there exist $T_\pm = T_\pm(u_0, u_1) > 0$ such that there exists a unique $u \in C([-T_- + \varepsilon, T_+ - \varepsilon]; \dot{H}^1 \times L^2) \cap L^8_{[-T_- + \varepsilon, T_+ - \varepsilon]} L^8_x$ for each $\varepsilon > 0$, solving the equation and the interval $(-T_-(u_0, u_1), T_+(u_0, u_1))$ is maximal with this property. If, say, $T_+(u_0, u_1) < \infty$, then $\|u\|_{L^8[0,T_+]L^8_x} = +\infty$ and given $t_n \uparrow T_+(u_0, u_1)$, $(u(t_n), \partial_t u(t_n))$ has no convergent subsequence in $\dot{H}^1 \times L^2$. (See [6] for instance for detailed proofs.) In the defocussing case, works of Struwe (in the radial case) and of Grillakis, Shatah–Struwe, Kapitansky and Bahouri–Shatah (in the non-radial case) show that for any data $(u_0, u_1) \in \dot{H}^1 \times L^2$, the solution exists globally in time and scatters and higher regularity is preserved. (See for instance [5] for precise references.)

We now turn to the focusing case. Here, H. Levine (1973) showed (by an "obstruction argument") that if $(u_0, u_1) \in H^1 \times L^2$, $E_-(u_0, u_1) < 0$, then $T_\pm < \infty$. Also, recall that $W(x) = (1 + \frac{|x|^2}{3})^{-1/2}$ solves the elliptic equation $\Delta u + u^5 = 0$ in $\mathbb{R}^3$, $W \in \dot{H}^1(\mathbb{R}^3)$. It is also the unique (modulo obvious invariances) minimizer in the Sobolev embedding

$$\|u\|_{L^6} \leq C_3 \|\nabla u\|_{L^2},$$

where C_3 is the best constant, which is explicitly known (Talenti). Also, W is the unique (modulo obvious invariances) non-negative solution of the elliptic equation (Gidas–Ni–Nirenberg) and the only radial $\dot{H}^1$ solution (Pohozaev). (See [5], for instance, for the precise references.)

From the above two facts, simple calculations (see [6]) give that $\|\nabla W\|^2_{L^2} = \frac{1}{C_3^3}$ and $E_-((W, 0)) = \frac{1}{3C_3^3}$. Since W is a solution to the elliptic equation, it also solves the focusing (NLW) for all time, but it does not scatter. Thus, we may have global existence, but no scattering.

I will now recall the work of Kenig–Merle [6].

Theorem 1 [6] *If* $(u_0, u_1) \in \dot{H}^1 \times L^2$, $E_-(u_0, u_1) < E_-((W, 0))$, *then*:

(i) *If* $\|u_0\|_{\dot{H}^1} < \|W\|_{\dot{H}^1}$, *then solution exists for all times and scatters.*
(ii) *If* $\|u_0\|_{\dot{H}^1} > \|W\|_{\dot{H}^1}$, *then* $T_\pm(u_0, u_1) < \infty$.

Also, the case $E_-(u_0, u_1) < E_-((W, 0))$, $\|u_0\|_{\dot{H}^1} = \|W\|_{\dot{H}^1}$ *is impossible.*

The proof of this result follows from the concentration-compactness/rigidity theorem method developed in a series of works by Kenig–Merle. See [5] for a survey of the method and results that have been obtained using it.

The paper [6] in fact gave a strengthening of Theorem 1. (See also [7] where some of the details of the proof of the following theorem were fully developed.)

Theorem 2 [6] *If* $\limsup_{t \to T^+(u_0, u_1)} \|\nabla u(t)\|^2_{L^2} + \|\partial_t u(t)\|^2_{L^2} < \|\nabla W\|^2$, *then* $T^+(u_0, u_1) = +\infty$ *and* u *scatters as* $t \to \infty$.

Theorem 2 is a strengthening of Theorem 1(i) because the hypothesis $E_-((u_0, u_1)) < E_-((W, 0))$, $\|u_0\|_{\dot{H}^1} < \|W\|_{\dot{H}^1}$ imply the hypothesis of Theorem 2, as was shown in [6]. It turns out that Theorem 2 is sharp. In fact, Krieger, Schlag and Tataru [8] constructed, for each $\eta_0 > 0$ a radial solution with $T_+ < \infty$ and such that

$$\sup_{t\in[0,T_+)} \|\nabla u(t)\|_{L^2}^2 + \|\partial_t u(t)\|_{L^2}^2 \leq \|\nabla W\|_{L^2}^2 + \eta_0.$$

Moreover, the solutions they constructed have the following description:

$$u(t) = \frac{1}{\lambda^{\frac{1}{2}}(t)} W\left(\frac{x}{\lambda(t)}\right) + \varepsilon(t),$$

where

$$\lim_{t\to T^+} \int_{|x|<T_+-t} |\nabla\varepsilon(t)|^2 + \int_{|x|<T_+-t} |\partial_t\varepsilon(t)|^2 + \int_{|x|<T_+-t} \varepsilon^6(t)\, dt = 0, \qquad (1)$$

where $\lambda(t) = (T_+ - t)^{1+\nu}$ and $\nu > \frac{1}{2}$. (The restriction $\nu > \frac{1}{2}$ is technical and it is expected that the construction can be pushed to $\nu > 0$.)

Our next topic is a general study of the so-called "type II blow-up solutions", i.e., solutions for which $T_+ < \infty$ and $\sup_{t\in[0,T_+)} \|\nabla u(t)\|_{L^2}^2 + \|\partial_t u(t)\|_{L^2}^2 < \infty$. These results are joint work with Duyckaerts and Merle, [2] and [3].

Let $x_0 \in \mathbb{R}^3$. We will say that x_0 is regular if for each $\varepsilon > 0$, there exists $R > 0$ such that for all $t \in [0, T_+)$ we have

$$\int_{|x-x_0|<R} |\nabla u|^2 + (\partial_t u)^2 + \frac{|u|^2}{|x-x_0|^2} < \varepsilon.$$

If x_0 is not regular, we say that it is singular.

We let $S = \{\text{all singular points}\}$.

Theorem 3 [2] *Assume that u is a type II blow-up solution. Then there exists $N \in \mathbb{N}$ and N distinct points $x_1, \ldots, x_N \in \mathbb{R}^3$ so that $S = \{x_1, \ldots, x_N\}$. Moreover, there exists $(v_0, v_1) \in \dot{H}^1 \times L^2$ so that $(u(t), \partial_t u(t))\ t \rightharpoonup T_+\ (v_0, v_1)$ weakly in $\dot{H}^1 \times L^2$. Also, if $\varphi \in C_0^\infty(\mathbb{R}^3)$, $\varphi \equiv 1$ near each x_R, we have*

$$\lim_{t\to T_+} \|(1-\varphi)[u(t) - v_0]\|_{\dot{H}^1} + \|(1-\varphi)[\partial_t u(t) - v_1]\|_{L^2} = 0.$$

Moreover, for each $1 \leq k \leq N$, we have

$$\limsup_{t\to T_+} \int_{|x-x_k|<|t-T_+|} |\nabla u(x,t)|^2 + (\partial_t u(x,t))^2 \geq \int |\nabla W|^2,$$

and

$$\liminf_{t\to T_+} \int_{|x-x_k|<|t-T_+} |\nabla u(x,t)|^2 + (\partial_t u(x,t))^2 \geq \frac{2}{3} \int |\nabla W|^2.$$

If v is the solution of (NLW) with $(v(T_+), \partial_t v(T_+)) = (v_0, v_1)$, we call v the regular part of u at the blow-up time T_+ and $a = u - v$ the singular part of u. Note that the above result, together with the finite speed of propagation shows that

$$\operatorname{supp} a \subset \bigcup_{k=1}^{N} \{(x,t) : |x - x_k| \leq |t - T_+|\}.$$

From now on we assume, without loss of generality, that $T_+ = 1$ and that our solution u satisfies

$$\sup_{t_0<t<1} \|\nabla u(t)\|_{L^2}^2 + \|\partial_t u(t)\|_{L^2}^2 \leq \|\nabla W\|_{L^2}^2 + \eta_0, \tag{2}$$

where η_0 is small. Note that, as a consequence of the above theorem, $S = \{x_1\}$. We will assume, without loss of generality that $x_1 = 0$. Our first result is that, in the radial case, the Krieger–Schlag–Tataru construction has a "converse" and W is the universal blow-up profile.

Theorem 4 [2] *Let u be a radial type II blow-up solution as above, with $T_+ = 1$, η_0 small. Then there exists a smooth positive function $\lambda(t)$ on* $(0,1)$, *so that* $\lim_{t\uparrow 1} \frac{\lambda(t)}{(1-t)} = 0$ *and as $t \to 1$, we have*

$$(u(t), \partial_t v(t)) - (v_0, v_1) - \pm\left(\frac{W(\frac{x}{\lambda(t)})}{\lambda(t)^{\frac{1}{2}}}, 0\right) \underset{t\to 1}{\to} 0$$

in $\dot{H}^1 \times L^2$.

A crucial ingredient in the proof of the result is a characterization of compact radial solutions. In the proof the Theorem 1, using the concentration-compactness/rigidity theorem method, Kenig–Merle [6] had shown the following rigidity theorem

Theorem 5 [6] *Let u be a solution to* (NLW) *with $E_-((u_0, u_1)) < E_-((W, 0))$, $\|\nabla u_0\|_{L^2} < \|\nabla W\|_{L^2}$, so that there exist $\lambda(t) \in \mathbb{R}^+$, $x(t) \in \mathbb{R}^3$, $t \in (-T_-, T_+)$ so that*

$$K = \left\{\left(\frac{1}{\lambda(t)^{\frac{1}{2}}} u\left(\frac{x - x(t)}{\lambda(t)}, t\right), \frac{1}{\lambda(t)^{\frac{3}{2}}} \partial_t u\left(\frac{x - x(t)}{\lambda(t)}, t\right)\right), t \in (-T_-, T_+)\right\}$$

has compact closure in $\dot{H}^1 \times L^2$. Then, $(u_0, u_1) = (0, 0)$.

Now, when we restrict our attention to radial solutions we have a strengthening of this rigidity theorem, without any size limitation

Theorem 6 [2] *Let u be a non-zero radial solution of* (NLW)*. Assume that there exists $\lambda(t) \in \mathbb{R}^+$, $t \in (-T_-, T_+)$ so that*

$$K = \left\{\left(\frac{1}{\lambda(t)^{\frac{1}{2}}} u\left(\frac{x}{\lambda(t)}, t\right), \frac{1}{\lambda(t)^{\frac{3}{2}}} \partial_t u\left(\frac{x}{\lambda(t)}, t\right)\right), t \in (-T_-, T_+)\right\}$$

is so that $\overline{K}$ is compact in $\dot{H}^1 \times L^2$. Then there exists $\lambda_0 > 0$ so that $u(x,t) = \pm\frac{1}{\lambda_0^{1/2}} W(\frac{x}{\lambda_0})$.

This result is fundamental for the proof of Theorem 4. It depends heavily on the proof of Theorem 5 and on the characterization of the dynamics of solutions with the property $E_-((u_0,u_1)) = E_-((W,0))$, obtained by Duyckaerts–Merle in [4]. In the non-radial case Theorem 6 is not correct because of the existence of non-radial solutions in $\dot{H}^1(\mathbb{R}^3)$ of the elliptic equation $\Delta u + u^5 = 0$, which was shown by Ding [1].

We now turn to results in the non-radial case, still under hypothesis (2). We now need to pay attention to the Lorentz invariance of solutions to (NLW). In this connection, a fundamental fact is that for solutions of (NLW) the momentum $\int \nabla u(t)\partial_t u(t)$ is constant in time.

To explain our result, we recall the following family of solutions to (NLW) obtained as Lorentz transformations of W. Fix l, $|l| < 1$. Then, let

$$\begin{aligned} W_l(x,t) &= W\left(\frac{x_1 - tl}{\sqrt{1-l^2}}, x_2, x_3\right) \\ &= \left(1 + \frac{(x_1-tl)^2}{3\sqrt{1-l^2}} + \frac{x_2^2 + x_3^2}{3}\right)^{-\frac{1}{2}}, \end{aligned}$$

which are solutions to (NLW).

Elementary calculations show

$$\begin{aligned} \int |\nabla W_l(t)|^2 + \int (\partial_t W_l(t))^2 &= \frac{3-l^2}{3\sqrt{1-l^2}} \int |\nabla W|^2, \\ E_-((W_l(0), \partial_t W_l(0))) &= \frac{1}{\sqrt{1-l^2}} E_-((W,0)), \\ \int \nabla W_l(0) \cdot \partial_t W_l(0) &= -l E_-(W_l(0), \partial_t W_l(0))\mathbf{e}, \mathbf{e} = (1,0,0). \end{aligned}$$

Our "universal profile" result, in the non-radial case is

Theorem 7 [3] *Let u be a type II blow-up solution, with $T_+ = 1$, verifying (2) with η_0 small. Then, after a rotation and a translation of $\mathbb{R}^3$, there exists a small real parameter l and smooth functions $\lambda(t) > 0$ on $(0,T_+)$, $x(t) \in \mathbb{R}^3$, $t \in (0,T_+)$ so that*

$$(u(t), \partial_t u(t)) - (v_0, v_1) - \pm\left(\frac{W_l(\frac{x-x(t)}{\lambda(t)}, 0)}{\lambda(t)^{\frac{1}{2}}}, \frac{\partial_t W_l(\frac{x-x(t)}{\lambda(t)}, 0)}{\lambda(t)^{\frac{3}{2}}}\right) \underset{t\to 1}{\to} 0,$$

in $\dot{H}^1 \times L^2$. Moreover,

$$\lim_{t\to 1} \frac{\lambda(t)}{(1-t)} = 0, \qquad \lim_{t\to 1} \frac{x(t)}{1-t} = \mathbf{e}, \qquad |l| \le c\sqrt{\eta_0}.$$

A fundamental role in the proof of this result is played by the following rigidity theorem

Theorem 8 [3] *Let u be a non-zero solution of* (NLW) *such that there exist* $\lambda(t) > 0, x(t) \in \mathbb{R}^3, t \in (-T_-, T_+)$ *so that*

$$K = \left\{ \left(\frac{1}{\lambda(t)^{\frac{1}{2}}} u\left(\frac{x - x(t)}{\lambda(t)}, t\right), \frac{1}{\lambda(t)^{\frac{3}{2}}} \partial_t u\left(\frac{x - x(t)}{\lambda(t)}, t\right) \right) \right\}$$

has compact closure in $\dot{H}^1 \times L^2$. *Assume further that*

$$\sup_{t \in (-T_-, T_+)} \int |\nabla u(t)|^2 + (\partial_t u(t))^2 < 2 \int |\nabla W|^2.$$

Then $T_\pm = \pm\infty$, *and there exist* $l \in (-1, 1)$, *a rotation* R *of* $\mathbb{R}^3$, $\lambda_0 > 0$, $x_0 \in \mathbb{R}^3$ *so that*

$$u(x,t) = \pm \frac{W_l(\frac{R(x) - x_0}{\lambda_0}, \frac{t}{\lambda_0})}{\lambda_0^{\frac{1}{2}}}.$$

Acknowledgements The author was supported in part by NSF grant DMS-0968472.

References

1. Ding, W.Y.: On a conformally invariant elliptic equation on $\mathbb{R}^n$. Commun. Math. Phys. **107**(2), 331–335 (1986)
2. Duyckaerts, T., Kenig, C., Merle, F.: Universality of blow-up profile for small radial type II blow-up solution of the energy-critical wave equation. J. Eur. Math. Soc. (to appear)
3. Duyckaerts, T., Kenig, C., Merle, F.: Universality of the blow-up profile for small type II blow-up solutions of the energy-critical wave equation: the non-radial case. Preprint
4. Duyckaerts, T., Merle, F.: Dynamics of threshold solutions for energy-critical wave equation. Int. Math Res. Pap., IMRP, ArtID rnp002, 67 pages (2008)
5. Kenig, C.: Recent developments on the global behavior to critical nonlinear dispersive equations. In: Proc. of the ICM, Hyderabad, India (2010, to appear)
6. Kenig, C., Merle, F.: Global well-posedness, scattering and blow-up for the energy-critical focusing non-linear wave equation. Acta Math. **201**(2), 147–212 (2008)
7. Kenig, C., Merle, F.: Scattering for $H^{1/2}$ bounded solutions to the cubic defocusing NLS in 3 dimensions. Trans. Am. Math. Soc. **362**, 1937–1962 (2010)
8. Kricger, J., Schlag, W., Tataru, D.: Slow blow-up solutions for the $H^1(\mathbb{R}^3)$ critical focusing semi-linear wave equation in $\mathbb{R}^3$. Duke Math. J. **147**, 1–53 (2009)

A Simple Energy Pump for the Surface Quasi-geostrophic Equation

Alexander Kiselev and Fedor Nazarov

Abstract We consider the question of growth of high order Sobolev norms of solutions of the conservative surface quasi-geostrophic equation. We show that if $s > 0$ is large then for every given A there exists initial data with a norm that is small in H^s such that the H^s norm of corresponding solution at some time exceeds A. The idea of the construction is quasilinear. We use a small perturbation of a stable shear flow. The shear flow can be shown to create small scales in the perturbation part of the flow. The control is lost once the nonlinear effects become too large.

1 Introduction

In this paper, we consider the surface quasi-geostrophic equation

$$\partial_t \theta = (u \cdot \nabla)\theta, \qquad \theta(x,0) = \theta_0(x), \tag{1}$$

$u = \nabla^{\perp}(-\Delta)^{-1/2}\theta$, set on the torus $\mathbb{T}^2$ (which is equivalent to working with periodic initial data in $\mathbb{R}^2$). Observe that the structure of the SQG equation is similar to the 2D Euler equation written for vorticity, but the velocity is less regular in the SQG case ($u = \nabla^{\perp}(-\Delta)^{-1}\theta$ for the 2D Euler). The SQG equation comes from atmospheric science, and can be derived via formal asymptotic expansion (assuming small Rossby and Ekman numbers) from a larger system of 3D Navier–Stokes equations in a rotating frame coupled with temperature equation through gravity induced buoyancy force (see [4, 8]). The equation (1) describes evolution of the potential temperature on the surface, and its solution can be used to determine the main order approximation for the solution of the full three dimensional problem.

In mathematical literature, the SQG equation was introduced for the first time by Constantin, Majda and Tabak in [1], where a parallel between the structure of

A. Kiselev (✉) · F. Nazarov
Department of Mathematics, University of Wisconsin, Madison, WI 53706, USA
e-mail: kiselev@math.wisc.edu

F. Nazarov
e-mail: nazarov@math.wisc.edu

H. Holden, K.H. Karlsen (eds.), *Nonlinear Partial Differential Equations*,
Abel Symposia 7, DOI 10.1007/978-3-642-25361-4_9,

the conservative SQG equation and 3D Euler equation was drawn. Numerical experiments carried out in [1] showed steep growth of the gradient of solution in the saddle point scenario for the initial data, and suggested the possibility of singularity formation in finite time. Subsequent numerical experiments [7] suggested the solutions stay regular. Later, Cordoba [2] ruled out singularity formation in the scenario suggested by [1]. Despite significant effort by many researchers, whether blow up for the solutions of (1) can happen in a finite time remains open. Moreover, there are no examples that exhibit just infinite growth in time for some high order Sobolev norm. This paper is a step towards better understanding of this phenomenon.

Before stating the main result, we would like to compare the situation with what is known for two-dimensional Euler equation, which in vorticity form coincides with (1) but the velocity is given by $u = \nabla^{\perp}(-\Delta)^{-1}\theta$. The global existence of smooth solutions is known in this case, and there is an upper bound on the gradient and higher order Sobolev norms of θ that is double exponential in time (see, e.g. [5]). However the examples with actual growth are much weaker—the best current result is just superlinear in time (Denisov [3], with earlier works by Nadirashvili [6] and Yudovich [9, 10] giving linear or weaker rates of growth). It may appear that the SQG equation being more singular, it should be easier to prove infinite growth in this case. However to prove infinite in time growth, one needs to produce an example of "stable instability", a controllable mechanism of small scale production. This control is more difficult for the SQG than for two-dimensional Euler equation.

Let us denote H^s the usual scale of Sobolev spaces on $\mathbb{T}^2$. The main purpose of this short note is to show that the identically zero solution is strongly unstable in H^s for any s sufficiently large. Namely, we will prove the following

Theorem 1 *Assume that s is sufficiently large ($s \geq 11$ will do). Given any $A > 0$, there exists θ_0 such that $\|\theta_0\|_{H^s} \leq 1$, but the corresponding solution of (1) satisfies*

$$\limsup_{t\to\infty} \|\theta(\cdot,t)\|_{H^s} \geq A. \tag{2}$$

Remark 1

1. In our example, the initial data θ_0 will be simply a trigonometric polynomial with a few nonzero harmonics. Its size will be well controlled in any H^s norm.
2. From the argument, it will be clear that it is not difficult to derive a lower bound on time when the bound (2) is achieved. This time scales as a certain power of A.
3. The arguments of Denisov [3] can also be used to produce similar result, with better control of constants and time—but in a different scenario. Denisov considers perturbation of an explicitly given saddle point flow. In this note, we will consider a technically simpler but less singular case of a shear flow.

2 The Proof

We shall view the solutions $\theta(x,t)$ of (1) as sequences of Fourier coefficients $\hat{\theta}_k$, $k = (k_1, k_2) \in \mathbb{Z}^2$. On the Fourier side, after symmetrization, our solution satisfies the following equation.

$$\frac{d}{dt}\hat{\theta}_k = \frac{1}{2}\sum_{l+m=k}(l \wedge m)\left(\frac{1}{|l|} - \frac{1}{|m|}\right)\hat{\theta}_l\hat{\theta}_m, \qquad \hat{\theta}_k(0) = (\hat{\theta}_0)_k. \tag{3}$$

Here $l \wedge m = l_1 m_2 - l_2 m_1$.

Our initial data θ_0 will be just a simple trigonometric polynomial p, given by $\hat{p}_e = \hat{p}_{-e} = 1$, $\hat{p}_g = \hat{p}_{g+e} = \hat{p}_{-g} = \hat{p}_{-g-e} = \tau$ where $e = (1,0)$, $g = (0,2)$, and $\tau = \tau(A) > 0$ is a small parameter to be chosen later. Then it follows from (3) that the solution is an even real-valued function with $\hat{\theta}_0 = 0$ for all times. Moreover, $\hat{\theta}_k(t) \equiv 0$ whenever k_2 is odd.

We have two easy to check conservation laws: $\sum_k \hat{\theta}_k(t)^2 = 2 + 4\tau^2$ and $\sum_k \frac{\hat{\theta}_k^2(t)}{|k|} = 2 + 2\tau^2(\frac{1}{2} + \frac{1}{\sqrt{5}})$. After subtraction, we obtain that

$$\sum_k \hat{\theta}_k(t)^2\left(1 - \frac{1}{|k|}\right) = \left(3 - \frac{2}{\sqrt{5}}\right)\tau^2,$$

for all $t \geq 0$. Since $\hat{\theta}_{\pm g/2}(t) = 0$, this implies

$$\sum_{k \neq \pm e} \hat{\theta}_k(t)^2 \leq \left(\frac{\sqrt{2}}{\sqrt{2}-1}\right)\left(3 - \frac{2}{\sqrt{5}}\right)\tau^2 \leq 10\tau^2$$

for all times. Then the first conservation law also implies $\hat{\theta}_e(t) \in (1 - 8\tau^2, 1 + 2\tau^2) \subset (1/2, 2)$ for all times, provided that τ is sufficiently small.

Consider the quadratic form

$$\mathscr{J}(\hat{\theta}) = \sum_{k \in \mathbb{Z}^2_+} \Phi(k)\hat{\theta}_k\hat{\theta}_{k+e}.$$

We have

$$\begin{aligned}
&\frac{d}{dt}\mathscr{J}(\hat{\theta}) \\
&= \frac{1}{2}\sum_{k \in \mathbb{Z}^2_+}\Phi(k)\Bigg[\hat{\theta}_k \sum_{l+m=x+e, l,m \neq \pm e}(l \wedge m)\left(\frac{1}{|l|} - \frac{1}{|m|}\right)\hat{\theta}_l\hat{\theta}_m \\
&\quad + \hat{\theta}_{k+e}\sum_{l+m=k, l,m \neq \pm e}(l \wedge m)\left(\frac{1}{|l|} - \frac{1}{|m|}\right)\hat{\theta}_l\hat{\theta}_m\Bigg] \\
&\quad + \frac{1}{2}\hat{\theta}_e\sum_{k \in \mathbb{Z}^2_+}(e \wedge k)\Phi(k)\Bigg[\left(1 - \frac{1}{|k|}\right)\hat{\theta}_k^2 - \left(1 - \frac{1}{|k+2e|}\right)\hat{\theta}_k\hat{\theta}_{k+2e} \\
&\quad - \left(1 - \frac{1}{|k+e|}\right)\hat{\theta}_{k+e}^2 + \left(1 - \frac{1}{|k-e|}\right)\hat{\theta}_{k+e}\hat{\theta}_{k-e}\Bigg] \\
&\equiv \sigma + \Sigma,
\end{aligned}$$

where σ denotes the first sum and Σ the second. Since for $l+m=k$, we have $|(l\wedge m)(\frac{1}{|l|}-\frac{1}{|m|})|\leqslant 2|k|$ and $|k+e|\asymp|k|$ for $k\in\mathbb{Z}^2_+$, we conclude that

$$|\sigma|\leqslant\left(\sum_{k\in\mathbb{Z}^2_+}|k||\Phi(k)||\hat\theta_k|\right)\left(\sum_{l\neq\pm e}\hat\theta_l^2\right)\leqslant C\tau^2\sum_{k\in\mathbb{Z}^2_+}|k||\Phi(k)||\hat\theta_k|.$$

On the other hand, Σ can be rewritten as

$$\begin{aligned}\hat\theta_e\sum_{k_2>0}k_2\sum_{k_1\in\mathbb{Z}}\frac14\times\Bigg[&(\Phi(k-e)-\Phi(k-2e))\left(1-\frac{1}{|k-e|}\right)\hat\theta^2_{k-e}\\ &+(\Phi(k+e)-\Phi(k))\left(1-\frac{1}{|k+e|}\right)\hat\theta^2_{k+e}\\ &+2\left\{\Phi(k)\left(1-\frac{1}{|k-e|}\right)-\Phi(k-e)\left(1-\frac{1}{|k+e|}\right)\right\}\hat\theta_{k-e}\hat\theta_{k+e}\Bigg].\end{aligned}$$

Now let $\Phi(k)=k_1+\frac12$. We get the sum of quadratic forms with the coefficients

$$1-\frac{1}{\sqrt{(k_1-1)^2+k_2^2}},\quad 1-\frac{1}{\sqrt{(k_1+1)^2+k_2^2}}$$

at the squares and

$$\left(k_1+\frac12\right)\left(1-\frac{1}{\sqrt{(k_1-1)^2+k_2^2}}\right)-\left(k_1-\frac12\right)\left(1-\frac{1}{\sqrt{(k_1+1)^2+k_2^2}}\right)$$

at the double product.

A straightforward computation shows that when $k_1=0$, this form is degenerate and when $k_1\neq 0$, it is strictly positive definite and dominates $\frac{c}{|k|^3}(\hat\theta^2_{k-e}+\hat\theta^2_{k+e})$.

Using that fact that $\hat\theta_e(t)\geq 1/2$ for all times, we obtain

$$\Sigma\geqslant c\sum_{k\in\mathbb{Z}^2_+}\frac{\hat\theta_k^2}{|k|^3}.$$

Now there are several possibilities.

(A) At some time t, we will have $\sum_{k\in\mathbb{Z}^2_+}|k|^2|\hat\theta_k|\geqslant\sum_{k\in\mathbb{Z}^2_+}|k||\Phi(k)||\hat\theta_k|\geqslant\tau^{1/2}$.
Observe that

$$\sum_{k\in\mathbb{Z}^2_+}|k|^2|\hat\theta_k|\leqslant\left(\sum_{k\in\mathbb{Z}^2_+}\hat\theta_k^2\right)^{1/3}\left(\sum_{k\in\mathbb{Z}^2_+}|k|^{21}\hat\theta_k^2\right)^{1/6}\left(\sum_{k\in\mathbb{Z}^2_+}|k|^{-3}\right)^{1/2}.$$

Then, since $\sum_{k\in\mathbb{Z}^2_+}\hat\theta_k^2\leqslant 10\tau^2$, we get that the H^{11} norm of the solution gets large: $\|\theta\|_{H^{11}}\geq C\tau^{-1/6}$.

(B) The case (A) never occurs but $\sum_{k\in\mathbb{Z}^2_+} |k|^{-3}\hat{\theta}_k^2$ becomes comparable with $\tau^{5/2}$. Note that until this moment $\mathscr{J}(\hat{\theta})$ increases from its initial value about τ^2. Also, $\mathscr{J}(\hat{\theta}) \leqslant \sum_{k\in\mathbb{Z}^2_+} |k|\hat{\theta}_k^2$.

Thus, in this case, we use

$$\sum_{k\in\mathbb{Z}^2_+} |k|\hat{\theta}_k^2 \leqslant \left(\sum_{k\in\mathbb{Z}^2_+} |k|^{-3}\hat{\theta}_k^2\right)^{5/6} \left(\sum_{k\in\mathbb{Z}^2_+} |k|^{21}\hat{\theta}_k^2\right)^{1/6}$$

and, again, it follows that the H^{11} norm becomes large: $\|\theta\|_{H^{11}} \geq C\tau^{-\frac{1}{12}}$.

At last, if neither (A), nor (B) occur, then $\mathscr{J}(\hat{\theta})$ grows without bound and the $H^{1/2}$-norm gets large eventually. Now given A just choose τ sufficiently small and Theorem 1 follows.

Acknowledgements Research of AK has been supported in part by the NSF-DMS grant 1104415. Research of FN has been partially supported by the NSF-DMS grant 0800243.

References

1. Constantin, P., Majda, A., Tabak, E.: Formation of strong fronts in the 2D quasi-geostrophic thermal active scalar. Nonlinearity **7**, 1495–1533 (1994)
2. Cordoba, D.: Nonexistence of simple hyperbolic blow up for the quasi-geostrophic equation. Ann. Math. **148**, 1135–1152 (1998)
3. Denisov, S.: Infinite superlinear growth of the gradient for the two-dimensional Euler equation. Discrete Contin. Dyn. Syst., Ser. A **23**, 755–764 (2009)
4. Held, I., Pierrehumbert, R., Garner, S., Swanson, K.: Surface quasi-geostrophic dynamics. J. Fluid Mech. **282**, 1–20 (1995)
5. Majda, A., Bertozzi, A.: Vorticity and Incompressible Flow. Cambridge University Press, Cambridge (2002)
6. Nadirashvili, N.S.: Wandering solutions of the two-dimensional Euler equation. Funkc. Anal. Prilozh. **25**, 70–71 (1991) (in Russian). Translation in Funct. Anal. Appl. **25** (1991), 220–221 (1992)
7. Ohkitani, K., Yamada, M.: Inviscid and inviscid-limit behavior of a SQG flow. Phys. Fluids **9**, 876–882 (1997)
8. Pedlosky, J.: Geophysical Fluid Dynamics Springer, New York (1987)
9. Yudovich, V.I.: The loss of smoothness of the solutions of the Euler equation with time. Dinamika Sploshn. Sredy, Nestacionarnye Problemy Gidrodinamiki **16**, 71–78 (1974) (in Russian)
10. Yudovich, V.I.: On the loss of smoothness of the solutions of the Euler equations and the inherent instability of flows of an ideal fluid. Chaos **10**, 705–719 (2000)

On the Formation of Trapped Surfaces

Sergiu Klainerman

Abstract In a recent important breakthrough D. Christodoulou (The Formation of Black Holes in General Relativity. Monographs in Mathematics. Eur. Math. Soc., Zurich, 2009) has solved a long standing problem of General Relativity of evolutionary formation of trapped surfaces in the Einstein-vacuum space-times. He has identified an open set of regular initial conditions on a finite outgoing null hypersurface leading to a formation a trapped surface in the corresponding vacuum space-time to the future of the initial outgoing hypersurface and another incoming null hypersurface with the prescribed Minkowskian data. He also gave a version of the same result for data given on part of past null infinity. His proof is based on an inspired choice of the initial condition, an ansatz which he calls *short pulse*, and a complex argument of propagation of estimates, consistent with the ansatz, based, largely, on the methods used in the global stability of the Minkowski space (Christodoulou and Klainerman in The Global Nonlinear Stability of the Minkowski Space. Princeton Mathematical Series, vol. 41, 1993). Once such estimates are established in a sufficiently large region of the space-time the actual proof of the formation of a trapped surface is quite straightforward.

Christodoulou's result has been significantly simplified and extended in my joint works with I. Rodnianski (Klainerman and Rodnianski in On the formation of trapped surfaces, Acta Math. 2011, in press) and (Klainerman and Rodnianski in Discrete Contin. Dyn. Syst. 28(3):1007–1031, 2010). In this note I will give a short survey of these results.

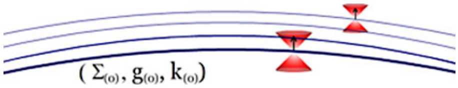
1 Introduction

I start with a short introduction to the initial value problem for the Einstein field equations which is meant to put in context the results discussed later. An initial data set consists of a 3 dimensional manifold $\Sigma_{(0)}$, a complete Riemannian metric $g_{(0)}$, a symmetric 2-tensor $k_{(0)}$, and a well specified set of initial conditions corresponding to the matter-fields under consideration. These have to be restricted to a well known

S. Klainerman (✉)
Department of Mathematics, Princeton University, Princeton, NJ 08544, USA
e-mail: seri@math.princeton.edu

H. Holden, K.H. Karlsen (eds.), *Nonlinear Partial Differential Equations*, Abel Symposia 7, DOI 10.1007/978-3-642-25361-4_10,

Fig. 1



set of constraint equations. A Cauchy development of an initial data set is a globally hyperbolic space-time $(\mathscr{M}, g)$, verifying the Einstein field equations,

$$R_{\alpha\beta} - \frac{1}{2} R g_{\alpha\beta} = T_{\alpha\beta}, \qquad \text{(EFE)}$$

and an embedding $i : \Sigma \to \mathscr{M}$ such that $i_*(g_{(0)}), i_*(k_{(0)})$ are the first and second fundamental forms of $i(\Sigma_{(0)})$ in $\mathscr{M}$. In what follows I will mostly restrict the discussion to the Einstein vacuum equations, i.e. the case when the energy momentum tensor vanishes identically and the equations take the purely geometric form,

$$R_{\alpha\beta} = 0. \qquad \text{(EVE)}$$

I also restrict our attention to asymptotically flat initial data sets, i.e. outside a sufficiently large compact set K, $\Sigma_{(0)} \setminus K$ is diffeomorphic to the complement of the unit ball in $\mathbb{R}^3$ and admits a system of coordinates in which $g_{(0)}$ is asymptotically euclidean and $k_{(0)}$ vanishes at appropriate order. The most primitive question asked about the initial value problem, solved in a satisfactory way, for very large classes of evolution equations, is that of local existence and uniqueness of solutions. For the Einstein equations this type of result was first established by Y.C. Bruhat [1] with the help of wave coordinates.[1] According to this result any smooth initial data set admits a unique, smooth, local (up to an isometry) *globally hyperbolic*[2] Cauchy development. In the case of nonlinear systems of differential equations the local existence and uniqueness result leads, through a straightforward extension argument, to a global result concerning the maximal time interval of existence. If this interval is bounded the solution must become infinite at its upper boundary. The formulation of the same type of result for the Einstein equations is a little more subtle; something similar was achieved in [2].

Theorem 1 (Bruhat–Geroch) *For each smooth initial data set there exists a unique*, *smooth*, *maximal*, *future*, *globally hyperbolic development* (*MFGHD*).

Thus any construction, obtained by an evolutionary approach from a specific initial data set, must be necessarily contained in its maximal development MFGHD. This may be said to solve the problem of global[3] existence and uniqueness in General Relativity; all further questions, one could say, concern the qualitative proper-

[1] These allow one to cast the Einstein vacuum equations in the form of a system of nonlinear wave equations.

[2] Any past directed, in-extendable causal curve of the development intersects Σ_0.

[3] A proper definition of global solutions in GR requires a special discussion concerning the proper time of causal geodesics.

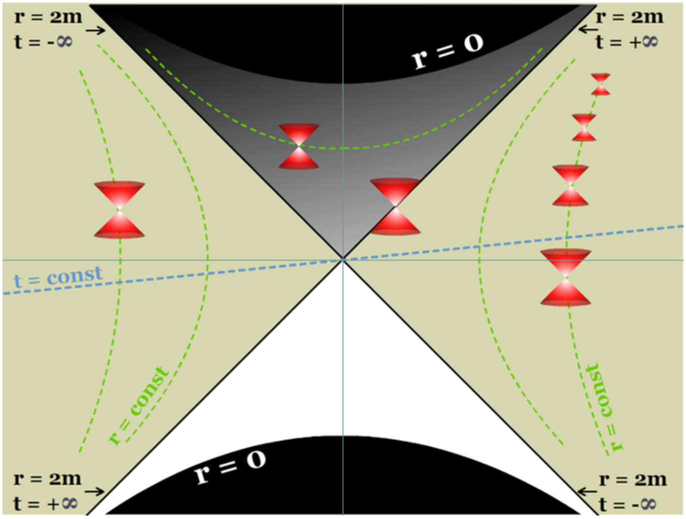


Fig. 2 Kruskal's maximally extended Schwarzschild space-time. Note the two disconnected external regions, $r > 2m$, the *black* and *white holes* and the curvature singularity at $r = 0$. Note the behavior of *light cones* at the event horizon, $r = 2m$

ties of these maximal developments. The central issue becomes that of existence and character of singularities.

1.1 Special Solutions

We recall that EVE admits a remarkable family of explicit, stationary, solutions given by the two parameter family of Kerr solutions among which one distinguishes the Schwarzschild family of solutions, of mass $m > 0$,

$$g_S = -\left(1 - \frac{2m}{r}\right) dt^2 + \left(1 - \frac{2m}{r}\right)^{-1} dr^2 + r^2\, d\sigma_{\mathbb{S}^2}. \tag{1}$$

Though the metric seems singular at $r = 2m$ it turns out that one can glue together two regions $r > 2m$ and two regions $r < 2m$ of the Schwarzschild metric to obtain a metric which is smooth along $\mathscr{H} = \{r = 2m\}$, see [6], called the Schwarzschild horizon. The portion of $r < 2m$ to the future of the hypersurface $t = 0$ is a black hole whose future boundary $r = 0$ is singular. The region $r > 2m$, called the domain of outer communication, is free of singularities.

The Schwarzschild family is included in a larger two parameter family of solutions $\mathscr{K}(a, m)$ discovered by Kerr. A given Kerr space-time, with $0 \le a < m$ has a well defined domain of outer communication $r > r_+ := m + (m^2 - a^2)^{1/2}$. In

Boyer–Lindquist coordinates, well adapted to $r > r_+$ the Kerr metric has the form,

$$g_K = -\frac{\Delta - a^2 \sin^2\theta}{\Sigma} dt^2 - \frac{2a\sin^2\theta(r^2 + a^2 - \Delta)}{\Sigma} dt\, d\phi + \frac{(r^2+a^2)^2 - \Delta a^2 \sin^2\theta}{\Sigma} \sin^2 d\phi^2 + \frac{\Sigma}{\Delta} dr^2 + \Sigma\, d\theta^2$$

with $\Sigma = r^2 + a^2\cos^2\theta$, $\Delta = r^2 + a^2 - 2mr$. As in the Schwarzschild case, the exterior Kerr metric extends smoothly across the Kerr event horizon, $\mathscr{H} = \{r = r_+\}$. It can be shown that the future and past sets of any point in the domain of outer communication intersects any time-like curve, passing through points of arbitrary large values of r, in finite time as measured relative to proper time along the curve. This fact is violated by points in the region $r \le r_+$, which defines the *black hole* region of the space-time. Thus physical signals which initiate at points in $r \le r_+$ cannot be registered by far away observers. The extended Kerr is singular only at $r = 0$. Thus the singularities in Kerr cannot have any effect on the domain of outer communication which is completely smooth.

1.2 Cosmic Censorship

In general we expect maximal developments to be incomplete, with singular boundaries. The unavoidable presence of singularities, for sufficiently large initial data sets, as well as the analysis of explicit examples (such as Schwarzschild and Kerr) have led Penrose to formulate two fundamental conjectures, concerning the character of general solutions to the Einstein equations. Here I restrict my discussion only to the so called weak cosmic censorship conjecture (WCC), which is the only one relevant to the issue of stability of Kerr. To understand the statement of (WCC) consider the different behavior of null rays in Schwarzschild and Minkowski space-times. In Minkowski space light originating at any point $p = (t_0, x_0)$ propagates, towards future, along the null rays of the null cone $t - t_0 = |x - x_0|$. Any free observer in $\mathbb{R}^{1+3}$, following a straight time-like line, will necessarily meet this light cone in finite time, thus experiencing the event p. On the other hand, any point p in the trapped region $r < 2m$ of the Schwarzschild space, is such that all null rays initiating at p remain trapped in the region $r < 2m$. In particular events causally connected to the singularity at $r = 0$ cannot influence events in the domain of outer communication $r > 2m$, which is thus entirely free of singularities. The same holds true in any Kerr solution with $0 \le a < m$.

WCC is an optimistic extension of this fact to the future developments of general, asymptotically flat initial data. The desired conclusion of the conjecture is that any such development, with the possible exception of a non-generic set of initial conditions, has the property that any *sufficiently distant observer* will never encounter singularities or any other effects propagating from them. To make this more precise one needs define what a sufficiently distant observer means. This is typically done by

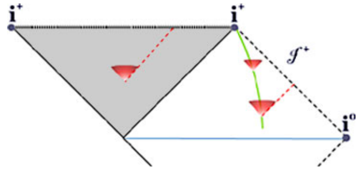


Fig. 3 Behavior of null geodesics in the domain of outer communication by contrast to those in a *black hole*

introducing the notion of future null infinity $\mathscr{I}^+$ which, roughly speaking, provides end points for the null geodesics which propagate to asymptotically large distances. The future null infinity is constructed by conformally embedding the physical space-time $(\mathscr{M}, g)$ under consideration to a larger space-time[4] $(\bar{\mathscr{M}}, \bar{g})$, $\bar{g} = \Omega^2 g$ in $\mathscr{M}$, with a null boundary $\mathscr{I}^+$ (where $\Omega = 0$, $d\Omega \neq 0$).

Definition 1 The future null infinity $\mathscr{I}^+$ is said to be complete if any future null geodesics along it can be indefinitely extended relative an affine parameter.

Conjecture 1 (WCC) *Generic asymptotically flat initial data have maximal future developments possessing a complete future null infinity.*

1.3 Penrose Singularity Theorem

The fundamental notion of a trapped surface has also been introduced by R. Penrose, see [10, 11], in connection to his famous singularity theorem.

Theorem 2 (Penrose) *If the manifold support of an initial data set is non-compact and contains a closed trapped surface the corresponding maximal future development is incomplete. The result holds true in the presence of any matter-fields which verify the positive energy conditions, i.e. if for any null vector L,*

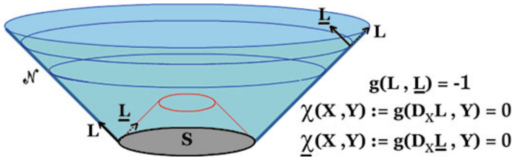
$$\mathrm{Ric}(L, L) \geq 0. \tag{2}$$

The notion of a trapped surface $S \subset \Sigma$, can be rigorously defined in terms of a local condition on S. More precisely, at any point $p \in S$, let $L, \underline{L}$ by a pair of null vectors orthogonal to S and normalized by the condition $g(L, \underline{L}) = -1$. We then define the null second fundamental forms $\chi, \underline{\chi}$ as in Fig. 4, where X, Y are arbitrary vectorfields tangent to S, and their traces $\mathrm{tr}\,\chi$, $\mathrm{tr}\,\underline{\chi}$ with respect to the induced metric on S. The surface is said to be trapped if both $\mathrm{tr}\,\chi$ and $\mathrm{tr}\,\underline{\chi}$ are negative at all points of S. In normal situation only the trace of the null second fundamental form in the *incoming* direction $\underline{L}$ is negative.

The flat initial data set (whose development is the Minkowski space) have, of course, no such surfaces. On the other hand, for the Schwarzschild initial data set,

[4] Note however that the boundary of this extended space-time is not smooth, generically.

Fig. 4



i.e. the one whose development is Schwarzschild, any surface $r = r_0$, with $r_0 < 2m$ is trapped. Of course, the Schwarzschild metric has a genuine singularity at $r = 0$, where the curvature tensor becomes infinite. This is a lot stronger than just saying that space-time is incomplete. All Kerr solutions, with the exception of the flat Minkowski space itself, have trapped surfaces.

Despite its obvious importance the Penrose singularity theorem gives only a faint glimpse on the nature of singularities in GR. It leaves unanswered the following obvious questions.

1. What is the significance of the uniformity condition, i.e. the condition that the outgoing expansion $\operatorname{tr}\chi$ is strictly negative, at all points of S? Can this be relaxed?
2. Can trapped surfaces form in evolution? Can they form in vacuum?
3. Once a trapped surface has formed the theorem only tells us that the corresponding maximal future development must be incomplete. It does not provide any information concerning the nature of singularities.

Before the major new advance of Christodoulou in [4] the only rigorous results about formation of trapped surfaces where obtained by Christodoulou himself for the spherical symmetric Einstein equations coupled with a scalar field, see [3].

Remark 1 In view of the Penrose singularity theorem and assuming WCC to be true, formation of a trapped surface implies formation of a black hole. It is for this reason that Christodoulou calls his result "formation of black holes" rather then simply formation of trapped surfaces. Of course, WCC remains way out of reach in this context.

To state the result of [4], as well as that of [8, 9] we need to introduce some notation.

2 Heuristic Argument

2.1 Double Null Foliations

We consider a region $\mathscr{D} = \mathscr{D}(u_*, \underline{u}_*)$ of a vacuum space-time (M, g) spanned by a double null foliation generated by the optical functions $(u, \underline{u})$ increasing towards the future, $0 \le u \le u_*$ and $0 \le \underline{u} \le \underline{u}_*$. We denote by H_u the outgoing null hypersurfaces generated by the level surfaces of u and by $\underline{H}_{\underline{u}}$ the incoming null hypersurfaces generated level hypersurfaces of $\underline{u}$. We write $S_{u,\underline{u}} = H_u \cap \underline{H}_{\underline{u}}$ and denote by $H_u^{(\underline{u}_1,\underline{u}_2)}$, and $\underline{H}_{\underline{u}}^{(u_1,u_2)}$ the regions of these null hypersurfaces defined by $\underline{u}_1 \le \underline{u} \le \underline{u}_2$ and respectively $u_1 \le u \le u_2$. Let $L, \underline{L}$ be the geodesic vectorfields associated to the two foliations and define,

$$\frac{1}{2}\Omega^2 = -g(L, \underline{L})^{-1}. \tag{3}$$

Observe that the flat value[5] of Ω is 1. As well known, our space-time slab $\mathscr{D}(u_*, \underline{u}_*)$ is completely determined (for small values of $u_*, \underline{u}_*$) by data along the null, characteristic, hypersurfaces H_0, $\underline{H}_0$ corresponding to $\underline{u} = 0$, respectively $u = 0$. Following [4] we assume that our data is trivial along $\underline{H}_0$, i.e. assume that H_0 extends for $\underline{u} < 0$ and the space-time (M, g) is Minkowskian for $\underline{u} < 0$ and all values of $u \ge 0$. Moreover, we can construct our double null foliation such that $\Omega = 1$ along H_0, i.e.,

$$\Omega(0, \underline{u}) = 1, \quad 0 \le \underline{u} \le \underline{u}_*. \tag{4}$$

Throughout this paper we work with the normalized null pair (e_3, e_4),

$$e_3 = \Omega \underline{L}, \qquad e_4 = \Omega L, \qquad g(e_3, e_4) = -2.$$

Given a 2-surfaces $S(u, \underline{u})$ and $(e_a)_{a=1,2}$ an arbitrary frame tangent to it we define the Ricci coefficients,

$$\Gamma_{(\lambda)(\mu)(\nu)} = g(e_{(\lambda)}, D_{e_{(\nu)}} e_{(\mu)}), \quad \lambda, \mu, \nu = 1, 2, 3, 4. \tag{5}$$

These coefficients are completely determined by the following components,

$$\begin{aligned}
\chi_{ab} &= g(D_a e_4, e_b), & \underline{\chi}_{ab} &= g(D_a e_3, e_b),\\
\eta_a &= -\frac{1}{2} g(D_3 e_a, e_4), & \underline{\eta}_a &= -\frac{1}{2} g(D_4 e_a, e_3),\\
\omega &= -\frac{1}{4} g(D_4 e_3, e_4), & \underline{\omega} &= -\frac{1}{4} g(D_3 e_4, e_3),\\
\zeta_a &= \frac{1}{2} g(D_a e_4, e_3)
\end{aligned} \tag{6}$$

where $D_a = D_{e_{(a)}}$. We also introduce the null curvature components,

$$\begin{aligned}
\alpha_{ab} &= R(e_a, e_4, e_b, e_4), & \underline{\alpha}_{ab} &= R(e_a, e_3, e_b, e_3),\\
\beta_a &= \frac{1}{2} R(e_a, e_4, e_3, e_4), & \underline{\beta}_a &= \frac{1}{2} R(e_a, e_3, e_3, e_4),\\
\rho &= \frac{1}{4} R(Le_4, e_3, e_4, e_3), & \sigma &= \frac{1}{4} {}^*R(e_4, e_3, e_4, e_3).
\end{aligned} \tag{7}$$

[5]Note that our normalization for Ω differ from that of [7].

Here ${}^*R$ denotes the Hodge dual of R. We denote by ∇ the induced covariant derivative operator on $S(u,\underline{u})$ and by ∇_3, ∇_4 the projections to $S(u,\underline{u})$ of the covariant derivatives D_3, D_4, see precise definitions in [7]. Observe that,

$$\begin{aligned}\omega &= -\frac{1}{2}\nabla_4(\log\Omega), & \underline{\omega} &= -\frac{1}{2}\nabla_3(\log\Omega),\\ \eta_a &= \zeta_a + \nabla_a(\log\Omega), & \underline{\eta}_a &= -\zeta_a + \nabla_a(\log\Omega).\end{aligned} \tag{8}$$

The connection coefficients Γ verify equations which have, very roughly, the form,

$$\begin{aligned}\nabla_4\Gamma &= R + \nabla\Gamma + \Gamma\cdot\Gamma,\\ \nabla_3\Gamma &= R + \nabla\Gamma + \Gamma\cdot\Gamma.\end{aligned} \tag{9}$$

Similarly the Bianchi identities for the null curvature components verify, also very roughly,

$$\begin{aligned}\nabla_4 R &= \nabla R + \Gamma\cdot R,\\ \nabla_3 R &= \nabla R + \Gamma\cdot R.\end{aligned} \tag{10}$$

Among these equations we note the following two, which play an essential role in Christodoulou's argument for the formation of trapped surfaces,

$$\nabla_4 \operatorname{tr}\chi + \frac{1}{2}(\operatorname{tr}\chi)^2 = -|\hat{\chi}|^2 - 2\omega\operatorname{tr}\chi, \tag{11}$$

$$\nabla_3\hat{\chi} + \frac{1}{2}\operatorname{tr}\underline{\chi}\,\hat{\chi} = \nabla\widehat{\otimes}\eta + 2\underline{\omega}\hat{\chi} - \frac{1}{2}\operatorname{tr}\chi\,\underline{\hat{\chi}} + \eta\widehat{\otimes}\eta. \tag{12}$$

2.2 *Heuristic Argument*

We start by making some important simplifying assumptions. As mentioned above we assume that our data is trivial along $\underline{H}_0$, i.e. assume that H_0 extends for $\underline{u} < 0$ and the space-time (M, g) is Minkowskian for $\underline{u} < 0$ and all values of $u \geq 0$. We introduce a small parameter $\delta > 0$ and restrict the values of $\underline{u}$ to $0 \leq \underline{u} \leq \delta$, i.e. $\underline{u}_* = \delta$.

We also make the following additional assumptions, assumed to hold in *the entire slab* $\mathscr{D}(u,\delta)$. We denote by $r = r(u,\underline{u})$ the radius of the 2-surfaces $S = S(u,\underline{u})$, i.e. $|S(u,\underline{u})| = 4\pi r^2$. We denote by r_0 the value of r for $S(0,0)$, i.e. $r_0 = r(0,0)$.

- For small δ, $u, \underline{u}$ are comparable with their standard values in flat space, i.e. $u \approx \frac{t-r+r_0}{2}$, $\underline{u} \approx \frac{t+r-r_0}{2}$. We also assume that $\Omega \approx 1$, $\frac{dr}{du} \approx -1$.
- Assume that $\operatorname{tr}\underline{\chi}$ is close to its value in flat space, i.e. $\operatorname{tr}\underline{\chi} \approx -\frac{2}{r}$.
- Assume that the term $E = \nabla\widehat{\otimes}\eta + 2\underline{\omega}\hat{\chi} - \frac{1}{2}\operatorname{tr}\chi\,\underline{\hat{\chi}} + \eta\widehat{\otimes}\eta$ on the right hand side of (12) is sufficiently small and can be neglected in a first approximation. Assume also that we can neglect the term $\operatorname{tr}\chi\omega$ on the right hand side of (11).

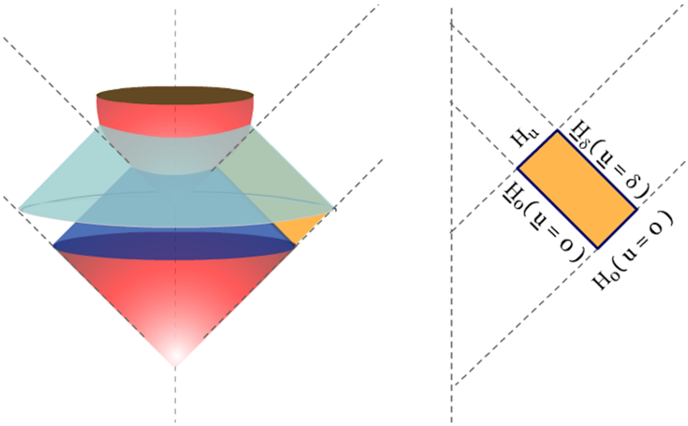


Fig. 5 The *colored region* on the *right* represents the domain $\mathscr{D}(u, \underline{u})$, $0 \le \underline{u} \le \delta$. The same picture is represented, more realistically on the *left*. The *lower red region* on the *left* is the flat portion of H_0, $u = 0$, while the *upper red region*, corresponding to a large values of u, is trapped starting with $\underline{u} = \delta$

Given these assumptions we can rewrite (11),

$$\frac{d}{d\underline{u}} \operatorname{tr} \chi \lesssim -|\hat{\chi}|^2$$

or, integrating,

$$\begin{aligned} \operatorname{tr} \chi(u, \underline{u}) &\lesssim \operatorname{tr} \chi(u, 0) - \int_0^{\underline{u}} |\hat{\chi}|(u, \underline{u}')^2 d\underline{u}' \\ &= \frac{2}{r(u, 0)} - \int_0^{\underline{u}} |\hat{\chi}(u, \underline{u}')|^2 d\underline{u}'. \end{aligned} \tag{13}$$

Multiplying (12) by $\hat{\chi}$ we deduce,

$$\frac{d}{du}|\hat{\chi}|^2 + \operatorname{tr} \underline{\chi} |\hat{\chi}|^2 = \hat{\chi} \cdot E$$

or, in view of our assumptions for $\operatorname{tr} \underline{\chi}$, and $\frac{dr}{du}$

$$\begin{aligned} \frac{d}{du}(r^2|\hat{\chi}|^2) &= r^2 \frac{d}{du}|\hat{\chi}|^2 + 2r\frac{dr}{du}|\hat{\chi}|^2 = r^2|\hat{\chi}|^2\left(-\operatorname{tr} \underline{\chi} + \frac{2}{r}\frac{dr}{du}\right) + r^2 \hat{\chi} \cdot E \\ &= r^2|\hat{\chi}|^2\left(-\left(\operatorname{tr} \underline{\chi} + \frac{2}{r}\right) + \frac{2}{r}\left(1 + \frac{dr}{du}\right)\right) + r^2\hat{\chi} \cdot E := F \end{aligned}$$

i.e.

$$r^2|\hat{\chi}|^2(u, \underline{u}) = r^2(0, \underline{u})|\hat{\chi}|^2(0, \underline{u}) + \int_0^u F(u', \underline{u}) du'.$$

Therefore, as $\int_0^u |F|$ is negligible in $\mathscr{D}$, we deduce

$$r^2|\hat\chi|^2(u,\underline{u}) \approx r^2(0,\underline{u})|\hat\chi|^2(0,\underline{u}).$$

We now freely prescribe $\hat\chi$ along the initial hypersurface $H_0^{(0,\delta)}$, i.e.

$$\hat\chi(0,\underline{u}) = \hat\chi_0(\underline{u}) \tag{14}$$

for some traceless 2 tensor $\hat\chi_0$. We deduce,

$$|\hat\chi|^2(u,\underline{u}) \approx \frac{r^2(0,\underline{u})}{r^2(u,\underline{u})}|\hat\chi_0|^2(\underline{u})$$

or, since $|\underline{u}| \le \delta$ and $r(u,\underline{u}) = r_0 + \underline{u} - u$,

$$|\hat\chi|^2(u,\underline{u}) \approx \frac{r_0^2}{(r_0-u)^2}|\hat\chi_0|^2(\underline{u}).$$

Thus, returning to (13),

$$\operatorname{tr}\chi(u,\underline{u}) \le \frac{2}{r_0-u} - \frac{r_0^2}{(r_0-u)^2}\int_0^{\underline{u}} |\hat\chi_0|^2(\underline{u}')d\underline{u}' + \text{error}.$$

Hence, for small δ, the necessary condition to have $\operatorname{tr}\chi(u,\underline{u}) \le 0$ is,

$$\frac{2(r_0-u)}{r_0^2} < \int_0^\delta |\hat\chi_0|^2.$$

Analyzing (11) along H_0 we easily deduce that the condition for the initial hypersurface H_0 not to contain trapped hypersurfaces is,

$$\int_0^\delta |\hat\chi_0|^2 < \frac{2}{r_0}$$

i.e. we are led to prescribe $\hat\chi_0$ such that,

$$\frac{2(r_0-u)}{r_0^2} < \int_0^\delta |\hat\chi_0|^2 < \frac{2}{r_0}. \tag{15}$$

We thus expect, following Christodoulou, that trapped surfaces may form if (15) is verified.

Remark 2 Observe that the argument we have presented above can be easily localized to angular sectors. More precisely, if condition (15) holds on an angular sector Λ on the initial hypersurface H_0 (i.e. a neighborhood of a fixed null geodesic) and all our other assumptions hold in the space-time region spanned by the incoming null geodesics starting on Λ we expect to form a scar on $S(u,\delta)$ i.e., $\operatorname{tr}\chi < 0$ in an angular region of $S(u,\delta)$.

2.3 Short Pulse Data

To prove such a result however we need to check that all the assumptions we made above can be verified. To start with, the assumption (15) requires, in particular, an L^∞ upper bound of the form,

$$|\hat{\chi}_0| \lesssim \delta^{-1/2}.$$

If we can show that such a bound persist in $\mathscr{D}$ then, in order to control the error terms F we need, for some $c > 0$,

$$\begin{aligned} &\operatorname{tr}\underline{\chi} + \frac{2}{r} = O(\delta^c), \qquad \frac{dr}{du} + 1 = O(\delta^c), \qquad \eta = O(\delta^{-1/2+c}), \\ &\omega = O(\delta^{-1+c}), \qquad \nabla\eta = O(\delta^{-1/2+c}). \end{aligned} \tag{16}$$

Other bounds will be however needed as we have to take into account all null structure equations. We face, in particular, the difficulty that most null structure equations have curvature components as sources. Thus we are obliged to derive bounds not just for all Ricci coefficients $\chi, \omega, \eta, \underline{\eta}, \underline{\chi}, \underline{\omega}$ but also for all null curvature components $\alpha, \beta, \rho, \sigma, \underline{\beta}, \underline{\alpha}$. In his work [4] Christodoulou has been able to derive such estimates starting with an ansatz (which he calls short pulse) for the initial data $\hat{\chi}_0$. More precisely he assumes, in addition to the triviality of the initial data along $\underline{H}_0$, that $\hat{\chi}_0$ verifies, relative to coordinates $\underline{u}$ and transported coordinates ω along H_0, (i.e. transported with respect to $\frac{d}{d\underline{u}}$),

$$\hat{\chi}_0(\underline{u}, \omega) = \delta^{-1/2} f_0(\delta^{-1}\underline{u}, \omega) \tag{17}$$

where f_0 is a fixed traceless, symmetric S-tangent two tensor along H_0. This ansatz is consistent with the following more general condition, for sufficiently large number of derivatives N and sufficiently small $\delta > 0$,

$$\delta^{1/2+k}\|\nabla_4^k \nabla^m \hat{\chi}_0\|_{L^2(0,\underline{u})} < \infty, \qquad 0 \le k+m \le N,\ 0 \le \underline{u} \le \delta. \tag{18}$$

Notation Here $\|\cdot\|_{L^2(u,\underline{u})}$ denotes the standard L^2 norm for tensor-fields on $S(u,\underline{u})$. Whenever there is no possible confusion we will also denote these norms by $\|\cdot\|_{L^2(S)}$. We shall also denote by $\|\cdot\|_{L^2(H)}$ and $\|\cdot\|_{L^2(\underline{H})}$ the standard L^2 norms along the null hypersurfaces $H = H_u$ and $\underline{H} = \underline{H}_{\underline{u}}$.

Remark 3 In [4] Christodoulou also includes weights, depending on $|u|$, in his estimates. These allow him to derive not only a local result but also one with data at past null infinity. In our work here we only concentrate on the local result, for $|u| \lesssim 1$, and thus drop the weights.

Assumption (18), together with the null structure equations (9) and null Bianchi equations (10) leads to the following estimates for the null curvature components, along the initial null hypersurface H_0,

$$\delta\|\alpha\|_{L^2(H_0)} + \|\beta\|_{L^2(H_0)} + \delta^{-1/2}\|(\rho,\sigma)\|_{L^2(H_0)} + \delta^{-3/2}\|\underline{\beta}\|_{L^2(H_0)} < \infty. \tag{19}$$

Consistent with (18), the angular derivatives of $\alpha, \beta, \rho, \sigma, \beta$ obey the same scaling as in (19) while each ∇_4 derivative costs an additional power of δ,

$$\begin{aligned} &\delta\|\nabla\alpha\|_{L^2(H_0)} + \|\nabla\beta\|_{L^2(H_0)} + \delta^{-\frac{1}{2}}\|\nabla(\rho,\sigma)\|_{L^2(H_0)} \\ &\quad + \delta^{-3/2}\|\nabla\underline{\beta}\|_{L^2(H_0)} < \infty, \\ &\delta^2\|\nabla_4\alpha\|_{L^2(H_0)} + \delta\|\nabla_4\beta\|_{L^2(H_0)} + \delta^{1/2}\|\nabla_4(\rho,\sigma)\|_{L^2(H_0)} \\ &\quad + \delta^{-1/2}\|\nabla_4\underline{\beta}\|_{L^2(H_0)} < \infty. \end{aligned} \tag{20}$$

Moreover, one can derive estimates for the Ricci coefficients, in various norms, weighted by appropriated powers of δ. Note that if one were to neglect the quadratic terms in (10) than the expected scaling behavior in δ would have been,

$$\delta\|\alpha\|_{L^2(H_0)} + \|\beta\|_{L^2(H_0)} + \delta^{-1}\|(\rho,\sigma)\|_{L^2(H_0)} + \delta^{-2}\|\underline{\beta}\|_{L^2(H_0)} < \infty.$$

Most of the body of work in [4] is to prove that these estimates can be propagated in the entire space-time region $\mathscr{D}(u_*, \delta)$, with u_* of size one and δ sufficiently small, and thus fulfill the necessary conditions for the formation of a trapped surface along the lines of the heuristic argument presented above. The proof of such estimates, which follows the main outline of the proof of stability of Minkowski space, as in [5] and [7], requires a step by step analysis to make sure that all estimates are consistent with the assigned powers of δ. This task is made particularly taxing in view of the fact that there are many nonlinear interferences which have to be tracked precisely.

2.4 New Initial Conditions

In [8] we embed the short-pulse ansatz of Christodoulou into a more general set of initial conditions, based on a different underlying scaling. The new scaling, which we incorporate into our basic norms, allows us to conceptualize the separation between the linear and nonlinear terms in the null Bianchi and null structure equations and explain the favorable appearance of additional positive powers of δ in the nonlinear error terms mentioned above. Though the initial conditions required to include Christodoulou's data do not quite satisfy this scaling, the generated anomalies are fewer and thus much easier to track.

We start with the observation that a natural alternative to (17) which comes to mind, related to the familiar parabolic scaling on null hyperplanes in Minkowski space, is

$$\hat{\chi}_0(\underline{u}, \omega) = \delta^{-1/2} f_0(\delta^{-1}\underline{u}, \delta^{-1/2}\omega). \tag{21}$$

This does not quite make sense in our framework of compact 2-surfaces $S(u, \underline{u})$, unless of course one is willing to consider the initial data $\hat{\chi}_0(\underline{u}, \omega)$ supported in the angular sector ω of size $\delta^{\frac{1}{2}}$. Such a support assumption would be however in contradiction with the lower bound in (15) required to be satisfied for *each* $\omega \in \mathbb{S}^2$.

The following interpretation of (21) (compare with (18)) makes sense however,

$$\delta^{k+\frac{m}{2}} \sup_{0\le \underline{u}\le \delta} \|\nabla_4^k \nabla^m \hat{\chi}_0\|_{L^2(0,\underline{u})} < \infty, \quad 0 \le k+m \le N. \tag{22}$$

Just as in the derivation of (19) we can use null structure equations (9) and null Bianchi equations (10) to derive, from (22),

$$\begin{aligned}
\delta^{1/2}\|\alpha\|_{L^2(H_0)} + \|\beta\|_{L^2(H_0)} + \delta^{-1/2}\|(\rho,\sigma)\|_{L^2(H_0)} + \delta^{-1}\|\underline{\beta}\|_{L^2(H_0)} &< \infty,\\
\delta\|\nabla\alpha\|_{L^2(H_0)} + \delta^{1/2}\|\nabla\beta\|_{L^2(H_0)} + \|\nabla(\rho,\sigma)\|_{L^2(H_0)} + \delta^{-1/2}\|\nabla\underline{\beta}\|_{L^2(H_0)} &< \infty,\\
\delta^{3/2}\|\nabla_4\alpha\|_{L^2(H_0)} + \delta\|\nabla_4\beta\|_{L^2(H_0)} + \delta^{1/2}\|\nabla_4(\rho,\sigma)\|_{L^2(H_0)} + \|\nabla_4\underline{\beta}\|_{L^2(H_0)} &< \infty.
\end{aligned} \tag{23}$$

We refer to these conditions, consistent with the null parabolic scaling, as δ-*coherent* assumptions. Observe that, unlike in the Christodoulou's case, each ∇ derivative costs a $\delta^{-1/2}$. It turns out that proving the propagation of such estimates can be done easily and systematically without the need of the step by step procedure mentioned earlier. In fact one can show, in this case, that all error terms, generated in the process of the energy estimates are either quadratic in the curvature and can be easily taken care by Gronwall or, if cubic, they must come with a factor of $\delta^{1/2}$ and therefore can be all absorbed for small values of δ.

The main problem with the ansatz (21), as with initial conditions (22), however, is that it is inconsistent with the formation of trapped surfaces requirements discussed above. One can only hope to show that the expansion scalar $\mathrm{tr}\,\chi$ along H_u, at $S(u, \underline{u})$, for some $u \approx 1$, will become negative[6] only in a small angular sector of size $\delta^{1/2}$. This is because, consistent with (23), condition (15) may only be satisfied in such a sector.

To obtain a trapped surface, i.e. to be able to treat initial data which verify Christodoulou's uniform assumption (15) we are forced to abandon the ansatz formulation of the characteristic initial data problem for the Einstein-vacuum equations and replace with a hierarchy of bounds, which "interpolate" between the regular δ-coherent assumptions (23) and the estimates (19)–(20) following from Christodoulou's short pulse ansatz.

[6] We could call such a region locally trapped, or a pre-scar.

At the level of curvature the new assumptions correspond to:

$$\begin{aligned}\delta\|\alpha\|_{L^2(H_0)}+\|\beta\|_{L^2(H_0)}+\delta^{-1/2}\|(\rho,\sigma)\|_{L^2(H_0)}+\delta^{-1}\|\underline{\beta}\|_{L^2(H_0)}&<\infty,\\ \delta\|\nabla\alpha\|_{L^2(H_0)}+\delta^{1/2}\|\nabla\beta\|_{L^2(H_0)}+\|\nabla(\rho,\sigma)\|_{L^2(H_0)}+\delta^{-1/2}\|\nabla\underline{\beta}\|_{L^2(H_0)}&<\infty,\\ \delta^2\|\nabla_4\alpha\|_{L^2(H_0)}+\delta\|\nabla_4\beta\|_{L^2(H_0)}+\delta^{1/2}\|(\nabla_4\rho,\nabla_4\sigma)\|_{L^2(H_0)}+\|\nabla_4\underline{\beta}\|_{L^2(H_0)}&<\infty.\end{aligned} \tag{24}$$

Observe that, by comparison with (23), the only anomalous terms are $\|\alpha\|_{L^2(H_0)}$ and $\|\nabla_4\alpha\|_{L^2(H_0)}$.

3 Main Results of [8]

3.1 Signature and Scaling

The results and proofs of [8] are intimately tied with a natural scaling which we introduce below.

Signature To every null curvature component $\alpha,\beta,\rho,\sigma,\underline{\beta},\underline{\alpha}$, null Ricci coefficients components $\chi,\zeta,\eta,\underline{\eta},\omega,\underline{\omega}$, and metric γ we assign a signature according to the following rule:

$$\mathrm{sgn}(\phi)=1\cdot N_4(\phi)+\frac{1}{2}\cdot N_a(\phi)+0\cdot N_3(\phi)-1 \tag{25}$$

where $N_4(\phi),N_3(\phi),N_a(\phi)$ denote the number of times e_4, respectively e_3 and $(e_a)_{a=1,2}$, which appears in the definition of ϕ. Thus,

$$\begin{aligned}&\mathrm{sgn}(\alpha)=2,\qquad \mathrm{sgn}(\beta)=1+1/2,\qquad \mathrm{sgn}(\rho,\sigma)=1,\\ &\mathrm{sgn}(\underline{\beta})=1/2,\qquad \mathrm{sgn}(\underline{\alpha})=0.\end{aligned}$$

Also,

$$\begin{aligned}&\mathrm{sgn}(\chi)=\mathrm{sgn}(\omega)=1,\qquad \mathrm{sgn}(\zeta,\eta,\underline{\eta})=1/2,\\ &\mathrm{sgn}(\underline{\chi})=\mathrm{sgn}(\underline{\omega})=\mathrm{sgn}(\gamma)=0.\end{aligned}$$

Consistent with this definition we have, for any given null component ϕ,

$$\mathrm{sgn}(\nabla_4\phi)=1+\mathrm{sgn}(\phi),\qquad \mathrm{sgn}(\nabla\phi)=\frac{1}{2}+\mathrm{sgn}(\phi),\qquad \mathrm{sgn}(\nabla_3\phi)=\mathrm{sgn}(\phi).$$

Also, based on our convention,

$$\mathrm{sgn}(\phi_1\cdot\phi_2)=\mathrm{sgn}(\phi_1)+\mathrm{sgn}(\phi_2). \tag{26}$$

Remark 4 All terms in a given null structure or null Bianchi identity have the same overall signature. Remark also that the definition (25) applies only to the null components of the curvature tensor and Ricci coefficients and not to their tensor products (i.e. not to null decompositions of tensor products).

We now introduce a notion of scale for any quantity ϕ which has a signature $\mathrm{sgn}(\phi)$, in particular for our basic null curvature quantities $\alpha, \beta, \rho, \sigma, \underline{\beta}, \underline{\alpha}$ and null Ricci coefficients components $\chi, \zeta, \eta, \underline{\eta}, \omega, \underline{\omega}$. This scaling plays a fundamental role in our work.

Definition 2 For an arbitrary horizontal tensor-field ϕ, with a well defined signature $\mathrm{sgn}(\phi)$, we set:

$$\mathrm{sc}(\phi) = -\,\mathrm{sgn}(\phi) + \frac{1}{2}. \tag{27}$$

Observe that $\mathrm{sc}(\nabla_L \phi) = \mathrm{sc}(\phi) - 1$, $\mathrm{sc}(\nabla \phi) = \mathrm{sc}(\phi) - \frac{1}{2}$, $\mathrm{sc}(\nabla_{\underline{L}} \phi) = \mathrm{sc}(\phi)$. For a given product of two horizontal tensor-fields we have,

$$\mathrm{sc}(\phi_1 \cdot \phi_2) = \mathrm{sc}(\phi_1) + \mathrm{sc}(\phi_2) - \frac{1}{2}. \tag{28}$$

3.2 Scale Invariant Norms

For any horizontal tensor-field ψ with scale $\mathrm{sc}(\psi)$ we define the following scale invariant norms along the null hypersurfaces $H = H_u^{(0,\delta)}$ and $\underline{H} = \underline{H}_{\underline{u}}^{(0,1)}$,

$$\|\psi\|_{\mathscr{L}^2_{(sc)}(H)} = \delta^{-\,\mathrm{sc}(\psi)-1} \|\psi\|_{L^2(H)}, \qquad \|\psi\|_{\mathscr{L}^2_{(sc)}(\underline{H})} = \delta^{-\,\mathrm{sc}(\psi)-\frac{1}{2}} \|\psi\|_{L^2(\underline{H})}. \tag{29}$$

We also define the scale invariant norms on the 2 surfaces $S = S_{u,\underline{u}}$,

$$\|\psi\|_{\mathscr{L}^p_{(sc)}(S)} = \delta^{-\,\mathrm{sc}(\psi)-\frac{1}{p}} \|\psi\|_{L^p(S)}. \tag{30}$$

In particular,

$$\|\psi\|_{\mathscr{L}^2_{(sc)}(S)} = \delta^{-\,\mathrm{sc}(\psi)-\frac{1}{2}} \|\psi\|_{L^2(S)}, \qquad \|\psi\|_{\mathscr{L}^\infty_{(sc)}(S)} = \delta^{-\,\mathrm{sc}(\psi)} \|\psi\|_{L^\infty(S)}.$$

Observe that we have,

$$\begin{aligned} \|\psi\|^2_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})} &= \delta^{-1} \int_0^{\underline{u}} \|\psi\|^2_{\mathscr{L}^2_{(sc)}(u,\underline{u}')}\, d\underline{u}', \\ \|\psi\|^2_{\mathscr{L}^2_{(sc)}(\underline{H}_{\underline{u}}^{(0,u)})} &= \int_0^{u} \|\psi\|^2_{\mathscr{L}^2_{(sc)}(u',\underline{u})}\, du'. \end{aligned} \tag{31}$$

We denote the scale invariant L^∞ norm in $\mathscr{D}$ by $\|\psi\|_{\mathscr{L}^\infty_{(sc)}}$.

Remark 5 Observe that the noms above are scale invariant if we take into account the scales of the L^2 noms along H and $\underline{H}$, given by,

$$\mathrm{sc}(\| \, \|_{L^2(H_u^{0,\delta})}) = 1, \qquad \mathrm{sc}(\| \, \|_{L^2(\underline{H}_{\underline{u}}^{0,1})}) = \frac{1}{2}, \qquad \mathrm{sc}(\| \, \|_{L^p(S)}) = \frac{1}{p}.$$

Moreover, they are consistent to the following convention,

$$\nabla_4 \sim \delta^{-1}, \qquad \nabla \sim \delta^{-\frac{1}{2}}, \qquad \nabla_3 \sim 1.$$

In view of (28) all standard product estimates in the usual L^p spaces translate into product estimates in $\mathscr{L}_{(sc)}$ spaces with a gain of $\delta^{1/2}$. Thus, for example,

$$\|\psi_1 \cdot \psi_2\|_{\mathscr{L}^2_{(sc)}(S)} \lesssim \delta^{1/2} \|\psi_1\|_{\mathscr{L}^\infty_{(sc)}(S)} \cdot \|\psi_2\|_{\mathscr{L}^2_{(sc)}(S)} \tag{32}$$

or,

$$\|\psi_1 \cdot \psi_2\|_{\mathscr{L}^2_{(sc)}(H)} \lesssim \delta^{1/2} \|\psi_1\|_{\mathscr{L}^\infty_{(sc)}(H)} \cdot \|\psi_2\|_{\mathscr{L}^2_{(sc)}(H)}.$$

Remark 6 If f is a scalar function constant along the surfaces $S(u, \underline{u}) \subset \mathscr{D}$, we have

$$\|f \cdot \psi\|_{\mathscr{L}^p_{(sc)}(S)} \lesssim \|\psi\|_{\mathscr{L}^p_{(sc)}(S)}$$

or, if f is also bounded on H,

$$\|f \cdot \psi\|_{\mathscr{L}^2_{(sc)}(H)} \lesssim \|\psi\|_{\mathscr{L}^2_{(sc)}(H)}.$$

This remark applies in particular to the constant $\mathrm{tr}\, \underline{\chi}_0 = \frac{4}{2r_0 + \underline{u} - u}$.

We are ready to introduce our main curvature and Ricci coefficient norms:[7]

$$\mathscr{R}_0(u, \underline{u}) := \delta^{1/2} \|\alpha\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})} + \|(\beta, \rho, \sigma, \underline{\beta})\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})},$$

$$\mathscr{R}_1(u, \underline{u}) := \delta^{1/2} \|\nabla_4 \alpha\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})} + \|\nabla(\alpha, \beta, \rho, \sigma, \underline{\beta})\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})},$$

$$\underline{\mathscr{R}}_0(u, \underline{u}) := \delta^{1/2} \|\beta\|_{\mathscr{L}^2_{(sc)}(\underline{H}_{\underline{u}}^{(u,0)})} + \|(\rho, \sigma, \underline{\beta}, \underline{\alpha})\|_{\mathscr{L}^2_{(sc)}(\underline{H}_{\underline{u}}^{(0,u)})},$$

$$\underline{\mathscr{R}}_1(u, \underline{u}) := \|\nabla_3 \underline{\alpha}\|_{\mathscr{L}^2_{(sc)}(\underline{H}_{\underline{u}}^{(u,0)})} + \|\nabla(\beta, \rho, \sigma, \underline{\beta}, \underline{\alpha})\|_{\mathscr{L}^2_{(sc)}(\underline{H}_{\underline{u}}^{(0,u)})}.$$

Remark 7 All curvature norms are scale invariant except for the anomalous $\|\alpha\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})}$, $\|\nabla_4 \alpha\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})}$ and $\|\beta\|_{\mathscr{L}^2_{(sc)}(\underline{H}_{\underline{u}}^{(u,0)})}$. By abuse of language, in a given context, we refer to α, respectively β, as anomalous.

[7] We use the short hand notation $\|(\beta, \rho, \sigma, \underline{\beta})\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})} = \|\beta\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})} + \|\rho\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})} + \|\sigma\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})} + \cdots$.

The Ricci coefficient norms are,

$$
\begin{aligned}
{}^{(S)}\mathscr{O}_{0,\infty}(u,\underline{u}) &= \|(\chi,\omega,\eta,\underline{\eta},\widetilde{\operatorname{tr}\underline{\chi}},\hat{\underline{\chi}},\underline{\omega})\|_{\mathscr{L}^{\infty}_{(sc)}(S)},\\
{}^{(S)}\mathscr{O}_{0,4}(u,\underline{u}) &= \delta^{1/4}\big(\|\hat{\chi}\|_{\mathscr{L}^{4}_{(sc)}(S)}+\|\hat{\underline{\chi}}\|_{\mathscr{L}^{4}_{(sc)}(S)}\big),\\
&\quad + \|(\operatorname{tr}\chi,\omega,\eta,\underline{\eta},\widetilde{\operatorname{tr}\underline{\chi}},\underline{\omega})\|_{\mathscr{L}^{4}_{(sc)}(S)},\\
{}^{(S)}\mathscr{O}_{1,4}(u,\underline{u}) &= \|\nabla(\chi,\omega,\eta,\underline{\eta},\widetilde{\operatorname{tr}\underline{\chi}},\hat{\underline{\chi}},\underline{\omega})\|_{\mathscr{L}^{4}_{(sc)}(S)},\\
{}^{(S)}\mathscr{O}_{1,2}(u,\underline{u}) &= \|\nabla(\chi,\omega,\eta,\underline{\eta},\widetilde{\operatorname{tr}\underline{\chi}},\hat{\underline{\chi}},\underline{\omega})\|_{\mathscr{L}^{2}_{(sc)}(S)},\\
{}^{(H)}\mathscr{O}(u,\underline{u}) &= \|\nabla^{2}(\chi,\omega,\eta,\underline{\eta},\widetilde{\operatorname{tr}\underline{\chi}},\hat{\underline{\chi}},\underline{\omega})\|_{\mathscr{L}^{2}_{(sc)}(H_u^{(0,\underline{u})})}.
\end{aligned}
$$

Remark 8 All quantities are scale invariant except for $\hat{\chi}$, $\hat{\underline{\chi}}$ in the $\mathscr{L}^4_{(sc)}(S)$ norm.

3.3 Main Results

We define the initial data quantity,

$$
\mathscr{I}^{(0)} = \sup_{0\le\underline{u}\le\delta} \mathscr{I}^{(0)}(\underline{u}) \tag{33}
$$

where, with the notation convention in (18),

$$
\begin{aligned}
\mathscr{I}^{(0)}(\underline{u}) &= \delta^{1/2}\|\hat{\chi}_0\|_{L^\infty(0,\underline{u})} + \sum_{0\le k\le 2}\delta^{1/2}\|(\delta\nabla_4)^k\hat{\chi}_0\|_{L^2(0,\underline{u})}\\
&\quad + \sum_{0\le k\le 1}\sum_{1\le m\le 4}\delta^{1/2}\|(\delta^{1/2}\nabla)^{m-1}(\delta\nabla_4)^k\nabla\hat{\chi}_0\|_{L^2(0,\underline{u})}.
\end{aligned}
$$

Our main assumption, replacing Christodoulou's ansatz, is

$$
\mathscr{I}^{(0)} < \infty. \tag{34}
$$

We show that, under this assumption and for sufficiently small $\delta>0$, the space-time slab $\mathscr{D}(u,\delta)$ can be extended for values of $u\ge 1$, with precise estimates for all Ricci coefficients of the double null foliation and null components of the curvature tensor. We can then show, by a slight modification of this assumption together with Christodoulou's lower bound assumption on $\int_0^\delta|\hat{\chi}_0|^2$ (see (14), (15) in [4]), that a trapped surface must form in $\mathscr{D}(u\approx 1,\delta)$. As in the case of [4] most of the work is required to prove the semi global result concerning the double null foliation. Once this is established the actual formation of trapped surfaces result is proved by following the heuristic argument outlined above. In addition we show that a small modification of the regular δ-coherence assumption leads to the formation of a pre-scar.

The first result follows from analyzing assumption (33) on the initial hypersurface H_0.

Proposition 1 *In view of our initial assumption* (34) *we have, for sufficiently small* $\delta > 0$, *along* H_0,

$$\mathscr{R}^{(0)} + \mathscr{O}^{(0)} \lesssim \mathscr{I}^{(0)}.$$

Theorem 3 (Main Theorem) *Assume that* $\mathscr{R}^{(0)} \lesssim \mathscr{I}^{(0)}$ *for an arbitrary constant* $\mathscr{I}^{(0)}$. *Then, there exists a sufficiently small* $\delta > 0$ *such that,*

$$\mathscr{R} + \underline{\mathscr{R}} + \mathscr{O} \lesssim \mathscr{I}^{(0)}. \tag{35}$$

Theorem 4 *Assume that, in addition to* (33), *we also have, for* $2 \le k \le 4$

$$\|(\delta^{\frac{1}{2}}\nabla)^k \hat{\chi}_0\|_{L^2(0,\underline{u})} \le \varepsilon \tag{36}$$

for a sufficiently small parameter ε *such that* $0 < \delta \ll \varepsilon$. *Assume also that* $\hat{\chi}_0$ *verifies* (15). *Then, for* $\delta > 0$ *sufficiently small, a trapped surface must form in the slab* $\mathscr{D}(u \approx 1, \delta)$.

Our next and final result in [8] concerns the formation of a *pre-scar* in an angular sector of size $\delta^{\frac{1}{2}}$.

Theorem 5 *Let* ε *be a small parameter such that* $0 < \delta \ll \varepsilon$. *Assume that the initial data* $\hat{\chi}_0$ *satisfies*

$$\delta^{1/2}\|\hat{\chi}_0\|_{L^\infty} + \sum_{0\le k\le 1}\sum_{0\le m\le 4} \varepsilon\|(\varepsilon^{-1}\delta^{\frac{1}{2}}\nabla)^m(\delta\nabla_4)^k\hat{\chi}_0\|_{L^2(0,\underline{u})} < \infty$$

and that the lower bound in (15) *is verified in angular sector* $\omega \in \Lambda$ *of size* $\delta^{\frac{1}{2}}$. *Then, for* $\delta > 0$ *sufficiently small, a pre-scar must form in the slab* $\mathscr{D}(u \approx 1, \delta)$, *i.e. the expansion scalar* $\operatorname{tr}\chi(u, \underline{u}, \omega)$ *becomes strictly negative for some values of* $u \approx 1$, $\underline{u} = \delta$ *and all* $\omega \in \Lambda$.

4 Main Results of [9], I

The results in [9] are better expressed with respect to different notion of scale invariant norms as discussed above. More precisely, given any horizontal tensor-field

ψ with signature $\mathrm{sgn}(\psi)$ we define the following scale invariant norms along the null hypersurfaces $H = H_u^{(0,\delta)}$ and $\underline{H} = \underline{H}_{\underline{u}}^{(0,1)}$,

$$\|\psi\|_{\mathscr{L}^2_{(sc)}(H)} = \delta^{\mathrm{sgn}(\psi)-1}\|\psi\|_{L^2(H)}, \qquad \|\psi\|_{\mathscr{L}^2_{(sc)}(\underline{H})} = \delta^{\mathrm{sgn}(\psi)-\frac{1}{2}}\|\psi\|_{L^2(\underline{H})}. \quad (37)$$

We also define the scale invariant norms on the 2 surfaces $S = S_{u,\underline{u}}$,

$$\|\psi\|_{\mathscr{L}^p_{(sc)}(S)} = \delta^{\mathrm{sgn}(\psi)-\frac{1}{p}}\|\psi\|_{L^p(S)}. \quad (38)$$

We have,

$$\begin{aligned} \|\psi\|^2_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})} &= \delta^{-1}\int_0^{\underline{u}} \|\psi\|^2_{\mathscr{L}^2_{(sc)}(u,\underline{u}')} d\underline{u}', \\ \|\psi\|^2_{\mathscr{L}^2_{(sc)}(\underline{H}_{\underline{u}}^{(0,u)})} &= \int_0^{u} \|\psi\|^2_{\mathscr{L}^2_{(sc)}(u',\underline{u})} du'. \end{aligned} \quad (39)$$

We denote the scale invariant L^∞ norm in $\mathscr{D}$ by $\|\psi\|_{\mathscr{L}^\infty_{(sc)}}$.

Remark 9 As mentioned above, these norms are different than those of [8], discussed in the previous section. Indeed in [8] the scale invariant norms were based on the definition of the scale of a horizontal component of scale $\mathrm{sc}(\psi) = -\,\mathrm{sgn}(\psi) + \frac{1}{2}$. The norms introduced here would correspond to a new definition of scale give by $\mathrm{sc}(\psi) = -\,\mathrm{sgn}(\psi)$. To distinguish between them we denote the old scaling by $\dot{\mathrm{sc}}$. Thus, for example,

$$\|\psi\|_{\mathscr{L}^p_{(sc)}(S)} = \delta^{-1/2}\|\psi\|_{\mathscr{L}^p_{(\dot{sc})}(S)}.$$

Remark 10 With the new scale invariant norms introduced here we have,

$$\|\psi_1 \cdot \psi_2\|_{\mathscr{L}^2_{(sc)}(S)} \lesssim \|\psi_1\|_{\mathscr{L}^\infty_{(sc)}(S)} \cdot \|\psi_2\|_{\mathscr{L}^2_{(sc)}(S)} \quad (40)$$

or,

$$\|\psi_1 \cdot \psi_2\|_{\mathscr{L}^2_{(sc)}(H)} \lesssim \|\psi_1\|_{\mathscr{L}^\infty_{(sc)}(H)} \cdot \|\psi_2\|_{\mathscr{L}^2_{(sc)}(H)}.$$

These differ from the situation in [8] where the corresponding estimates (with (sc) replaced by ($\dot{\mathrm{sc}}$)) had an additional power of $\delta^{1/2}$ on the right.

Curvature Norms We introduce our main curvature norms

$$
\begin{aligned}
\mathscr{R}_0(u,\underline{u}) &:= \|\alpha\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})} + \mathscr{R}_0'(u,\underline{u}'),\\
\mathscr{R}_0'(u,\underline{u}') &:= \varepsilon^{-1}\|(\beta,\rho,\sigma,\underline{\beta},K)\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})},\\
\mathscr{R}_1(u,\underline{u}) &:= \|\nabla_4\alpha\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})} + \mathscr{R}_1'(u,\underline{u}),\\
\mathscr{R}_1'(u,\underline{u}) &:= \varepsilon^{-1}\|\nabla(\alpha,\beta,\rho,\sigma,\underline{\beta},K)\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})},\\
\underline{\mathscr{R}}_0(u,\underline{u}) &:= \|\beta\|_{\mathscr{L}^2_{(sc)}(\underline{H}_{\underline{u}}^{(u,0)})} + \underline{\mathscr{R}}_0'(u,\underline{u}'),\\
\underline{\mathscr{R}}_0'(u,\underline{u}) &= \varepsilon^{-1}\|(\rho,\sigma,\underline{\beta},\underline{\alpha},K)\|_{\mathscr{L}^2_{(sc)}(\underline{H}_{\underline{u}}^{(0,u)})},\\
\underline{\mathscr{R}}_1(u,\underline{u}) &:= \|\nabla_3\underline{\alpha}\|_{\mathscr{L}^2_{(sc)}(\underline{H}_{\underline{u}}^{(u,0)})} + \underline{\mathscr{R}}_1'(u,\underline{u}),\\
\underline{\mathscr{R}}_1'(u,\underline{u}) &:= \varepsilon^{-1}\|\nabla(\beta,\rho,\sigma,\underline{\beta},\underline{\alpha},K)\|_{\mathscr{L}^2_{(sc)}(\underline{H}_{\underline{u}}^{(0,u)})}.
\end{aligned}
\tag{41}
$$

Also,

$$
\mathscr{R} = \mathscr{R}_0 + \mathscr{R}_1, \qquad \underline{\mathscr{R}} = \underline{\mathscr{R}}_0 + \underline{\mathscr{R}}_1. \tag{42}
$$

Remark 11 We have included the Gauss curvature K with the null components. Since $K = -\rho + \frac{1}{2}\hat{\chi}\cdot\underline{\hat{\chi}} - \frac{1}{4}\operatorname{tr}\chi\operatorname{tr}\underline{\chi}$ we easily deduce that,

$$
\varepsilon^{-1}\|K\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})} \lesssim \varepsilon^{-1}\|\rho\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})} + \big(1 + (\varepsilon^{-2}\delta)^{\frac{1}{2}}\big)\,{}^{(S)}\mathscr{O}_{0,\infty}\,{}^{(S)}\mathscr{O}_{0,2}.
$$

Remark 12 All curvature norms above have a factor of ε^{-1} in front of them except for $\|\alpha\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})}$, $\|\nabla_4\alpha\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})}$ and $\|\beta\|_{\mathscr{L}^2_{(sc)}(\underline{H}_{\underline{u}}^{(u,0)})}$. These correspond exactly to the *anomalous* curvature norms of [8].

To rectify the anomaly of α we introduce, as in [8], an additional scale-invariant norm,

$$
\mathscr{R}_0^{(\varepsilon)}[\alpha](u,\underline{u}) := \sup_{{}^{(\varepsilon)}H\subset H} \varepsilon^{-1}\|\alpha\|_{\mathscr{L}^2_{(sc)}({}^{(\varepsilon)}H)},
$$

where ${}^{(\varepsilon)}H$ is a piece of the hypersurface $H = H_u^{(0,\delta)}$ obtained by evolving an angular disc $S_\varepsilon \subset S_{u,0}$ of radius ε relative to our transported coordinates. We define the initial quantity $\mathscr{R}^{(0)}$ by,

$$
\mathscr{R}^{(0)} = \sup_{0\le\underline{u}\le\delta}\big(\mathscr{R}(0,\underline{u}) + \mathscr{R}_0^{(\varepsilon)}[\alpha](0,\underline{u})\big). \tag{43}
$$

4.1 Connection Coefficients Norms

We introduce the Ricci coefficient norms, with the supremum taken over all surfaces $S = S(u', \underline{u}'), 0 \le u' \le u,\ 0 \le \underline{u}' \le \underline{u}$,

$$\begin{aligned}
{}^{(S)}\mathscr{O}_{0,\infty}(u,\underline{u}) &= \varepsilon^{-1} \sup_S \|(\hat{\chi}, \omega, \eta, \underline{\eta}, \widetilde{\operatorname{tr}\underline{\chi}}, \hat{\underline{\chi}}, \underline{\omega})\|_{\mathscr{L}^\infty_{(sc)}(S)},\\
{}^{(S)}\mathscr{O}_{0,2}(u,\underline{u}) &= \sup_S \big(\|\hat{\chi}\|_{\mathscr{L}^2_{(sc)}(S)} + \|\hat{\underline{\chi}}\|_{\mathscr{L}^2_{(sc)}(S)}\big) + {}^{(S)}\mathscr{O}'_{0,2}(u,\underline{u}),\\
{}^{(S)}\mathscr{O}'_{0,2}(u,\underline{u}) &= \varepsilon^{-1} \sup_S \|(\operatorname{tr}\chi, \omega, \eta, \underline{\eta}, \widetilde{\operatorname{tr}\underline{\chi}}, \underline{\omega})\|_{\mathscr{L}^2_{(sc)}(S)},\\
{}^{(S)}\mathscr{O}_{0,4}(u,\underline{u}) &= \varepsilon^{-1/2} \sup_S \big(\|\hat{\chi}\|_{\mathscr{L}^4_{(sc)}(S)} + \|\hat{\underline{\chi}}\|_{\mathscr{L}^4_{(sc)}(S)}\big) + {}^{(S)}\mathscr{O}'_{0,4}(u,\underline{u}),\\
{}^{(S)}\mathscr{O}'_{0,4}(u,\underline{u}) &= \varepsilon^{-1} \sup_S \|(\operatorname{tr}\chi, \omega, \eta, \underline{\eta}, \widetilde{\operatorname{tr}\underline{\chi}}, \underline{\omega})\|_{\mathscr{L}^4_{(sc)}(S)},\\
{}^{(S)}\mathscr{O}_{1,4}(u,\underline{u}) &= \varepsilon^{-1} \sup_S \|\nabla(\chi, \omega, \eta, \underline{\eta}, \widetilde{\operatorname{tr}\underline{\chi}}, \hat{\underline{\chi}}, \underline{\omega})\|_{\mathscr{L}^4_{(sc)}(S)},\\
{}^{(S)}\mathscr{O}_{1,2}(u,\underline{u}) &= \varepsilon^{-1} \sup_S \|\nabla(\chi, \omega, \eta, \underline{\eta}, \widetilde{\operatorname{tr}\underline{\chi}}, \hat{\underline{\chi}}, \underline{\omega})\|_{\mathscr{L}^2_{(sc)}(S)},\\
{}^{(H)}\mathscr{O}(u,\underline{u}) &= \varepsilon^{-1} \|\nabla^2(\chi, \omega, \eta, \underline{\eta}, \widetilde{\operatorname{tr}\underline{\chi}}, \hat{\underline{\chi}}, \underline{\omega})\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})}
\end{aligned} \tag{44}$$

and,

$$\mathscr{O} = {}^{(S)}\mathscr{O}_{0,2} + {}^{(S)}\mathscr{O}_{0,4} + {}^{(S)}\mathscr{O}_{0,\infty} + {}^{(S)}\mathscr{O}_{1.4} + {}^{(H)}\mathscr{O}. \tag{45}$$

Remark 13 Note that the only norms which do not contain powers of ε^{-1} are the $\mathscr{L}^2_{(sc)}(S)$ norms of $\hat{\chi}$ and $\hat{\underline{\chi}}$. This anomaly is also manifest in the $\mathscr{L}^4_{(sc)}(S)$ norms of the same quantities. These are precisely the same quantities which were anomalous in [8], with respect to the $\dot{sc}$ scaling.

To cure the above anomaly we define the auxiliary norms,

$${}^{(S)}\mathscr{O}^{(\varepsilon)}_{0,4}(u,\underline{u}) = \varepsilon^{-1} \sup_S \sup_{S_\varepsilon \subset S} \|(\hat{\chi}, \hat{\underline{\chi}})\|_{\mathscr{L}^4_{(sc)}(S_\varepsilon)}$$

with S_ε—an angular subset of S of size ε relative to our transported coordinates.

Finally we define the initial data quantity:

$$\mathscr{O}^{(0)} = \sup_{0 \le \underline{u} \le \delta} \big(\mathscr{O}(0,\underline{u}) + {}^{(S)}\mathscr{O}^{(\varepsilon)}_{0,4}(0,\underline{u})\big). \tag{46}$$

4.2 Initial Conditions

Define the main initial data quantity,

$$
\begin{aligned}
\mathscr{I}^{(0)}(\underline{u}) = &\sum_{0\le k\le 2} \|\nabla_4^k \hat{\chi}_0\|_{\mathscr{L}^2_{(sc)}(0,\underline{u})} \\
&+ \varepsilon^{-1}\Big(\|\hat{\chi}_0\|_{\mathscr{L}^\infty_{(sc)}(0,\underline{u})} + \sum_{0\le k\le 1}\sum_{1\le m\le 4} \|\nabla^{m-1}\nabla_4^k \nabla\hat{\chi}_0\|_{\mathscr{L}^2_{(sc)}(0,\underline{u})}\Big)
\end{aligned}
\tag{47}
$$

or, in the natural norms,

$$
\begin{aligned}
\mathscr{I}^{(0)}(\underline{u}) = &\sum_{0\le k\le 2} \delta^{k+1/2} \|\nabla_4^k \hat{\chi}_0\|_{L^2(0,\underline{u})} \\
&+ \varepsilon^{-1}\Big(\delta\|\hat{\chi}_0\|_{L^\infty(0,\underline{u})} + \sum_{0\le k\le 1}\sum_{1\le m\le 4} \delta^{\frac{m+1}{2}+k} \|\nabla^{m-1}\nabla_4^k \nabla\hat{\chi}_0\|_{L^2(0,\underline{u})}\Big).
\end{aligned}
$$

4.3 Main Propagation Result

The first result establishes the boundedness of the initial curvature and Ricci coefficient scale invariant norms $\mathscr{R}^{(0)}$, $\mathscr{O}^{(0)}$ in terms of $\mathscr{I}^{(0)}$.

Proposition 2 *Assume that the initial data along $\underline{H}_0$ is flat and that $\mathscr{I}^{(0)} < \infty$ along $H_0^{(0,\delta)}$. Then, for $\delta^{1/2}\varepsilon^{-1}$ and $\varepsilon > 0$ sufficiently small we have, with C a fixed super-linear polynomial*

$$
\mathscr{R}^{(0)} + \mathscr{O}^{(0)} \lesssim \mathscr{I}^{(0)} + C(\mathscr{I}^{(0)}).
$$

Also, starting with $\mathscr{R}^{(0)} < \infty$ and $\delta^{1/2}\varepsilon^{-1}$, ε sufficiently small, we have, with C a fixed super-linear polynomial,

$$
\mathscr{O}^{(0)} \lesssim \mathscr{R}^{(0)} + C(\mathscr{R}^{(0)}). \tag{48}
$$

We can now state our main propagation result.

Theorem 6 (Main Theorem I) *Under the assumption $\mathscr{R}^{(0)} < \infty$, if $\delta^{1/2}\varepsilon^{-1}$ and ε are sufficiently small then, for $0 \le u \le 1$, $0 \le \underline{u} \le \delta$, with C a fixed super-linear polynomial,*

$$
(\mathscr{R} + \underline{\mathscr{R}} + \mathscr{O})(u,\underline{u}) \lesssim \mathscr{R}^{(0)} + C(\mathscr{R}^{0}).
$$

Remark 14 The results presented extend all the results of [8], discussed in the previous section. Indeed, to derive the results of Propositions 2.5, Theorems 2.6, and 2.7 there, it suffices to choose $\varepsilon = \mu\delta^{1/2}$ with μ sufficiently small.

Remark 15 The additional smallness assumption on $\delta^{1/2}\varepsilon^{-1}$ is due to the lower order terms which appear in some of the calculus inequalities presented in the next section.

5 Main results of [9], II

Relying on the results of Theorem 6 we state a second result of [9] concerning the formation of pre-scars. Throughout this section we assume that the assumptions and conclusions of Theorem 6 hold true.

5.1 Local Scale Invariant Norms

Consider a partition of $S_0 = S(0,0)$ into angular sectors Λ of a given size $|\Lambda|$. Let ${}^{(\Lambda)}f_{(0)}$ be a partition of unity associated to this partition. They can be extend trivially, first along $\underline{H}_0$ and then along each H_u, to be constant along the corresponding null generators. In particular we have,

$$\nabla_L {}^{(\Lambda)}f = 0, \qquad {}^{(\Lambda)}f|_{\underline{H}_0} = {}^{(\Lambda)}f_{(0)}. \tag{49}$$

Then, under the assumptions and conclusions of Theorem 6 we can easily deduce,

Lemma 1 *We have*,

$$\sum_{\Lambda} {}^{(\Lambda)}f = 1. \tag{50}$$

Also,

$$|\nabla^{(\Lambda)}f|_{L^\infty} \lesssim |\Lambda|^{-1}, \qquad |\nabla_{\underline{L}}^{(\Lambda)}f|_{L^\infty} \lesssim \varepsilon\delta^{1/2}|\Lambda|^{-1} \tag{51}$$

or, in scale invariant norms (*assigning to* f *signature* 0),

$$|\nabla^{(\Lambda)}f|_{\mathscr{L}^\infty_{(sc)}} \lesssim \delta^{1/2}|\Lambda|^{-1}, \qquad |\nabla_{\underline{L}}{}^{(\Lambda)}f|_{\mathscr{L}^\infty_{(sc)}} \lesssim \varepsilon\delta^{1/2}|\Lambda|^{-1}.$$

We now introduce the localized curvature norms,

$$
\begin{aligned}
{}^{(\Lambda)}\mathscr{R}_0(u,\underline{u}) &:= \|{}^{(\Lambda)}f\alpha\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})} + {}^{(\Lambda)}\mathscr{R}_0'(u,\underline{u}'),\\
{}^{(\Lambda)}\mathscr{R}_0'(u,\underline{u}) &:= \varepsilon^{-1}\|{}^{(\Lambda)}f(\beta,\rho,\sigma,\underline{\beta},K)\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})},\\
{}^{(\Lambda)}\mathscr{R}_1(u,\underline{u}) &:= \|{}^{(\Lambda)}f\nabla_4\alpha\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})} + {}^{(\Lambda)}\mathscr{R}_1'(u,\underline{u}),\\
{}^{(\Lambda)}\mathscr{R}_1'(u,\underline{u}) &:= \varepsilon^{-1}\|{}^{(\Lambda)}f\nabla(\alpha,\beta,\rho,\sigma,\underline{\beta},K)\|_{\mathscr{L}^2_{(sc)}(H_u^{(0,\underline{u})})},\\
{}^{(\Lambda)}\underline{\mathscr{R}}_0(u,\underline{u}) &:= \|{}^{(\Lambda)}f\beta\|_{\mathscr{L}^2_{(sc)}(\underline{H}_{\underline{u}}^{(u,0)})} + \underline{\mathscr{R}}_0'(u,\underline{u}'),\\
{}^{(\Lambda)}\underline{\mathscr{R}}_0'(u,\underline{u}) &= \varepsilon^{-1}\|{}^{(\Lambda)}f(\rho,\sigma,\underline{\beta},\underline{\alpha},K)\|_{\mathscr{L}^2_{(sc)}(\underline{H}_{\underline{u}}^{(0,u)})},\\
{}^{(\Lambda)}\underline{\mathscr{R}}_1(u,\underline{u}) &:= \|{}^{(\Lambda)}f\nabla_3\underline{\alpha}\|_{\mathscr{L}^2_{(sc)}(\underline{H}_{\underline{u}}^{(u,0)})} + \underline{\mathscr{R}}_1'(u,\underline{u}),\\
\underline{\mathscr{R}}_1'(u,\underline{u}) &:= \varepsilon^{-1}\|{}^{(\Lambda)}f\nabla(\beta,\rho,\sigma,\underline{\beta},\underline{\alpha},K)\|_{\mathscr{L}^2_{(sc)}(\underline{H}_{\underline{u}}^{(0,u)})}
\end{aligned}
\tag{52}
$$

and,

$$
\begin{aligned}
{}^{[\Lambda]}\mathscr{R}_0(u,\underline{u}) &:= \sup_\Lambda {}^{(\Lambda)}\mathscr{R}_0, &\quad {}^{[\Lambda]}\mathscr{R}_1(u,\underline{u}) &:= \sup_\Lambda {}^{(\Lambda)}\mathscr{R}_1,\\
{}^{[\Lambda]}\underline{\mathscr{R}}_0(u,\underline{u}) &:= \sup_\Lambda {}^{(\Lambda)}\underline{\mathscr{R}}_0, &\quad {}^{[\Lambda]}\underline{\mathscr{R}}_1(u,\underline{u}) &:= \sup_\Lambda {}^{(\Lambda)}\underline{\mathscr{R}}_1
\end{aligned}
\tag{53}
$$

with the supremum taken with respect to all elements of the partition and,

$$
{}^{[\Lambda]}\mathscr{R} = {}^{[\Lambda]}\mathscr{R}_0 + {}^{[\Lambda]}\mathscr{R}_1, \qquad {}^{[\Lambda]}\underline{\mathscr{R}} = {}^{[\Lambda]}\underline{\mathscr{R}}_0 + {}^{[\Lambda]}\underline{\mathscr{R}}_1. \tag{54}
$$

5.2 Angular Localized Curvature Estimates

Using a variation of our main energy estimates, with an additional angular localization, we can prove the following.

Theorem 7 *Under the assumptions and conclusions of Theorem 6, if in addition* $\delta^{\frac{1}{2}}|\Lambda|^{-1}$ *is sufficiently small, then, for* $0 \le u \le 1$, $0 \le \underline{u} \le \delta$,

$$
({}^{[\Lambda]}\mathscr{R} + {}^{[\Lambda]}\underline{\mathscr{R}})(u,\underline{u}) \lesssim {}^{[\Lambda]}\mathscr{R}^{(0)}.
$$

Moreover,

$$
({}^{(\Lambda)}\mathscr{R} + {}^{(\Lambda)}\underline{\mathscr{R}})(u,\underline{u}) \lesssim {}^{(\Lambda)}\mathscr{R}^{(0)} + \delta^{\frac{1}{2}}|\Lambda|^{-1\,[\Lambda]}\mathscr{R}^{(0)}. \tag{55}
$$

Remark 16 By the standard domain of dependence argument the energy estimate cannot fully localized to individual sectors ${}^{(\Lambda)}H_u$ and ${}^{(\Lambda)}\underline{H}_{\underline{u}}$ contained in the support of the function ${}^{(\Lambda)}f$. This explains the need for the supremum in Λ in the definition of the ${}^{[\Lambda]}\mathscr{R}$, ${}^{[\Lambda]}\underline{\mathscr{R}}$ norms for the first part of the theorem. The second part of the theorem gives a bound for each sector individual Λ with the second term on the right hand side of (55) accounting for the defect of localization.

5.3 Emerging Scars

Definition 3 We say that the data $\mathscr{R}^{(0)}$ is uniformly distributed on the scale $\delta^{\frac{1}{2}}\varpi^{-1}$ if there exists a partition $\{\Lambda\}$ such that $|\Lambda| \approx \delta^{\frac{1}{2}}\varpi^{-1}$ and

$$ {}^{[\Lambda]}\mathscr{R}^{(0)} \lesssim \delta^{\frac{1}{2}}\varpi^{-1}\mathscr{R}^{(0)}. \tag{56} $$

Our second main result of this paper is the following.

Theorem 8 (Main Theorem II) *Assume that, in additions to the conditions of validity of Theorem 6, the data $\mathscr{R}^{(0)}$ is uniformly distributed on the scale $\delta^{\frac{1}{2}}\varpi^{-1}$ for some constant $\varpi \ll 1$ and $\varepsilon\varpi^{-1}$ sufficiently small. Let Λ be a fixed angular sector of size $|\Lambda| = q^{-1}\delta^{\frac{1}{2}}$ with $q = \varepsilon\varpi^{-1}$ sufficiently small. Then, if*

$$ \inf_{\theta\in\Lambda}\int_0^\delta |\hat{\chi}_0|^2(\underline{u},\theta)\,d\underline{u} > \frac{2(r_0-u)}{r_0^2} \tag{57} $$

the Λ-angular section ${}^{(\Lambda)}S_{u,\delta}$ of the surface $S_{u,\delta}$ must be trapped, i.e. $\operatorname{tr}\chi < 0$ *there.*

Alternatively, if for some constant $c > 0$ independent of $\delta, \varepsilon, q, \varpi$,

$$ \sup_{\theta\in\Lambda}\int_0^\delta |\hat{\chi}_0|^2(\underline{u},\theta)\,d\underline{u} < \frac{2(r_0-u)}{r_0^2} - c \tag{58} $$

then $\operatorname{tr}\chi > 0$ *throughout the angular sector ${}^{(\Lambda)}S_{u,\delta}$.*

Remark 17 Observe that the parameters $\delta, \varepsilon, \varpi$ in Theorem 8 verify the conditions:

$$ 0 < \delta^{1/2} < \varepsilon < \varpi < 1, \qquad \delta^{1/2}\varepsilon^{-1} \ll 1, \qquad q = \varepsilon\varpi^{-1} \ll 1. $$

References

1. Bruhat, Y.C.: Théorème d'existence pour certains systémes d'equations aux dérivées partielles non-linéaires. Acta Math. **88**, 141–225 (1952)
2. Bruhat, Y.C., Geroch, R.: Global aspects of the Cauchy problem in GR. Commun. Math. Phys. **14**, 329–335 (1969)

3. Christodoulou, D.: The formation of black holes and singularities in spherically symmetric gravitational collapse. Commun. Pure Appl. Math. **44**(3), 339–373 (1991)
4. Christodoulou, D.: The Formation of Black Holes in General Relativity. Monographs in Mathematics. Eur. Math. Soc., Zurich (2009)
5. Christodoulou, D., Klainerman, S.: The Global Nonlinear Stability of the Minkowski Space. Princeton Mathematical Series, vol. 41 (1993)
6. Hawking, S., Ellis, G.F.R.: The Large Scale Structure of Spacetime. Cambridge University Press, Cambridge (1973)
7. Klainerman, S., Nicolo, F.: The Evolution Problem in General Relativity. Progress in Mathematical Physics. Birkhäuser, Basel (2002)
8. Klainerman, S., Rodnianski, I.: On the formation of trapped surfaces. Acta Math. (2011, in press)
9. Klainerman, S., Rodnianski, I.: On emerging scarred surfaces for the Einstein vacuum equations. Discrete Contin. Dyn. Syst. **28**(3), 1007–1031 (2010)
10. Penrose, R.: Gravitational collapse: the role of general relativity. Nuovo Cimento **1**, 252–276 (1969)
11. Penrose, R.: Singularities and time-asymmetry. In: Hawking, S., Israel, W. (eds.) General Relativity: An Einstein Centenary Survey. Cambridge University Press, Cambridge (1979)

Surface Relaxation Below the Roughening Temperature: Some Recent Progress and Open Questions

Robert V. Kohn

Abstract We discuss two recent projects concerning the evolution of a crystal surface below the roughening temperature. One addresses the evolution of a monotone one-dimensional step train (joint work with Hala Al Hajj Shehadeh and Jonathan Weare). The other addresses the finite-time flattening predicted by a fourth-order PDE model (joint work with Yoshikazu Giga). For each project we begin with a discussion of the mathematical model; then we summarize the recent results, the main ideas behind them, and some related open problems.

1 Introduction

The surface of a crystal below its roughening temperature consists of steps and terraces, with facets at the peaks and valleys. The steps interact and collide, and if the average slope is zero then surface relaxes to a perfectly flat state (a single facet). This process can be modeled at the atomic scale using a kinetic Monte Carlo scheme. But Monte Carlo is limited to relatively small length and time scales. To understand and simulate the macroscopic consequences of relaxation, it is therefore attractive to use either

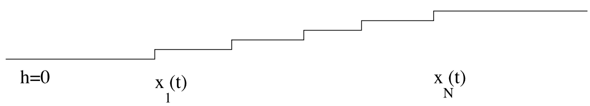
- *step evolution laws*, which ignore atomic-scale fluctuations and track instead the positions of steps; or
- *partial differential equations*, which model the evolution on a coarser, continuum scale.

For a given physical system there is often considerable uncertainty which approach is best. Therefore it is important to understand the models' qualitative predictions. Exploration of this issue leads to a wide array of mathematical questions, many of which remain open. The present article summarizes recent progress on two problems of this type:

R.V. Kohn (✉)
Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10023, USA
e-mail: kohn@cims.nyu.edu

H. Holden, K.H. Karlsen (eds.), *Nonlinear Partial Differential Equations*, Abel Symposia 7, DOI 10.1007/978-3-642-25361-4_11,

Fig. 1 A monotone one-dimensional step train



(a) work with H. Al Hajj Shehadeh and J. Weare on the asymptotic self-similarity of a monotone one-dimensional step train in the attachment-detachment-limited (ADL) regime [1], and
(b) work with Y. Giga on the finite-time flattening predicted by a fourth-order PDE describing relaxation by surface diffusion in the diffusion-limited (DL) regime [8].

We also highlight a number of open questions of an essentially mathematical character, which we think may be ripe for progress. For broader surveys of the physics of crystal growth and relaxation we refer to the recent monographs [18, 23] and the review article [12].

2 A Monotone One-Dimensional Step Train in the ADL Regime

For a two-dimensional surface with peaks and valleys, the analysis of step motion is complicated because (i) each step is a curve with a different shape, and (ii) the steps collide at peaks and valleys. Therefore it is natural to start with a simpler problem: a one-dimensional step train connecting two semi-infinite terraces as in Fig. 1. We take the convention that there are N steps, each of height $1/N$. The semi-infinite terrace to the left of x_1 has height 0, and the one to the right of x_N has height 1.

After discussing the steps' motion law, we shall explain why the evolution is asymptotically self-similar as $t \to \infty$, and we'll discuss the anticipated behavior in the continuum limit $N \to \infty$.

2.1 The Step Equations

Ozdemir and Zangwill [22] were, it seems, the first to identify the step velocities in this setting by specializing the framework of Burton, Cabrera, and Frank [3] (a special case was considered a few years earlier by Rettori and Villain [24]). Their calculation is reviewed e.g. in [19] and in Appendix A of [1]. Away from the extremes, the velocity of the ith step is

$$\dot{x}_i = c\left(\frac{\mu_{i+1} - \mu_i}{(x_{i+1} - x_i) + 2D/k} - \frac{\mu_i - \mu_{i-1}}{(x_i - x_{i-1}) + 2D/k}\right) \tag{1}$$

where c is a (dimensional) constant,

$$\mu_i = (x_{i+1} - x_i)^{-3} - (x_i - x_{i-1})^{-3}, \tag{2}$$

and D/k is the ratio of the terrace diffusion constant D to a kinetic coefficient k with the dimensions of length/time. For the second step ($i=2$) and the second-last step ($i=N-1$) the velocity is still given by (1), but the definitions of μ_1 and μ_{N-1} are

$$\mu_1 = (x_2 - x_1)^{-3}, \qquad \mu_{N-1} = -(x_N - x_{N-1})^{-3}. \tag{3}$$

For the first and last step, the velocities are

$$\dot{x}_1 = c\frac{\mu_2 - \mu_1}{(x_2 - x_1) + 2D/k}, \qquad \dot{x}_N = -c\frac{\mu_N - \mu_{N-1}}{(x_N - x_{N-1}) + 2D/k}. \tag{4}$$

Note that (3) is consistent with (2) if we take $x_0 = -\infty$ and $x_{N+1} = \infty$. The formulas for $\dot{x}_1$ and $\dot{x}_N$ have just one term rather than two, because the extreme steps receive diffusing atoms only from one side.

2.2 A Convenient Reorganization

As Ozdemir and Zangwill pointed out, it is convenient to consider the discrete slopes

$$u_i = \frac{1/N}{x_{i+1} - x_i}, \quad 1 \le i \le N-1, \tag{5}$$

and to view the second order difference operator

$$\Delta_i \xi = N^2(\xi_{i+1} - 2\xi_i + \xi_{i-1}) \tag{6}$$

as the discrete Laplacian with respect to height. Then the step equations (1)–(4) reduce by algebraic manipulation to

$$\dot{u}_i = -cNu_i^2 \Delta_i \left[\frac{u}{1 + \frac{2DN}{k}u}\Delta u^3\right], \quad 1 \le i \le N-1, \tag{7}$$

with the conventions

$$u_0 = u_N = 0 \quad \text{and} \quad \Delta_0 u^3 = \Delta_N u^3 = 0. \tag{8}$$

2.3 The DL and ADL Regimes

The ratio D/k has the dimensions of length. So does $1/N$, since it is the height of a single step. Therefore DN/k is a dimensionless ratio. When it is large, diffusion across terraces is rapid, and the attachment or detachment of atoms at steps sets the timescale of step motion; this is called the *attachment-detachment limited* (ADL) regime. Conversely, when DN/k is small, attachment or detachment at steps is rapid, and it is the diffusion of atoms across terraces that sets the timescale of step motion; this is *diffusion-limited* (DL) regime. Evidently, the step motion laws reduce in these two limits (after an appropriate rescaling of time) to

$$\dot{u}_i = -u_i^2 \Delta_i[\Delta(u^3)] \quad \text{in the ADL regime, and} \tag{9}$$

$$\dot{u}_i = -u_i^2 \Delta_i[u\Delta(u^3)] \quad \text{in the DL regime.} \tag{10}$$

2.4 Self-similar Asymptotics in the ADL Regime

Ozdemir and Zangwill observed numerically that as $t \to \infty$, solutions of the ADL evolution law (9) are asymptotically self-similar. Our recent paper [1] provides the first rigorous analysis, showing that:

- There is a unique positive self-similar solution of (9), which by definition has the form

$$u_i(t) = t^{-1/4}\phi_i \tag{11}$$

  with $(\phi_1, \dots, \phi_{N-1})$ independent of t.
- A solution of (9) with positive initial data remains positive for all t, and is asymptotically self-similar in the sense that

$$t^{1/4}u_i(t) \to \phi_i \quad \text{as } t \to \infty. \tag{12}$$

- The self-similar profile $(\phi_1, \dots, \phi_{N-1})$ is the unique positive minimizer of

$$SE_N[w] = \sum_{i=1}^{N-1} -\frac{1}{8}w_i^2 + \frac{1}{6}[\Delta_i(w^3)]^2 \tag{13}$$

  subject to $w_0 = w_N = 0$. Moreover, the graph of ϕ_i^3 is concave in the sense that $\Delta_i(\phi^3) < 0$ for $1 \le i \le N-1$.

The key to these results is the observation that the ADL evolution law (9) is L^2 steepest descent for

$$E_N = \sum_{i=1}^{N-1} \frac{1}{6}(\Delta_i u^3)^2 \tag{14}$$

subject to $u_0 = u_N = 0$. Changing to "similarity variables," i.e. writing

$$u_i(t) = t^{-1/4}w_i(s) \quad \text{with } s = \log t, \tag{15}$$

one finds that the slope evolution (9) is equivalent to

$$\frac{dw_i}{ds} = \frac{1}{4}w_i - w_i^2 \Delta_i \Delta(w^3), \tag{16}$$

which amounts to L^2 steepest descent for the functional SE_N defined by (13). Now, SE_N has a unique positive critical point, because if we write $v_i = w_i^3$ then each of its two terms is a convex function of v. In addition, one can show (using energy-type inequalities) that the solution "in similarity variables" $w_i(s)$ remains strictly positive and bounded for all time. Asymptotic self-similarity follows, since the ω-limit set of a steepest descent consists of critical points of its Lyapunov functional.

2.5 The Continuum Limit of the ADL Law

Recall from (6) that Δ_i is the centered finite-difference Laplacian with respect to height. Therefore the continuum analogue of the ADL evolution law (9) is (at least formally)

$$u_t = -u^2(u^3)_{hhhh} \quad \text{for } 0 < h < 1, \tag{17}$$

with boundary conditions $u(t,0) = u(t,1) = 0$ and $(u^3)_{hh}(t,0) = (u^3)_{hh}(t,1) = 0$. This PDE is degenerate-parabolic wherever u approaches 0, which must happen at least at the endpoints $h = 0, 1$. We do not know whether (17) has a well-defined, unique solution (to which our discrete evolution should presumably converge as $N \to \infty$).

The continuum self-similar solution is, however, more accessible: since the discrete self-similar solution minimizes the discrete "energy" defined by (13), the continuum self-similar solution should minimize its continuous analogue

$$SE[w] = \int_0^1 -\frac{1}{8}w^2 + \frac{1}{6}[(w^3)_{hh}]^2 \, ds \tag{18}$$

subject to $w(0) = w(1) = 0$. We studied this variational problem in [1], showing that

(i) The functional SE has a unique positive minimizer ϕ. Moreover, ϕ^3 is $C^{3,\alpha}$ on the closed interval [0, 1], and it vanishes linearly at each endpoint in the sense that $(\phi^3)_h(0) = -(\phi^3)_h(1) > 0$. In addition, ϕ^3 is a concave function of h.
(ii) The discrete self-similar solution ϕ_N converges to the continuous one ϕ as $N \to \infty$, with (roughly speaking) $\|(\phi_N^3)_{hh} - (\phi^3)_{hh}\|_{L^2} \le CN^{-5/6}$.

The arguments for (i) are parallel to those used in the discrete setting. The $C^{3,\alpha}$ regularity of ϕ^3 comes from the Euler–Lagrange equation (the continuum analogue of (16) specialized to a steady state)

$$(\phi^3)_{hhhh} = \frac{1}{4}\phi^{-1},$$

combined with the concavity of ϕ^3 and its linear behavior near $h = 0, 1$. The arguments for (ii) use the convexity of SE when viewed as a function of $v = w^3$.

We suppose the continuum limit of the time-dependent evolution should resemble that of the similarity solution; in particular, we suppose the solution of (17) should have the property that u^3 is $C^{3,\alpha}$ up to the boundary, vanishing linearly near $h = 0, 1$. Since u gives the slope as a function of height and time, one must integrate to get the height as a function of space and time. The length of the region occupied by the steps is

$$x|_{h=1} - x|_{h=0} = \int_0^1 x_h \, dh = \int_0^1 \frac{1}{u} \, dh,$$

which is finite provided (as proved for the similarity solution) $1/u$ is integrable on [0, 1]. Thus, the continuum limit should consist of a monotone, asymptotically self-similar profile connecting two semi-infinite facets.

2.6 Some Open Problems

We have concentrated on the ADL evolution law (9), because the analysis in [1] is restricted to it. But solutions of the DL evolution law (10) also appear numerically to be self-similar [22]. The decay rate is $t^{-1/5}$, and the DL evolution "in similarity variables" is obtained by substituting $u_i(t) = t^{-1/5} v_i(s)$ with $s = \log t$ into the diffusion-limited evolution law (10). This leads to the DL analogue of (16):

$$\frac{dv_i}{ds} = \frac{1}{5} v_i - v_i^2 \Delta_i [v \Delta(v^3)]. \tag{19}$$

Unfortunately, we do not know a steepest-descent interpretation of the DL evolution. Therefore the methods we used for the ADL case seem not to be applicable in the DL setting. (The recent paper by Nakamura and Margetis [16] examines the DL self-similar solution using an entirely different technique, but does not address its stability.)

Our paper [1] considers only a monotone step train connecting two semi-infinite terraces as in Fig. 1. But there are also other examples involving monotone step trains where the step dynamics appears to be asymptotically self-similar. This was pointed out by Israeli, Jeong, Kandel, and Weeks [11], who considered several cases including

(a) *relaxation of an infinite bunch*—directly analogous to the example discussed above, but with a terrace only on one side (Fig. 2, right); and
(b) *reconstruction-driven faceting*—in which two infinite step trains of the same sign are repelled by an energetically-preferred terrace (Fig. 2, left).

The simulations and analysis in [11] (which studies the ADL regime and evaporation-condensation dynamics) show convincingly that the solutions are asymptotically self-similar. We wonder whether the methods of [1] might suffice to give a proof, and perhaps to give a variational characterization of the similarity solution.

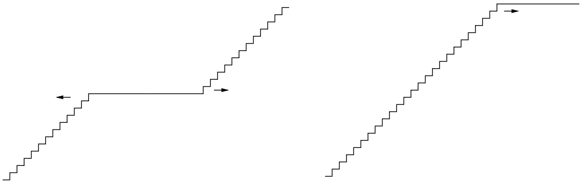


Fig. 2 Initial conditions for two examples considered in [11]. *Left*: reconstruction-driven faceting. *Right*: relaxation of an infinite bunch

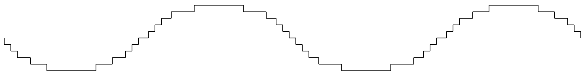


Fig. 3 In a periodic step-train, steps collide at peaks and valleys

A monotone step train has the advantage that its steps don't collide, so the maximum and minimum heights don't change with time. The relaxation of a rough profile—for example a periodic one, like Fig. 3—is fundamentally different, because collision of steps is the crucial mechanism by which the maximum height decreases and the minimum increases, leading eventually to a single flat facet. But perhaps the self-similarity of monotone profiles can be used to understand this process. Indeed, Ozdemir and Zangwill argue in [22] that each monotone region will be close to a self-similar solution, and they use this idea to predict the times when steps collide and the overall flattening rate. Such an argument was explored further by Israeli and Kandel [10], who permitted steps of opposite sign to interact. We wonder whether these arguments can be made rigorous. A natural first step might be to consider the flattening of an infinite groove—another example from [11]. Similar issues arise in the radial setting [5, 9]. The paper [10] also raises an interesting modeling question: can one determine, by an atomic-scale or Burton–Cabrera–Frank-type argument, how steps of opposite sign should interact?

This section has focused on one-dimensional step models. Some authors have questioned the physical validity of such models, arguing that two-dimensional fluctuations near a peak or valley cannot be ignored. A good discussion of this topic can be found in [28].

3 A Thermodynamic PDE Model

The approach discussed in Sect. 2 starts with a step motion law, and considers its continuum limit by taking $N \to \infty$. There is also another, more phenomenological viewpoint, in which the continuum surface height solves a fourth-order PDE of the form

$$h_t = -\operatorname{div}\left[M(\nabla h)\nabla \operatorname{div}\left(\beta_1 \frac{\nabla h}{|\nabla h|} + \beta_3 |\nabla h| \nabla h\right)\right] \tag{20}$$

with $\beta_1, \beta_3 > 0$. For spatially periodic structures the boundary condition should of course be periodic. For spatially complex but statistically homogeneous structures it is again natural to use periodic boundary conditions, with a large enough period cell. Therefore we shall focus in the following discussion on the periodic setting (though the paper [8] also considers some other cases). On the facets, where $\nabla h = 0$, the PDE (20) requires interpretation; we shall comment on this below. We always assume that the initial data has mean value 0; since h_t is a divergence, the mean remains 0 for all t.

As time increases, the surface should relax to a perfectly flat state (a single facet). The *rate* at which it flattens can often be observed experimentally. Therefore it is natural to ask what the models predict in this regard. One-dimensional simulations with sinusoidal initial data and $M(\nabla h) = 1$ (the DL regime, see below) indicate [7, 26] that when $\beta_1 > 0$, the solution becomes identically zero at a finite time T^*, and $h_{\max} \sim (T^* - t)\lambda^{-3}$ where λ is the period size. Some analytical results supporting this conclusion can be found in [21, 25].

After briefly summarizing the logic behind (20) we shall discuss our recent work [8], which addresses flattening behavior of the PDE when $M = 1$ and $\beta_1 = 1$. Our main result is an estimate for the flattening time, which shows (at the level of an inequality) that the general case (for any initial data, in any space dimension) is similar to the 1D, sinusoidal setting.

Numerical results are also available in the ADL setting, and when $\beta_1 = 0$. We shall discuss them briefly in Sect. 3.4.

3.1 Background

The PDE (20) is, roughly speaking, obtained by applying a "thermodynamic" viewpoint to the singular surface energy

$$E[h] = \int \beta_1 |\nabla h| + \frac{1}{3}\beta_3 |\nabla h|^3. \tag{21}$$

The thermodynamic viewpoint, which dates back at least to Mullins' work on grain boundary grooving [20], has long been accepted in connection with motion by surface diffusion above the roughening temperature.

When $\beta_1 > 0$ the surface energy density has a conical singularity at $\nabla h = 0$. This reflects the hypothesis that the horizontal orientation is a *facet*. The second term in the energy density is cubic rather than quadratic, to achieve consistency with the 1D step motion law discussed in Sect. 2. (We explain this consistency below.)

It is instructive to consider the one-dimensional, small-slope version of the isoperimetric problem:

$$\begin{aligned} &\min \int_0^1 \beta_1 |h_x| + \frac{1}{3}\beta_3 |h_x|^3 \, dx, \\ &\text{subject to} \quad h(0) = h(1) = 0 \quad \text{and} \quad \int_0^1 h(x)\, dx = \text{const.} \end{aligned} \tag{22}$$

The solution has curved sides and a flat facet at the top, as shown in Fig. 4. The edges of the facet are free boundaries, in the sense that their locations are determined by energy minimization.

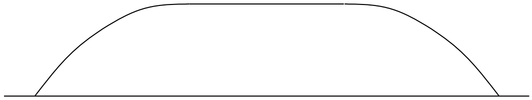


Fig. 4 Solution of the isoperimetric problem (22)

We are interested in dynamics, not statics. Conservation of mass says

$$h_t + \operatorname{div} j = 0 \tag{23}$$

where j is the surface current. The thermodynamic viewpoint says

$$j = M(\nabla h)\nabla \mu \tag{24}$$

where $M(\nabla h)$ is a suitable "mobility" and μ (the "chemical potential") is minus the first variation of E with respect to h. When E has the form (21), (23) and (24) combine to give (20).

What is $M(\nabla h)$? In one space dimension, the answer can be found by coarse-graining the step velocity law (1) for a monotone step train [19, 22]. Since $h(x_i(t), t)$ is constant in time, the step velocity is $\dot{x}_i = -h_t/h_x$. To find the continuum version of the right hand side of (1), we observe that

$$\frac{\mu_{i+1} - \mu_i}{(x_{i+1} - x_i) + 2D/k} = \left(1 + \frac{2D}{k\delta x}\right)^{-1} \frac{\delta\mu}{\delta x} \tag{25}$$

where $\delta x = x_{i+1} - x_i$ and $\delta\mu = \mu_{i+1} - \mu_i$. With the approximations

$$\frac{\delta h}{\delta x} \approx h_x, \qquad \frac{\delta\mu}{\delta x} \approx \mu_x$$

and remembering that $\delta h = 1/N$, the right hand side of (25) becomes $M(h_x)\mu_x$ with

$$M(h_x) = \left(1 + \frac{2DN}{k} h_x\right)^{-1}.$$

Proceeding in the same spirit, one finds that the continuum version of μ (defined by (2)) is a constant times $h_x^{-1}(h_x^3)_x = \frac{3}{2}(h_x^2)_x$, and the continuum limit of the step velocity law becomes (after a suitable change of variables in time)

$$h_t = -(M(h_x)(h_x^2)_{xx})_x.$$

This is simply (20) with $\beta_3 = 1$. (Notice that in one space dimension, the term in the PDE associated with the singular part of the energy $\beta_1|h_x|$ vanishes when $h_x > 0$.)

This calculation is the basis of the widespread view that the mobility in (20) should be

$$M_{\mathrm{DL}}(\nabla h) = 1 \quad \text{and} \quad M_{\mathrm{ADL}} = |\nabla h|^{-1} \tag{26}$$

in the DL and ADL limits (when $DN/k \to 0$ or ∞ respectively). Actually, the situation is more complicated [15]: in the 2+1-dimensional setting where $h = h(x_1, x_2)$

the ADL mobility obtained by coarse-graining the step motion law is anisotropic, with eigenvalues $|\nabla h|^{-1}$ and 1, and eigendirections normal and tangent to the steps. In the DL case, however, such complications are absent: the calculation in [15] supports the view that $M_{\mathrm{DL}} = 1$.

The PDE (20) is somewhat formal, since we seem to be dividing by zero when $\nabla h = 0$—which is expected to occur, since (like the solution of the isoperimetric problem) the evolving height $h(x,t)$ is expected to have facets at its local minima and maxima. The PDE looks local, but actually it is nonlocal: the velocity of the facet edge depends on the size and shape of the facet (see [6] for a survey on singular diffusions like (20)).

When $M = 1$ a fully rigorous understanding of this evolution is available, using the functional analysis of steepest-descent evolution, and the convexity of (21). Briefly: h evolves by H^{-1} steepest-descent for E, where H^{-1} is dual to the space of mean-value-zero periodic functions endowed with the H^1 norm (see [8] for additional details and references to the literature). The evolution defined this way is the same as the one obtained by regularizing the singular term (e.g. replacing $|\nabla h|$ by $(|\nabla h|^2 + \varepsilon^2)^{1/2}$ then taking the limit $\varepsilon \to 0$). It is also the same as the one obtained by implicit time-stepping in the limit $\Delta t \to 0$.

3.2 Finite-Time Flattening when $M = 1$*: Heuristics*

Near the flattening time we expect ∇h to be small, so that $|\nabla h|^3 \ll |\nabla h|$. Therefore we expect behavior similar to that obtained when $\beta_3 = 0$. Setting $\beta_1 = 1$ for simplicity, we are left to consider the H^{-1} steepest descent for $\int |\nabla h|\, dx$, which formally solves

$$h_t = -\Delta \operatorname{div}\left(\frac{\nabla h}{|\nabla h|}\right).$$

This evolution law has two scale invariances:

$$t \to \lambda t, \quad h \to \lambda h, \quad x \to x \quad \text{and} \quad t \to \lambda^4 t, \quad h \to \lambda h, \quad x \to \lambda x. \tag{27}$$

The first invariance shows that in its dependence on the initial data, the extinction time $T^* = T^*(h_0)$ is positively homogeneous of degree one, i.e. $T^*(\lambda h_0) = \lambda T^*(h_0)$ for $\lambda > 0$. This suggests a flattening time estimate of the form

$$T^* \leq C\|h_0\|_X \tag{28}$$

for some function space X. The second invariance in (27) determines how the constant in (28) depends on the size of the period cell. In particular, it places restrictions on X if we want the estimate to be scale-invariant (e.g. to be valid—with a fixed constant—when the period cell is a cube of any size).

Since any time can be viewed as the initial time, the heuristic estimate can be written

$$\|h(t)\|_X \geq C^{-1}(T^* - t) \quad \text{for } t < T^*.$$

Thus it captures (at the level of an inequality) the intuition that the solution should flatten linearly in time.

3.3 Finite-Time Flattening: Rigorous Results

The central accomplishment of [8] is a family of flattening time estimates for

$$h_t = -\Delta \operatorname{div}\left(\frac{\nabla h}{|\nabla h|} + \beta_3 |\nabla h| \nabla h\right) \tag{29}$$

that are more or less of the form (28). Notice that (29) is our original equation (20) with $M = 1$ and $\beta_1 = 1$.

The argument is relatively easy at the formal level. Multiplying (29) by $(-\Delta)^{-1} h$ and integrating by parts gives

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}\|h\|^2_{H^{-1}}(t) &= \int h_t (-\Delta)^{-1} h \, dx \\
&= \int h \operatorname{div}\left(\frac{\nabla h}{|\nabla h|} + \beta_3 |\nabla h| \nabla h\right) dx \\
&= -\int |\nabla h| + \beta_3 |\nabla h|^3 \, dx \\
&\le -\int |\nabla h| \, dx.
\end{aligned} \tag{30}$$

For a non-scale-invariant estimate in dimension $n \le 4$, we can finish by observing that

$$\|h\|_{H^{-1}} \le C \int |\nabla h| \, dx \tag{31}$$

for periodic functions with mean value 0, with a constant that depends on the period. Combining this with (30) gives

$$\frac{d}{dt}\|h\|_{H^{-1}} \le -C \quad \text{for } t < T^*.$$

Since the right hand side is independent of t, we conclude that

$$T^*(h_0) \le C \|h_0\|_{H^{-1}} \tag{32}$$

for any (periodic) initial condition h_0.

We are, however, especially interested in scale-invariant estimates, because the PDE is often used to simulate problems with complex or random initial data. In such settings, spatial periodicity is used numerically (with a sufficiently large period cell) as a scheme for minimizing "finite-size effects." The key, of course, is to replace (31) by a scale-invariant estimate. We showed in [8] that

$$\|h\|_{H^{-1}} \le C_* \|(-\Delta)^{-1} h\|_{W^{-1,p}}^{1-\theta} \left(\int |\nabla h|\, dx \right)^{\theta} \tag{33}$$

for mean-value-zero periodic functions in $\mathbb{R}^n$, provided $1 \le n \le 4$, $1 \le p \le \infty$, and $1/2 \le \theta \le 1$ are related by

$$1 + \frac{n}{2} = \theta(n-1) + (1-\theta)\left(3 + \frac{n}{p}\right). \tag{34}$$

Moreover, this estimate is scale-invariant, in the sense that while the constant C_* depends on the shape of the period cell, it does not depend on its size.

To make use of (33) we must control the growth of $\|(-\Delta)^{-1} h\|_{W^{-1,p}}$. This is again easy at the formal level: since $\|\cdot\|_{W^{-1,p}}$ is a norm, we have

$$\frac{d}{dt}\|(-\Delta)^{-1} h\|_{W^{-1,p}} \le \|(-\Delta)^{-1} h_t\|_{W^{-1,p}}.$$

From the PDE (29) we have

$$\|(-\Delta)^{-1} h_t\|_{W^{-1,p}} = \|z\|_{W^{-1,p}}$$

with

$$z = \operatorname{div}(\nabla h/|\nabla h|) + \operatorname{div}(\beta_3 |\nabla h| \nabla h) = z_1 + z_2.$$

Since $\nabla h/|\nabla h|$ has magnitude less than or equal to 1 pointwise, we have

$$\|z_1\|_{W^{-1,p}} \le 1$$

if the unit cell has volume 1. As for z_2, we have

$$\|z_2\|_{W^{-1,p}} \le \beta_3 \||\nabla h|^2\|_{L^p},$$

which is controlled by the initial energy provided $p \le 3/2$, using the fact that $E = \int |\nabla h| + \frac{1}{3}\beta_3 |\nabla h|^3$ is decreasing in time. Our article [8] makes these arguments rigorous and uses them to show that

$$T^* \le \frac{1}{2-(1/\theta)} C_*^{1/\theta} \|(-\Delta)^{-1} h_0\|_{W^{-1,p}}^{(1/\theta)-1} \|h_0\|_{H^{-1}}^{2-(1/\theta)} \tag{35}$$

for periodic solutions in $\mathbb{R}^n$ when $1 \le n \le 4$, $1 \le p \le 3/2$ and $1/2 < \theta \le 1$ are related by (34). The constant C_* in (35) is the one from (33); in particular, it is scale-invariant. In the physical case $n = 2$, the relationship (34) reduces to $(2 + \frac{2}{p})\theta = 1 + \frac{2}{p}$. Any p in the range $1 \le p \le 3/2$ is permitted, and $\theta(p)$ ranges between $\theta(1) = 3/4$ and $\theta(3/2) = 7/10$.

3.4 Some Open Problems

We have focused on the DL setting, where $M = 1$ and (20) becomes (29). In this setting the evolution is H^{-1} steepest descent for the surface energy (21). As we explained above, however, this is just a special case; physically, it is equally natural to consider $M(\nabla h) = (1 + c|\nabla h|)^{-1}$ where c is any constant. Alas, when M is not constant even the well-posedness of the PDE seems unclear, since the evolution is no longer (to the best of our knowledge) steepest descent for (21) in any Hilbert space.

The ADL setting (when $M = |\nabla h|^{-1}$) has received substantial attention in the physics literature, both because some experiments are in this regime [4, 27] and because heuristic arguments like those of Sect. 3.2 are available in this case. Numerical simulation with sinusoidal initial data shows that $h_{\max} \sim \lambda^{-1}(T^* - t)^{1/2}$, where λ is the spatial period [4, 27]. Interestingly, when the 1D, ADL evolution was simulated in [4] with more complex initial data modeling the decay of nanoripples on a Cu(001) surface, the result was different: the amplitude $\|h\|_{L^2}$ decayed with a linear rather than square-root law.

We expect solutions of (20) to flatten in finite time only when $\beta_1 > 0$. The behavior when $\beta_1 = 0$ has been considered heuristically and numerically. The expected behavior (see e.g. [4]) is that $h \to 0$ as $t \to \infty$, with $\|h\| \sim \tau/t$ in the DL setting and $\|h\| \sim e^{-t/\tau}$ in the ADL setting, where τ is a suitably defined characteristic time. Two-dimensional simulations with $\beta_1 = 0$ and the tensor mobility suggested by [15] reveal an interesting shape transition, in which biperiodic (but asymmetric) initial data become essentially one-dimensional [2].

Our discussion about the form of M suggests that the PDE discussed in this section represents the continuum limit of the step model considered in Sect. 2. This conclusion is correct (at least formally) for a monotone step train separating two semi-infinite facets (see Appendix B of [1]). However it seems false in 1D for mean-value-zero initial data [10, 11], and it is definitely false in the radial context with an infinite cone as initial data [17]. The step and PDE models are consistent away from facets, but they treat the facets differently, providing (it seems) different evolution laws for the free boundary marking the edge of the facet. This does not mean that either model is wrong—but it raises the question, for a given physical system, which is more appropriate. Some insight is provided by [14] and [29], which derive equations like (20) by heuristically coarse-graining certain kinetic Monte Carlo models.

We have mentioned some numerical studies of (20). Those in [4, 26, 27] use an interesting nonlinear Galerkin scheme whose numerical convergence has not yet been proved. However the convergence of a finite-element scheme for the DL equation (29) was recently examined in [13].

Acknowledgements This research was supported by NSF through grant DMS-0807347.

References

1. Al Hajj Shehadeh, H., Weare, J., Kohn, R.V.: The evolution of a crystal surface: Analysis of a one-dimensional step train connecting two facets in the ADL regime. Physica D **240**, 1771–1784 (2011)
2. Bonito, A., Nochetto, R., Quah, J., Margetis, D.: Self-organization of decaying surface corrugations: a numerical study. Phys. Rev. E **79**, 050601 (2009)
3. Burton, W.K., Cabrera, N., Frank, F.C.: The growth of crystals and the equilibrium structure of their surfaces. Philos. Trans. R. Soc. Lond. Ser. A **243**, 299–358 (1951)
4. Chan, W.L., Ramasubramaniam, A., Shenoy, V.B., Chason, E.: Relaxation kinetics of nanoripples on Cu(001) surface. Phys. Rev. B **70**, 245403 (2004)
5. Fok, P.W., Rosales, R.R., Margetis, D.: Facet evolution on supported nanostructures: the effect of finite height. Phys. Rev. B **78**, 235401 (2008)
6. Giga, M.-H., Giga, Y.: Very singular diffusion equations: second and fourth order problems. Jpn. J. Ind. Appl. Math. **27**, 323–345 (2010)
7. Hager, J., Spohn, H.: Self-similar morphology and dynamics of periodic surface profiles below the roughening transition. Surf. Sci. **324**, 365–372 (1995)
8. Giga, Y., Kohn, R.V.: Scale-invariant extinction time estimates for some singular diffusion equations. Discrete Contin. Dyn. Syst., Ser. A **30**, 509–535 (2011)
9. Israeli, N., Kandel, D.: Profile of a decaying crystalline cone. Phys. Rev. B **62**, 5946–5962 (1999)
10. Israeli, N., Kandel, D.: Decay of one-dimensional surface modulations. Phys. Rev. B **62**, 12707–13717 (2000)
11. Israeli, N., Jeong, H.C., Kandel, D., Weeks, J.D.: Dynamics and scaling of one-dimensional surface structures. Phys. Rev. B **61**, 5698–5706 (2000)
12. Jeong, H.C., Williams, E.D.: Steps on surfaces: experiment and theory. Surf. Sci. Rep. **34**, 171–294 (1999)
13. Kohn, R.V., Versieux, H.: Numerical analysis of a steepest-descent model for surface relaxation below the roughening temperature. SIAM J. Numer. Anal. **48**, 1781–1800 (2010)
14. Krug, J., Dobbs, H.T., Majaniemi, S.: Adatom mobility for the solid-on-solid model. Z. Phys. B **97**, 281–291 (1995)
15. Margetis, D., Kohn, R.V.: Continuum relaxation of interacting steps on crystal surfaces in $2+1$ dimensions. Multiscale Model. Simul. **5**, 729–758 (2006)
16. Margetis, D., Nakamura, K.: From crystal steps to continuum laws: behavior near large facets in one dimension. Physica D **240**, 1100–1110 (2011)
17. Margetis, D., Fok, P.W., Aziz, M.J., Stone, H.A.: Continuum theory of nanostructure decay via a microscale condition. Phys. Rev. Lett. **97**, 096102 (2006)
18. Michely, T., Krug, J.: Islands, Mounds, and Atoms. Patterns and Processes in Crystal Growth Far from Equilibrium. Springer, Berlin (2004)
19. Murty, M.V.R.: Morphological stability of nanostructures. Phys. Rev. B **62**, 17004–17011 (2000)
20. Mullins, W.W.: Theory of thermal grooving. J. Appl. Phys. **28**, 333–339 (1956)
21. Odisharia, I.V.: Simulation and analysis of the relaxation of a crystalline surface. PhD thesis, New York University (2006)
22. Ozdemir, M., Zangwill, A.: Morphological equilibration of a corrugated crystalline surface. Phys. Rev. B **42**, 5013–5024 (1990)
23. Pimpinelli, A., Villain, J.: Physics of Crystal Growth. Cambridge University Press, Cambridge (1998)
24. Rettori, A., Villain, J.: Flattening of grooves on a crystal surface: a method of investigation of surface roughness. J. Phys. I (France) **49**, 257–267 (1988)
25. Spohn, H.: Surface dynamics below the roughening transition. J. Phys. I (France) **3**, 69–81 (1993)
26. Shenoy, V.B., Ramasubramaniam, A., Freund, L.B.: A variational approach to nonlinear dynamics of nanoscale surface modulations. Surf. Sci. **529**, 365–383 (2003)

27. Shenoy, V.B., Ramasubramaniam, A., Ramanarayan, H., Tambe, D.T., Chan, W.-L., Chason, E.: Influence of step-edge barriers on the morphological relaxation of nanoscale ripples on crystal surfaces. Phys. Rev. Lett. **92**, 256101 (2004)
28. Tang, L.-H.: Flattening of grooves: from step dynamics to continuum theory. In: Duxbury, P.M., Pence, T.J. (eds.) Dynamics of Crystal Surfaces and Interfaces, pp. 169–184. Plenum Press, New York (1997)
29. Vvedensky, D.D., Zangwill, A., Luse, C.N., Wilby, M.R.: Stochastic equations of motion for epitaxial growth. Phys. Rev. E **48**, 852–862 (1993)

Climate Science, Waves and PDEs for the Tropics

Andrew J. Majda

Abstract A reader's guide to recent applied mathematics development in multiscale modeling in the tropics is provided here including the mathematical theory of precipitation fronts as well as singular limits with variable coefficients in the fast variables.

1 Introduction

One of the grand challenges of contemporary science is a comprehensive predictive model for the atmosphere and coupled climate system. This is one of the most difficult multiscale problems in contemporary science because there is an incredible range of strongly interacting anisotropic nonlinear processes over many spatiotemporal scales; contemporary comprehensive computer models, GCMs, are currently incapable of adequately resolving or parameterizing many of these interactions on time scales appropriate for seasonal prediction as well as climate change projections. An overview for mathematicians can be found in the recent article by the author [33].

Basic questions which drive climate research are the prediction of the weather from 1 to 14 days, the prediction of climate variations on seasonal to yearly time scales and finally, climate change projections on decadal and centennial time scales as well as quantifying the uncertainty associated with these predictions. One of the striking recent observational discoveries is the profound impact of variations in the tropics on all of these problems. The primary issue in the influence of the tropics occurs through the interaction and organization of clouds into clusters, super clusters, and planetary scale dynamics, an inherently fully nonlinear multiscale process. For climate change, water vapor is the most important greenhouse gas and the microphysical processes in clouds are a key mechanism for radiative feedback. In fact,

A.J. Majda (✉)
Department of Mathematics and Center for Atmosphere/Ocean Science (CAOS), Courant Institute of Mathematical Sciences (CIMS), New York University, 251 Mercer Street, New York, NY 10012, USA
e-mail: jonjon@cims.nyu.edu

H. Holden, K.H. Karlsen (eds.), *Nonlinear Partial Differential Equations*, Abel Symposia 7, DOI 10.1007/978-3-642-25361-4_12,

only a 4% change in average cloudiness would overwhelm the effects of CO_2 in climate change.

A new perspective on several of the issues discussed above for climate dynamics has been developed through the paradigm of modern applied mathematics, where rigorous and asymptotic multiscale mathematical theory, the development of prototype model problems and novel computational strategies all interact simultaneously in understanding these complex scientific problems.

2 Multi-scale Models in the Tropics and the Madden–Julian Oscillation

The dominant component of intraseasonal variability in the tropics is the 40- to 50-day tropical intraseasonal oscillation, often called the Madden–Julian oscillation (MJO) after its discoverers [30, 31]. In the troposphere, the MJO is an equatorial planetary-scale wave envelope of complex multiscale convective processes. It begins as a standing wave in the Indian Ocean and propagates eastward across the western Pacific Ocean at a speed of roughly $\approx$5 m/s [68]. The planetary scale circulation anomalies associated with the MJO significantly affect monsoon development, intraseasonal predictability in mid-latitudes, and the development of the El Nino Southern Oscillation (ENSO) in the Pacific Ocean, which is one of the most important components of seasonal prediction [28, 68].

Despite the widespread importance of the MJO, present day computer general circulation models (GCMs) typically have poor representations of it [29]. A growing body of evidence suggests that this poor performance of GCMs is due to the inadequate treatment of interactions of organized tropical convection on multiple spatiotemporal scales [29, 52]. Such hierarchical organized structures that generate the MJO as their envelope are the focus of current observational initiatives and modeling studies [52], and there is a general lack of theoretical understanding of these processes and the MJO itself.

A large number of theories attempting to explain the MJO through mechanisms such as evaporation-wind feedback [8, 53], boundary layer frictional convective instability [66], stochastic linearized convection [61], radiation instability [58], and the planetary-scale linear response to moving heat sources [7]. While they all provide some insight into the mechanisms of the MJO, these theories are all at odds with the observational record in various crucial ways [28, 68], and it is therefore likely that none of them captures the fundamental physical mechanisms of the MJO. Nevertheless, they are all interesting theories that contribute to our understanding of certain aspects of the MJO. Other insight has been gained through the study of MJO-like waves in multi-cloud model simulations [23, 47] and in super-parameterization computer simulations [13–15, 51], which appear to capture many of the observed features of the MJO by accounting for smaller scale convective structures within the MJO envelope. The role of convective momentum transport from synoptic scale waves in producing key features of the MJOs planetary scale envelope has also been elucidated by multi-scale asymptotic models [2, 3, 5, 38, 45]. Despite all of the interesting contributions listed above, no theory for the MJO has yet been generally accepted, and the problem of explaining the MJO has recently been called the search for the Holy Grail of tropical atmospheric dynamics [58].

Although theory and simulation of the MJO remain difficult challenges, they are guided by the generally accepted, fundamental features of the MJO (i.e., the MJOs skeleton) on intraseasonal/planetary scales, which have been identified relatively clearly in observations:

I. peculiar dispersion relation of $d\omega/dk \approx 0$ [59, 60, 67],
II. slow phase speed of roughly 5 m/s [16, 17, 49], and
III. horizontal quadrupole vortex structure [16, 17, 49].

The goal of a recent article [44] is to design the simplest dynamical model that captures and predicts the intraseasonal/planetary scale features of the MJOs skeleton in I–III, and to recover these features robustly throughout the parameter space of the model.

Geophysical flows are a rich source of fascinating problems for applied mathematicians involving complex multi-scale nonlinear systems, where energy cascades upward from the small scales to the large scales through anisotropic processes involving vortices and gravity waves. On the other hand, the improved parameterization of unresolved features of moist tropical convection is a central challenge in current computer models for long range ensemble forecasting of weather and short term climate with large worldwide societal impact [52]. The reason for this is the observed multi-scale features of organized coherent tropical convection across a wide range of scales varying from tens of kilometers and a few hours to the planetary scale of order 40,000 km on intraseasonal time scales with significant energy transfer across these scales [19, 50, 54, 67]. Recent processing of observational data [50] suggests the statistical self-similarity of tropical convection from the smallest, shortest scales to organized mesoscale convective systems [18] to convective clusters to equatorial synoptic-scale superclusters to planetary/intraseasonal oscillations. For this reason, it is interesting to develop systematic multiscale asymptotic models [25–27, 34, 42] for the nonlinear cascade across scales in the tropics, and the author has done this recently for the self-similar behavior from the microscales to mesoscales to planetary/intraseasonal scales [35, 36]. Such quantitative models are useful for quantifying the observed multiscale behavior in, for example, tropical intraseasonal oscillations [2–5, 38, 45].

3 The Dynamics of Equatorial Waves: Singular Limits with Fast Variable Coefficients

Geophysical flows are a rich source of novel problems for applied mathematics and the contemporary theory of partial differential equations (PDE) ([34] and references therein). The reason for this is that many physically important geophysical flows involve complex nonlinear interaction over multi-scales in both time and space so developing simplified reduced models which are simpler yet capture key physical phenomena is of central importance [12, 34, 56, 57]. In mid-latitudes, the fact that the rotational Coriolis terms are bounded away from zero leads to a strict temporal frequency scale separation between slow potential vorticity dynamics and fast gravity waves; this physical fact leads to new theorems justifying the quasi-geostrophic mid-latitude dynamics even with general unbalanced initial data for both rapidly rotating shallow water equations and completely stratified flows [6, 9, 10, 34, 39]. The strategy in the above proofs is to adapt the classical framework of Klainerman and Majda for singular limits [24, 34] together with the important generalizations by Schochet [62, 63], which allow for fast wave averaging, to the dispersive systems of geophysical flows; it is well known that these theories require constant symmetric hyperbolic coefficients for the fast wave dynamics in order to obtain higher derivative estimates on the solution.

At the equator, the tangential projection of the Coriolis force from rotation vanishes identically so that there is no longer a time scale separation between potential vortical flows and gravity waves. This has profound consequences physically that allow the tropics to behave as a waveguide with extremely warm surface temperatures. The resulting behavior profoundly influences longer term mid-latitude weather prediction and climate change through hurricanes, monsoons, El Nino, and global teleconnections with the mid-latitude atmosphere. How this happens through detailed physical mechanisms is one of the most important contemporary problems in the atmosphere-ocean science community with a central role played by nonlinear interactive heating involving the interaction of clouds, moisture, and convection [1, 2, 11, 37, 38, 46, 57, 64]. The variable coefficient degeneracy of the Coriolis term at the equator alluded to earlier leads to both important new physical effects as well as fascinating new mathematical phenomena and PDEs [1, 2, 11, 37, 38, 42, 46]. Chapter 9 of Ref. [34] provides an introduction to these topics for mathematicians while Ref. [11] introduces and studies the simplest physical equatorial models with moisture. In this equatorial context, the new multi-scale reduced dynamical PDE models are relatively recent in origin [42]. Thus, the need for additional PDE theory is very important for these disciplinary problems and this is the main topic of recent research [40, 41, 48].

4 Precipitation Fronts: A Novel Hyperbolic Free Boundary Problem in Several Space Variables

Precipitation fronts are the boundaries between the zones of extremely moist air (with constant precipitation) such as over the Indonesian marine continent, the Indian ocean, and Western Pacific, and the zones of extremely dry air in the tropics and subtropics that occur over areas such as the Galapagos islands at the equator or the Arabian peninsula in the subtropics. An important practical question in contemporary meteorology for long range weather prediction and climate change projections is what determines the boundaries of the precipitating fronts as well as their evolution in time. Such assessments are performed, for example, by the Intergovernmental Panel for Climate Change (IPCC) by running extremely complex general computer models called GCMs. An important practical issue with the GCMs is how they treat moisture and what type of moisture waves do they produce at large scales compared with those in nature. This is very subtle. Although the GCMs have millions of variables and run on the largest supercomputers, they still have grid spacings of order 50 km to 200 km. In addition many complex physical processes need to be parametrized often by ad hoc recipes guided by physical intuition. These issues are discussed in detail in [11, 55, 65], and the references therein.

A novel mathematical theory of precipitating fronts was put forward in [11, 55, 65] to address the above issues in idealized tropical climate models consisting of a shallow water system for the temperature, T, and velocity, $u = (u; v)$, coupled with an equation for the moisture or humidity, q, through a relaxation source term P representing the depletion of moisture through precipitation and the condensational heating of the atmosphere in making clouds and depending on the type of moisture parametrization. (See [11] for a detailed derivation.)

A novel point of view for atmospheric science developed in [11, 55, 65] is to formally take the zero relaxation limit, $\varepsilon \to 0$, and to study the type of the emerging precipitation fronts in order to get analytic insight into the behavior at positive ε. This procedure shows formally that, in the limit $\varepsilon \to 0$, precipitation fronts are free boundaries where $U = (u; T; q)$ is continuous across them but ∇U, which formally solves the hyperbolic system derived, has jumps satisfying the Rankine–Hugoniot–type shock conditions [32]. These considerations were utilized in [11] to build three distinct wave families, namely drying, slow moistening, and fast moistening precipitating fronts, with the last two families violating Lax's shock conditions. Nevertheless, careful numerical experiments demonstrate (see [11, 20, 21]) and additional mathematical theory (see [65]) established, at positive ε, the robust reliability of all three wave types as well as interesting half smooth traveling waves. Finally numerical simulations (see [55]), again at ε positive, confirm a theory for reflection and transmission of waves impinging on precipitation zones.

Given all the above mentioned formal results it is extremely interesting to pass to the zero relaxation limit and to prove rigorously the existence and uniqueness of suitable weak solutions for the limiting problem. This is the main topic of the recent paper [43].

There is (see [11]) an interesting more complex version of moisture dynamics involving coupling with the barotropic model. Obtaining higher order energy estimates for this problem is a hard unsolved problem. The models discussed here are excellent ones for understanding the precipitation fronts at large scales in GCMs. The actual behavior as observed in nature is captured in a much more realistic fashion by more complex multi-cloud models [22, 23] with large scale instability.

References

1. Biello, J.A., Majda, A.J.: The effect of meridional and vertical shear on the interaction of equatorial baroclinic and barotropic Rossby waves. Stud. Appl. Math. **112**(4), 341–390 (2004)
2. Biello, J.A., Majda, A.J.: A new multiscale model for the Madden–Julian oscillation. J. Atmos. Sci. **62**, 694–1721 (2005)
3. Biello, J.A., Majda, A.J.: Modulating synoptic scale convective activity and boundary layer dissipation in the IPESD models of the Madden–Julian oscillation. Dyn. Atmos. Ocean. **42**, 152–215 (2006)
4. Biello, J.A., Majda, A.J.: Intraseasonal multi-scale moist dynamics of the tropical troposphere. Commun. Math. Sci. **8**(2), 519–540 (2010)
5. Biello, J.A., Majda, A.J., Moncrieff, M.W.: Meridional momentum flux and superrotation in the multi-scale IPESD MJO model. J. Atmos. Sci. **64**, 1636–1651 (2007)
6. Bourgeois, A.J., Beale, J.T.: Validity of the quasigeostrophic model for large-scale flow in the atmosphere and ocean. SIAM J. Math. Anal. **25**(4), 1023–1068 (1994)
7. Chao, W.C.: On the origin of the tropical intraseasonal oscillation. J. Atmos. Sci. **44**, 1940–1949 (1987)
8. Emanuel, K.A.: An air-sea interaction model of intraseasonal oscillations in the tropics. J. Atmos. Sci. **44**, 2324–2340 (1987)
9. Embid, P., Majda, A.J.: Averaging over fast gravity waves for geophysical flows with arbitrary potential vorticity. Commun. Partial Differ. Equ. **21**(3–4), 619–658 (1996)
10. Embid, P.F., Majda, A.J.: Low Froude number limiting dynamics for stably stratified flow with small or finite Rossby numbers. Geophys. Astrophys. Fluid Dyn. **87**(1–2), 1–50 (1998)
11. Frierson, D.M.W., Majda, A.J., Pauluis, O.M.: Large scale dynamics of precipitation fronts in the tropical atmosphere: a novel relaxation limit. Commun. Math. Sci. **19**, 591–626 (2004)
12. Gill, A.E.: Atmosphere-Ocean Dynamics. Academic Press, New York (1982)
13. Grabowski, W.W.: Coupling cloud processes with the large-scale dynamics using the cloud-resolving convection parameterization (CRCP). J. Atmos. Sci. **58**, 978–997 (2001)
14. Grabowski, W.W.: MJO-like coherent structures: sensitivity simulations using the cloud-resolving convection parameterization (CRCP). J. Atmos. Sci. **60**, 847–864 (2003)
15. Grabowski, W.W., Moncrieff, M.W.: Moisture-convection feedback in the tropics. Q. J. R. Meteorol. Soc. **130**, 3081–3104 (2004)
16. Hendon, H.H., Salby, M.L.: The life cycle of the Madden–Julian oscillation. J. Atmos. Sci. **51**, 2225–2237 (1994)
17. Hendon, H.H., Liebmann, B.: Organization of convection within the Madden–Julian oscillation. J. Geophys. Res. **99**, 8073–8084 (1994)
18. Houze, R.A.: Mesoscale convective systems. Rev. Geophys. **42**, RG4003 (2004)
19. Johnson, R.H., Hamilton, P.J.: The relationship of surface pressure features to the precipitation and airflow structure of an intense midlatitude squall line. Mon. Weather Rev. **116**, 1444–1473 (1988)
20. Khouider, B., Majda, A.J.: A non-oscillatory balanced scheme for an idealized tropical climate model. Part I: Algorithm and validation. Theor. Comput. Fluid Dyn. **19**, 331–354 (2005)
21. Khouider, B., Majda, A.J.: A non-oscillatory balanced scheme for an idealized tropical climate model. Part II: Nonlinear coupling and moisture effects. Theor. Comput. Fluid Dyn. **19**, 355–375 (2005)
22. Khouider, B., Majda, A.J.: A simple multi-cloud model for convectively coupled tropical waves. Part I: Linear analysis. J. Atmos. Sci. **63**, 1308–1323 (2006)
23. Khouider, B., Majda, A.J.: A simple multicloud parameterization for convectively coupled tropical waves. Part II. Nonlinear simulations. J. Atmos. Sci. **64**, 381–400 (2007)
24. Klainerman, S., Majda, A.J.: Compressible and incompressible fluids. Commun. Pure Appl. Math. **35**(5), 629–651 (1982)
25. Klein, R.: An applied mathematical view of meteorological modeling. In: Hill, J.M., Moore, R. (eds.) Applied Mathematics Entering the 21st Century: Invited Talks from the ICIAM 2003 Congress. Proceedings in Appl. Math., vol. 116, 227–289. SIAM, Philadelphia (2004)

26. Klein, R., Angew, Z.: Asymptotic analyses for atmospheric flows and the construction of asymptotically adaptive numerical methods. Math. Mech. **80**, 765–777 (2000)
27. Klein, R., Majda, A.J.: Systematic multiscale models for deep convection on mesoscales. Theor. Comput. Fluid Dyn. **20**, 525–551 (2006)
28. Lau, W.K.M., Waliser, D.E. (eds.): Intraseasonal Variability in the Atmosphere Ocean Climate System. Springer, Berlin (2005)
29. Lin, J.L., et al.: Tropical intraseasonal variability in 14 IPCC AR4 climate models. Part I. Convective signals. J. Climate **19**, 2665–2690 (2006)
30. Madden, R., Julian, P.: Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. J. Atmos. Sci. **28**, 702–708 (1971)
31. Madden, R.A., Julian, P.R.: Observations of the 40–50-day tropical oscillation–a review. Mon. Weather Rev. **122**, 814–837 (1994)
32. Majda, A.J.: Compressible Fluid Flow and Systems of Conservation Laws in Several Space Variables. Applied Mathematical Sciences, vol. 53. Springer, New York (1984)
33. Majda, A.J.: Real world turbulence and modern applied mathematics. In: Mathematics: Frontiers and Perspectives, pp. 137–151. Am. Math. Soc., Providence (2000)
34. Majda, A.J.: Introduction to PDEs and Waves for the Atmosphere and Ocean. Courant Lecture Notes in Mathematics, vol. 9. AMS/CIMS, New York (2003)
35. Majda, A.J.: Multiscale models with moisture and systematic strategies for superparameterization. J. Atmos. Sci. **64**, 2726–2734 (2007)
36. Majda, A.J.: New multiscale models and self-similarity in tropical convection. J. Atmos. Sci. **64**, 1393–1404 (2007)
37. Majda, A.J., Biello, J.A.: The nonlinear interaction of barotropic and equatorial baroclinic Rossby waves. J. Atmos. Sci. **60**(15), 1809–1821 (2003)
38. Majda, A.J., Biello, J.A.: A multiscale model for the intraseasonal oscillation. Proc. Natl. Acad. Sci. USA **101**, 4736–4741 (2004)
39. Majda, A.J., Embid, P.: Averaging over fast gravity waves for geophysical flow with unbalanced initial data. Theor. Comput. Fluid Dyn. **11**, 155–169 (1998)
40. Majda, A.J., Dutrifoy, A.: The dynamics of equatorial long waves: a singular limit with fast variable coefficients. Commun. Math. Sci. **4**(2), 375–397 (2006)
41. Majda, A.J., Dutrifoy, A.: Fast wave averaging for the equatorial shallow water equations. Commun. Partial Differ. Equ. **32**(10), 1617–1642 (2007)
42. Majda, A.J., Klein, R.: Systematic multiscale models for the tropics. J. Atmos. Sci. **60**, 393–408 (2003)
43. Majda, A.J., Souganidis, P.: The existence and uniqueness of weak solutions for precipitation fronts: a novel hyperbolic free boundary problem in several space variables. Commun. Pure Appl. Math. **63**(10), 1351–1361 (2010)
44. Majda, A.J., Stechmann, S.N.: The skeleton of tropical intraseasonal oscillations. Proc. Natl. Acad. Sci. USA **106**(21), 8417–8422 (2009)
45. Majda, A.J., Stechmann, S.N.: A simple dynamical model with features of convective momentum transport. J. Atmos. Sci. **66**, 373–392 (2009)
46. Majda, A.J., Rosales, R.R., Tabak, E.G., Turner, C.V.: Interaction of large-scale equatorial waves and dispersion of Kelvin waves through topographic resonances. J. Atmos. Sci. **56**(24), 4118–4133 (1999)
47. Majda, A.J., Stechmann, S.N., Khouider, B.: Madden–Julian oscillation analog and intraseasonal variability in a multicloud model above the equator. Proc. Natl. Acad. Sci. USA **104**, 9919–9924 (2007)
48. Majda, A.J., Dutrifoy, A., Schochet, S.: A simple justification of the singular limit for equatorial shallow-water dynamics. Commun. Pure Appl. Math. **62**, 322–333 (2009)
49. Maloney, E.D., Hartman, D.L.: Frictional moisture convergence in a composite life cycle of the Madden–Julian oscillation. J. Climate **11**, 2387–2403 (1998)
50. Mapes, B.E., Tulich, S., Lin, J.L., Zuidema, P.: The mesoscales convection life cycle: building block or prototype for large-scale tropical waves? Dyn. Atmos. Ocean. **42**, 3–29 (2006)
51. Moncrieff, M.W.: Analytic representation of the large-scale organization of tropical convection. J. Atmos. Sci. **61**, 1521–1538 (2004)

52. Moncrieff, M.W., Shapiro, M., Slingo, J., Molteni, F.: Collaborative research at the intersection of weather and climate. WMO Bull. **56**, 204–211 (2007)
53. Neelin, J.D., Held, I.M., Cook, K.H.: Evaporation-wind feedback and low-frequency variability in the tropical atmosphere. J. Atmos. Sci. **44**, 2341–2348 (1987)
54. Nakazawa, T.: Tropical super clusters within intraseasonal variations over the western Pacific. J. Meteorol. Soc. Jpn. **66**, 823–839 (1988)
55. Pauluis, O., Frierson, D.M.W., Majda, A.J.: Precipitation fronts and the reflection and transmissions of tropical disturbances. Q. J. R. Meteorol. Soc. **134**, 913–930 (2008)
56. Pedlosky, J.: Geophysical Fluid Dynamics. Springer, New York (1979)
57. Philander, S.G.: El Nino, La Nina, and The Southern Oscillation. Academic Press, San Diego (1990)
58. Raymond, D.J.: A new model of the Madden–Julian oscillation. J. Atmos. Sci. **58**, 2807–2819 (2001)
59. Roundy, P., Frank, W.: A climatology of waves in the equatorial region. J. Atmos. Sci. **61**, 2105–2132 (2004)
60. Salby, M., Hendon, H.: Intraseasonal behavior of clouds, temperature, and motion in the tropics. J. Atmos. Sci. **51**, 2207–2224 (1994)
61. Salby, M.L., Garcia, R.R., Hendon, H.H.: Planetary-scale circulations in the presence of climatological and wave-induced heating. J. Atmos. Sci. **51**, 2344–2367 (1994)
62. Schochet, S.: Resonant nonlinear geometric optics for weak solutions of conservation laws. J. Differ. Equ. **113**(2), 473–504 (1994)
63. Schochet, S.: Fast singular limits of hyperbolic PDEs. J. Differ. Equ. **114**(2), 476–512 (1994)
64. Smith, R.K. (ed.): The Physics and Parametrization of Moist Atmospheric Convection. NATO Advanced Study Institute Series C. Mathematical and Physical Sciences, vol. 505. Kluwer Academic, Norwell (1997)
65. Stechmann, S.N., Majda, A.J.: The structure of precipitation fronts for finite relaxation time. Theor. Comput. Fluid Dyn. **20**, 377–404 (2006)
66. Wang, B., Rui, H.: Dynamics of the coupled moist Kelvin–Rossby wave on an equatorial beta-plane. J. Atmos. Sci. **47**, 397–413 (1990)
67. Wheeler, M., Kiladis, G.N.: Convectively coupled equatorial waves: analysis of clouds and temperature in the wave number-frequency domain. J. Atmos. Sci. **56**, 374–399 (1999)
68. Zhang, C.: Madden–Julian oscillation. Rev. Geophys. **43**, G2003 (2005)

On the Propagation of Oceanic Waves Driven by a Strong Macroscopic Flow

Isabelle Gallagher, Thierry Paul, and Laure Saint-Raymond

Abstract In this work we study oceanic waves in a shallow water flow subject to strong wind forcing and rotation, and linearized around an inhomogeneous (non-zonal) stationary profile. This extends the study (Cheverry et al. in Semiclassical and spectral analysis of oceanic waves, Duke Math. J., accepted), where the profile was assumed to be zonal only and where explicit calculations were made possible due to the 1D setting.

Here the diagonalization of the system, which allows to identify Rossby and Poincaré waves, is proved by an abstract semi-classical approach. The dispersion of Poincaré waves is also obtained by a more abstract and more robust method using Mourre estimates. Only some partial results however are obtained concerning the Rossby propagation, as the two dimensional setting complicates very much the study of the dynamical system.

Keywords Semiclassical analysis · Microlocal analysis · Mourre estimates · Geophysical flows

1 Introduction

This paper is a continuation of [2] (see also [1]) so before discussing the matter of this paper (in Sect. 2) let us review the contents of that work. We shall start by

I. Gallagher
Institut de Mathématiques de Jussieu, UMR 7586, Université Paris VII, 175, rue du Chevaleret, 75013 Paris, France
e-mail: Isabelle.Gallagher@math.jussieu.fr

T. Paul
CNRS and Centre de mathématiques Laurent Schwartz, École polytechnique, 91128 Palaiseau cedex, France
e-mail: thierry.paul@math.polytechnique.fr

L. Saint-Raymond (✉)
Université Paris VI and DMA École Normale Supérieure, 45 rue d'Ulm, 75230 Paris cedex 05, France
e-mail: Laure.Saint-Raymond@ens.fr

H. Holden, K.H. Karlsen (eds.), *Nonlinear Partial Differential Equations*, Abel Symposia 7, DOI 10.1007/978-3-642-25361-4_13,

recalling briefly the model, then we shall explain the methods and results obtained in [2] and discuss their limitations.

1.1 The Model

The goal of [2] is to understand, through the study of a toy model, the persistence of oceanic eddies observed long past by physicists among which [4–6, 11, 12], who gave heuristic arguments to explain their formation due both to wind forcing and to convection by a macroscopic current.

The ocean is considered in this toy model as an incompressible, inviscid fluid with free surface submitted to gravitation and wind forcing, and we further make the following classical assumptions: we assume that the density of the fluid is homogeneous $\rho = \rho_0 =$ constant, that the pressure law is given by the hydrostatic approximation $p = \rho_0 g z$, and that the motion is essentially horizontal and does not depend on the vertical coordinate. This leads to the so-called shallow water approximation [8].

For the sake of simplicity, the effects of the interaction with the boundaries are not discussed and the model is purely horizontal with the longitude x_1 and the latitude x_2 both in $\mathbf{R}$.

The evolution of the water height h and velocity v is then governed by the shallow-water equations with Coriolis force

$$\begin{aligned} \partial_t(\rho_0 h) + \nabla\cdot(\rho_0 h v) = 0, \\ \partial_t(\rho_0 h v) + \nabla\cdot(\rho_0 h v\otimes v) + \omega(\rho_0 h v)^\perp + \rho_0 g h\nabla h = \rho_0 h\tau \end{aligned} \tag{1}$$

where ω denotes the vertical component of the Earth rotation vector Ω, $v^\perp := (-v_2, v_1)$, g is the gravity and τ is the—stationary—forcing responsible for the macroscopic flow. The vertical component of the Earth rotation is therefore $\Omega\sin(x_2/R)$, where R is the radius of the Earth; note that it is classical in the physical literature to consider the linearization of ω (known as the betaplane approximation) $\omega(x_2) = \Omega x_2/R$. We consider general functions ω in the sequel, with some restrictions that are be made precise later.

We consider small fluctuations (η, u) around the stationary solution $(\bar h, \bar v)$ satisfying

$$\bar h = \text{constant}, \qquad \nabla\cdot(\bar v\otimes\bar v) + \omega\bar v^\perp = \tau, \qquad \operatorname{div}\bar v = 0.$$

In [2] the study is restricted to the case of a shear flow, in the sense that $\bar v(x) = (\bar v_1(x_2), 0)$, with $\bar v_1$ a smooth, compactly supported function. Some orders of magnitude and scalings allow to transform the previous system into the following one:

$$\begin{aligned} \partial_t\eta + \frac{1}{\varepsilon}\nabla\cdot u + \bar u\cdot\nabla\eta + \varepsilon^2\nabla\cdot(\eta u) = 0, \\ \partial_t u + \frac{1}{\varepsilon^2} b u^\perp + \frac{1}{\varepsilon}\nabla\eta + \bar u\cdot\nabla u + u\cdot\nabla\bar u + \varepsilon^2 u\cdot\nabla u = 0 \end{aligned} \tag{2}$$

where $b := \omega/|\Omega|$ and with ε of the order of both Fr^2 and $\mathrm{Ro}^{1/2}$, where the Froude number Fr and the Rossby number Ro are nondimensional parameters measuring respectively the influence of gravity and of the Coriolis force. For the large scale motions under consideration, we have typically $\mathrm{Fr}^2 := (v_0\ell_0)/(t_0 g\delta h) \sim 0.1$ and $\mathrm{Ro} := 1/(t_0|\Omega|) \sim 0.01$, where $t_0 \sim 10^6$ s (~ 0.38 months) and $\ell_0 \sim 10^4$ km are the typical time and length scales, while $\delta h = (h - \bar{h})/\eta \ \sim 1$ m is the typical height fluctuation, and $v_0 \sim 0.1\ \mathrm{m\,s^{-1}}$ is the typical velocity fluctuation: we write $u = (v - \bar{v})/v_0$.

1.2 Methods and Results in [2]

Most of the analysis in [2] concerns the linear version of (2), namely the following system:

$$\varepsilon^2 i\partial_t \mathbf{v} + A(x_2, \varepsilon D, \varepsilon)\mathbf{v} = 0, \quad \mathbf{v} = (v_0, v_1, v_2), \tag{3}$$

where $D := \frac{1}{i}\partial$, and the linear propagator is given by

$$A(x_2, \varepsilon D, \varepsilon) := i \begin{pmatrix} \varepsilon\bar{u}_1\varepsilon\partial_1 & \varepsilon\partial_1 & \varepsilon\partial_2 \\ \varepsilon\partial_1 & \varepsilon\bar{u}_1\varepsilon\partial_1 & -b(x_2) + \varepsilon^2\bar{u}_1' \\ \varepsilon\partial_2 & b(x_2) & \varepsilon\bar{u}_1\varepsilon\partial_1 \end{pmatrix}.$$

The first step of the analysis consists in diagonalizing (approximately) the system (3). The computation of a kind of **characteristic polynomial** associated with (3), in symbolic form, allows to construct three symbols the quantization of which provides three scalar propagators (this will be explained more explicitly below).

Two of those propagators, called Poincaré propagators, are then proved to satisfy dispersive estimates; that result relies on a spectral analysis (usual semi-classical theory does not operate here due to the very large time scales at play) using **global quantum normal forms**, which requires that b has at most one, nondegenerate critical value and which also uses very much the fact that the motion is translation-invariant in x_1. A stationary phase argument on the spectral decomposition of any solution to the Poincaré propagation gives the result: Poincaré modes exit any compact set in finite time.

The last propagator is the Rossby one, which is one order of magnitude (in ε) smaller than the Poincaré modes. This allows to analyse the propagation by semi-classical analysis tools. In particular the precise study of the dynamical system associated with those waves, which is an **integrable system** due to translation invariance in x_1, allows to derive a condition on the initial microlocalization of the solution which guarantees that the Rossby waves are trapped for all times in a compact set.

Those results on the linear system (3) can finally be transposed to the original system (2) due to the high power of ε in front of the nonlinearity, and due to the semi-classical setting, which allows to exhibit vector fields which almost-commute with the linear operator $A(x_2, \varepsilon D, \varepsilon)$.

1.3 Limitations of the Methods of [2]

The restriction which is the most used in the analysis described briefly in the previous paragraph is the fact that the stationary flow $\bar{u}$ is a shear flow of the type $\bar{u} = (\bar{u}_1(x_2), 0)$. Indeed

- It allows to Fourier-transform in the direction x_1, which makes the diagonalization procedure much easier;
- It simplifies the spectral analysis of Poincaré waves, again due to the Fourier transform (in particular the dual variable ξ_1 is fixed during the propagation, and there is a wave-like behavior in x_1);
- It allows the Rossby dynamical system to be integrable, which is a tremendous help in the analysis.

An additional restriction in the previous arguments is that in order to prove the dispersion of Poincaré waves, the rotation amplitude b should have at most one, nondegenerate critical value: this allows to use a Bohr–Sommerfeld quantization argument to compute the eigenvalues of the Poincaré operator. This assumption on b is not really restrictive from the physical point of view. On the other hand, it is important for physical reasons to consider 2D convection flows.

1.4 On the Nonlinear Term

As explained above, most of the analysis in [2] is concerned with the linear system (3). In order to transpose the linear results to the nonlinear setting, one uses the following arguments (along with the fact that the coupling is vanishing when ε goes to zero):

- Uniform existence which is obtained via an almost-commutation result;
- Bilinear estimates in anisotropic semi-classical spaces;
- A Gronwall lemma, which requires an $L^\infty(\mathbf{R}^2)$ bound on the linear solution. This is not known in general, due to the bad Sobolev embeddings in semi-classical settings, so the nonlinear result is proved for vanishing couplings only.

It is important to notice that none of those three steps require that $\bar{u}$ is a shear flow. In the whole of this paper we shall therefore only focus on the linear equation, and leave to the reader the transposition to the nonlinear equation, using the above steps.

2 Main Result of This Paper and Strategy of the Proof

2.1 The Model

In this paper we shall be concerned with the linear system

$$\varepsilon^2 i\partial_t \mathbf{v} + A(x, \varepsilon D, \varepsilon)\mathbf{v} = 0, \quad \mathbf{v} = (v_0, v_1, v_2), \tag{4}$$

where the linear propagator is given by

$$A(x,\varepsilon D,\varepsilon) := i\begin{pmatrix} \varepsilon\bar{u}\cdot\varepsilon\nabla & \varepsilon\partial_1 & \varepsilon\partial_2 \\ \varepsilon\partial_1 & \varepsilon\bar{u}\cdot\varepsilon\nabla+\varepsilon^2\partial_1\bar{u}_1 & -b+\varepsilon^2\partial_2\bar{u}_1 \\ \varepsilon\partial_2 & b+\varepsilon^2\partial_1\bar{u}_2 & \varepsilon\bar{u}\cdot\varepsilon\nabla+\varepsilon^2\partial_2\bar{u}_2 \end{pmatrix}. \tag{5}$$

We shall assume throughout the paper that b is smooth, with a symbol-like behavior: for all $\alpha\in\mathbf{N}$, there is a constant C_α such that for all $x_2\in\mathbf{R}$,

$$|b^{(\alpha)}(x_2)| \le C_\alpha\big(1+b^2(x_2)\big)^{\frac{1}{2}}. \tag{6}$$

We shall further assume that

$$\lim_{|x_2|\to\infty} b^2(x_2) = \infty,$$

and that b^2 has only nondegenerate critical points.

We shall also suppose that the initial data is microlocalized in some compact set $\mathscr{C}$ of $T^*\mathbf{R}^2$ satisfying

$$\mathscr{C}\cap\{\xi_1^2+\xi_2^2+b^2(x_2)=0\}=\emptyset \tag{7}$$

or actually rather

$$\mathscr{C}\cap\{\xi_1=0\}=\emptyset. \tag{8}$$

We shall prove that assumption (7) is propagated by the flow, while (8) is propagated only by the Poincaré component. We recall (see for instance [2], Appendix B) that a function f is microlocalized in a compact set $\mathscr{C}$ of $T^*\mathbf{R}^2$ if for any (x_0,ξ_0) in the complement of $\mathscr{C}$ in $\mathbf{R}^4$ (we shall identify $T^*\mathbf{R}^2$ to $\mathbf{R}^4$ in the following), there is a smooth function χ_0, bounded as well as all its derivatives and equal to one at (x_0,ξ_0), satisfying

$$\|\mathrm{Op}_\varepsilon^W(\chi_0)u_0\|_{L^2(\mathbf{R}^2)} = O(\varepsilon^\infty), \tag{9}$$

where $\mathrm{Op}_\varepsilon^W$ denotes the Weyl quantization:

$$\mathrm{Op}_\varepsilon^W(\chi_0)u_0(x) := \frac{1}{(2\pi\varepsilon)^4}\int e^{i(x-y)\cdot\xi/\varepsilon}\chi_0\Big(\frac{x+y}{2},\xi\Big)u_0(y)\,dy\,d\xi. \tag{10}$$

We also recall that (9) means that for any $N\in\mathbf{N}$, there are ε_0 and C such that

$$\forall\varepsilon\in]0,\varepsilon_0],\quad \|\mathrm{Op}_\varepsilon^W(\chi_0)u_0\|_{L^2(\mathbf{R}^2)}\le C\varepsilon^N.$$

In the following, to simplify some formulations, we shall denote by $(\mu)\,\mathrm{Supp}_\star f$ the projection of the (micro)support of f onto the $\star=0$ axis, where $\star$ represents an element of $\{x_1,x_2,\xi_1,\xi_2\}$.

2.2 Statement of the Main Result and Organization of the Paper

Let us state the main theorem proved in this paper.

Theorem 1 *Let* $\mathbf{v}_{\varepsilon,0}$ *be a family of initial data, microlocalized in a compact set* $\mathscr{C}$ *satisfying Assumption* (8). *For any parameter* $\varepsilon > 0$, *denote by* $\mathbf{v}_\varepsilon$ *the associate solution to* (4). *Then for all* $t \geq 0$ *one can write* $\mathbf{v}_\varepsilon(t)$ *as the sum of a "Rossby" vector field and a "Poincaré" vector field*: $\mathbf{v}_\varepsilon(t) = \mathbf{v}_\varepsilon^R(t) + \mathbf{v}_\varepsilon^P(t)$, *satisfying the following properties*:

1. $\mu\,\mathrm{Supp}\,\mathbf{v}_\varepsilon^R(t)$ *and* $\mu\,\mathrm{Supp}\,\mathbf{v}_\varepsilon^P(t)$ *satisfy* (7) *for all times.*
2. *For any compact set* Ω *in* $\mathbf{R}^2$, *one has*

$$\forall t > 0, \quad \|\mathbf{v}_\varepsilon^P(t)\|_{L^2(\Omega)} = O(\varepsilon^\infty).$$

3. $\mu\,\mathrm{Supp}_{x_2}\,\mathbf{v}_\varepsilon^R(t)$ *lies in a bounded subset of* $\mathbf{R}$ *uniformly in time.*

Compared to [2], the main difficulties are due to the presence of an x_1-dependent underlying flow $\bar{u}$. The diagonalization of the system (exhibiting Rossby and Poincaré-type waves, with very different qualitative features) must be revised, and obtained in a less explicit way. Moreover the proof of (2) in Theorem 1, namely the dispersion of Poincaré waves can also not be proved in the same way (note that it is not assumed here that b^2 has at most one nondegenerate critical value). Finally the trapping of Rossby waves seems much harder to obtain since the underlying dynamical system decouples no more; the behavior of the Rossby waves is therefore much less precise than in [2].

Let us explain our strategy here, compared with that in [2] described above.

The Diagonalization

The construction of the Rossby and Poincaré modes is not as direct as in [2] due to the lack of translation invariance in x_1. We choose therefore to follow a more abstract way to recover those modes in Sect. 3, which relies on **semi-classical analysis, and normal forms** (instead of explicit computations as in [2]). Finding the propagators associated with those modes requires a microlocalization assumption of the type (7), in order for the eigenvalues of the matrix of principal symbols to be well separated (see for instance [10] for a related result). The diagonalization result is therefore in this paragraph conditional to the fact that the solution to the propagation equation is correctly microlocalized (that corresponds to Point 1 of Theorem 1).

Dispersion of Poincaré Waves and Propagation of the Nondegeneracy Assumption (8)

In order to prove (2) in Theorem 1 we again rely on a more abstract, and more efficient method than that followed in [2]. It is based on Mourre estimates and on

Assumption (8) on the initial data: we start by proving, by a semi-classical argument, that after a very short time (of the order of ε) the support in x_1 of the solution escapes the support of $\bar{u}$. Then we use **Mourre estimates** to prove that the solution remains outside the support of $\bar{u}$ for all times, and actually escapes any compact set in x_1 in finite time (to prove this last point we use the fact that the equation reduces to a translation-invariant equation in x_1 since the support of the solution is outside the support in x_1 of $\bar{u}$). This allows finally to check that the nondegeneracy assumption (8) does hold for all times. This analysis is achieved in Sect. 4.

Study of Rossby Waves and Propagation of the Nondegeneracy Assumption (7)

In Sect. 5 we first prove that the nondegeneracy assumption (7) does hold during the propagation of Rossby waves. That is due to semi-classical analysis, by the study of the dynamical system associated with those waves. The study of that system is also the key to the proof of Point (3), which is also proved in Sect. 5.

3 Reduction to Scalar Propagators

In this section we shall construct three operators T_+, T_- and T_R diagonalizing $A(x, \varepsilon D, \varepsilon)$.

We shall start by proving a general diagonalization result, and at the end we shall apply the general result to our context.

Before stating the general result, let us give some notation. A semi-classical symbol is a function $a = a(x, \xi; \varepsilon)$ defined on $\mathbf{R}^{2d} \times]0, \varepsilon_0]$ for some $\varepsilon_0 > 0$, which depends smoothly on (x, ξ) and such that for any $\alpha \in \mathbf{N}^{2d}$ and any compact set $\mathscr{K} \subset \mathbf{R}^{2d}$, there is a constant C such that for any $((x, \xi), \varepsilon) \in \mathscr{K} \times]0, \varepsilon_0]$,

$$|\partial^\alpha a((x, \xi), \varepsilon)| \leq C.$$

We shall consider the Weyl quantization of such symbols, as recalled in (10): for all u is in $\mathscr{D}(\mathbf{R}^d)$,

$$\mathrm{Op}_\varepsilon^W(a)u(x) := \frac{1}{(2\pi\varepsilon)^d} \int e^{i(x-y)\cdot\xi/\varepsilon} a\left(\frac{x+y}{2}, \xi\right) u(y)\, dy\, d\xi.$$

We shall denote the principal symbol of a pseudo-differential operator A by $\sigma_p(A)$. We shall say that a pseudodifferential operator $\mathrm{Op}_\varepsilon^W(a)$ is supported in a set $\mathscr{K}$ if for any smooth function χ equal to one in a neighborhood of $\mathscr{K}$ one has $a\chi = a$.

Finally we shall say that a matrix is pseudodifferential if each of its entries is a pseudodifferential operator.

Let us first prove the following general result.

Theorem 2 *Let $\mathscr{K}$ be a compact subset of $\mathbf{R}^{2d}$, and consider an $N \times N$ hermitian pseudodifferential matrix $A_\varepsilon = A(x, \varepsilon D, \varepsilon)$, supported in $\mathscr{K}$. Assume that*

- *the (matrix) principal symbol of $A(x, \varepsilon D, 0)$, denoted by $\mathscr{A}_0$, is diagonalizable, in the sense that there are some unitary and diagonal matrices of symbols, $\mathscr{U}$ and $\mathscr{D}$, such that*

$$\mathscr{U}^{-1} \mathscr{A}_0 \mathscr{U} = \mathscr{D},$$

- *the eigenvalues $(\delta_1(x,\xi), \dots, \delta_N(x,\xi))$ satisfy*

$$\forall i \neq j, \qquad \inf_{(x,\xi)\in\mathscr{K}} |\delta_i(x,\xi) - \delta_j(x,\xi)| \geq C > 0. \tag{11}$$

Then there exists a family of unitary and diagonal pseudodifferential operators V_ε and D_ε supported in $\mathscr{K}$, such that:

$$V_\varepsilon^* A_\varepsilon V_\varepsilon = D_\varepsilon + O(\varepsilon^\infty), \qquad V_\varepsilon^* V_\varepsilon = I + O(\varepsilon^\infty). \tag{12}$$

Moreover one has

$$D_\varepsilon = D_0 + \varepsilon D_1 + O(\varepsilon^2), \tag{13}$$

where $D_0 = \mathrm{Op}_\varepsilon^W(\mathscr{D})$ and the principal symbol of D_1 is given by

$$\mathscr{D}_1 = \sigma_p(D_1) = \mathrm{diag}\left(\sigma_p\left(\widetilde{\Delta}_1 - \frac{D_0 I_1 + I_1 D_0}{2}\right)\right)$$

with the notations

$$\begin{aligned} \widetilde{\Delta}_1 &= \frac{1}{\varepsilon}\big(\mathrm{Op}_\varepsilon^W(\mathscr{U}^*) A_\varepsilon \,\mathrm{Op}_\varepsilon^W(\mathscr{U}) - D_0\big), \\ I_1 &= \frac{1}{\varepsilon}\big(\mathrm{Op}_\varepsilon^W(\mathscr{U}^*)\, \mathrm{Op}_\varepsilon^W(\mathscr{U}) - I\big). \end{aligned} \tag{14}$$

More explicitly, let us denote by $a_{ij}(x,\xi)$ the matrix elements of $\mathscr{A}_1$, subsymbol of $A(x, \varepsilon D, \varepsilon)$ defined by:

$$\mathscr{A}_1 := \sigma_p(\partial_\varepsilon A)$$

and by $u_{nj}(x,\xi)$, $j = 1 \dots d$, the coordinates of any unit eigenvector of $\mathscr{A}_0(x,\xi)$ of eigenvalue $\delta_n(x,\xi)$. We have

$$\begin{aligned} (\mathscr{D}_1)_{nn} = &\sum_{j,k=1\dots d}\left(\Im(\overline{u_{jn}}\{a_{jk}, u_{kn}\}) + \frac{a_{jk}\{\overline{u_{jn}}, u_{kn}\}}{2i}\right) \\ &+ (\mathscr{U}^* \mathscr{A}_1 \mathscr{U})_{nn} + \frac{1}{2i}\sum_{j=1}^d \delta_n\{u_{jn}, \overline{u_{jn}}\}, \end{aligned} \tag{15}$$

where $\{f,g\} := \nabla_\xi f \nabla_x g - \nabla_x f \nabla_\xi g$ is the Poisson bracket on $T^*\mathbf{R}^d$.

Here and in all the sequel, we say that a pseudo-differential operator V is unitary if it satisfies

$$V^*V = I + O(\varepsilon^\infty).$$

The proof is divided into two parts: in Sect. 3.1 we present the formal construction and in Sect. 3.2 we show that the symbols of the various operators formally constructed are indeed symbols. Finally Sect. 3.3 is devoted to the case of the matrix given by (5).

3.1 The Formal Construction

The proof of Theorem 2 is a combination of semiclassical and perturbation methods. Let us start by defining

$$U_0 = \mathrm{Op}_\varepsilon^W(\mathscr{U}).$$

Elementary properties of the Weyl quantization imply then that $U_0^* A_\varepsilon U_0 = D_0 + O(\varepsilon)$.

The following proposition shows that one can construct a unitary pseudodifferential operator U_∞ such that

$$U_\infty^* A_\varepsilon U_\infty = D_0 + O(\varepsilon).$$

Lemma 3.1 *Let U be a pseudodifferential matrix such that $U^*U = I + \varepsilon I_1$, where I is the identity. Then one can find $V \sim \sum_{k=0}^\infty \varepsilon^k V_k$ such that*

$$(U + \varepsilon V)^*(U + \varepsilon V) = I + O(\varepsilon^\infty). \tag{16}$$

Proof Let us denote $V_0 := -\frac{1}{2} U I_1$. On easily checks that $(U + \varepsilon V_0)^*(U + \varepsilon V_0) = I + O(\varepsilon^2)$. Indeed

$$\begin{aligned}(U + \varepsilon V_0)^*(U + \varepsilon V_0) &= U^*U - \frac{\varepsilon}{2}(I_1 U^*U + U^*U I_1) + O(\varepsilon^2)\\ &= I + \varepsilon I_1 - \varepsilon I_1 + O(\varepsilon^2).\end{aligned}$$

Then one concludes by iteration. □

That lemma allows to define the pseudo-differential operator of (semiclassical) order 0

$$\Delta_1 = \frac{1}{\varepsilon}(U_\infty^* A_\varepsilon U_\infty - D_0),$$

where U_∞ is a unitary operator.

Now our aim is to find a unitary operator V_∞ (up to $O(\varepsilon^\infty)$) such that

$$(U_\infty V_\infty)^* A_\varepsilon (U_\infty V_\infty) = D_\infty + O(\varepsilon^\infty),$$

where $D_\infty = D_0 + \varepsilon D_1 + \cdots$ is a diagonal matrix satisfying the conclusions of the theorem.

We shall write $V_\infty = e^{i\varepsilon W}$, with W selfadjoint (so V_∞ thus constructed is automatically unitary). We recall that if W is a pseudodifferential operator, then so is $e^{i\varepsilon W}$ (simply by writing $e^{i\varepsilon W} \sim \sum_0^\infty \frac{(i\varepsilon)^k}{k!} W^k$). We look for W under the form $W \sim \sum_0^\infty \varepsilon^k W_k$, and compute the W_k recursively. Since

$$V_\infty^*(D_0 + \varepsilon\Delta_1)V_\infty = (D_0 + \varepsilon\Delta_1) + i\varepsilon[(D_0 + \varepsilon\Delta_1), W] + \frac{(i\varepsilon)^2}{2}[[(D_0 + \varepsilon\Delta_1), W], W] + \cdots$$

we see that, if W_1 satisfies

$$i[D_0, W_1] + \Delta_1 = D_1 + O(\varepsilon), \quad D_1 \text{ diagonal}, \tag{17}$$

then we have that

$$e^{-i\varepsilon W_1}(D_0 + \varepsilon\Delta_1)e^{i\varepsilon W_1} = D_0 + \varepsilon D_1 + \varepsilon^2\Delta_2, \tag{18}$$

where Δ_2 is a zero order pseudodifferential operator. The following lemma is a typical normal form type result, and is crucial for the following.

Lemma 3.2 *Let D_0 be a diagonal pseudodifferential matrix whose principal symbol $\mathscr{D}_0$ has a spectrum satisfying* (11) *and let Δ_1 be a pseudodifferential matrix.*

Then there exist two pseudodifferential matrices W and D_1, with D_1 diagonal, such that:

$$[D_0, W] + \Delta_1 = D_1 + \varepsilon\widetilde{\Delta}_2, \tag{19}$$

where $\widetilde{\Delta}_2$ is a pseudodifferential matrix of order 0.

Moreover the principal symbol of D_1 is the diagonal part of the principal symbol of Δ_1: we have $\sigma_p(D_1) = \operatorname{diag}\sigma_p(\Delta_1)$.

Proof By the nondegeneracy condition of the spectrum of $\mathscr{D}_0$ we know, by standard arguments (see [13] for instance), that there exists a matrix $\mathscr{W}_0$ and a diagonal one $\mathscr{D}_1$ such that

$$[\mathscr{D}_0, \mathscr{W}_0] + \mathscr{D}_{1,0} = \mathscr{D}_1,$$

where $\mathscr{D}_{1,0}$ is the principal symbol of Δ_1. Indeed it is enough to take $\mathscr{D}_1$ as the diagonal part of $\mathscr{D}_{1,0}$ and

$$(\mathscr{W}_0(x,\xi))_{i,j} = \frac{(\mathscr{D}_{1,0}(x,\xi))_{i,j}}{\delta_i(x,\xi) - \delta_j(x,\xi)} \tag{20}$$

and notice that the Weyl quantization of $\mathscr{W}_0$ satisfies (19). ☐

By Lemma 3.2 we know that there exists W_1 satisfying (17). Writing

$$e^{-i\varepsilon W_1}(D_0+\varepsilon\Delta_1)e^{i\varepsilon W_1} = D_0+\varepsilon(\Delta_1+[D_0,W_1])+\varepsilon^2(\Delta_2-\tilde{\Delta}_2),$$

we get immediately (18). It is easy to get convinced that all the W_k will satisfy recursively an equation of the form

$$[D_0,W_k]+\Delta_k = D_k+O(\varepsilon),$$

which can be solved by Lemma 3.2.

The expression for the principal symbol of D_1 follows by construction and the following well known lemma (see [9] for instance):

Lemma 3.3 *Let a and b two symbols. Then the principal symbol of* $\mathrm{Op}_\varepsilon^W(a)\,\mathrm{Op}_\varepsilon^W(b)$ *is ab and its subprincipal symbol is* $\frac{1}{2i}\{a,b\}$.

In order to derive (15) we have to compute the subprincipal symbol of the diagonal part of the right-hand side of (14), that is, for each $n=1\dots d$ and using Lemma 3.3,

$$\sum_{jk}\mathrm{Op}_\varepsilon^W(\overline{\mathscr{U}_{jn}})\,\mathrm{Op}_\varepsilon^W((\mathscr{A}_0+\varepsilon\mathscr{A}_1)_{jk})\,\mathrm{Op}_\varepsilon^W(\mathscr{U}_{kn})-\frac{1}{2i}\sum_{j=1}^{d}\delta_n\{\overline{u_{jn}},u_{jn}\},$$

since $\mathscr{U}$ is unitary.

The term $\varepsilon\mathscr{A}_1$ is obviously responsible for the second term in the right-hand side of (15). Using Lemma 3.3 and the distributivity of the Poisson bracket, we get the following expression for the first one:

$$\sum_{jk}\frac{1}{2i}\big(\{\overline{\mathscr{U}_{jn}},(\mathscr{A}_0)_{jk}\mathscr{U}_{kn}\}+\overline{\mathscr{U}_{jn}}\{(\mathscr{A}_0)_{jk},\mathscr{U}_{kn}\}\big)$$
$$=\sum_{jk}\frac{1}{2i}\big(\overline{\mathscr{U}_{jn}}\{(\mathscr{A}_0)_{jk},\mathscr{U}_{kn}\}+(\mathscr{A}_0)_{jk}\{\overline{\mathscr{U}_{jn}},\mathscr{U}_{kn}\}+\mathscr{U}_{kn}\{\overline{\mathscr{U}_{jn}},(\mathscr{A}_0)_{jk}\}\big).$$

Inverting j and k in half of the terms and noticing that, since $\mathscr{A}_0$ is Hermitian, $(\mathscr{A}_0)_{jk}=\overline{(\mathscr{A}_0)_{kj}}$, we get easily (15).

3.2 Symbolic Properties

With the hypothesis that both $\mathscr{D}$ and $\mathscr{U}$ are pseudodifferential matrices it is quite obvious that V_ε and D_ε are pseudodifferential matrices as well. Indeed the formal construction in the preceding section shows that the iterative process uses only three tools: multiplications of pseudodifferential operators, computation of subprincipal symbols and solving Eq. (19). For (19), the formula (20) used in the proof of

Lemma 3.2, together with the nondegeneracy condition (11) which shows clearly that $(\delta_i(x,\xi)-\delta_j(x,\xi))^{-1}$ is a symbol, implies that W_0 is a pseudodifferential operator.

Note that the microlocalization assumption is crucial in order that the expansions obtained by this iterative construction do define symbols. We have indeed no uniform control on the growth at infinity.

3.3 The Rossby–Poincaré Case

In the case of oceanic waves $A(x,\varepsilon D,\varepsilon)$ is given by (5):

$$A(x,\varepsilon D) := i \begin{pmatrix} \varepsilon\bar{u}\cdot\varepsilon\nabla & \varepsilon\partial_1 & \varepsilon\partial_2 \\ \varepsilon\partial_1 & \varepsilon\bar{u}\cdot\varepsilon\nabla+\varepsilon^2\partial_1\bar{u}_1 & -b+\varepsilon^2\partial_2\bar{u}_1 \\ \varepsilon\partial_2 & b+\varepsilon^2\partial_1\bar{u}_2 & \varepsilon\bar{u}\cdot\varepsilon\nabla+\varepsilon^2\partial_2\bar{u}_2 \end{pmatrix}.$$

Therefore

$$\mathscr{A}_0(x,\xi) := \begin{pmatrix} 0 & \xi_1 & \xi_2 \\ \xi_1 & 0 & -ib \\ \xi_2 & ib & 0 \end{pmatrix},$$

and

$$\mathscr{A}_1(x,\xi) := \begin{pmatrix} \bar{u}\cdot\xi & 0 & 0 \\ 0 & \bar{u}\cdot\xi & 0 \\ 0 & 0 & \bar{u}\cdot\xi \end{pmatrix} = \bar{u}\cdot\xi \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

A straightforward computation shows that the spectrum of $\mathscr{A}_0$ is

$$\left\{0, \sqrt{\xi_1^2+\xi_2^2+b^2(x_2)}, -\sqrt{\xi_1^2+\xi_2^2+b^2(x_2)}\right\}.$$

Microlocalization

The three eigenvalues of $\mathscr{A}_0$ are separated if and only if

$$\xi_1^2+\xi_2^2+b^2(x_2)\neq 0.$$

Therefore, considering a compact subset $\mathscr{K}$ of $\mathbf{R}^4$ such that

$$\mathscr{K}\cap\{(x_1,x_2,\xi_1,\xi_2)\,/\,\xi_1^2+\xi_2^2+b^2(x_2)=0\}=\emptyset$$

ensures that

- the eigenvalues do not cross, so that it is possible to get a unitary diagonalizing matrix with regular entries;
- the nondegeneracy condition (11) is satisfied.

In other words, $A(x, \varepsilon D, \varepsilon)$ satisfies the assumptions of Theorem 2 provided that one considers only its action on vector fields which are suitably microlocalized.

We assume of course that this microlocalization condition is satisfied by the initial datum, which is the condition (7).

Furthermore, we shall prove in the next two sections that the propagation by the scalar operators $T_\pm$ and T_R (to be defined now) preserves this suitable microlocalization, thus justifying a posteriori the diagonalization procedure for all times.

Computation of the Poincaré and Rossby Hamiltonians

The above computations show that one can define the two Poincaré Hamiltonians as follows:

$$\tau_\pm := \pm\sqrt{\xi_1^2 + \xi_2^2 + b^2(x_2)}$$

and we shall denote the associate operator constructed via Theorem 2 by $T_\pm$.

Now let us consider the Rossby Hamiltonian. In all this paragraph, for the sake of readability, we shall denote

$$\langle\xi\rangle_b := \sqrt{\xi_1^2 + \xi_2^2 + b^2(x_2)}.$$

An easy computation shows that a (normalized) eigenvector of $\mathscr{A}_0(x, \xi)$ of zero eigenvalue is

$$u_0 = \frac{1}{\langle\xi\rangle_b}\begin{pmatrix} b \\ i\xi_2 \\ -i\xi_1 \end{pmatrix}.$$

By Theorem 2, the Rossby Hamiltonian is then given by the formula

$$\tau_R = \sum_{j,k=1\dots3}\left(\Im(\overline{u_{j0}}\{a_{jk}, u_{k0}\}) + \frac{a_{jk}\{\overline{u_{j0}}, u_{k0}\}}{2i}\right) + \sum_{j,k=1\dots3}(\mathscr{A}_1)_{jk}\overline{u_{j0}}u_{k0}. \quad (21)$$

In order to compute the different Lie brackets, we start with a couple of simple remarks:

$$\{\xi_j, f\} = \partial_{x_j} f \quad \text{and} \quad \{b, f\} = -b'\partial_{\xi_2} f.$$

In particular, if f does not depend on x_1, then $\{\xi_1, f\} = 0$.

The contribution of the first term in the parenthesis in (21) is

$$\begin{aligned}&\sum_{j,k=1\dots3}(\overline{u_{j0}}\{a_{jk}, u_{k0}\}) \\ &\quad = \frac{b}{\langle\xi\rangle_b}\left\{\xi_2, \frac{-i\xi_1}{\langle\xi\rangle_b}\right\} + \frac{i\xi_2}{\langle\xi\rangle_b}\left\{ib, \frac{-i\xi_1}{\langle\xi\rangle_b}\right\} + \frac{i\xi_1}{\langle\xi\rangle_b}\left(\left\{\xi_2, \frac{b}{\langle\xi\rangle_b}\right\} + \left\{ib, \frac{i\xi_2}{\langle\xi\rangle_b}\right\}\right)\end{aligned}$$

$$= \frac{-ib\xi_1}{\langle\xi\rangle_b}\partial_{x_2}\frac{1}{\langle\xi\rangle_b} - \frac{i\xi_2\xi_1 b'}{\langle\xi\rangle_b}\partial_{\xi_2}\frac{1}{\langle\xi\rangle_b} + \frac{i\xi_1}{\langle\xi\rangle_b}\partial_{x_2}\frac{b}{\langle\xi\rangle_b} + \frac{i\xi_1 b'}{\langle\xi\rangle_b}\partial_{\xi_2}\frac{\xi_2}{\langle\xi\rangle_b}$$
$$= \frac{2i\xi_1 b'}{\langle\xi\rangle_b^2}.$$

Using the distributivity of the Poisson brackets, we get the contribution of the second term in a very similar way

$$\sum_{j,k=1\ldots 3} \frac{a_{jk}\{\overline{u_{j0}}, u_{k0}\}}{2}$$
$$= \xi_1\left\{\frac{b}{\langle\xi\rangle_b}, \frac{i\xi_2}{\langle\xi\rangle_b}\right\} - \xi_2\left\{\frac{b}{\langle\xi\rangle_b}, \frac{i\xi_1}{\langle\xi\rangle_b}\right\} + ib\left\{\frac{i\xi_1}{\langle\xi\rangle_b}, \frac{i\xi_2}{\langle\xi\rangle_b}\right\}$$
$$= \xi_1\left(\frac{ib}{\langle\xi\rangle_b}\left\{\frac{1}{\langle\xi\rangle_b}, \xi_2\right\} + \frac{i\xi_2}{\langle\xi\rangle_b}\left\{b, \frac{1}{\xi_b}\right\} + \frac{i}{\langle\xi\rangle_b^2}\{b, \xi_2\}\right) - \frac{i\xi_2\xi_1}{\langle\xi\rangle_b}\left\{b, \frac{1}{\langle\xi\rangle_b}\right\}$$
$$- \frac{ib\xi_1}{\langle\xi\rangle_b}\left\{\frac{1}{\langle\xi\rangle_b}, \xi_2\right\}$$
$$= -\frac{ib'\xi_1}{\langle\xi\rangle_b^2}.$$

The computation of the second term of the right hand side of (21) is trivial since $\mathscr{A}_1$ is a multiple of the identity.

Adding the two previous expressions we get finally

$$\tau_R = \frac{\xi_1 b'}{\xi_1^2 + \xi_2^2 + b(x_2)^2} + \bar{u}\cdot\xi$$

and the associate operator will be denoted by T_R.

Remark 3.4 Since the elementary steps of the diagonalization process use only multiplications, computations of subprincipal symbols and solving normal forms equations, all the subsymbols of T_R and $T_\pm$ depend on x_1 only through $\bar{u}$ and its derivatives.

4 Study of the Poincaré Waves

In Sect. 3 we constructed two linear operators, called $T_\pm$, whose principal symbols are

$$\tau_\pm = \pm\sqrt{\xi_1^2 + \xi_2^2 + b^2(x_2)}.$$

We now want to study the propagation equation associated to those operators, namely the following linear equation, in $\mathbf{R} \times \mathbf{R}^2$

$$i\varepsilon^2 \partial_t \varphi_\pm = T_\pm \varphi_\pm, \quad \varphi_{\pm|t=0} = \varphi_\pm^0 \tag{22}$$

where $\varphi_\pm^0$ are microlocalized in a compact set $\mathscr{C}$ satisfying Assumption (7). Before studying that equation we need to check that it makes sense, since a priori $T_\pm$ is only defined on vector fields microlocalized on such a compact set. This is achieved in the coming section, where we check that the separation of eigenvalues (11) required in the statement of Theorem 2 holds because $\sqrt{\xi_1^2 + \xi_2^2 + b^2(x_2)}$ remains bounded away from zero during the propagation.

Then we shall show that the solutions to these equations exit any compact set in finite time (Point (2) of Theorem 1).

4.1 Microlocalization

Let us prove the following result, which provides the first part of Point (1) in Theorem 1 and allows to make sense of Eq. (22) for all times.

Proposition 4.1 *Under the assumptions of Theorem 1, the operators $T_\pm$ are self-adjoint, and the function $\varphi(t) = e^{i\frac{t}{\varepsilon^2}T_\pm}\varphi_\pm^0$ are such that $\mu\,\mathrm{Supp}\,\varphi_\pm(t)$ satisfies (7) for all times.*

Proof The proof of that result relies on a spectral argument. Due to the form of the principal symbols of $T_\pm$ recalled above, the operators $T_\pm$ are self-adjoint (see [9]). We can therefore define two families $(\psi_n^\pm)_{n\in\mathbf{N}}$ of eigenvectors of $T_\pm$ in $L^2(\mathbf{R}^2)$ and two sequences of eigenvalues $\lambda_n^\pm$ such that if the initial data writes

$$\varphi_\pm^0(x) = \sum_n c_n^{\pm,0}\psi_n^\pm(x),$$

then

$$\varphi_\pm(t,x) = \sum_n e^{i\frac{\lambda_n^\pm t}{\varepsilon^2}} c_n^{\pm,0}\psi_n^\pm(x).$$

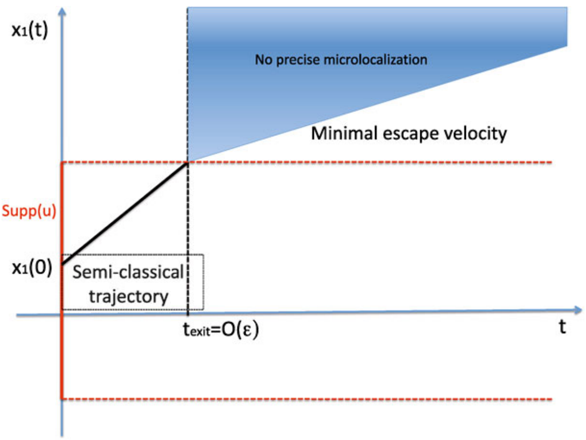
Since the eigenfunctions $\psi_n^\pm$ are microlocalized on the energy surfaces of the Poincaré Hamiltonians (see for instance [9]; Proposition 2.9.6), the result follows. □

4.2 Dispersion

In this paragraph we shall prove Point (2) of Theorem 1. The strategy is the following.

In Sect. 4.2 we prove using semi-classical analysis that for a very short time, the solutions to (22) remain microlocalized in a compact set satisfying assumption (8), and such that $\mu\,\mathrm{Supp}_{x_1}\,\varphi_\pm$ become disjoint from $\mathrm{Supp}_{x_1}\,\bar{u}$. Section 4.2 is then devoted to the long-time behavior of the solution, and Mourre estimates allow to prove that the solution exits any compact set after some time, and that it remains microlocalized far from $\xi_1 = 0$.

The result of the analysis carried out in this paragraph is that the behavior of $\mu\,\mathrm{Supp}_{x_1}\,\varphi_\pm$ is as depicted in the following figure.



Short time behavior

The aim of this paragraph is to prove the following result. It shows that the solutions of (22) exit the support of $\bar{u}$ after a time $t_{\mathrm{exit}}\varepsilon$, for $|t_{\mathrm{exit}}|$ large enough (independent of ε). We only state the forward in time result: the backwards result is identical, up to changing the sign of time. We shall further restrict the analysis to T_+ since the argument for T_- is identical, up to some sign changes.

Proposition 4.2 *Let φ^0 be a function, microlocalized in a compact set $\mathscr{C}$ satisfying Assumption (8), and let φ be the associate solution of (22). Let $[u_-, u_+]$ be a closed interval of $\mathbf{R}$ containing $\mathrm{Supp}_{x_1}\,\bar{u}$. There exists a constant $t_{\mathrm{exit}} > 0$ such that for any $\varepsilon \in]0, 1[$, the function $\varphi(\varepsilon t_{\mathrm{exit}}, \cdot)$ is microlocalized in a compact set $\mathscr{K}$ such that the projection of $\mathscr{K}$ onto the x_1-axis does not intersect $[u_-, u_+]$. Moreover $\mu\,\mathrm{Supp}_{\xi_1}\,\varphi$ is unchanged.*

More precisely, if $\mu\,\mathrm{Supp}_{\xi_1}\,\varphi^0 \subset \mathbf{R}^+ \setminus \{0\}$, then $\mu\,\mathrm{Supp}_{x_1}\,\varphi(\varepsilon t_{\mathrm{exit}}, \cdot) \subset]u_+, +\infty[$, and if $\mu\,\mathrm{Supp}_{\xi_1}\,\varphi^0 \subset \mathbf{R}^- \setminus \{0\}$, then $\mu\,\mathrm{Supp}_{x_1}\,\varphi(\varepsilon t_{\mathrm{exit}}, \cdot) \subset]-\infty, u_-[$.

Proof Define the function $\psi(s) := \varphi(\varepsilon s)$. Then (22) reads

$$i\varepsilon\partial_s\psi = T_+\psi, \quad \psi_{|s=0} = \varphi^0, \tag{23}$$

and any result proved on ψ on $[0, \mathscr{T}]$ will yield the same result for φ on $[0, \mathscr{T}\varepsilon]$. Notice that (23) is written in a semi-classical setting, so by the propagation of the microsupport theorem (see for instance [9], Theorem 4.3.7), the microsupport of ψ is propagated by the bicharacteristics, which are the integral curves of the principal symbol. Recall that the principal symbol of T_+ is

$$\tau_+(\xi_1, x_2, \xi_2) = \sqrt{\xi_1^2 + \xi_2^2 + b^2(x_2)}$$

and the bicharacteristics are given by the following set of ODEs:

$$\begin{cases} \dot{x}^t = \nabla_\xi \tau_+(\xi_1^t, x_2^t, \xi_2^t), & x^0 = (x_1^0, x_2^0), \\ \dot{\xi}^t = -\nabla_x \tau_+(\xi_1^t, x_2^t, \xi_2^t), & \xi^0 = (\xi_1^0, \xi_2^0). \end{cases}$$

Notice that τ_+ is independent of x_1, so $\dot{\xi}_1^t$ is identically zero and therefore $\xi_1^t \equiv \xi_1^0$. So for all $s \geq 0$, the microlocal support in ξ_1 of $\psi(s)$ remains unchanged, and in particular is far from $\xi_1 = 0$. Moreover one has

$$\dot{x}_1^t = \frac{\xi_1^0}{\sqrt{(\xi_1^0)^2 + (\xi_2^t)^2 + b^2(x_2^t)}}.$$

Now we recall that the bicharacteristic curves lie on energy surfaces, meaning that on each bicharacteristic, $\tau_+(\xi_1^0, x_2^t, \xi_2^t)$ is a constant. That implies that $(\xi_2^t)^2 + b^2(x_2^t)$ is a constant on each bicharacteristic, so that for all times,

$$\dot{x}_1^t \equiv \frac{\xi_1^0}{\sqrt{(\xi_1^0)^2 + (\xi_2^0)^2 + b^2(x_2^0)}}.$$

If $\xi_1^0 > 0$, then x_1 is propagated to the right and eventually escapes to the right of the support in x_1 of $\bar{u}$, whereas if $\xi_1^0 < 0$, the converse (to the left) occurs. Proposition 4.2 is proved. □

Long Time Behavior

The aim of this paragraph is to prove the following result, which again is only proved for positive times for simplicity.

Proposition 4.3 *Under the assumptions of Proposition 4.2, let φ^+ be the solution of (22) associated with the data $\varphi(\varepsilon t_{\mathrm{exit}}, \cdot)$. Then $\mu\,\mathrm{Supp}_{x_1}\,\varphi^+(t)$ does not intersect $\mathrm{Supp}_{x_1}\,\bar{u}$ for $t \geq \varepsilon t_{\mathrm{exit}}$, and $\mu\,\mathrm{Supp}_{\xi_1}\,\varphi^+(t)$ remains unchanged for $t \geq \varepsilon t_{\mathrm{exit}}$. Finally $\mu\,\mathrm{Supp}_{x_1}\,\varphi^+(t)$ exits any compact set in x_1 as soon as $t > \varepsilon t_{\mathrm{exit}}$.*

Proof Before going into the proof, we shall simplify the analysis by only studying the case of T_+ (the case T_- is obtained by identical arguments), and we shall only deal with the case when the support in ξ_1 of the data lies in the positive half space. The other case is obtained similarly.

The proof is based on Mourre's theory which we shall now briefly recall, and we refer to [7] and [3] for all details. Let us consider two self-adjoint operators H and A on a Hilbert space $\mathscr{H}$. We make the following assumptions:

1. The intersection of the domains of A and H is dense in the domain $\mathscr{D}(H)$ of H.
2. $t \mapsto e^{itA}$ maps $\mathscr{D}(H)$ to itself, and for all $\varphi^0 \in \mathscr{D}(H)$,

$$\sup_{t\in[0,1]} \|He^{itA}\varphi^0\| < \infty.$$

3. The operator $i[H, A]$ is bounded from below and closable, and the domain $\mathscr{D}(B_1)$ where iB_1 is its closure, contains $\mathscr{D}(H)$. More generally for all $n \in \mathbf{N}$ the operator $i[iB_n, A]$ is bounded from below and closable and the domain $\mathscr{D}(B_{n+1})$ of its closure iB_{n+1} contains $\mathscr{D}(H)$, and finally B_{n+1} extends to a bounded operator from $\mathscr{D}(H)$ to its dual.
4. There exists $\theta > 0$ and an open interval Δ of $\mathbf{R}$ such that if E_Δ is the corresponding spectral projection of H, then

$$E_\Delta i[H, A]E_\Delta \geq \theta E_\Delta. \tag{24}$$

Note that Assumptions (1–3) can be replaced by the fact that $[f(H), A]$ and all commutator iterates are bounded for any smooth, compactly supported function f (see [3]).

Under those assumptions, for any integer $m \in \mathbf{N}$ and for any $\theta' \in]0, \theta[$, there is a constant C such that

$$\|\chi_-(A - a - \theta' t)e^{-iHt}g(H)\chi_+(A - a)\| \leq Ct^{-m}$$

where $\chi_\pm$ is the characteristic function of $\mathbf{R}^\pm$, g is any smooth compactly supported function in Δ, and the above bound is uniform in $a \in \mathbf{R}$.

Let us apply this theory to our situation. We consider Eq. (22) with data $e^{i\frac{t_{\text{exit}}}{\varepsilon}T_+}\varphi_+^0$, and let us define the operator T_+^0 as the operator T_+ where $\bar{u}$ has been chosen identically zero. We shall start by studying the equation

$$i\varepsilon^2\partial_t\widetilde{\varphi} = T_+^0\widetilde{\varphi}, \qquad \widetilde{\varphi}_{|t=\varepsilon t_{\text{exit}}} = e^{i\frac{t_{\text{exit}}}{\varepsilon}T_+}\varphi_+^0, \tag{25}$$

for which we shall prove Proposition 4.3. Then we shall prove that the solution $\widetilde{\varphi}$ actually solves the original equation (22) with the same data $e^{i\frac{t_{\text{exit}}}{\varepsilon}T_+}\varphi_+^0$ at $t = \varepsilon t_{\text{exit}}$ up to $O(\varepsilon^\infty)$, because its support in x_1 lies outside the support of $\bar{u}$ and because the symbolic expansion of T_+ depends on x_1 only through $\bar{u}$ and its derivatives (see Remark 3.4).

So let us start by applying Mourre's theory to (25). Let us write the projection of $\mathscr{K}$ onto the ξ_1-axis as included in $[d_0, d_1]$ with $0 < d_0 < d_1 < \infty$. We recall that on the support of $e^{i\frac{t_{\text{exit}}}{\varepsilon}T_+}\varphi_+^0$, x_1 remains to the right of the support of $\bar{u}$. Then we apply the theory to $H = T_+^0$ and to $A = x_1$ (the pointwise multiplication). Assumptions (1) to (3) are easy to check, in particular because this is a semiclassical setting, so only the principal symbols need to be considered. Similarly finding a lower bound for $E_\Delta i[T_+^0, x_1]E_\Delta$ boils down to computing the Poisson bracket $\{\tau_+, x_1\}$ where

$$\{f, g\} = \nabla_\xi f \cdot \nabla_x g - \nabla_x f \cdot \nabla_\xi g,$$

and one finds

$$\{\tau_+, x_1\} = \frac{\xi_1}{\sqrt{\xi_1^2 + \xi_2^2 + b^2(x_2)}}. \tag{26}$$

Since T_+^0 has constant coefficients in x_1, ξ_1 is preserved, so in particular for all times one has

$$\mu\, \mathrm{Supp}_{\xi_1}\, \widetilde{\varphi}(t) \subset [d_0, d_1].$$

One can furthermore choose for Δ an interval of $\mathbf{R}$ of the type $]D_0, D_1[$ where the constants D_0 and D_1 are chosen so that for any $(x, \xi) \in \mathscr{K}$, one has

$$D_0 < \sqrt{\xi_1^2 + \xi_2^2 + b^2(x_2)} < D_1. \tag{27}$$

As the microlocal supports of the eigenfunctions of T_+^0 lie on energy surfaces, we know that the solution to (25) will remain in E_Δ for all times.

Now let us apply the results of [7] and [3]. By Lemma 3.3, (26), (27) and the assumption on ξ_1 written above, we have that

$$E_\Delta i[H, A]E_\Delta \geq \varepsilon \frac{d_0}{D_1} E_\Delta,$$

so (24) holds with $\theta = \varepsilon d_0/D_1$. It follows that the solution $e^{i\frac{(t-\varepsilon t_{\text{exit}})}{\varepsilon^2}T_+^0}(e^{i\frac{t_{\text{exit}}}{\varepsilon}T_+}\varphi_+^0)$ to (25) has a support in x_1 such that

$$x_1 > u_+ + \frac{d_0}{D_1}\frac{t}{\varepsilon}$$

which proves the result for (25).

Since $\mu\, \mathrm{Supp}_{x_1}\, e^{i\frac{(t-\varepsilon t_{\text{exit}})}{\varepsilon^2}T_+^0}(e^{i\frac{t_{\text{exit}}}{\varepsilon}T_+}\varphi_+^0)$ does not cross $\mathrm{Supp}_{x_1}\, \bar{u}$, one has actually

$$e^{i\frac{(t-\varepsilon t_{\text{exit}})}{\varepsilon^2}T_+^0}\left(e^{i\frac{t_{\text{exit}}}{\varepsilon}T_+}\varphi_+^0\right) = e^{i\frac{(t-\varepsilon t_{\text{exit}})}{\varepsilon^2}T_+}\left(e^{i\frac{t_{\text{exit}}}{\varepsilon}T_+}\varphi_+^0\right) \quad \text{in } L^2$$

locally uniformly in t (see Proposition 5.3 of the Appendix). The proposition follows. □

5 Propagation of the Rossby Waves

5.1 Semiclassical Transport Equations and Microlocalization

Because of the scaling of the Rossby Hamiltonian (which is smaller than the Poincaré Hamiltonians by one order of magnitude), on the times scales considered here the propagation of energy by Rossby waves is described by the Hamiltonian dynamics

$$\frac{dx_i}{dt} = \frac{\partial \tau_R}{\partial \xi_i}, \qquad \frac{d\xi_i}{dt} = -\frac{\partial \tau_R}{\partial x_i},$$

which can be written explicitly

$$\begin{aligned} \frac{dx_1}{dt} &= b'(x_2)\frac{\langle \xi \rangle_b^2 - 2\xi_1^2}{\langle \xi \rangle_b^4} + \bar{u}_1(x), \\ \frac{dx_2}{dt} &= -2b'(x_2)\frac{\xi_1\xi_2}{\langle \xi \rangle_b^4} + \bar{u}_2(x), \\ \frac{d\xi_1}{dt} &= -\partial_1\bar{u}_1(x)\xi_1 - \partial_1\bar{u}_2(x)\xi_2, \\ \frac{d\xi_2}{dt} &= \xi_1\frac{2b(b')^2 - b''\langle \xi \rangle_b^2}{\langle \xi \rangle_b^4} - \partial_2\bar{u}_1(x)\xi_1 - \partial_2\bar{u}_2(x)\xi_2 \end{aligned} \tag{28}$$

where we recall that $\langle \xi \rangle_b = \sqrt{\xi_1^2 + \xi_2^2 + b^2(x_2)}$. In order for the dynamics to be well defined and also in order to justify the diagonalization process, we need the quantity $\langle \xi \rangle_b$ to remain bounded from below for all times.

Proposition 5.1 *Let $\mathscr{C}$ be some compact subset of $\mathbf{R}^4$ such that*

$$\mathscr{C} \cap \{(x_1, x_2, \xi_1, \xi_2)/\xi_1^2 + \xi_2^2 + b^2(x_2) = 0\} = \emptyset.$$

Then the bicharacteristics $t \mapsto (x(t), \xi(t))$ of the Rossby Hamiltonian starting from any point $(x_1^0, x_2^0, \xi_1^0, \xi_2^0)$ of $\mathscr{C}$ are defined globally in time, and $\forall t \in \mathbf{R}$,

$$\inf_{(x_1^0, x_2^0, \xi_1^0, \xi_2^0) \in \mathscr{C}} (\xi_1(t)^2 + \xi_2(t)^2 + b^2(x_2(t))) > 0.$$

Proof As b', b'', u and Du are Lipschitz, by the Cauchy–Lipschitz theorem the system of ODEs (28) has a unique maximal solution. In order to prove that this solution is defined globally, it is enough to prove that the time derivative of this solution is uniformly bounded. This comes from assumption (6) giving an upper bound on $b'/(1 + b^2(x_2))^{\frac{1}{2}}$ and $b''/(1 + b^2(x_2))^{\frac{1}{2}}$, and from the lower bound on $\langle \xi \rangle_b$ to be established now. The crucial assumption here is the fact that b' and b do not vanish simultaneously.

So let us suppose that $\langle \xi \rangle_b$ vanishes, and consider the first time t^* at which $\langle \xi \rangle_b(t^*) = 0$. Assume to start with that $x(t^*)$ lies outside the support of $\bar{u}$. Then there is a small amount of time (t^-, t^*), $t^- < t^*$, on which $x(t)$ remains outside the support of $\bar{u}$. So on the interval (t^-, t^*), ξ_1 is a constant hence remains zero, and an inspection of the ODEs then shows that on (t^-, t^*), x_2 and ξ_2 are also constant, hence $\langle \xi \rangle_b(t) = 0$ which is impossible by definition of t^*.

Now let us assume that $x(t^*)$ does not lie outside the support of $\bar{u}$, where t^* is still the first time t^* at which $\langle \xi \rangle_b(t^*) = 0$, assuming such a time exists. We shall prove that

$$|\langle \xi \rangle_b(t)| \lesssim (t^* - t)^{\frac{1}{2}}, \quad t \to t^*. \tag{29}$$

Indeed we have clearly

$$\begin{aligned} \frac{1}{2}\frac{d}{dt}\langle \xi \rangle_b^2 &= b(x_2)b'(x_2)\frac{dx_2}{dt} + \xi_1\frac{d\xi_1}{dt} + \xi_2\frac{d\xi_2}{dt} \\ &= b(x_2)b'(x_2)\bar{u}_2(x) - \partial_1\bar{u}(x)\cdot\xi - \partial_2\bar{u}(x)\cdot\xi + \frac{b''(x_2)\xi_1\xi_2}{\langle \xi \rangle_b^2} \end{aligned} \tag{30}$$

so in particular we find that $\frac{d}{dt}\langle \xi \rangle_b^2$ is bounded as t goes to t^*, hence (29) holds. Moreover along a trajectory of the Rossby Hamiltonian, τ_R is conserved, and we have

$$\frac{dx_1}{dt} = \frac{b'(x_2)}{\langle \xi \rangle_b^2} - \frac{2(\tau_R - u(x)\cdot\xi)^2}{b'(x_2)} + u_1(x).$$

Since b' and b do not vanish simultaneously, this in turn implies that there is a constant C such that as t goes to t^*,

$$\left|\frac{dx_1}{dt}\right| \geq \frac{C}{t^* - t}.$$

In particular there is a time $t < t^*$ at which the trajectory has escaped the support of $\bar{u}$, which is contrary to our assumption. This concludes the proof of the proposition. □

5.2 Dynamics Outside from the Support of $\bar{u}$

Using the fact that $\bar{u}$ has compact support, and simple properties of the Rossby dynamics in the absence of zonal flow, we can prove the following result.

Proposition 5.2 *Let $\mathscr{C}$ be some compact subset of $\mathbf{R}^4$ such that*

$$\mathscr{C} \cap \{(x_1, x_2, \xi_1, \xi_2)/\xi_1^2 + \xi_2^2 + b^2(x_2) = 0\} = \emptyset.$$

Then the bicharacteristics of the Rossby Hamiltonian starting from any point of $\mathscr{C}$ are bounded in x_2:

$$\forall t \in \mathbf{R}, \quad \sup_{(x_1^0, x_2^0, \xi_1^0, \xi_2^0) \in \mathscr{C}} |x_2(t)| < \infty.$$

Proof

- Let us start by describing the dynamics in the absence of zonal flow: ξ_1 is then an invariant of the motion, so that the dynamics in (x_2, ξ_2) can be decoupled. Furthermore, as the energy surfaces are compact

$$\tau = \frac{\xi_1 b'(x_2)}{\xi_1^2 + \xi_2^2 + b^2(x_2)}$$

the motion along x_2 is periodic (with infinite period for homoclinic and heteroclinic orbits).

The motion along x_1 is then determined by the equation

$$\frac{dx_1}{dt} = b'(x_2) \frac{\langle \xi \rangle_b^2 - 2\xi_1^2}{\langle \xi \rangle_b^4} \cdot$$

It is trapped if and only if the average of the right-hand side over one period is zero. Outside from saddle points, this quantity depends continuously on ξ_1, so that we expect the initial data leading to trapped trajectories to belong to a manifold of codimension 1. This can be proved rigorously if b^2 has only one nondegenerate critical points (see [2]).

- Let us now turn to the influence of the zonal flow. We will first check that the only possible escape direction is again x_1. Indeed the energy surfaces corresponding to $\tau_R \neq 0$ are bounded in the x_2 direction: as $x_2 \to \pm\infty$,

$$\frac{b'(x_2)\xi_1}{\langle \xi \rangle_b^2} + \bar{u}(x) \cdot \xi \to 0.$$

Consider now a trajectory on the energy level $\tau_R = 0$, and some point of this trajectory (y_1, y_2, ξ_1, ξ_2) such that $y_2 \notin \mathrm{Supp}_{x_2} \bar{u}$. One has

$$b'(y_2)\xi_1 = 0.$$

  – If $b'(y_2) = 0$, then

$$\frac{dx_1}{dt} = \frac{dx_2}{dt} = \frac{d\xi_1}{dt} = \frac{d\xi_2}{dt} = 0.$$

  The uniqueness in Cauchy–Lipschitz theorem implies then that the trajectory is nothing else than a fixed point, and therefore in particular is bounded.

– If $\xi_1 = 0$, then

$$\frac{dx_2}{dt} = \frac{d\xi_1}{dt} = \frac{d\xi_2}{dt} = 0 \quad \text{and} \quad \frac{dx_1}{dt} = b'(y_2)\frac{\xi_2^2 + b^2(y_2) - \xi_1^2}{\langle \xi \rangle_b^4},$$

meaning that the trajectory is a uniform straight motion along x_1. In particular, it is bounded in the x_2-direction.

Finally, we conclude that trajectories on the energy level $\tau_R = 0$ are either trapped in the support $\mathrm{Supp}_{x_2} \bar{u}$, or trivial in the x_2-direction. □

5.3 *Perspectives*

As recalled in the introduction, it is generally believed that in the situation depicted in this paper (a flow around a large macroscopic current) Rossby waves are trapped. However due to the 2-dimensional setting (compared to the work in [2]) the trapping in the x_1 direction seems difficult to prove, outside some specific cases studied in the previous paragraph. One way to be convinced of the trapping phenomenon should be by implementing the dynamical system numerically. It should be pointed out however that actually in order to get physically relevant predictions for the oceanic eddies, one should consider 3D models, or at least 2D models involving the influence of stratification. The methods presented here seem to be robust and should be extended to such complex models, up to again the study of the Hamiltonian system describing the Rossby dynamics.

Acknowledgements The authors are grateful to J.-F. Bony and N. Burq for introducing them to Mourre estimates, and for interesting discussions. They also thank C. Cheverry for pointing out a mistake in a previous version of the paper. I. Gallagher and L. Saint-Raymond are partially supported by the French Ministry of Research grant ANR-08-BLAN-0301-01.

Appendix: A Comparison Result

For the sake of completeness, we state here the result which shows the stability of the propagation under an $O(\varepsilon^\infty)$ error on the propagator. This result has been used in the proof of the diagonalization when comparing A and $T_\pm$, T_R, and in the proof of dispersion when comparing $T_\pm$ and $T_\pm^0$.

Proposition 5.3 *Let A_ε and $\tilde{A}_\varepsilon$ be two pseudo-differential operators such that*

- *iA_ε is hermitian in $L^2(\mathbf{R}^d)$,*
- *$A_\varepsilon - \tilde{A}_\varepsilon = O(\varepsilon^\infty)$ microlocally on $\Omega \subset \mathbf{R}^{2d}$.*

Let $\tilde{\varphi}$ be a solution to

$$i\partial_t \tilde{\varphi} + \tilde{A}_\varepsilon \tilde{\varphi} = 0$$

microlocalized in Ω, and φ be the solution to

$$i\partial_t \varphi + A_\varepsilon \varphi = 0$$

with the same initial data. Then, for all $N \in \mathbf{N}$,

$$\sup_{t \leq \varepsilon^{-N}} \|\varphi(t) - \tilde{\varphi}(t)\|_{L^2(\mathbf{R}^d)} = O(\varepsilon^\infty).$$

Proof The proof is based on a simple energy inequality and is completely straightforward. We have

$$\begin{aligned} \frac{d}{dt}\|\varphi - \tilde{\varphi}\|^2_{L^2(\mathbf{R}^d)} &= 2\langle iA_\varepsilon \varphi - i\tilde{A}_\varepsilon \tilde{\varphi} | \varphi - \tilde{\varphi}\rangle \\ &= 2\langle (iA_\varepsilon - i\tilde{A}_\varepsilon)\tilde{\varphi} | \varphi - \tilde{\varphi}\rangle \\ &\leq 2\|(A_\varepsilon - \tilde{A}_\varepsilon)\tilde{\varphi}\|_{L^2(\mathbf{R}^d)} \|\varphi - \tilde{\varphi}\|_{L^2(\mathbf{R}^d)}. \end{aligned}$$

This leads to

$$\|\varphi(t) - \tilde{\varphi}(t)\|^2_{L^2(\mathbf{R}^d)} = O(\varepsilon^\infty) t,$$

which concludes the proof. □

References

1. Cheverry, C., Gallagher, I., Paul, T., Saint-Raymond, L.: Trapping Rossby waves. C. R. Math. **347**(2), 879–884 (2009)
2. Cheverry, C., Gallagher, I., Paul, T., Saint-Raymond, L.: Semiclassical and spectral analysis of oceanic waves. Duke Math. J. (accepted)
3. Hunziker, W., Sigal, I.M., Soffer, A.: Minimal escape velocities. Commun. Partial Differ. Equ. **24**(11–12), 2279–2295 (1999)
4. Gill, A.E.: Atmosphere-Ocean Dynamics. International Geophysics Series, vol. 30 (1982)
5. Gill, A.E., Longuet-Higgins, M.S.: Resonant interactions between planetary waves. Proc. R. Soc. Lond. A **299**, 120–140 (1967)
6. Greenspan, H.P.: The Theory of Rotating Fluids. Cambridge Monographs on Mechanics and Applied Mathematics (1969)
7. Jensen, A., Mourre, E., Perry, P.: Multiple commutator estimates and resolvent smoothness in quantum scattering theory. Ann. Inst. Henri Poincaré, Sect. A **41**(2), 207–225 (1984)
8. Majda, A.: Introduction to PDEs and Waves for the Atmosphere and Ocean. Courant Lecture Notes in Mathematics, vol. 9. New York University, Courant Institute of Mathematical Sciences/Am. Math. Soc., New York/Providence (2003)
9. Martinez, A.: An Introduction to Semiclassical and Microlocal Analysis. Springer, Berlin (2002)
10. Panati, G., Spohn, H., Teufel, S.: Space-adiabatic perturbation theory. Adv. Theor. Math. Phys. **7**, 145–204 (2003)
11. Pedlosky, J.: Geophysical Fluid Dynamics. Springer, Berlin (1979)
12. Pedlosky, J.: Ocean Circulation Theory. Springer, Berlin (1996)
13. Reed, M., Simon, B.: Methods of Modern Mathematical Physics. Academic Press, San Diego (1979)

Hierarchical Construction of Bounded Solutions of $\operatorname{div} U = F$ in Critical Regularity Spaces

Eitan Tadmor and Changhui Tan

Abstract We implement the hierarchical decomposition introduced in (Tadmor in Hierarchical construction of bounded solutions in critical regularity spaces, arXiv:1003.1525v2), to construct uniformly bounded solutions of the problem $\operatorname{div} U = F$, where the two-dimensional data is in the critical regularity space, $F \in L^2_\#(\mathbb{T}^2)$. Criticality in this context, manifests itself by the lack of linear mapping, $F \in L^2_\#(\mathbb{T}^2) \mapsto U \in L^\infty(\mathbb{T}^2)$, (Bourgain and Brezis in J. Am. Math. Soc. 16(2):393–426, 2003). Thus, the intriguing aspect here is that although the problem is linear, the construction of its uniformly bounded solutions is not.

1 Introduction

We are concerned with the construction of uniformly bounded solutions, $U \in L^\infty(\mathbb{T}^2, \mathbb{R}^2)$ of the equation

$$\operatorname{div} U = F, \qquad F \in L^2_\#(\mathbb{T}^2), \tag{1}$$

where $L^2_\#(\mathbb{T}^2)$ is the space of L^2 integrable functions over the 2-dimensional torus $\mathbb{T}^2$ with zero mean.

The existence of uniformly bounded solutions of (1) follows from the closed range theorem together with Gagliardo–Nirenberg inequality, [1]. Moreover, Bourgain and Brezis [1] proved that any mapping, $F \in L^2_\# \mapsto U \in L^\infty(\mathbb{T}^2)$, must be *nonlinear*: thus, the intriguing aspect here is that although (1) is linear, the construction of its uniformly bounded solutions for $L^2_\#$-data is not.

E. Tadmor (✉)
Department of Mathematics, Center of Scientific Computation and Mathematical Modeling (CSCAMM) and Institute for Physical sciences and Technology (IPST), University of Maryland, College Park, MD 20742-4015, USA
e-mail: tadmor@cscamm.umd.edu

C. Tan
Center of Scientific Computation and Mathematical Modeling (CSCAMM), University of Maryland, College Park, MD 20742-4015, USA
e-mail: ctan@cscamm.umd.edu

H. Holden, K.H. Karlsen (eds.), *Nonlinear Partial Differential Equations*, Abel Symposia 7, DOI 10.1007/978-3-642-25361-4_14,

It follows, in particular, that the classical Helmholtz solution of (1), $U_{\mathrm{Hel}} = \nabla\Delta^{-1}F$, cannot be a uniformly bounded solution for *all* $F \in L^2_\#$. Indeed, $F \in L^2_\#$ implies that $U_{\mathrm{Hel}} \in H^1(\mathbb{T}^2)$, but since H^1 is not a subset of L^∞, Helmholtz solution need not be uniformly bounded. The following concrete counterexample due to L. Nirenberg, [1, Remark 7], demonstrates this type of unboundedness: fix $\theta \in (0, 1/2)$, let $\zeta(r)$ be a smooth cut-off function supported near the origin, and set

$$F = \Delta v, \quad v(x, y) := x|\log r|^\theta \zeta(r), \ r = \sqrt{x^2 + y^2}. \tag{2}$$

In this case, $F \in L^2_\#(\mathbb{T}^2)$, but the Helmholtz solution, $U_{\mathrm{Hel}} = \nabla\Delta^{-1}F = \nabla v$, has a fractional logarithmic growth at the origin.

Inspired by the hierarchical decompositions which were introduced in [8, 9] in the context of image processing, Tadmor [7] utilized such decompositions as a constructive procedure to solve (1): the solution is given in terms of *hierarchical decomposition*, $U_{\mathrm{Bdd}} = \sum \mathbf{u}_j$, where the $\{\mathbf{u}_j\}$'s can be computed recursively as the following minimizers,

$$\mathbf{u}_{j+1} = \arg\min_{\mathbf{u}} \left\{ \|\mathbf{u}\|_{L^\infty} + \lambda_1 2^j \left\| F - \mathrm{div}\left(\sum_{k=1}^{j} \mathbf{u}_k \right) - \mathrm{div}\,\mathbf{u} \right\|^2_{L^2} \right\}, \quad j = 0, 1, \ldots. \tag{3}$$

Here, λ_1 is any sufficiently large parameter, $\lambda_1 > 1/(2\|F\|_{BV})$, which guarantees that the hierarchical decomposition starts with a non-trivial solution of (3), consult (20) below.

In this paper, we propose a numerical approach to solve the minimization problem (3), which in turn generates the uniformly bounded hierarchical solution of problem (1).

We begin, in Sect. 2, by quoting the hierarchical construction proposed in [7]. In Sect. 3 we analyze the minimization problem (3) in terms of its corresponding dual problem. This dual problem amounts to a nonlinear PDE which governs the residual $r := f - \mathrm{div}\,\mathbf{u}$, where f stands for $F - \mathrm{div}(\sum \mathbf{u}_k)$. As a final step, we introduce a procedure to recover the desired minimizer $\mathbf{u}$ from its residual r. In Sect. 4 we discuss the numerical solution of the governing PDE: it is solved by an iterative procedure which avoids significantly large errors in the recovering stage. In Sect. 5, we report on our computations which compare the bounded hierarchical solution, U_{Bdd}, vs. the unbounded Helmholtz solution, U_{Hel}. Finally, in Sect. 6, we introduce a new construction of bounded solutions for (1), based on two-step solution of the form,

$$U_{\mathrm{2step}} = \mathbf{u}_1 + \nabla\Delta^{-1}r_1, \quad [\mathbf{u}_1, r_1] = \arg\min_{\mathrm{div}\,\mathbf{u}+r=F} \left\{ \|\mathbf{u}\|_{L^\infty} + \lambda_1\|r\|^2_{L^2} \right\}. \tag{4}$$

This two-step solution consists of one hierarchical decomposition step, $\mathbf{u}_1$ followed by one Helmholtz step, which are shown to yield a uniformly bounded solution of (1).

2 Hierarchical Solution of div $U = F \in L^2_\#(\mathbb{T}^2)$

Our starting point for the construction of a uniformly bounded solution of (1), $U \in L^\infty(\mathbb{T}^2, \mathbb{R}^2)$, is a decomposition of F,

$$F = \operatorname{div} \mathbf{u}_1 + r_1, \quad F \in L^2_\#(\mathbb{T}^2) := \left\{ g \in L^2(\mathbb{T}^2) \,\middle|\, \int_{\mathbb{T}^2} g(x)\,dx = 0 \right\}, \tag{5a}$$

where $[\mathbf{u}_1, r_1]$ is a minimizing pair of the functional,

$$[\mathbf{u}_1, r_1] = \underset{\operatorname{div}\mathbf{u}+r=F}{\arg\min} \left\{ \|\mathbf{u}\|_{L^\infty} + \lambda_1 \|r\|^2_{L^2} \right\}. \tag{5b}$$

Here, λ_1 is a fixed parameter at our disposal where we distinguish between two cases, consult (20) below. If $\lambda_1 \le \frac{1}{2\|F\|_{BV}}$, then the minimizer of (5b) is the trivial one, $\mathbf{u}_1 \equiv 0$, $r_1 = F$; otherwise, by choosing λ_1 large enough, $\lambda_1 > \frac{1}{2\|F\|_{BV}}$, then (5b) admits a non-trivial minimizer, $[\mathbf{u}_1, r_1]$, which is characterized by a residual satisfying $\|r_1\|_{BV} = \frac{1}{2\lambda_1}$. By Gagliardo–Nirenberg isoperimetric inequality, e.g., [11, §2.7], there exists $\beta > 0$ such that

$$\|g\|_{L^2} \le \beta \|g\|_{BV}, \quad \int_{\mathbb{T}^2} g(x)\,dx = 0. \tag{6}$$

It follows that r_1 is L^2-bounded:

$$\|r_1\|_{L^2} \le \beta \|r_1\|_{BV} = \frac{\beta}{2\lambda_1}. \tag{7}$$

Moreover, since F has a zero mean so does the residual r_1. We conclude that the residual $r_1 \in L^2_\#(\mathbb{T}^2)$, and we can therefore implement the same variational decomposition of F in (5a), (5b), and use it to decompose r_1. To this end, we use the same variational statement, $\{\|\mathbf{u}\|_{L^\infty} + \lambda_2 \|r\|^2_{L^2}\}$, with a new parameter, $\lambda = \lambda_2 > \lambda_1$,

$$r_1 = \operatorname{div} \mathbf{u}_2 + r_2, \quad [\mathbf{u}_2, r_2] = \underset{\operatorname{div}\mathbf{u}+r=r_1}{\arg\min} \left\{ \|\mathbf{u}\|_{L^\infty} + \lambda_2 \|r\|^2_{L^2} \right\}. \tag{8}$$

Borrowing the terminology from our earlier work on image processing [8, 9], the decomposition (8) has the effect of "zooming" on the residual r_1, and it is here that we use the refined scale $\lambda_2 > \lambda_1$. Combining (8) with (5a) we obtain $F = \operatorname{div} U_2 + r_2$ with $U_2 := \mathbf{u}_1 + \mathbf{u}_2$, which is viewed as an improved *approximate solution* of (1). Indeed, the "zooming" effect $\lambda_2 > \lambda_1$ implies that U_2 has a smaller residual $\|r_2\|_{BV} = 1/(2\lambda_2)$ compared with $\|r_1\|_{BV} = 1/(2\lambda_1)$ in (7). In particular,

$$\|r_2\|_{L^2} \le \beta \|r_2\|_{BV} = \frac{\beta}{2\lambda_2}.$$

This process can be repeated: if $r_j \in L^2_\#(\mathbb{T}^2)$ is the residual at step j, then we decompose it

$$r_j = \operatorname{div} \mathbf{u}_{j+1} + r_{j+1}, \tag{9a}$$

where $[\mathbf{u}_{j+1}, r_{j+1}]$ is a minimizing pair of

$$[\mathbf{u}_{j+1}, r_{j+1}] = \underset{\operatorname{div}\mathbf{u}+r=r_j}{\arg\min} \left\{\|\mathbf{u}\|_{L^\infty} + \lambda_{j+1}\|r\|^2_{L^2}\right\}, \quad j = 0, 1, \ldots. \tag{9b}$$

For $j = 0$, the decomposition (9a), (9b) is interpreted as (5a) by setting $r_0 := F$. Note that the recursive decomposition (9a) depends on the invariance that the residuals $r_j \in L^2_\#(\mathbb{T}^2)$: indeed, if r_j has a zero mean then so does r_{j+1}, and since by (20) the minimizer r_{j+1} has a bounded variation, $r_{j+1} \in L^2_\#(\mathbb{T}^2)$. The iterative process depends on a sequence of increasing scales, $\lambda_1 < \lambda_2 < \cdots < \lambda_{j+1}$, which are yet to be determined.

The telescoping sum of the first k steps in (9a) yields an improved approximate solution, $U_k := \sum_{j=1}^k \mathbf{u}_j$:

$$F = \operatorname{div} U_k + r_k, \quad \|r_k\|_{L^2} \leq \beta\|r_k\|_{BV} = \frac{\beta}{2\lambda_k} \downarrow 0,\ k = 1, 2, \ldots. \tag{10}$$

The key question is whether the U_k's remain uniformly bounded, and it is here that we use the freedom in choosing the scaling parameters λ_k: comparing the minimizing pair $[\mathbf{u}_{j+1}, r_{j+1}]$ of (9b) with the trivial pair $[\mathbf{u} \equiv 0, r_j]$, we find

$$\begin{aligned} &\|\mathbf{u}_{j+1}\|_{L^\infty} + \lambda_{j+1}\|r_{j+1}\|^2_{L^2} \leq \|0\|_{L^\infty} + \lambda_{j+1}\|r_j\|^2_{L^2}, \\ &r_j = \operatorname{div}\mathbf{u}_{j+1} + r_{j+1} = \operatorname{div}(0) + r_j. \end{aligned}$$

It remains to upper-bound the energy norm of the r_j's: for $j = 0$ we have $r_0 = F$; for $j > 0$, (10) implies that $\|r_j\|_{L^2} \leq \beta/(2\lambda_j)$. We end up with

$$\|\mathbf{u}_{j+1}\|_{L^\infty} + \lambda_{j+1}\|r_{j+1}\|^2_{L^2} \leq \lambda_{j+1}\|r_j\|^2_{L^2} \leq \begin{cases} \lambda_1\|F\|^2_{L^2}, & j = 0, \\ \frac{\beta^2\lambda_{j+1}}{4\lambda_j^2}, & j = 1, 2, \ldots. \end{cases} \tag{11}$$

We conclude that by choosing a sufficiently fast increasing λ_j's such that $\sum_j \lambda_{j+1}\lambda_j^{-2} < \infty$, then the approximate solutions $U_k = \sum_1^k \mathbf{u}_j$ form a Cauchy sequence in L^∞ whose limit, $U = \sum_1^\infty \mathbf{u}_j$, satisfies the following.

Theorem 2.1 [7] *Fix β such that (6) holds. Then, for any given $F \in L^2_\#(\mathbb{T}^2)$, there exists a uniformly bounded solution of* (1),

$$\operatorname{div} U = F, \quad \|U\|_{L^\infty} \leq 2\beta\|F\|_{L^2}.$$

The solution U is given by $U = \sum_{j=1}^\infty \mathbf{u}_j$, where the $\{\mathbf{u}_j\}$'s are constructed recur-

sively as minimizers of

$$[\mathbf{u}_{j+1}, r_{j+1}] = \underset{\operatorname{div}\mathbf{u}+r=r_j}{\arg\min} \left\{\|\mathbf{u}\|_{L^\infty} + \lambda_1 2^j \|r\|_{L^2}^2\right\}, \quad r_0 := F, \ \lambda_1 = \frac{\beta}{\|F\|_{L^2}}. \tag{12}$$

Proof Set $\lambda_j = \lambda_1 2^{j-1}$, $j = 1, 2, \ldots$, then, $\|U_k - U_\ell\|_{L^\infty} \lesssim 2^{-k}$, $k > \ell \gg 1$. Let U be the limit of the Cauchy sequence $\{U_k\}$ then $\|U_j - U\|_{L^\infty} + \|\operatorname{div} U_j - F\|_{L^2} \lesssim 2^{-j} \to 0$, and since div has a closed graph on its domain $D := \{\mathbf{u} \in L^\infty : \operatorname{div}\mathbf{u} \in L^2(\mathbb{T}^2)\}$, it follows that $\operatorname{div} U = F$. By (11) we have

$$\|U\|_{L^\infty} \le \sum_{j=1}^{\infty} \|\mathbf{u}_j\|_{L^\infty} \le \lambda_1 \|F\|_{L^2}^2 + \frac{\beta^2}{4\lambda_1} \sum_{j=2}^{\infty} \frac{1}{2^{j-3}} = \lambda_1 \|F\|_{L^2}^2 + \frac{\beta^2}{\lambda_1}.$$

Here $\lambda_1 > \frac{1}{2\|F\|_{BV}}$ is a free parameter at our disposal: we choose $\lambda_1 := \beta/\|F\|_{L^2}$ which by (6) is admissible, $\lambda_1 = \frac{\beta}{\|F\|_{L^2}} > \frac{1}{2\|F\|_{BV}}$, and the result follows. □

Remark 2.1 (Energy decomposition) By squaring the refinement step (5a), $r_j = r_{j+1} + \operatorname{div}\mathbf{u}_{j+1}$, and using the characterization of $[\mathbf{u}_{j+1}, r_{j+1}]$ as an *extremal pair* (consult Remark 3.2 below), we find

$$\begin{aligned}\|r_j\|_{L^2}^2 - \|r_{j+1}\|_{L^2}^2 &= 2(r_{j+1}, \operatorname{div}\mathbf{u}_{j+1}) + \|\operatorname{div}\mathbf{u}_{j+1}\|_{L^2}^2 \\ &= \frac{1}{\lambda_{j+1}} \|\mathbf{u}_{j+1}\|_{L^\infty} + \|\operatorname{div}\mathbf{u}_{j+1}\|_{L^2}^2.\end{aligned}$$

A telescoping sum of the last equality yields the "energy decomposition"

$$\sum_{j=1}^{\infty} \frac{1}{\lambda_j} \|\mathbf{u}_j\|_{L^\infty} + \sum_{j=1}^{\infty} \|\operatorname{div}\mathbf{u}_j\|_{L^2(\mathbb{T}^2)}^2 = \|F\|_{L^2(\mathbb{T}^2)}^2. \tag{13}$$

Remark 2.2 We note that the constructive proof of Theorem 2.1 does not assume the existence of bounded solution for (14): it is deduced from the Gagliardo–Nirenberg inequality (6). The hierarchical construction of solutions for $\mathscr{L}U = F$, in the general setup of linear closed operators, $\mathscr{L} : \mathbb{B} \mapsto L^p_\#$, $1 < p < \infty$, with boundedly invertible duals $\mathscr{L}^*$, was proved in [7]. In [2], Bourgain and Brezis proved that (1) admits a bounded solution in the *smaller* space, $\mathbb{B} = L^\infty \cap H^1$. This requires a considerably more delicate argument, which could be justified by the refined dual estimate (compared with (6)), $\|g\|_{L^2(\mathbb{T}^2)} \lesssim \|\nabla g\|_{L^1 + H^{-1}(\mathbb{T}^2)}$. The proof of [2] is *constructive*: it is based on an intricate Littlewood–Paley decomposition, which cannot be readily implemented in actual computations.

3 Construction of Hierarchical Minimizers

3.1 The Minimization Problem

We rewrite each minimization step of the hierarchical decompositions (3) in the following form,

$$\bar{\mathbf{u}} = \arg\min_{\mathbf{u}:\mathbb{T}^2\to\mathbb{R}^2} \left\{ \|\mathbf{u}\|_{L^\infty} + \lambda \| f - \operatorname{div} \mathbf{u}\|_{L^2}^2 \right\}, \qquad \|\mathbf{u}\|_{L^\infty} := \operatorname*{ess\,sup}_{x,y} \sqrt{u_1^2 + u_2^2}. \tag{14}$$

Here, f is an L^2 function with zero mean which stands for $F - \operatorname{div}(\sum_{k=1}^{j} \mathbf{u}_k)$ in (3), and λ stands for the dyadic scales, $\lambda_1 2^j$, $j = 0, 1, \ldots$.

3.2 The Dual Problem

To circumvent the difficulty of handling the L^∞ norm in (14), we concentrate on the dual problem associated with (14). We let $\mathcal{N}(\mathbf{u}) = \|\mathbf{u}\|_{L^\infty} : V \mapsto \bar{\mathbb{R}}$, $\mathcal{E}(p) = \|f - p\|_{L^2}^2 : Y \mapsto \bar{\mathbb{R}}$, and $\Lambda = \operatorname{div} : V \mapsto Y$ with $V = L^\infty(\mathbb{T}^2)$ and $Y = L^2(\mathbb{T}^2)$. By duality theorem, [4, §3, Remark 4.2], the variational problem (14),

$$(\mathcal{P}): \qquad \inf_{u \in V} [\mathcal{N}(u) + \mathcal{E}(\Lambda u)]$$

is equivalent to its dual problem

$$(\mathcal{P}^*): \qquad \sup_{p^* \in Y^*} [-\mathcal{N}^*(\Lambda^* p^*) - \mathcal{E}^*(-p^*)];$$

moreover, if $\bar{u}$ and $\bar{p}^*$ are solutions of $(\mathcal{P})$ and $(\mathcal{P}^*)$ respectively, then $\Lambda^* \bar{p}^* \in \partial\mathcal{N}(\bar{u})$, and $-\bar{p}^* \in \partial\mathcal{E}(\Lambda\bar{u})$. Here, $\mathcal{N}^*, \mathcal{E}^*$ are conjugate functions of $\mathcal{N}, \mathcal{E}$, expressed in terms of the usual L^2 pairing $\langle \mathbf{w}_1, \mathbf{w}_2 \rangle := \int_{\mathbb{T}^2} \mathbf{w}_1 \cdot \mathbf{w}_2 \, dx$,

$$\begin{aligned}
\mathcal{N}^*(\mathbf{u}^*) &= \sup_{\mathbf{u}} \left\{ \langle \mathbf{u}, \mathbf{u}^* \rangle - \|\mathbf{u}\|_{L^\infty} \right\} \\
&= \sup_{\mathbf{u}} \left\{ \|\mathbf{u}\|_{L^\infty} \|\mathbf{u}^*\|_{L^1} - \|\mathbf{u}\|_{L^\infty} \right\} = \chi_{\{\|\mathbf{u}^*\|_{L^1} \le 1\}} = \begin{cases} 0, & \text{if } \|\mathbf{u}^*\|_{L^1} \le 1, \\ +\infty, & \text{otherwise;} \end{cases} \\
\mathcal{E}^*(p^*) &= \sup_{p} \left\{ \langle p, p^* \rangle - \lambda \|f - p\|_{L^2}^2 \right\} \\
&= \sup_{p} \left\{ -\lambda \langle p, p \rangle + \langle p^* + 2\lambda f, p \rangle - \lambda \langle f, f \rangle \right\} = \left\langle f + \frac{1}{4\lambda} p^*, p^* \right\rangle,
\end{aligned}$$

and $\Lambda^* = -\nabla$ is the dual operator of Λ.

We end up with the dual ($\mathcal{P}^*$) problem

$$\inf_{\{p^*:\|\nabla p^*\|_{L^1}\le 1\}} \left\langle \frac{1}{4\lambda} p^* - f, p^* \right\rangle$$

or

$$\inf_{p^*} \sup_{\mu\ge 0} \left[\left\langle \frac{1}{4\lambda} p^* - f, p^* \right\rangle + \mu(\|\nabla p^*\|_{L^1} - 1) \right]. \tag{15}$$

Moreover, $-\bar{p}^* \in \partial\mathcal{E}(\Lambda\bar{u})$, meaning that $p^* = 2\lambda r$, where r is the residual, $r = f - \operatorname{div}\mathbf{u}$. So, we can express the dual problem (15) in terms of r,

$$\bar{r} = \arg\min_r \sup_{\mu\ge 0} L(r,\mu), \qquad L(r,\mu) := \lambda\langle r - 2f, r\rangle + \mu\left(\|\nabla r\|_{L^1} - \frac{1}{2\lambda} \right), \tag{16}$$

where $\bar{r} := f - \operatorname{div}\bar{\mathbf{u}}$, is the residual corresponding to the optimal minimizer $\bar{u}$.

Since $L(\cdot,\mu)$ is convex and $L(r,\cdot)$ is concave and, for $r \in BV$ continuous, we can apply the minimax theorem, e.g., [4, §6], which allows us to interchange the infimum and supremum in (16), yielding

$$\sup_{\mu\ge 0} \min_r \left[\lambda\langle r - 2f, r\rangle + \mu\left(\|\nabla r\|_{L^1} - \frac{1}{2\lambda} \right) \right]. \tag{17}$$

The dual problem, (17), can be solved in two steps. An inner minimization problem

$$r_\mu = \arg\min_r \left[\lambda\langle r - 2f, r\rangle + \mu\left(\|\nabla r\|_{L^1} - \frac{1}{2\lambda} \right) \right]. \tag{18a}$$

Here, for any given $\mu \ge 0$, there exists a unique $r = r_\mu$ such that (μ, r_μ) is a saddle point of L. The optimal $\mu = \mu^*$ is determined by an outer maximization problem,

$$\mu^* = \arg\min_{\mu\ge 0} \big[P(\mu) + \mu Q(\mu) \big],$$

$$P(\mu) := \lambda\langle r_\mu - 2f, r_\mu\rangle, \ \ Q(\mu) := \|\nabla r_\mu\|_{L^1} - \frac{1}{2\lambda}. \tag{18b}$$

Once μ^* is found, then $\bar{r} = r_{\mu^*}$ is the optimal residual which is sought as the solution of (16).

3.3 The Outer Maximization Problem

We begin by characterizing the maximizer, $\mu = \mu^*$, of the outer problem (18b). Fix μ: since r_μ minimizes $L(r,\mu)$ we have

$$P(\mu) + \mu Q(\mu) \le P(\nu) + \mu Q(\nu).$$

Similarly, $P(\nu)+\nu Q(\nu)\le P(\mu)+\nu Q(\mu)$. Sum the last two inequalities to get, $(\mu-\nu)[Q(\mu)-Q(\nu)]\le 0$, which yields that $Q(\cdot)$ is non-increasing.

Let μ^* be a maximizer of (18b). Then $\forall\mu\ge 0$,

$$P(\mu)+\mu Q(\mu)\le P(\mu^*)+\mu^* Q(\mu^*)\le P(\mu)+\mu^* Q(\mu),$$

which implies $(\mu^*-\mu)Q(\mu)\ge 0$. We distinguish between two cases.

Case #1 $\mu^*>0$. We have $Q(\mu)\le 0$ if $\mu>\mu^*$ and $Q(\mu)\ge 0$ if $0\le\mu<\mu^*$. We conclude that μ^* is determined as a root of $Q(\cdot)$,

$$Q(\mu^*)=0, \quad \text{i.e.} \quad \|\nabla r_{\mu^*}\|_{L^1}=\frac{1}{2\lambda}. \tag{19}$$

Case #2 $\mu^*=0$. In this case, r_0 minimizes $\langle r-2f,r\rangle$, namely, $r_0=f$. This corresponds to the trivial minimizer of (14), $\bar{\mathbf{u}}\equiv 0$, which is the case we want to avoid. Case #2 happens when $Q(0)\le 0$, i.e.

$$\mu^*\leftrightarrow\|\nabla r_0\|_{L^1}-\frac{1}{2\lambda}\le 0\leftrightarrow\|\nabla f\|_{L^1}\le\frac{1}{2\lambda}.$$

So, to make sure that we pick a non-trivial minimizer, $\bar{\mathbf{u}}\not\equiv 0$, we must pick a sufficiently large λ such that

$$\lambda>\frac{1}{2\|f\|_{BV}}\leftrightarrow\bar{\mathbf{u}}\not\equiv 0, \qquad \|\bar{r}\|_{BV}=\frac{1}{2\lambda}. \tag{20}$$

This coincides with the same lower bound on λ's which yield non-trivial minimizers, asserted in [7, Lemma 5.3].

3.4 The Inner Minimization Problem

We return to the inner minimization problem (18a). Fix $\mu=\mu^*$. The Euler–Lagrange equations characterizing minimizers of (18a) are

$$2\lambda(r_{\mu^*}-f)-\mu^*\operatorname{div}\left(\frac{\nabla r_{\mu^*}}{|\nabla r_{\mu^*}|}\right)=0. \tag{21}$$

Take the L^2-inner product of (21) with r_{μ^*} to get

$$2\lambda\langle r_{\mu^*}-f,r_{\mu^*}\rangle-\mu^*\left\langle\operatorname{div}\left(\frac{\nabla r_{\mu^*}}{|\nabla r_{\mu^*}|}\right),r_{\mu^*}\right\rangle=0.$$

Using (19) (and in the non-periodic case, the Neumann boundary condition $\nabla r_{\mu^*}\cdot\mathbf{n}=0$), we find

$$\left\langle\operatorname{div}\left(\frac{\nabla r_{\mu^*}}{|\nabla r_{\mu^*}|}\right),r_{\mu^*}\right\rangle=-\left\langle\frac{\nabla r_{\mu^*}}{|\nabla r_{\mu^*}|},\nabla r_{\mu^*}\right\rangle=-\int_{\mathbb{T}^2}|\nabla r_{\mu^*}|\,dx=-\frac{1}{2\lambda}.$$

This yields, $\mu^* = 4\lambda^2\langle f - r_{\mu^*}, r_{\mu^*}\rangle$, and the governing equation (21) for the optimal residual, $\bar{r} = r_{\mu^*}$, amounts to

$$(\bar{r} - f) - 2\lambda\langle f - \bar{r}, \bar{r}\rangle \operatorname{div}\left(\frac{\nabla\bar{r}}{|\nabla\bar{r}|}\right) = 0. \tag{22}$$

Remark 3.1 This system has two solutions: one solution, $\bar{r} = f$, corresponds to the trivial case, $\bar{\mathbf{u}} \equiv 0$. The other is the target solution, i.e., the optimal residual $\bar{r}$ for (16). We will discuss numerical algorithms to solve system (22) in Sect. 4.

3.5 *From r to $\mathbf{u}$: Recovering the Uniformly Bounded Solution*

So far, we identified the residual, $\bar{r} = f - \operatorname{div}\bar{\mathbf{u}}$, corresponding to the uniformly bounded solution $\bar{\mathbf{u}}$ of (14). To recover $\bar{\mathbf{u}}$ itself, we substitute $\bar{r} - f = -\operatorname{div}\bar{\mathbf{u}}$ as the first term of (22), and get

$$\operatorname{div}\left(\bar{\mathbf{u}} - 2\lambda\langle\bar{r} - f, \bar{r}\rangle\frac{\nabla\bar{r}}{|\nabla\bar{r}|}\right) = 0. \tag{23}$$

Therefore, we can recover a solution $\bar{\mathbf{u}}$ of (14),

$$\bar{\mathbf{u}} = 2\lambda\langle\bar{r} - f, \bar{r}\rangle\frac{\nabla\bar{r}}{|\nabla\bar{r}|}. \tag{24}$$

Observe that this $\bar{\mathbf{u}}$ is indeed uniformly bounded:

$$\|\bar{\mathbf{u}}\|_{L^\infty} = 2\lambda|\langle\bar{r} - f, \bar{r}\rangle| < \infty. \tag{25}$$

Remark 3.2 The explicit expression of $\bar{\mathbf{u}}$ in (24) shows that $[\bar{\mathbf{u}}, \bar{r}]$ forms an extremal pair, [5, Theorem 4], [9, Theorem 2.3], [7, Theorem 5.1], in the sense of achieving an equality in the duality inequality of pairing $\operatorname{div}\bar{\mathbf{u}}$ and $\bar{r}$:

$$|\langle\operatorname{div}\bar{\mathbf{u}}, \bar{r}\rangle| = \|\bar{\mathbf{u}}\|_{L^\infty}\frac{1}{2\lambda} = \|\bar{\mathbf{u}}\|_{L^\infty}\|\nabla\bar{r}\|_{L^1}.$$

4 Numerical Algorithms for the Hierarchical Solution

We solve problem (1) using its hierarchical decomposition. In each iteration, we solve the minimization problem (14). Each iteration consists of three stages:

Stage 1. Find the non-trivial solution, r_j, of Euler–Lagrange equations (22) with $\lambda = \lambda_j$ and $f = f_j$;
Stage 2. Recover $\mathbf{u}_j$ from r_j using Eq. (24);
Stage 3. Update $\lambda_{j+1} \leftarrow 2\lambda_j$, $f_{j+1} \leftarrow r_j$.

Initially, we set λ_1 sufficiently large so that $\lambda_1 > (2\|F\|_{BV})^{-1}$, and $f_1 := f$. The iterations terminate when $\|f_j\|_{L^2}$ is sufficiently small. The final solution U for (1) is given by the sum of all $\mathbf{u}_j$'s.

4.1 Numerical Discretization for the PDE System

We begin with *regularization*: to avoid the singularity in (18a) when $|\nabla r| = 0$, a standard approach is to regularize the problem using a small parameter $\varepsilon > 0$,

$$r_{\mu,\varepsilon} = \arg\min_r \left\{ \lambda\langle r - 2f, r\rangle + \mu\left(\int_{\mathbb{T}^2} \sqrt{\varepsilon^2 + |\nabla r|^2}\, dx\, dy - \frac{1}{2\lambda}\right)\right\}. \tag{26}$$

At stage 1 of each regularized iteration, we find the minimizer $r = r_{\mu^*,\varepsilon}$. The corresponding Euler–Lagrange equations of the regularized problem read,

$$(r - f) - 2\lambda\langle f - r, r\rangle \cdot \operatorname{div}\left(\frac{\nabla r}{\sqrt{\varepsilon^2 + |\nabla r|^2}}\right) = 0. \tag{27}$$

In the non-periodic case, these equations are augmented with Neumann boundary condition, $\nabla r \cdot \mathbf{n} = 0$.

To solve (27), we cover $\mathbb{T}^2$ with a computational grid with cell size h. Let D_{+x}, D_{-x} and D_{0x} be the usual forward, backward and centered divided difference operator on x, namely, $D_{\pm x} r_{i,j} = \pm(r_{i\pm1,j} - r_{i,j})/h$, $D_{0x} r_{i,j} = (r_{i+1,j} - r_{i-1,j})/2h$. Similarly, we can define $D_{\pm y}$ and D_{0y}. A straightforward discretization of (27) yields,

$$\begin{aligned}
r_{i,j} &= f_{i,j} - K(r)\cdot D_{-x}\left[\frac{1}{\sqrt{\varepsilon^2 + (D_{+x}r_{i,j})^2 + (D_{0y}r_{i,j})^2}} D_{+x} r_{i,j}\right] \\
&\quad - K(r)\cdot D_{-y}\left[\frac{1}{\sqrt{\varepsilon^2 + (D_{0x}r_{i,j})^2 + (D_{+y}r_{i,j})^2}} D_{+y} r_{i,j}\right] \\
&= f_{i,j} - \frac{K(r)}{h^2}\left[\frac{r_{i+1,j} - r_{i,j}}{\sqrt{\varepsilon^2 + (D_{+x}r_{i,j})^2 + (D_{0y}r_{i,j})^2}}\right. \\
&\quad \left. - \frac{r_{i,j} - r_{i-1,j}}{\sqrt{\varepsilon^2 + (D_{+x}r_{i-1,j})^2 + (D_{0y}r_{i-1,j})^2}}\right] \\
&\quad - \frac{K(r)}{h^2}\left[\frac{r_{i,j+1} - r_{i,j}}{\sqrt{\varepsilon^2 + (D_{0x}r_{i,j})^2 + (D_{+y}r_{i,j})^2}}\right. \\
&\quad \left. - \frac{r_{i,j} - r_{i,j-1}}{\sqrt{\varepsilon^2 + (D_{0x}r_{i,j-1})^2 + (D_{+y}r_{i,j-1})^2}}\right].
\end{aligned} \tag{28}$$

Here, $K(r) := 2\lambda\langle r - f, r\rangle$, which is approximated using any appropriate numerical quadrature.

4.2 Computing the Residuals r by Implicit Iterations

We use implicit iteration method to solve the nonlinear system (28),

$$
\begin{aligned}
r_{i,j}^{(n+1)} = f_{i,j} - \frac{K(r^{(n)})}{h^2}\Bigg[& \frac{r_{i+1,j}^{(n+1)} r_{i,j}^{(n+1)}}{\sqrt{\varepsilon^2 + (D_{+x} r_{i,j}^{(n)})^2 + (D_{0y} r_{i,j}^{(n)})^2}} \\
& - \frac{r_{i,j}^{(n+1)} - r_{i-1,j}^{(n+1)}}{\sqrt{\varepsilon^2 + (D_{+x} r_{i-1,j}^{(n)})^2 + (D_{0y} r_{i-1,j}^{(n)})^2}}\Bigg] \\
& - \frac{K(r^{(n)})}{h^2}\Bigg[\frac{r_{i,j+1}^{(n+1)} - r_{i,j}^{(n+1)}}{\sqrt{\varepsilon^2 + (D_{0x} r_{i,j}^{(n)})^2 + (D_{+y} r_{i,j}^{(n)})^2}} \\
& - \frac{r_{i,j}^{(n+1)} - r_{i,j-1}^{(n+1)}}{\sqrt{\varepsilon^2 + (D_{0x} r_{i,j-1}^{(n)})^2 + (D_{+y} r_{i,j-1}^{(n)})^2}}\Bigg],
\end{aligned} \tag{29}
$$

subject to initial condition which we set to be $r^{(0)} = f/2$.

Remark 4.1 Recall that $K(r)$ is continuous, and $K(\bar{r}) < 0$ while $K(f) = 0$. To avoid the convergence of $r^{(n)}$ to the trivial solution, $\bar{r} = f$ (mentioned in remark (3.1)), we set $r^{(0)}$ small enough, $K(r^{(0)}) < K(\bar{r}) < K(f)$, so that $r^{(n)}$ is expected to reach the non-trivial solution $\bar{r}$, rather than f. As $\arg\min_r K(r) = f/2$, a good choice of the initial condition of the iteration is $r^{(0)} = f/2$.

In the non-periodic case, we also need to apply Neumann boundary condition $\nabla r \cdot \mathbf{n} = 0$. To this end, we mirror r at the boundary, meaning $r_{0,j} = r_{2,j}$, $r_{N+1,j} = r_{N-1,j}$, etc., where the size of the grid is $N \times N$. So we only need to add the weight of the outer points to their corresponding inner points.

In summary, at the nth iteration amounts to an $N \times N$ linear system, $A(r^{(n)})\tilde{r}^{(n+1)} = \tilde{f}$, for the discretized nodes, $\{r^{(n+1)}\}$. Here, A is a sparse matrix with at most 5 non-zero entries every row or column, whose values depend on $r^{(n)}$.

4.3 Recovering $\mathbf{u}$ from r and Control of Errors

After we get a non-trivial solution r at stage 1, we move to stage 2 to recover $\mathbf{u}$ by (24). Normally, we apply centered divided difference operator on r to compute the discrete gradient, ∇r. However, this will cause a significant error of the solution $\mathbf{u}$.

For example, consider $u^1_{i,j} = K \cdot \frac{r_{i+1,j}-r_{i-1,j}}{2h\sqrt{\varepsilon^2+|\nabla r_{i,j}|^2}}$. Suppose the error for r in stage 1 is $e(r)$. Then, at points (x, y) such that $|\nabla r(x, y)| \approx 0$, the error for u^1 is of order $Ke(r)/(h\varepsilon)$. Therefore, dividing by $h\varepsilon$ with $\varepsilon \approx 0$, the error bound of u^1 can be significantly amplified at stage 2 of recovering $\mathbf{u}$, even if we obtain a sufficiently small $e(r)$ at stage 1. This amplification will get worse as we refine the mesh and h becomes smaller.

In order to get a reliable solution for $\mathbf{u}$, we cannot carry out stage 2 independent of the discretization stencil of stage 1. To this end, let

$$u^{1,(n+1)}_{i+1/2,j} = \frac{K^{(n)}}{h} \cdot \frac{r^{(n+1)}_{i+1,j} - r^{(n+1)}_{i,j}}{\sqrt{\varepsilon^2 + (D_{+x} r^{(n)}_{i,j})^2 + (D_{0y} r^{(n)}_{i,j})^2}}, \tag{30a}$$

$$u^{2,(n+1)}_{i,j+1/2} = \frac{K^{(n)}}{h} \cdot \frac{r^{(n+1)}_{i,j+1} - r^{(n+1)}_{i,j}}{\sqrt{\varepsilon^2 + (D_{0x} r^{(n)}_{i,j})^2 + (D_{+y} r^{(n)}_{i,j})^2}}. \tag{30b}$$

We then have

$$r_{i,j} = f_{i,j} - \frac{u^1_{i+1/2,j} - u^1_{i-1/2,j}}{h} - \frac{u^2_{i,j+1/2} - u^2_{i,j-1/2}}{h}.$$

The last two terms represent a numerical discretization of $\operatorname{div}\mathbf{u}$. Therefore, we use (30a), (30b) to recover $\mathbf{u}$ from the residual $r = f - \operatorname{div}\mathbf{u}$ calculated at (29).

5 Hierarchical Solution vs. Helmholtz Solution

We apply our algorithm for the hierarchically constructed uniformly bounded solution for the example of $F \in L^2_\#$ defined at (2) with

$$\mathbb{T}^2 = [-1, 1] \times [-1, 1], \qquad \theta = 1/3, \qquad \zeta(r) \begin{cases} = e^{-\frac{1}{1-r^2}}, & |r| < 1, \\ \equiv 0, & |r| \geq 1. \end{cases} \tag{31}$$

We concentrate on the first component of the solution U, denoted by U^1. Figure 1 shows Helmholtz solution, U^1_{Hel}, which slowly diverges at the origin. Figure 2 provides the hierarchical solution U^1_{Bdd} which remains uniformly bounded.

The computed hierarchical solution $\|U^{1,N}_{\text{Bdd}}\|_{L^\infty}/\|F^N\|_{L^2}$ remains uniformly bounded when N increases ($U^{1,N}_{\text{Bdd}}$ stands for the first component of hierarchical solution with grid size $N \times N$). In contrast, Table 1 illustrates the (slow) growth of the ratio $\|U^{1,N}_{\text{Hel}}\|_{L^\infty}/\|F^N\|_{L^2}$.

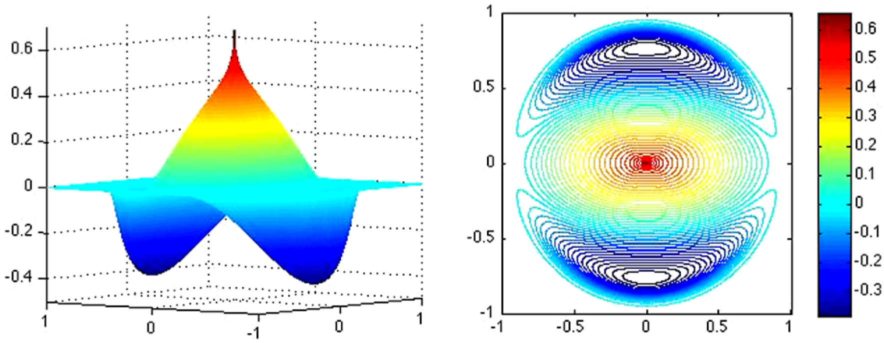


Fig. 1 Helmholtz solution U^1_{Hel} of example (2), (31)

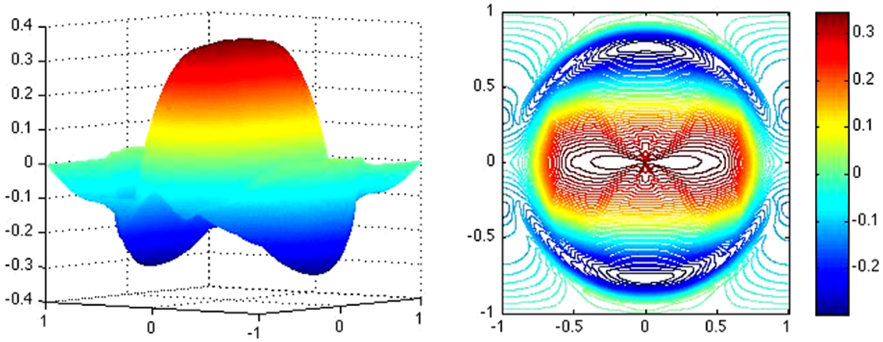


Fig. 2 Hierarchical solution U^1_{Bdd} of (2), (31)

Table 1 L^∞ norm of numerical solutions for different grids: Helmholtz vs. hierarchical construction

The $N \times N$ grid	50×50	100×100	200×200	400×400	800×800
$\frac{\Vert U^{1,N}_{\text{Hel}}\Vert_{L^\infty}}{\Vert F^N\Vert_{L^2}}$	0.2295	0.2422	0.2540	0.2650	0.2752
$\frac{\Vert U^{1,N}_{\text{Bdd}}\Vert_{L^\infty}}{\Vert F^N\Vert_{L^2}}$	0.1454	0.1451	0.1455	0.1458	0.1451

6 Hierarchical Solution Meets Helmholtz Solution

The hierarchical solution is uniformly bounded. However, as observed in Fig. 2, the hierarchical solution U^1_{Bdd} is oscillatory outside the support of F. As each step of the hierarchical decomposition relies on the previous steps, these oscillations will grow throughout the iterations. To limit their effect, we introduce a new, two-step method to construct bounded solutions of (1). It consists of one hierarchical decomposition step, whose residual is treated using Helmholtz decomposition:

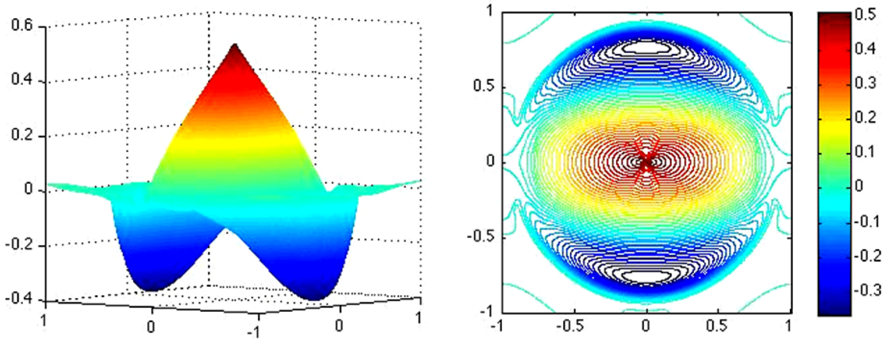


Fig. 3 Two-step solution, $U^1_{2\ \text{step}}$

Table 2 The two-step solution of (2), (31) for different grids

The $N \times N$ grid	50×50	100×100	200×200	400×400	800×800
$\frac{\|U^{1,N}_{2\ \text{step}}\|_{L^\infty}}{\|F^N\|_{L^2}}$	0.2096	0.2128	0.2144	0.2151	0.2154

Step 1. Solve minimization problem

$$\mathbf{u}_1 := \arg\min_{\mathbf{u}} \left\{ \|\mathbf{u}\|_{L^\infty} + \lambda_1 \|F - \operatorname{div} \mathbf{u}\|^2_{L^2} \right\}. \tag{32a}$$

Step 2. Find the Helmholtz solution for $\operatorname{div} \mathbf{u}_r = r_1$, i.e.

$$\mathbf{u}_r := \nabla \Delta^{-1} r_1, \quad r_1 = F - \operatorname{div} \mathbf{u}_1. \tag{32b}$$

Clearly, the two-step solution, $U_{2\ \text{step}} = \mathbf{u}_1 + \mathbf{u}_r$, satisfies $\operatorname{div} U = F$. Furthermore, it is uniformly bounded.

Proposition 6.1 *The two-step solution,* $U_{2\ \text{step}} = \mathbf{u}_1 + \mathbf{u}_r$ *given in* (32a), (32b) *is a uniformly bounded solution of* (1).

Proof Clearly, $\mathbf{u}_1$, as the first iteration of the hierarchical solution, is uniformly bounded. Next, $\mathbf{u}_r = \nabla \Delta^{-1} r_1 = (-\frac{1}{2\pi} \frac{\mathbf{x}}{|\mathbf{x}|^2}) \star r_1$. The Newtonian potential, $(-\frac{1}{2\pi} \frac{\mathbf{x}}{|\mathbf{x}|^2})$, belongs to the Lorentz space $L^{2,\infty}$. The residual, r_1 is BV-bounded and hence, [3, 10], $r_1 \in BV \subset L^{2,1}$. By Hölder's inequality for Lorentz spaces, [6, 10], $\mathbf{u}_r$ and therefore $U_{2\ \text{step}}$, are uniformly bounded. □

From Proposition 6.1, we know that $U_{2\ \text{step}}$ is also a solution of (1). As the minimization problem is solved only once, we expect fewer oscillations in $U_{2\ \text{step}}$ than U_{Bdd}.

Figure 3 shows the two-step solution of the example in Sect. 5. From the contour plot, we observe fewer oscillations than the hierarchical solution U_{Bdd}. Yet, the solution is not as smooth as U_{Bdd} at the origin. Table 2 reports that the ratio $\|U^{1,N}_{2\text{ step}}\|_{L^\infty}/\|F^N\|_{L^2}$ is also stable when N is large. This verifies the uniformly boundedness of the two-step solution.

Acknowledgements Research was supported in part by NSF grants DMS10-08397 and FRG07-57227.

References

1. Bourgain, J., Brezis, H.: On the equation div $Y = f$ and application to control of phases. J. Am. Math. Soc. **16**(2), 393–426 (2003)
2. Bourgain, J., Brezis, H.: New estimates for elliptic equations and Hodge type systems. J. Eur. Math. Soc. (JEMS) **9**(2), 277–315 (2007)
3. Cohen, A.: private communication
4. Ekeland, I., Témam, R.: Convex Analysis and Variational Problems. SIAM, Philadelphia (1999)
5. Meyer, Y.: Oscillating Patterns in Image Processing and Nonlinear Evolution Equations. University Lecture Series, vol. 22. AMS, Providence (2002)
6. O'Neil, R.: Convolution operators and $L(p,q)$ spaces. Duke Math. J. **30**, 129–143 (1963)
7. Tadmor, E.: Hierarchical construction of bounded solutions in critical regularity spaces. arXiv:1003.1525v2
8. Tadmor, E., Nezzar, S., Vese, L.: A multiscale image representation using hierarchical (BV, L^2) decomposition. Multiscale Model. Simul. **2**(4), 554–579 (2004)
9. Tadmor, E., Nezzar, S., Vese, L.: Hierarchical decomposition of images with applications to deblurring, denoising and segmentation. Commun. Math. Sci. **6**(2), 281–307 (2008)
10. Tartar, L.: Lorentz Spaces and Applications. Lecture Notes. Carnegie-Mellon University Press, Pittsburg (1989)
11. Ziemer, W.: Weakly Differentiable Functions. Graduate Texts in Math., vol. 120. Springer, Berlin (1989)

Nonlinear Diffusion with Fractional Laplacian Operators

Juan Luis Vázquez

Abstract We describe two models of flow in porous media including nonlocal (long-range) diffusion effects. The first model is based on Darcy's law and the pressure is related to the density by an inverse fractional Laplacian operator. We prove existence of solutions that propagate with finite speed. The model has the very interesting property that mass preserving self-similar solutions can be found by solving an elliptic obstacle problem with fractional Laplacian for the pair pressure-density. We use entropy methods to show that these special solutions describe the asymptotic behavior of a wide class of solutions.

The second model is more in the spirit of fractional Laplacian flows, but nonlinear. Contrary to usual Porous Medium flows (PME in the sequel), it has infinite speed of propagation. Similarly to them, an L^1-contraction semigroup is constructed and it depends continuously on the exponent of fractional derivation and the exponent of the nonlinearity.

1 Nonlinear Diffusion and Fractional Diffusion

Since the work by Einstein [39] and Smoluchowski [62] at the beginning of the last century (cf. also Bachelier [9]), we possess an explanation of diffusion and Brownian motion in terms of the heat equation, and in particular of the Laplace operator. This explanation has had an enormous success both in Mathematics and Physics. In the decades that followed, the Laplace operator has been often replaced by more general types of so-called elliptic operators with variable coefficients, and later by nonlinear differential operators; a huge body of theory is now available, both for the evolution equations [50] and for the stationary states, described by elliptic equations of different kinds [42, 49].

In recent years there has been a surge of activity focused on the use of so-called fractional diffusion operators to replace the standard Laplace operator (and the other kinds of elliptic operators with variable coefficients), with the aim of further extending the theory by taking into account the presence of so-called long range

J.L. Vázquez (✉)
Dpto. de Matemáticas, Universidad Autónoma de Madrid, 28049 Madrid, Spain
e-mail: juanluis.vazquez@uam.es

H. Holden, K.H. Karlsen (eds.), *Nonlinear Partial Differential Equations*, Abel Symposia 7, DOI 10.1007/978-3-642-25361-4_15,

interactions. The new operators do not act by pointwise differentiation but by a global integration with respect to a very singular kernel; in that way the nonlocal character of the process is represented. The paradigm of such operators is the so-called fractional Laplacian, $(-\Delta)^{\sigma/2}$, is defined as follows through Fourier transform: if g is a function in the Schwartz class and $(-\Delta)^{\sigma/2}g = h$, then

$$\widehat{h}(\xi) = |\xi|^{\sigma}\widehat{g}(\xi). \tag{1}$$

If $0 < \sigma < 2$ we can also use the integral representation

$$(-\Delta)^{\sigma/2}g(x) = C_{N,\sigma}\,\mathrm{P.V.}\int_{\mathbb{R}^N} \frac{g(x)-g(z)}{|x-z|^{N+\sigma}}\,dz,$$

where P.V. stands for principal value and $C_{N,\sigma} = \frac{2^{\sigma-1}\sigma\Gamma((N+\sigma)/2)}{\pi^{N/2}\Gamma(1-\sigma/2)}$ is a normalization constant, see for example [51, 63]. Note that $C_{N,\sigma} \approx \sigma$ as $\sigma \to 0$ and $C_{N,\sigma} \approx 2-\sigma$ as $\sigma \to 2$. This allows to recover in the limits respectively the identity or the standard Laplacian. The operators $(-\Delta)^{-\sigma/2}$, $0 < \sigma < 2$, are inverse the former ones and are now given by standard convolution expressions

$$(-\Delta)^{-\sigma/2}g(x) = C_{N,-\sigma}\int_{\mathbb{R}^N} \frac{g(z)}{|x-z|^{N-\sigma}}\,dz,$$

in terms of Riesz potentials. The basic reference for these operators are the books by Landkof [51] and Stein [63]. The interest in these operators has a long history in Probability since the fractional Laplacian operators of the form $(-\Delta)^{\sigma/2}$, $\sigma \in (0,2)$, are infinitesimal generators of stable Lévy processes [4, 14], see also [64]. Motivation from Mechanics appears in the famous Signorini problem (with $\alpha = 1/2$), cf. [22, 59]. And there are applications in Fluid Mechanics, cf. [24, 45] and the references therein. An extensive list of current applications is contained in the survey paper [37].

The systematic study of the corresponding PDE models is more recent and many of the results have arisen in the last decade. The linear or quasilinear elliptic theory has been actively studied recently in the works of Caffarelli and collaborators [6, 8, 27], Kassmann [44], Silvestre [60] and many others. The standard linear evolution equation involving fractional diffusion is

$$\frac{\partial u}{\partial t} + (-\Delta)^{\sigma/2}(u) = 0. \tag{2}$$

This is a model of so-called anomalous diffusion, a much studied topic in physics, probability and finance, see for instance [1, 47, 48, 55, 71, 72] and their references. The equation is solved with the aid of well-known Functional Analysis tools; for instance, it is proved that it generates a semigroup of ordered contractions in $L^1(\mathbb{R}^n)$.

Moreover, in this setting it has the integral representation

$$u(x,t) = \int_{\mathbb{R}^N} K_\sigma(x-z,t) f(z)\, dz, \tag{3}$$

where K_σ has Fourier transform $\widehat{K}_\sigma(\xi,t) = e^{-|\xi|^\sigma t}$. This means that, for $0 < \sigma < 2$, the kernel K_σ has the form $K_\sigma(x,t) = t^{-N/\sigma} F(|x|\, t^{-1/\sigma})$ for some profile function F that is positive and decreasing and behaves at infinity like $F(r) \sim r^{-(N+\sigma)}$, [20]. When $\sigma = 1$, F is explicit; if $\sigma = 2$ the function K_2 is the Gaussian heat kernel.

However, an integral representation of the evolution of the form (3) is not available in the nonlinear models coming from the applications, thus motivating our work to be described below.

1.1 Nonlinear Evolution Models

A feature of current research in the area of PDEs is the interest in nonlinear equations and systems. The present article is devoted to presenting the progress achieved in two different models for flow in porous media including nonlocal (long-range) diffusion effects, represented by fractional operators.

- The first model is based on the usual Darcy law, with the novelty that the pressure is related to the density by an inverse fractional Laplacian operator. We prove existence of solutions that propagate with finite speed. The model has the very interesting property that mass preserving self-similar solutions can be found by solving an elliptic obstacle problem with fractional Laplacian for the pair pressure-density. We then use entropy methods to show that the asymptotic behavior is described after renormalization by these solutions which play the role of the Barenblatt profiles of the standard porous medium model. This is a joint ongoing project with Luis Caffarelli, Univ. Texas, cf. [25, 26]. Regularity is studied in joint work with Luis Caffarelli and Fernando Soria, [29].

  As a limit case of this model, we obtain a variant of the equation for the evolution of vortices in superconductivity derived heuristically by Chapman–Rubinstein–Schatzman [32] and W. E [38] as the hydrodynamic limit of Ginzburg–Landau, and studied by Lin and Zhang [53], and Ambrosio and Serfaty [3]. Below I will report on progress in understanding this limit in collaboration with Sylvia Serfaty [58].
- The second model is more in the spirit of fractional Laplacian flows, but nonlinear. Contrary to standard PME flows [68] it has infinite speed of propagation. But similarly to them, an L^1-contraction semigroup is constructed and it depends continuously on the exponent of fractional derivation and the exponent of the nonlinearity. Joint work with Arturo de Pablo, Fernando Quirós and Ana Rodriguez, Madrid. Two papers contain the progress done so far, [35, 36]. On the other hand, I. Athanasopoulos and L. Caffarelli studied in [7] the continuity of the weak solutions in the framework of more general boundary heat control problems.

1.2 Traditional Porous Medium Equations

Many of the concepts and techniques we will use come from the now classical theory of nonlinear diffusion. The simplest model of nonlinear diffusion equation is maybe

$$u_t = \Delta u^m = \nabla \cdot (c(u)\nabla u) \tag{4}$$

where $c(u) \geq 0$ indicates density-dependent diffusivity, in this case

$$c(u) = mu^{m-1}.$$

This is valid in the typical case where $u \geq 0$. For functions u with possibly negative signs we must put $c(u) = m|u|^{m-1}$ and then the equation reads $u_t = \Delta(|u|^{m-1}u)$.

It is clear that for $m = 1$ we recover the *classical Heat Equation*, while for $m > 1$ the equation is degenerates at $u = 0$, which is important in many applications and means slow diffusion.

A Model for Gases in Porous Media The model arises from the consideration of a continuum, say, a fluid, represented by a *density* distribution $u(x,t) \geq 0$ that evolves with time following a *velocity field* $\mathbf{v}(\mathbf{x},\mathbf{t})$, according to the continuity equation

$$u_t + \nabla \cdot (u\mathbf{v}) = 0. \tag{5}$$

We assume next that $\mathbf{v}$ derives from a potential, $\mathbf{v} = -\nabla p$, as happens in fluids in porous media according to Darcy's law, and in that case p is the *pressure*. But potential velocity fields are found in many other applied instances, like Hele-Shaw cells, and other recent examples.

We still need a closure relation to relate u and p. In the case of gases in porous media, as modeled by Leibenzon and Muskat, the closure relation takes the form of a state law $p = f(u)$, where f is a nondecreasing scalar function, which is linear when the flow is isothermal, and a power, i.e., $f(u) = cu^{m-1}$ with $c > 0$ and $m > 1$, if it is adiabatic.

The linear relationship happens also in the simplified description of water infiltration in an almost horizontal soil layer according to Boussinesq. In both cases we get the standard porous medium equation, $u_t = c\Delta(u^2)$. See [68] for these and many other applications.

Fast Diffusion On the contrary, if $m < 1$ the equation becomes singular at $u = 0$ (i.e., $c(0) = +\infty$) which means Fast Diffusion. This equation has very different properties, like infinite speed of propagation and extinction in finite time; as m goes down to zero (or below) some quite uncommon and interesting features appear, like instantaneous extinction, [67].

General Models A more general model of nonlinear diffusion takes the divergence form

$$\partial_t H(u) = \nabla \cdot \vec{\mathscr{A}}(x, u, Du) + \mathscr{B}(x, t, u, Du) \tag{6}$$

with monotonicity conditions on H and $\nabla_p \vec{\mathscr{A}}(x, t, u, p)$ and structural conditions on $\vec{\mathscr{A}}$ and $\mathscr{B}$. This generality includes *Stefan Problems*, *p-Laplacian flows* (including $p = \infty$ and total variation flow $p = 1$) and many others, but this generality does not allow for a detailed theory see for instance [13].

Historical Mention and References Well-known work starting in Moscow with Zeldovich, Raizer [73] and Barenblatt [10] around 1950 and the first systematic theory by Oleinik et al. in 1958 [56], and then Kalashnikov, Aronson, Benilan, Brezis, Caffarelli, Crandall, Di Benedetto, Friedman, Kamin, Kenig, Peletier, Vázquez, and many others.

Let us now mention some topics and authors in the new century: the group Carrillo, Toscani, Dolbeault, Del Pino, Markowich, Otto, on entropies and gradient flow and functional inequalities; Daskalopoulos, Hamilton, Lee, Vázquez on concavity. Many works on Fast Diffusion flows and logarithmic diffusion, on p-Laplacian flows, with recent interest on L_∞ and L_1 Laplacians, and more.

Let us finally list some convenient general references. About the PME there is a comprehensive monograph by the author, "The Porous Medium Equation. Mathematical Theory", [68]. Earlier expositions are due to Peletier [57] and Aronson [5]. About estimates and scaling we refer to the book [67] which covers also many aspects of fast diffusion. The topic of asymptotic behavior has an enormous literature following the ideas of Lyapunov and Boltzmann. We have explained the proof of asymptotic convergence for the PME in two surveys, [65] for the Cauchy problem and [66] for the Dirichlet problem in a bounded domain. A more general survey on Nonlinear Diffusion is contained in the Proceedings of the International Congress of Mathematicians, ICM Madrid 2006 [69].

2 Nonlocal Diffusion Model of Porous Medium Type

The first diffusion model with nonlocal effects we will present here uses the beginning steps of the previous derivation of the equation for gases in porous media but differs in the *closure relation* between the density and the pressure that takes the form $p = \mathscr{K}(u)$, where $\mathscr{K}$ is a linear integral operator, which we assume in practice to be the inverse of a fractional Laplacian. Hence, p es related to u through a nonlocal operator $\mathscr{K}$ which in the prototype case is the fractional potential operator, $\mathscr{K} = (-\Delta)^{-s}$ with kernel

$$k(x, y) = c|x - y|^{-(n-2s)} \tag{7}$$

(i.e., a Riesz operator). We have $(-\Delta)^s p = u$. The diffusion model with nonlocal effects is thus given by the system

$$u_t = \nabla \cdot (u \nabla p), \quad p = \mathscr{K}(u) = (-\Delta)^{-s} u, \tag{8}$$

where u is a function of the variables (x,t) to be thought of as a density or concentration, and therefore nonnegative, while p is the pressure, which is related to u via a linear operator $\mathscr{K}$.

The problem is posed for $x \in \mathbb{R}^n$, $n \geq 1$, and $t > 0$, and we give initial conditions

$$u(x,0) = u_0(x), \quad x \in \mathbb{R}^n, \tag{9}$$

where u_0 is a nonnegative, bounded and integrable function in $\mathbb{R}^n$.

Precedents The interest in using *fractional Laplacians* in modeling diffusive processes has a wide literature, especially when one wants to model long-range diffusive interaction, and this interest has been activated by the recent progress in the mathematical theory as represented in [6, 8, 27], the thesis work by Silvestre [60], and many others.

A variant of the proposed model was studied by Lions and Mas-Gallic [54]. They study the *regularization of the velocity field* in the standard porous medium equation by means of a convolution kernel to get a system like ours, with a difference, namely that they assume the kernel to be smooth and integrable. Since the kernel of the fractional operator $(-\Delta)^s$ is $k(x,y) = |x-y|^{-(n-2s)}$, we are far away from that case, but it may serve as a regularization step below.

Modeling Dislocation Dynamics as a Continuum There is a model for such dislocation phenomena proposed by A.K. Head [43] that leads to our equation in one space dimension with $s = 1/2$. It is written in an integrated version as

$$v_t = |v_x| \Lambda^\alpha(v)$$

with $\Lambda^\alpha = (-\partial^2/\partial_x^2)^\alpha$. The model applies when $\alpha = 1$, and the dislocation density is $u = v_x$. This model has been recently studied by P. Biler, G. Karch, and R. Monneau, [19]. For the integrated version they introduce viscosity solutions à la Crandall–Evans–Lions. This version has the properties of uniqueness and comparison of solutions, which makes for a simpler mathematical analysis. The study of many-dimensional models for dislocations is a widely open matter.

Limit Cases

- If we take $s = 0$, then $\mathscr{K} =$ the identity operator, and we get the *standard porous medium equation*, whose behavior is well-known, as explained above.
- In the other end of the s interval, the case $s = 1$ is novel and interesting. We take $\mathscr{K} = -\Delta$ we get

$$u_t = \nabla u \cdot \nabla p - u^2, \quad -\Delta p = u. \tag{10}$$

In one dimension this leads to

$$u_t = u_x p_x - u^2, \quad p_{xx} = -u.$$

It is then convenient to introduce the intermediate variable $v = -p_x = \int u\,dx$. We have

$$v_t = up_x + c(t) = -v_x v + c(t).$$

For $c = 0$ this is the *Burgers equation* $v_t + vv_x = 0$ which generates shocks in finite time. Note that we may allow for u to have two signs.

Variants of this limit case in two space dimensions are used to model the evolution of vortices in superconductivity in [53] and [3], where u describes the vortex-density. The problem is sometimes posed in a bounded domain with appropriate (nonhomogeneous) boundary conditions. See Sect. 6 below.

Summing up, the equation we study for $0 < s < 1$ may be viewed as a sort of interpolation between the extreme cases. It has better regularity properties than $s = 1$ but is different in many properties from $s = 0$.

General Classes of Equations More ambitious mathematical theories are being considered. Thus, it could be assumed that $\mathscr{K}$ is an operator of integral type defined by convolution on all of $\mathbb{R}^n$, with the assumptions that is positive and symmetric. The fact the $\mathscr{K}$ is a homogeneous operator of degree $2s$, $0 < s < 1$, will be important in the proofs. An interesting variant would be $\mathscr{K} = (-\Delta + cI)^{-s}$. We are not exploring such extensions.

A Formal Analogue. Aggregation Equations Recent work of A. Bertozsi and collaborators has focused on aggregation models. One of them is formally the same as our porous medium equation

$$u_t = \nabla \cdot (u \nabla K \star u),$$

cf. [15–17]. However, the kernels that allow for aggregation phenomena are quite different, they are regular or in any case never very singular. A typical condition is: K radial and $\nabla K \in L^2(\mathbb{R}^n)$, and $\Delta K \in L^p(\mathbb{R}^n)$ with $p \in [2n/(n+2), 2]$, see [15]. Contrary to the theory we develop below, that model may lead to blow up in finite time. In [17] K is radially symmetric with a singularity at the origin of order $|x|^\sigma$ with $\sigma > 2 - n$, precisely outside of the fractional Laplacian range in which the nonlocal diffusion theory is set.

3 Mathematical Theory for the Model of Fractional Porous Medium Equation

The work that is presented next is explained in whole detail in the following papers [25, 26, 29]. The first deals with existence and basic propagation properties, the second about boundedness and regularity in the spirit of De Giorgi [33], and the third deals with asymptotic behavior through the associated obstacle problem and entropy dissipation methods.

3.1 Main Estimates

It is convenient to write the Fractional Porous Medium Equation (8) in the more general form $\partial_t u = \nabla \cdot (u\nabla \mathscr{K}(u))$. The equation is posed in the whole space $\mathbb{R}^n$ (work on the problem posed on bounded domains is in progress). We consider $\mathscr{K} = (-\Delta)^{-s}$ for some $0 < s < 1$ acting on Schwartz class functions defined in the whole space. It is a positive essentially self-adjoint operator. We also let $\mathscr{H} = \mathscr{K}^{1/2} = (-\Delta)^{-s/2}$. We take a fixed $s \in (0, 1)$. When necessary we indicate the dependence on s as follows: $\mathscr{K}_s$, $\mathscr{H}_s$.

We do at this stage formal calculations, assuming that $u \geq 0$ satisfies the required smoothness and integrability assumptions. This is to be justified by approximation. See whole details in [25].

- Conservation of mass

$$\frac{d}{dt}\int u(x,t)\,dx = 0. \tag{11}$$

- First energy estimate:

$$\frac{d}{dt}\int_{\mathbb{R}^n} u(x,t)\log u(x,t)\,dx = -\int_{\mathbb{R}^n} (\nabla u \cdot \nabla \mathscr{K} u)\,dx = -\int_{\mathbb{R}^n} |\nabla \mathscr{H} u|^2\,dx, \tag{12}$$

  where we use the fact that $\mathscr{K} = \mathscr{H}^2$, and $\mathscr{H}$ is a positive self-adjoint operator that commutes with the gradient.
- Second energy estimate

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}^n} |\mathscr{H} u(x,t)|^2\,dx &= \int_{\mathbb{R}^n} (\mathscr{H} u)(\mathscr{H} u)_t\,dx = \int_{\mathbb{R}^n} (\mathscr{K} u)\,u_t\,dx \\ &= \int_{\mathbb{R}^n} (\mathscr{K} u)\nabla\cdot(u\nabla \mathscr{K} u)\,dx \\ &= -\int_{\mathbb{R}^n} u|\nabla \mathscr{K} u|^2\,dx. \end{aligned}\tag{13}$$

- Conservation of positivity: $u_0 \geq 0$ implies that $u(t) \geq 0$ for all times.
- L^∞ estimate. We prove that the L^∞ norm does not increase in time.

Sketch of proof At a point of maximum of u at time $t = t_0$, say $x = 0$, we have

$$u_t = \nabla u \cdot \nabla P + u\Delta \mathscr{K}(u),$$

where $P = \mathscr{K}(u)$. The first term is zero, and for the second we have $-\Delta\mathscr{K} = L$ where $L = (-\Delta)^q$ with $q = 1 - s$ so that

$$\Delta \mathscr{K} u(0) = -Lu(0) = -c\int \frac{u(0)-u(y)}{|y|^{n+2(1-s)}}\,dy \leq 0.$$

This concludes the proof. □

- The L^p norm of the solution does not increase in time for all $1 \le p \le \infty$.
- INVARIANCE AND SCALING GROUP. The equation is clearly invariant under translations in space and time. More interesting is the observation that it is also invariant under a two-parameter scaling group. Thus, if $u(x,t)$ is a weak solution so is the rescaled function

$$\widetilde{u}(x,t) := Au(Bx, Ct) \tag{14}$$

for arbitrary constants $A, B > 0$ under the condition that $C = AB^{2-2s}$. This is based on the dimensional estimate $(\mathscr{K}\widetilde{u})(x,t) = AB^{-2s}\,(\mathscr{K}u)(Bx, Ct)$ and direct calculation on the equation.
- We did not find a clean comparison theorem, a form of the usual maximum principle is not proved, and there are counterexamples for $s > 1/2$ in all space dimensions. However, comparison of solutions is established in [19] for the integrated version in dimension $n = 1$ by techniques of viscosity solutions.

3.2 Finite Propagation. Solutions with Compact Support

One of the most important features of the porous medium equation and other related degenerate parabolic equations is the property of finite propagation, whereby compactly supported initial data $u_0(x)$ gives rise to solutions $u(x,t)$ that have the same property for all positive times, i.e., the support of $u(\cdot,t)$ is contained in a ball $B_{R(t)}(0)$ for all $t > 0$. One possible proof in the case of the PME is by constructing explicit weak solutions exhibiting that property (i.e., having a free boundary) and then using the comparison principle, that holds for that equation. Since we do not have such a general principle here, we have to devise a comparison method with a suitable family of "true supersolutions", which are in fact some quite excessive supersolutions. The technique has to be adapted to the peculiar form of the integral kernels involved in operator $\mathscr{K}_s$.

We begin with $n = 1$ for simplicity. We assume that our solution $u(x,t) \ge 0$ has bounded initial data $u_0(x) = u(x,t_0) \le M$ with compact support and is such that

$$u_0 \quad \text{is below the parabola} \quad a(x-b)^2, a, b > 0,$$

with graphs strictly separated. We may assume that u_0 is located under the left branch of the parabola. We take as comparison function

$$U(x,t) = a(Ct - (x-b))^2,$$

which is a traveling wave moving to the right with speed C that will be taken big enough. Then we argue at the first point and time where $u(x,t)$ touches the left branch of the parabola U from below. The key point is that if C is large enough such contact cannot exist. The formal idea is to write the equation as

$$u_t = u_x p_x + u p_{xx}$$

and observe that at the contact we have $u_t \geq U_t = 2aC(Ct - x + b)$, while $u_x = U_x = -2a(Ct - x + b)$, so the first can be made much bigger than the second by increasing C. The influence of p_x and p_{xx} as well as u is controlled, and then we conclude that the equation cannot hold if C is large enough. The argument can be translated for several dimensions. Here are the detailed results proved in [25].

Theorem 1 *Let* $0 < s < 1/2$ *and assume that* u *is a bounded solution of Eq.* (8) *with* $0 \leq u(x,t) \leq L$, *and* u_0 *lies below a function of the form*

$$U_0(x) = Ae^{-a|x|}, \quad A, a > 0. \tag{15}$$

If A *is large then there is a constant* $C > 0$ *that depends only on* (n, s, a, L, A) *such that for any* $T > 0$ *we will have the comparison*

$$u(x,t) \leq Ae^{Ct - a|x|} \quad \text{for all } x \in \mathbb{R}^n \text{ and all } 0 < t \leq T. \tag{16}$$

Theorem 2 *Let now* $1/2 \leq s < 1$. *Under the assumptions of the previous theorem the stated tail estimate works locally in time. The global statement must be replaced by the following*: *there exists an increasing function* $C(t)$ *such that*

$$u(x,t) \leq Ae^{C(t)t - a|x|} \quad \text{for all } x \in \mathbb{R}^n \text{ and all } 0 < t \leq T. \tag{17}$$

3.3 Instantaneous Boundedness and Regularity

- **Solutions are bounded in terms of data in** L^p, $1 \leq p \leq \infty$. This is a typical property of the heat semigroup and a wide class of parabolic equations with variable coefficients. The classical method of De Giorgi or Moser based on iterative techniques can be adapted to fractional diffusion in linear or nonlinear cases. This was done for instance by Caffarelli and Vasseur [24] by using the Caffarelli–Silvestre extension [23]. See also [11, 12]. Or we can use energy estimates based on the properties of the quadratic and bilinear forms associated to fractional operator, as done in [19]. For the equation and generality at hand, this is done in our paper [29] by the De Giorgi method.

Theorem *Let* u *be a weak solution the Initial Value Problem for the Fractional Porous Medium Equation* (8) *with data* $u_0 \in L^1(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$, *as constructed before. Then, there exists a positive constant* C *such that for every* $t > 0$

$$\sup_{x \in \mathbb{R}^n} |u(x,t)| \leq Ct^{-\alpha} \|u_0\|^{\gamma}_{L^1(\mathbb{R}^n)} \tag{18}$$

with $\alpha = n/(n + 2 - 2s)$, $\gamma = (2 - 2s)/(n + 2 - 2s)$. *The constant* C *depends only on* n *and* s.

- **Continuity**. Bounded weak solutions $u \geq 0$ of problem (8)–(9) are uniformly continuous on bounded sets of $s < 1$. Indeed, they are C^{α} continuous with a uniform modulus.

The proof done in [29] is lengthy and uses many techniques of the local regularity theory for elliptic and parabolic PDEs developed by Caffarelli and collaborators, and in particular some of the new ideas contained in Caffarelli–Chan–Vasseur [28]. The crucial point is to get a local version of the energy inequalities that can be iterated. It involves a delicate manipulation of the bilinear forms associated to the fractional operator, which amounts to knowing well the H^s spaces and then doing nonlinear versions of the embeddings and bounds.

4 Asymptotic Behavior for Standard PME Flow

In order to motivate the results for fractional diffusion, it is convenient to review the main results known for plain porous medium flow.

4.1 Barenblatt Profiles and Asymptotics

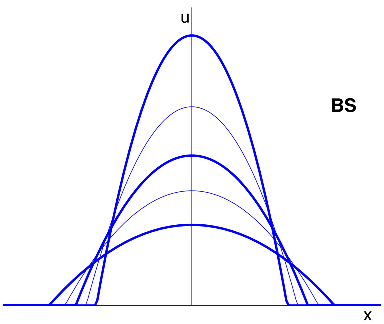
These profiles are the alternative to the Gaussian profiles that explain the asymptotic behavior in the heat equation flow. They are called *source-type solutions*. Here source means that $u(x,t) \to M\,\delta(x)$ as $t \to 0$. There exist explicit formulas for all $m > 1$ (1950, 52) [10, 73]:

$$\mathbf{B}(x,t;M) = t^{-\alpha}\mathbf{F}(x/t^{\beta}), \qquad \mathbf{F}(\xi) = (C - K\xi^2)_+^{1/(m-1)} \tag{19}$$

where the similarity exponents are smaller than in the Gaussian case:

$$\alpha = \frac{n}{2+n(m-1)} < \frac{n}{2}, \qquad \beta = \frac{1}{2+n(m-1)} < 1/2 \tag{20}$$

and the profile looks like



The difference with the Gaussian case is striking: the solution has no tail, but a compact support limited by a clearcut *free boundary* of propagation front. The solution has height $u = Ct^{-\alpha}$ and the *free boundary* at the distance $|x| = ct^{\beta}$.

We point for future reference the ideas of *Scaling law* and that of *anomalous diffusion* versus *Brownian motion* (where $\beta = 1/2$).

4.2 Nonlinear Central Limit Theorem

The standard porous medium flow has an asymptotic stabilization property that parallels the stabilization to the Gaussian profile embodied in the classical Central Limit Theorem. The choice of domain for such results is $\mathbb{R}^n$. Choice of data: $u_0(x) \in L^1(\mathbb{R}^n)$. We write the equation we can deal with as

$$u_t = \Delta(|u|^{m-1}u) + f. \tag{21}$$

We assume that $m > 1$. Let us put $f \in L^1(\mathbb{R}^n \times \mathbb{R}^+)$. Let $M = \int u_0(x)\,dx + \iint f\,dxdt$, called the total or final mass.

Theorem 3 *Let $B(x,t;M)$ be the Barenblatt solution with mass M equal to the asymptotic mass of u; u converges to B in the form*

$$\boxed{t^{\alpha}\|u(t) - B(t)\|_1 \to 0,} \tag{22}$$

as $t \to \infty$. Moreover, if $f = 0$ we have

$$\boxed{t^{\alpha}|u(x,t) - B(x,t)| \to 0} \tag{23}$$

uniformly in $x \in \mathbb{R}^n$ and for every $p \geq 1$ we have

$$\boxed{\|u(t) - B(t)\|_p = o(t^{-\alpha/p'}), \quad p' = p/(p-1).} \tag{24}$$

This is the main asymptotic theorem for the PME, proved in complete form by Vázquez in 2001, [65], expanding on the result by Friedman and Kamin, 1980, [40], where the authors took $u_0 \geq 0$, with compact support, and $f = 0$. I think it deserves the name of *Nonlinear Central Limit Theorem*. Note that the time weights are just the ones suggested by the size of the Barenblatt solutions, making the result precise.

Remarks

(1) When seeing the result from a numerical point of view, α and $\beta = \alpha/n = 1/(2+n(m-1))$ are the *zooming exponents*, just as in $B(x,t)$.
(2) The result is still true for $m \in (0,1)$ (Fast Diffusion) if $m > (n-2)/n$, see proof in [65], but not below the critical exponent $(n-2)/n$ where the situation is quite different. It has been studied by various authors in recent times in considerable detail, and general accounts are given in [21, 67].

(3) There are a number of improvements on this theorem, that were addressed around 2000. We will mention two: eventual geometry (Lee and Vázquez (2003), [52]) and establishing convergence rates. and explain only the latter, since it motivates the work on nonlinear fractional diffusion.

4.3 Calculation of Convergence Rates

This is the question of speed of convergence in formulas (22)–(24). The study was initiated by Carrillo and Toscani in 2000, [30], and there many interesting contributions (by Carrillo, Del Pino, Dolbeault, Markowich, McCann, Vázquez, and many others). Using entropy functional with entropy dissipation control you can prove decay rates when $\int u_0(x)|x|^2\,dx < \infty$ (finite variance):

$$\|u(t) - B(t)\|_1 = O(t^{-\delta}). \tag{25}$$

We would like to have $\delta = 1$. This problem is still open for $m > 2$.

The Entropy Method We rescale the function as $u(x,t) = r(t)^n \rho(y,s)$ with $x = yr(t)$ where $r(t) = c_M(t+1)^\beta$ is the Barenblatt radius at time $t+1$, and the "new time" is $s = \log(1+t)$. The PME becomes

$$\rho_s = \frac{1}{m}\Delta(\rho^m) + c\nabla(y\rho) = \operatorname{div}\left(\rho\left\{\nabla\rho^{m-1} + \frac{c}{2}\nabla y^2\right\}\right). \tag{26}$$

Then we define the **entropy** as

$$E(\rho)(s) := \int \left(\frac{1}{m}\rho^m + \frac{c}{2}\rho y^2\right) dy. \tag{27}$$

A key point is that the minimum of this entropy is identified as the entropy of the Barenblatt profile. Next, we calculate

$$\frac{dE}{ds} = -\int \rho|\nabla\rho^{m-1} + cy|^2\,dy = -D(\rho(\cdot,s)).$$

It is illuminating in this respect to notice that, when written in the variable ρ as a function of y and s, the self-similar Barenblatt solutions become stationary solutions $\overline{\rho}_M(y)$ of Eq. (26) (with inverted parabolic shape), and then it is easy to see that the dissipation $D(\overline{\rho}) = 0$, as befits a limit of an orbit according to the theory of Dynamical Systems. Moreover, it is shown that the minimum of this entropy along an orbit is the entropy of the stationary Barenblatt profile $\overline{\rho}_M$ with the same mass M. Moreover, by another round of (not so easy) time differentiation and manipulation we get along a rescaled orbit $\rho(\cdot,s)$ the expression

$$\frac{dD(\rho)}{ds} = -R(\rho) < 0,$$

and moreover we can prove that $R(\rho) \sim \lambda D(\rho)$ (so-called Bakry–Emery calculation, cf. [30]). We conclude exponential decay of D, and then of $E - E_{\min}$, in terms of the *new time* s, which in turn means power decay in the *real time* t.

4.4 Rates Through Entropies for Fast Diffusion

A large effort has been invested in making this machinery work for fast diffusion, $-\infty < m < 1$. The nice properties entropies have from the point of view of transport theory (cf. [70]) are lost soon, more precisely, when $m = (n-1)/n$. Indeed, the entropy of typical solutions is no more finite when the second moment is infinite, i.e., for $m = (n-1)/(n+1)$. The attractor of the evolution, i.e., the finite-mass Barenblatt solutions are lost for $m = (n-2)/n$.

The analysis for $m < (n-2)/n$ took a time to develop. A main feature is that solutions that decay reasonably at infinity will vanish completely in finite time, [67]. There is work by many authors: Blanchet, Bonforte, Carrillo, Dolbeault, Del Pino, Denzler, Grillo, McCann, Vázquez.... A rather definitive account is contained in a note just appeared in Proc. Natl. Acad. Sci. USA, [21]. See previously [34]. A quite rewarding mathematical feature of those analysis is the fact that functional inequalities play a crucial role in the asymptotic analysis, they are so to say "equivalent" to the form of asymptotic stabilization.

5 Asymptotic Behavior for the FPME

We now begin the study of the large time behavior of the proposed model of nonlocal diffusion (i.e., the FPME) following paper [26]. The first step is constructing the self-similar solutions that will serve as attractors.

5.1 Rescaling for the FPME

Inspired by the asymptotics of the standard porous medium equation, we define the rescaled (also called renormalized) flow through the transformation

$$u(x,t) = (t+1)^{-\alpha} v(x/(t+1)^{\beta}, \tau) \tag{28}$$

with new time $\tau = \log(1+t)$. We also put $y = x/(t+1)^{\beta}$ as rescaled space variable. In order to cancel the factors including t explicitly, we get the condition on the exponents

$$\alpha + (2-2s)\beta = 1. \tag{29}$$

Here we use the homogeneity of $\mathcal{K}$ in the form $(\mathcal{K}u)(x,t) = t^{-\alpha+2s\beta}(\mathcal{K}v)(y,\tau)$. From physical considerations we also impose the law that states conservation of (finite) mass, which amounts to the condition $\alpha = n\beta$, and in this way we arrive at the precise value for the exponents:

$$\beta = 1/(n+2-2s), \qquad \alpha = n/(n+2-2s). \tag{30}$$

Renormalized Flow We also arrive at the *nonlinear, nonlocal Fokker–Planck equation*

$$\boxed{v_\tau = \nabla_y \cdot (v(\nabla_y \mathscr{K}(v) + \beta y)).} \tag{31}$$

The transformation formula implies a transformation for the pressure of the form

$$p(u)(x,t) = (t+1)^{-\sigma} p(v)(x/(t+1)^{\beta}, \tau), \quad \text{with } \sigma = \alpha - 2s\beta = 1 - 2\beta < 1.$$

This last formula does not play a big role below. In all the above calculations the factor $(t+1)$ can be replaced by $t + t_0$ for any $t_0 > 0$, or even by plain t.

Stationary Renormalized Solutions It is important to concentrate on the stationary states of the new equation, i.e., on the solutions $V(y)$ of

$$\nabla_y \cdot (V \nabla_y (P + a|y|^2)) = 0, \quad \text{with } P = \mathscr{K}(V), \tag{32}$$

where $a = \beta/2$, and β is defined just above. Since we are looking for asymptotic profiles of the standard solutions of the FPME we also want $V \geq 0$ and integrable. The simplest possibility is integrating once to get

$$V \nabla_y (P + a|y|^2) = 0, \quad P = \mathscr{K}(V), \; V \geq 0. \tag{33}$$

The first equation gives an alternative choice that reminds us of the complementary formulation of the obstacle problems.

5.2 *Obstacle Problem. Barenblatt Solutions of New Type*

Indeed, if we solve the obstacle problem with fractional Laplacian we will obtain a unique solution $P(y)$ of the problem:

$$\begin{aligned} &P \geq \Phi, \quad V = (-\Delta)^s P \geq 0; \\ &\text{either } P = \Phi \;\text{ or }\; V = 0 \end{aligned} \tag{34}$$

with $0 < s < 1$. In order for solutions of (34) to be also solutions of (33) we have to choose as obstacle

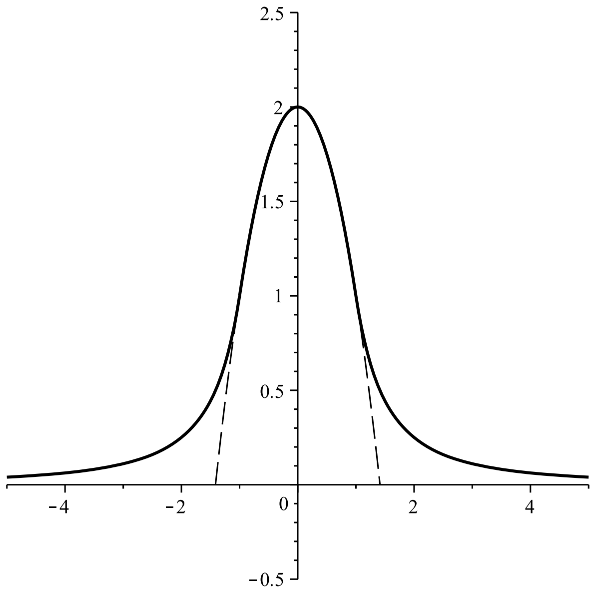
$$\Phi = C - a|y|^2, \tag{35}$$

where C is any positive constant and $a = \beta/2$. Note that $-\Delta\Phi = 2na = \alpha$. For uniqueness we also need the condition $P \to 0$ as $|y| \to \infty$. Fortunately, the corresponding theory had been developed by Caffarelli and collaborators, cf. [8, 27], and Silvestre's thesis [61]. The solution is unique and belongs to the space H^{-s} with pressure in H^s. Moreover, it is shown that the solutions have $P \in C^{1,s}$ and $V \in C^{1-s}$.

Note that for $C \le 0$ the solution is trivial, $P = 0$, $V = 0$, hence we choose $C > 0$. We also note the pressure is defined but for a constant, so that we could maybe take as pressure $\widehat{P} = P - C$ instead of P so that $\widehat{P} = 0$; but this does not simplify things since $P \to 0$ implies that $\widehat{P} \to -C$ as $|y| \to \infty$. Keeping thus the original proposal, we get a one parameter family of stationary profiles that we denote $V_C(y)$. These solutions of the obstacle problem produce correct weak solutions of the fractional PME equation with initial data a multiple of the Dirac delta for the density, in the form

$$U_C(x,t) = t^{-\alpha} V_C(|x|t^{-\beta}). \tag{36}$$

It is what we can call the source-type or Barenblatt solution for this problem, which is a profile $V \ge 0$. It is positive in the *contact set* of the obstacle problem, which has the form $\mathscr{C} = \{|y| \le R(C)\}$, and is zero outside, hence it has compact support. It is clear that R is smaller than the intersection of the parabola Φ with the axis $R_1 = (C/a)^{1/2}$. On the other hand, the rescaled pressure $P(|y|)$ is always positive and decays to zero as $|y| \to \infty$ according to fractional potential theory, cf. Stein [63]. The rate of decay of P as $|y| \to \infty$ turns out to be $P = O(|y|^{2s-n})$.



The solution of the obstacle problem with parabolic obstacle

Calculation of Density Profiles Biler, Karch and Monneau [19] studied the existence and stability of self-similar solutions in one space dimension. Recently, Biler, Imbert and Karch [18] obtain the explicit formula for a multi-dimensional self-similar solution in the form

$$U(x,t) = c_1 t^{-\alpha}(1 - x^2 t^{-2\alpha/n})_+^{1-s} \tag{37}$$

with $\alpha = n/(n+2-2s)$ as before. The derivation uses an important identity for fractional Laplacians which is found in Getoor [41]: $(-\Delta)^{\sigma/2}(1-y^2)_+^{\sigma/2} = K_{\sigma,n}$ if $\sigma \in (0,2]$. Here we must take $\sigma = 2(1-s)$. According to our previous calculations $\Delta P = -\alpha$ on the coincidence set, hence $c_1 = \alpha/K_{\sigma,n}$. Let us work a bit more: using the scaling (14) with $A = C$ and $B = 1$ we arrive at the following one-parameter family of self-similar solutions

$$U(x,t;C_1) = t^{-\alpha}(C_1 - k_1 x^2 t^{-2\alpha/n})_+^{1-s} \tag{38}$$

where $k_1 = c_1^{1/(1-s)}$ and $C_1 > 0$ is a free parameter that can be fixed in terms of the mass of the solution $M = \int U(x,t;C_1)\,dx$. This is the family of densities that corresponds to the pressures obtained above as solutions of the obstacle problem. All is quite similar to the formulas for the standard PME, [68]; note however that in the fractional case the pressure is not compactly supported but has a power tail at infinity, which points to the long-range effects.

5.3 Estimates for the Rescaled Problem. Entropy Dissipation

The next step is to prove that these profiles are attractors for the rescaled flow. We review the estimates of Sect. 3.1 above in order to adapt them to the rescaled equation (31).

There is no problem reproving mass conservation or positivity.

The first energy estimate becomes (recall that $\mathscr{H} = \mathscr{K}^{1/2}$)

$$\begin{aligned}\frac{d}{d\tau}\int v(y,\tau)\log v(y,\tau)\,dy &= -\int |\nabla \mathscr{H} v|^2\,dy - \beta\int \nabla v\cdot y\\ &= -\int |\nabla \mathscr{H} v|^2\,dy + \alpha\int v.\end{aligned}$$

We are going to base the proof of asymptotic behavior on the second energy estimate after an essential change. We define the *entropy* of the rescaled flow as

$$\mathscr{E}(v(\tau)) := \frac{1}{2}\int_{\mathbb{R}^n} (v\mathscr{K}(v) + \beta y^2 v)\,dy. \tag{39}$$

The entropy contains two terms. The first is

$$E_1(v(\tau)) := \int_{\mathbb{R}^n} v\mathscr{K}(v)\,dy = \int_{\mathbb{R}^n} |\mathscr{H} v|^2\,dy,$$

hence positive. The second is the moment $E_2(v(\tau)) = M_2(v(\tau)) := \int y^2 v\,dy$, also positive. By differentiation we get

$$\frac{d}{d\tau}\mathscr{E}(v) = -\mathscr{I}(v), \quad \mathscr{I}(v) := \int \left|\nabla\left(\mathscr{K}v + \frac{\beta}{2}y^2\right)\right|^2 v\,dy. \tag{40}$$

This means that whenever the initial entropy is finite, then $\mathscr{E}(v(\tau))$ is uniformly bounded for all $\tau > 0$, $\mathscr{I}(v)$ is integrable in $(0,\infty)$ and

$$\mathscr{E}(v(\tau)) + \int_0^\tau \int \left|\nabla\left(\mathscr{K}v + \frac{\beta}{2}y^2\right)\right|^2 v\,dy\,dt \le \mathscr{E}(v_0).$$

5.4 Convergence

The standard idea is to let $\tau \to \infty$ in the renormalized flow $v(\tau) = v(\cdot,\tau)$. The estimates we have just derived will be used here in the form of uniform bounds for the rescaled orbits in different norms and this will allow us to pass to the limit. Actually, since the entropy goes down there is a limit

$$E_* = \lim_{\tau\to\infty} \mathscr{E}(v(\tau)) \ge 0.$$

Notice that the family $v(\tau)$ is bounded in L^1_y uniformly in τ, and also vy^2 is bounded in L^1_y unif. in τ, and moreover $|\nabla\mathscr{H}(v(\tau))| \in L^2_y$ unif. in τ, we have that $v(\tau)$ is a compact family in $L^1(\mathbb{R}^n)$ (since there is local compactness by Nash–Sobolev embeddings and uniform mass control at infinity). It follows that there is a subsequence $\tau_j \to \infty$ that converges in L^1_y and almost everywhere to a limit $v_* \ge 0$. The mass of v_* is the same mass of u since the tail is uniformly small (tight convergence). One consequence is that the lim inf of the component $E_2(v(\tau_j))$ is equal or larger that $M_2(v_*)$ (by Fatou).

We also have $\mathscr{H}(v) \in L^2_y$ uniformly in t. The boundedness of $\nabla\mathscr{H}(v)$ in L^2_y implies the compactness of $\mathscr{H}(v)$ in space, so that it converges along a subsequence to $\mathscr{H}(v_*)$. This allows to pass to the limit in $E_1(v(\tau_j))$ and obtain a correct limit. After some more arguments detailed in [26] we get the consequence that for every $h > 0$ fixed

$$\int_{\tau_j}^{\tau_j+h} \int \left|\nabla\left(\mathscr{K}v + \frac{\beta}{2}y^2\right)\right|^2 v\,dy\,d\tau \to 0.$$

This implies that if $w(y,\tau) = \mathscr{K}v + \frac{\beta}{2}y^2$ and $w_h(y,\tau) = w(y,\tau+h)$, then w_h converges to a constant in space wherever v is not zero, and that constant must be $\mathscr{K}v_* + \frac{\beta}{2}y^2$ along the said subsequence, hence constant also in time.

Finally, after a rather delicate analysis, it is concluded that the limit is a solution of the Barenblatt obstacle problem. The final result is stated in [26] as follows

Theorem 4 *Let $u(x,t) \ge 0$ be a weak solution of Problem* (8)–(9) *with bounded and integrable initial data such that $u_0 \ge 0$ has finite entropy in the sense defined in formula* (39). *Let $v(y,\tau)$ be the corresponding rescaled solution. As $\tau \to \infty$ we have*

$$v(\cdot,\tau) \to V_C(y) \quad \text{in } L^1(\mathbb{R}^n) \text{ and also in } L^\infty(\mathbb{R}^n). \tag{41}$$

The constant C is determined by the rule of mass equality: $\int_{\mathbb{R}^n} v(y,\tau)\,dy = \int_{\mathbb{R}^n} V_C(y)\,dy$. In terms of function u, this translates into

$$\begin{aligned} &u(x,t) - U_C(x,t) \to 0 \quad \text{in } L^1(\mathbb{R}^n),\\ &t^\alpha |u(x,t) - U_C(x,t)| \to 0 \quad \text{uniformly in } x, \end{aligned} \tag{42}$$

both limits taken as $t \to \infty$.

Theorem 1.5 of [19] gives an equivalent asymptotic behavior result in $n = 1$, though the formulation is different.

6 Limits

- **The limit** $s \to 1$. I recall that the work by Lin and Zhang [53] on the dynamics of the Ginzburg–Landau vortices in the hydrodynamic limit arrives at equation for the density: $u_t + \nabla(u\,\nabla\Delta^{-1}u) = 0$ posed in dimension 2. The authors prove existence and uniqueness of positive L^∞ solutions (they also prove existence of positive-measure valued solutions). Existence with positive initial data of finite energy is also proven (for a slightly different model) by a gradient flow approach, in bounded domains of the plane by Ambrosio et al. in [3] and in $\mathbb{R}^2$ in [2].

  I report next on current work with S. Serfaty [58]. In general dimension $n \ge 2$, we obtain existence by taking the limit $s \to 1$ in the solutions for $s < 1$ constructed in [25], using of the estimates of Sect. 3.1, which are uniform in s.

  Uniqueness is reflected in the following result. *There exists at most a unique solution of Equation $u_t + \nabla(u\,\nabla\Delta^{-1}u) = 0$ in $L^\infty((0,T), L^\infty(\mathbb{R}^n))$, i.e. if two such solutions coincide at time* 0, *they are equal for all time $t > 0$.* Note this improves the result in [53] where they require u to be in a Zygmund class.

  On the other hand, the analysis of self-similarity is immediately adapted and leads to the self-similar solution

$$u(x,t) = \frac{1}{t+1} F(x/(t+1)^{1/n}), \qquad F(y) = \chi_{B_C(0)} \tag{43}$$

  and $C > 0$ is a free constant. We see immediately the analogy and the differences with the analysis of Sect. 5.2 for $s < 1$. Note in particular the solution is bounded, but not continuous.

  A further result consists of adapting the entropy analysis to prove that general bounded solutions with compactly supported data converge to one of the self-similar profiles as $t \to \infty$ up to rescaling.

- **The limit** $s \to 0$. Passing that the limit $s \to 0$ in a similar way does not offer special difficulties, thus arriving at the standard PME.

7 The Second Fractional Diffusion Model

Next we turn our attention to the nonlinear heat equation with fractional diffusion

$$\frac{\partial u}{\partial t} + (-\Delta)^{\sigma/2}(u^m) = 0. \tag{44}$$

Indeed, it is a whole family of such equations with exponents $\sigma \in (0,2)$ and $m > 0$. They can be seen as fractional-diffusion versions of the PME described above, [67, 68]. The classical Heat Equation is recovered in this model in the limit $\sigma = 2$ when $m = 1$, the PME when $m > 1$, the Fast Diffusion Equation when $m < 1$.

Equations of the form (44) are a natural choice of fractional diffusion, as an alternative to the model discussed in previous chapters. We will show that the present model leads to quite different properties. Interest in studying the nonlinear model we propose is two-fold: on the one hand, experts in the mathematics of diffusion want to understand the combination of fractional operators with porous medium type propagation. On the other hand, models of this kind arise in statistical mechanics when modeling for instance heat conduction with anomalous properties and one introduces jump processes into the modeling [46], see also [47, 48]. It is mentioned in heat control by [6]. The rigorous study of such nonlinear models has been delayed by mathematical difficulties in treating at the same time the nonlinearity and fractional diffusion.

7.1 Mathematical Problem and General Notions

Let us present the main features and results in the theory we have developed. To be specific, the theory of existence and uniqueness as well the main properties are studied by De Pablo, Quirós, Rodríguez, and Vázquez in [35, 36] for the Cauchy problem

$$\begin{cases} \dfrac{\partial u}{\partial t} + (-\Delta)^{\sigma/2}(|u|^{m-1}u) = 0, & x \in \mathbb{R}^n,\ t > 0, \\ u(x,0) = f(x), & x \in \mathbb{R}^n. \end{cases} \tag{45}$$

The notation $|u|^{m-1}u$ instead of u^m is used here to allow for solutions of two signs. We take initial data $f \in L^1(\mathbb{R}^n)$, which is a standard assumption in diffusion problems. As for the exponents, we consider the fractional exponent range $0 < \sigma < 2$, and take porous medium exponent $m > 0$. As we have said, in the limit $\sigma \to 2$ we want to recover the standard Porous Medium Equation (PME) $u_t - \Delta(|u|^{m-1}u) = 0$.

The papers contain a rather complete analysis of $u_t + (-\Delta)^{s/2}(|u|^{m-1}u) = 0$ for $x \in \mathbb{R}^n$, $0 < m < \infty$, $0 < s < 2$. A semigroup of weak energy solutions is constructed for every choice of m and σ, the smoothing effect C^α regularity work in most cases (if m is not near 0), and there is infinite propagation for all m and s.

The results can be viewed as a nonlinear interpolation between the extreme cases $\sigma = 2$: $u_t - \Delta(|u|^{m-1}u) = 0$, and $\sigma = 0$ which turns out to be a simple ODE: $u_t + |u|^{m-1}u = 0$. It is to be noted that the critical exponent $m_* := (n-\sigma)_+/n$ plays a role in the qualitative theory: the properties of the semigroup are more familiar when $m > m_*$. A similar exponent is well-known in the Fast Diffusion theory (putting $\sigma = 2$). Note that such exponent is not considered when $n = 1$ and $\sigma \geq 1$.

Preliminary Notions

If ψ and φ belong to the Schwartz class, the definition (1) of the fractional Laplacian together with Plancherel's theorem yield

$$\int_{\mathbb{R}^n} (-\Delta)^{\sigma/2}\psi\varphi = \int_{\mathbb{R}^n} |\xi|^\sigma \widehat{\psi}\widehat{\varphi} = \int_{\mathbb{R}^n} |\xi|^{\sigma/2}\widehat{\psi}|\xi|^{\sigma/2}\widehat{\varphi} = \int_{\mathbb{R}^n} (-\Delta)^{\sigma/4}\psi(-\Delta)^{\sigma/4}\varphi.$$

Therefore, if we multiply the equation in (45) by a test function φ and integrate by parts, we obtain

$$\int_0^T \int_{\mathbb{R}^n} u\frac{\partial\varphi}{\partial t}\,dx\,ds - \int_0^T \int_{\mathbb{R}^n} (-\Delta)^{\sigma/4}(|u|^{m-1}u)(-\Delta)^{\sigma/4}\varphi\,dx\,ds = 0. \quad (46)$$

This identity will be the base of our definition of a weak solution. The integrals in (46) make sense if u and u^m belong to suitable spaces. The right space for u^m is the fractional Sobolev space $\dot{H}^{\sigma/2}(\mathbb{R}^n)$, defined as the completion of $C_0^\infty(\mathbb{R}^n)$ with the norm

$$\|\psi\|_{\dot{H}^{\sigma/2}} = \left(\int_{\mathbb{R}^n} |\xi|^\sigma |\widehat{\psi}|^2\,d\xi\right)^{1/2} = \|(-\Delta)^{\sigma/4}\psi\|_{L^2}.$$

Definition A function u is a *weak solution* to Problem (45) if:

- $u \in L^1(\mathbb{R}^n \times (0,T))$ for all $T > 0$, $u^m \in L^2_{\text{loc}}((0,\infty); \dot{H}^{\sigma/2}(\mathbb{R}^n))$;
- identity (46) holds for every $\varphi \in C_0^1(\mathbb{R}^n \times (0,T))$;
- $u(\cdot,t) \in L^1(\mathbb{R}^n)$ for all $t > 0$, $\lim_{t\to 0} u(\cdot,t) = f$ in $L^1(\mathbb{R}^n)$.

A drawback of this definition is that there is no formula for the fractional Laplacian of a product or of a composition of functions. Moreover, we take no advantage in using compactly supported test functions since their fractional Laplacian loses this property. To overcome these and other difficulties, we will use the fact that our solution u is the trace of the solution of a *local* problem obtained by extending u^m to a half-space whose boundary is our original space. See also the paper by Cifani and Jakobsen [31] for an alternative L^1 theory dealing with a more general class of nonlocal porous medium equations, including strong degeneration and convection.

Extension Method In the particular case $\sigma = 1$ studied in [35], the problem is reformulated by means of the well-known representation of the half-Laplacian in terms of the Dirichlet–Neumann operator. This allowed us to transform the nonlocal problem into a local one (i.e., involving only derivatives and not integral operators). Of course, this simplification pays a prize, namely, introducing an extra space variable. The application of such an idea is not so simple when $\sigma \neq 1$; it involves a number of difficulties that we address in [36]. We have to use the characterization of the Laplacian of order σ, $(-\Delta)^{\sigma/2}$, $0 < \sigma < 2$, recently described by Caffarelli and Silvestre [23], in terms of the so-called σ-harmonic extension, which is the solution of an elliptic problem with a degenerate or singular weight.

Let us explain this extension in some more detail. If $g = g(x)$ is a smooth bounded function defined in $\mathbb{R}^n$, its σ-harmonic extension to the upper half-space $\mathbb{R}^{n+1}_+$, $v = \mathrm{e}(g)$, is the unique smooth bounded solution $v = v(x, y)$ to

$$\begin{cases} \nabla \cdot (y^{1-\sigma} \nabla v) = 0, & x \in \mathbb{R}^n, y > 0, \\ v(x, 0) = g(x), & x \in \mathbb{R}^n. \end{cases} \tag{47}$$

Then,

$$-\mu_\sigma \lim_{y \to 0^+} y^{1-\sigma} \frac{\partial v}{\partial y} = (-\Delta)^{\sigma/2} g(x), \tag{48}$$

where the precise constant, which does not depend on n, is $\mu_\sigma = \frac{2^{\sigma-1}\Gamma(\sigma/2)}{\Gamma(1-\sigma/2)}$, see [23]. Observe for future use that $\mu_\sigma \approx 2 - \sigma$ for $\sigma \to 2^-$, $\mu_\sigma \approx 1/\sigma$ for $\sigma \to 0^+$. In (47) the operator ∇ acts in all (x, y) variables, while in (48) $(-\Delta)^{\sigma/2}$ acts only on the $x = (x_1, \ldots, x_n)$ variables. In the sequel we denote

$$L_\sigma v \equiv \nabla \cdot (y^{1-\sigma} \nabla v), \qquad \frac{\partial v}{\partial y^\sigma} \equiv \mu_\sigma \lim_{y \to 0^+} y^{1-\sigma} \frac{\partial v}{\partial y}.$$

Notation The upper half-space, with points $\overline{x} = (x, y)$, $x \in \mathbb{R}^n$, $y > 0$, will be named Ω, and its boundary, which is identified to the original $\mathbb{R}^n$ with variable x, will be named Γ. Occasionally, Γ will be a bounded domain in $\mathbb{R}^n$ and then Ω will be the cylinder $\Gamma \times (0, \infty)$; those cases will be carefully indicated. Besides, we use the simplified notation u^m for data of any sign, instead of the actual "odd power" $|u|^{m-1}u$, and we will also use such a notation when m is replaced by $1/m$. The convention is not applied to any other powers.

Extended Problem. Weak Solutions With the above in mind, we rewrite problem (45) for $w = u^m$ as a quasi-stationary problem with a dynamical boundary condition

$$\begin{cases} L_\sigma w = 0 & \text{for } \overline{x} \in \Omega,\ t > 0, \\ \dfrac{\partial w}{\partial y^\sigma} - \dfrac{\partial w^{1/m}}{\partial t} = 0 & \text{for } x \in \Gamma,\ t > 0, \\ w(x, 0, 0) = f^m(x) & \text{for } x \in \Gamma. \end{cases} \tag{49}$$

This problem has been considered by Athanasopoulos and Caffarelli [7], who prove that any bounded weak solution is Hölder continuous if $m > 1$.

To define a weak solution of this problem we multiply formally the equation in (49) by a test function φ and integrate by parts to obtain

$$\int_0^T \int_\Gamma u \frac{\partial \varphi}{\partial t}\, dx\, ds - \mu_\sigma \int_0^T \int_\Omega y^{1-\sigma} \langle \nabla w, \nabla \varphi \rangle\, d\overline{x}\, ds = 0, \tag{50}$$

where $u = (\mathrm{Tr}(w))^{1/m}$ is the trace of w on Γ to the power $1/m$. This holds on the condition that φ vanishes for $t = 0$ and $t = T$, and also for large $|x|$ and y. We then introduce the energy space $X^\sigma(\Omega)$, the completion of $C_0^\infty(\Omega)$ with the norm

$$\|v\|_{X^\sigma} = \left(\mu_\sigma \int_\Omega y^{1-\sigma} |\nabla v|^2\, d\overline{x} \right)^{1/2}. \tag{51}$$

The trace operator is well defined in this space, see below.

Definition A pair of functions (u, w) is a *weak solution* to Problem (49) if:

- $w \in L^2_{\mathrm{loc}}((0, \infty); X^\sigma(\Omega))$, $u = (\mathrm{Tr}(w))^{1/m} \in L^1(\Gamma \times (0, T))$ for all $T > 0$;
- Identity (50) holds for every $\varphi \in C_0^1(\overline{\Omega} \times (0, T))$;
- $u(\cdot, t) \in L^1(\Gamma)$ for all $t > 0$, $\lim_{t \to 0} u(\cdot, t) = f$ in $L^1(\Gamma)$.

For brevity we will refer sometimes to the solution as only u, or even only w, when no confusion arises, since it is clear how to complete the pair from one of the components, $u = (\mathrm{Tr}(w))^{1/m}$, $w = \mathrm{e}(u^m)$.

Equivalence of Weak Formulations The key point of the above discussion is that the definitions of weak solution for our original nonlocal problem and for the extended local problem are equivalent. The main ingredient of the proof is that equation (48) holds in the sense of distributions for any $g \in \dot{H}^{\sigma/2}(\Gamma)$.

Proposition *A function u is a weak solution to Problem* (45) *if and only if* $(u, \mathrm{e}(u^m))$ *is a weak solution to Problem* (49).

Strong Solutions Weak solutions satisfy Eq. (45) in the sense of distributions. Hence, if the left hand side is a function, the right hand side is also a function and the equation holds almost everywhere. This fact allows to prove uniqueness and several other important properties, and hence motivates the following definition.

Definition We say that a weak solution u to Problem (45) is a strong solution if $u \in C([0, \infty) : L^1(\Gamma))$ as well as $\partial_t u$ and $(-\Delta)^{\sigma/2}(|u|^{m-1}u) \in L^1_{\mathrm{loc}}(\Gamma \times (0, \infty))$.

7.2 Main Results

Existence We prove existence of a suitable concept of (weak) solution for general L^1 initial data only in the restricted range $m > m_* \equiv (n-\sigma)_+/n$, which includes as a particular case the linear fractional heat equation, case $m = 1$. If $0 < m \le m_*$ (which implies that $0 < \sigma < 1$ if $n = 1$) we need to slightly restrict the data to obtain weak solutions.

Theorem 5 *If either $f \in L^1(\mathbb{R}^n)$ and $m > m_*$, or $f \in L^1(\mathbb{R}^n) \cap L^p(\mathbb{R}^n)$ with $p > p_*(m) = (1-m)n/\sigma$ and $0 < m \le m_*$, there exists a weak solution to Problem (45).*

Uniqueness We first prove uniqueness of weak solutions in the range $m \ge m_*$. If $0 < m < m_*$, we need to use the concept of strong solution, a concept that is standard in the abstract theory of evolution equations. This is no restriction in view of the next results proved in [36].

Theorem 6 *The solution given by Theorem 5 is a strong solution.*

We state the uniqueness result in its simplest version.

Theorem 7 *For every f and $m > 0$ there exists at most one strong solution to Problem (45).*

Qualitative Behavior The solutions to Problem (45) have some nice properties that are summarized here.

Theorem 8 *Assume f, f_1, f_2 satisfy the hypotheses of Theorem 5, and let u, u_1, u_2 be the corresponding strong solutions to Problem (45).*

(i) *If $m \ge m_*$, the mass $\int_{\mathbb{R}^n} u(x,t)\,dx$ is conserved.*
(ii) *If $0 < m < m_*$, then $u(\cdot,t)$ vanishes identically in a finite time.*
(iii) *An smoothing effect holds in the form:*

$$\|u(\cdot,t)\|_{L^\infty(\mathbb{R}^n)} \le C\, t^{-\gamma_p} \|f\|_{L^p(\mathbb{R}^n)}^{\delta_p} \tag{52}$$

with $\gamma_p = (m-1+\sigma p/n)^{-1}$, $\delta_p = \sigma p \gamma_p/n$, and $C = C(m,p,n,\sigma)$. This holds for all $p \ge 1$ if $m > m_*$, and only for $p > p_*(m)$ if $0 < m \le m_*$.
(iv) *Any L^p-norm of the solution, $1 \le p \le \infty$, is nonincreasing in time.*
(v) *There is an L^1-order-contraction property,*

$$\int_{\mathbb{R}^n} (u_1-u_2)_+(x,t)\,dx \le \int_{\mathbb{R}^n} (u_1-u_2)_+(x,0)\,dx.$$

(vi) *If $f \ge 0$ the solution is positive for all x and all positive times if $m \ge m_*$ (resp. for all x and all $0 < t < T$ if it vanishes in finite time T when $0 < m < m_*$).*
(vii) *If either $m \ge 1$ or $f \ge 0$, then $u \in C^\alpha(\mathbb{R}^n \times (0,\infty))$ for some $0 < \alpha < 1$.*

In the linear case $m = 1$ the above properties: conservation of mass, the smoothing effect with a precise decay rate, positivity and regularity, can be derived directly from the representation formula (3) and the properties of the kernel K_σ.

Continuous Dependence We show that the solution (i.e., the semigroup) depends continuously on the initial data and on both parameters m and σ, in particular in the nontrivial limit $\sigma \to 2$, that allows to recover the standard PME, $\partial_t u - \Delta |u|^{m-1}u = 0$, or the other end $\sigma \to 0$, for which we get the ODE: $\partial_t u + |u|^{m-1}u = 0$. Continuity will be true in general only in L^1_{loc}, unless we stay in the region of parameters where mass is conserved.

Theorem 9 *The strong solutions depend continuously in the norm of the space* $C([0,T] : L^1_{\mathrm{loc}}(\mathbb{R}^n))$ *on the parameters* m, σ, *and the initial data* f. *If moreover* $m \ge m_*$ *and* $0 < \sigma \le 2$, *convergence also holds in* $C([0,T] : L^1(\mathbb{R}^n))$.

8 Current and Future Work

A number of related models, issues and perspectives on elliptic and parabolic equations involving fractional Laplacians and more general integral operators is contained in L. Caffarelli's contribution in this volume.

Let me mention some of the many questions that need investigation in the models I have presented. (1) Study the optimal regularity of the solutions, (2) Study the regularity of the free boundary, (3) Study fine asymptotic behavior (asymptotics with rates) in the first model, or the whole asymptotic program in the second model (4) Study problems in bounded domains (current work with M. Bonforte and Y. Sire), (5) Decide conditions of uniqueness in the first model, (6) Decide conditions of comparison in the first model, (7) Write a performing numerical code, (8) Discuss the Stochastic Particle Models in the literature that involve long-range effects and anomalous diffusion parameters, (9) Study equations with more general long-range kernels in the spirit of the recent work of L. Caffarelli, L. Silvestre and collaborators, (10) Study equations and systems with convection effects, a wide and active topic involving difficult questions in Fluid Mechanics, that we will refrain from entering into since it deserves an exposition of its own.

Acknowledgements Author partially supported by Spanish Project MTM2008-06326-C02. The author is grateful to the referee and N. Guillén for carefully reading the manuscript.

References

1. Abe, S., Thurner, S.: Anomalous diffusion in view of Einstein's 1905 theory of Brownian motion. Physica A **356**(2–4), 403–407 (2005)
2. Ambrosio, L., Mainini, E., Serfaty, S.: Gradient flow of the Chapman–Rubinstein–Schatzman model for signed vortices. Ann. Inst. Henri Poincaré, Anal. Non Linéaire **28**(2), 217–246 (2011)

3. Ambrosio, L., Serfaty, S.: A gradient flow approach to an evolution problem arising in superconductivity. Commun. Pure Appl. Math. **61**(11), 1495–1539 (2008)
4. Applebaum, D.: Lévy Processes and Stochastic Calculus, 2nd edn. Cambridge Studies in Advanced Mathematics, vol. 116. Cambridge University Press, Cambridge (2009)
5. Aronson, D.G.: The porous medium equation. In: Nonlinear Diffusion Problems Montecatini Terme, 1985. Lecture Notes in Math., vol. 1224, pp. 1–46. Springer, Berlin (1986)
6. Athanasopoulos, I., Caffarelli, L.A.: Optimal regularity of lower dimensional obstacle problems. Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **310** (2004); Kraev. Zadachi Mat. Fiz. i Smezh. Vopr. Teor. Funkts. **35**(34), 49–66, 226; translation in J. Math. Sci. (N.Y.) **132**(3), 274–284 (2006)
7. Athanasopoulos, I., Caffarelli, L.A.: Continuity of the temperature in boundary heat control problem. Adv. Math. **224**(1), 293–315 (2010)
8. Athanasopoulos, I., Caffarelli, L.A., Salsa, S.: The structure of the free boundary for lower dimensional obstacle problems. Am. J. Math. **130**(2), 485–498 (2008)
9. Bachelier, L.: Théorie de la spéculation. Ann. Sci. Éc. Norm. Super. **3**(17), 21–86 (1900)
10. Barenblatt, G.I.: On self-similar motions of a compressible fluid in a porous medium. Akad. Nauk SSSR, Prikl. Mat. Meh. **16**, 679–698 (1952) (in Russian)
11. Bass, R.F., Kassmann, M.: Harnack inequalities for non-local operators of variable order. Trans. Am. Math. Soc. **357**(2), 837–850 (2005)
12. Bass, R.F., Levin, D.A.: Harnack inequalities for jump processes. Potential Anal. **17**(4), 375–388 (2002)
13. Bendahmane, M., Karlsen, K.H.: Renormalized entropy solutions for quasi-linear anisotropic degenerate parabolic equations. SIAM J. Math. Anal. **36**(2), 405–422 (2004)
14. Bertoin, J.: Lévy processes. Cambridge Tracts in Mathematics, vol. 121. Cambridge University Press, Cambridge (1996). ISBN: 0-521-56243-0
15. Bertozzi, A.L., Laurent, T.: The behavior of solutions of multidimensional aggregation equations with mildly singular interaction kernels. Chin. Ann. Math., Ser. B **30**(5), 463–482 (2009)
16. Bertozzi, A.L., Carrillo, J.L., Laurent, T.: Blow-up in multidimensional aggregation equations with mildly singular interaction kernels. Nonlinearity **22**(3), 683–710 (2009)
17. Bertozzi, A.L., Laurent, T., Rosado, J.: L^p theory for the multidimensional aggregation equation. Commun. Pure Appl. Math. **64**(1), 45–83 (2011)
18. Biler, P., Imbert, C., Karch, G.: Fractal porous media equation. arXiv:1001.0910
19. Biler, P., Karch, G., Monneau, R.: Nonlinear diffusion of dislocation density and self-similar solutions. Commun. Math. Phys. **294**(1), 145–168 (2010). MR2575479
20. Blumenthal, R.M., Getoor, R.K.: Some theorems on stable processes. Trans. Am. Math. Soc. **95**(2), 263–273 (1960)
21. Bonforte, M., Dolbeault, J., Grillo, G., Vázquez, J.L.: Sharp rates of decay of solutions to the nonlinear fast diffusion equation via functional inequalities. Proc. Natl. Acad. Sci. USA **107**(38), 16459–16464 (2010)
22. Caffarelli, L.A.: Further regularity for the Signorini problem. Commun. Partial Differ. Equ. **4**, 1067–1075 (1979)
23. Caffarelli, L.A., Silvestre, L.: An extension problem related to the fractional Laplacian. Commun. Partial Differ. Equ. **32**(7–9), 1245–1260 (2007)
24. Caffarelli, L.A., Vasseur, A.: Drift diffusion equations with fractional diffusion and the quasi-geostrophic equation. Ann. Math. **171**, 1903–1930 (2010)
25. Caffarelli, L.A., Vázquez, J.L.: Nonlinear porous medium flow with fractional potential pressure. Arch. Ration. Mech. Anal. **202**, 537–565 (2011). doi:10.1007/s00205-011-0420-4. arXiv:1001.0410v2.
26. Caffarelli, L.A., Vázquez, J.L.: Asymptotic behaviour of a porous medium equation with fractional diffusion. Discrete Contin. Dyn. Syst., Ser. A **29**(4), 1393–1404 (2011). A special issue “Trends and Developments in DE/Dynamics, Part III”
27. Caffarelli, L.A., Salsa, S., Silvestre, L.: Regularity estimates for the solution and the free boundary to the obstacle problem for the fractional Laplacian. Invent. Math. **171**(2), 425–461 (2008)

28. Caffarelli, L., Chan, C.-H., Vasseur, A.: Regularity theory for nonlinear integral operators. J. Am. Math. Soc. **24**, 849–869 (2011)
29. Caffarelli, L.A., Soria, F., Vázquez, J.L.: Regularity of solutions of the fractional porous medium flow (in preparation)
30. Carrillo, J.A., Toscani, G.: Asymptotic L^1-decay of solutions of the porous medium equation to self-similarity. Indiana Univ. Math. J. **49**, 113–141 (2000)
31. Cifani, S., Jakobsen, E.R.: Entropy solution theory for fractional degenerate convection-diffusion equations. Ann. Inst. Henri Poincaré, Anal. Non Linéaire **28**(3), 413–441 (2011)
32. Chapman, J.S., Rubinstein, J., Schatzman, M.: A mean-field model for superconducting vortices. Eur. J. Appl. Math. **7**(2), 97–111 (1996)
33. De Giorgi, E.: Sulla differenziabilità e l'analiticità delle estremali degli integrali multipli regolari. Mem. Accad. Sci. Torino, Cl. Sci. Fis. Mat. Nat. **3**, 25–43 (1957)
34. Denzler, J., McCann, R.: Phase transitions and symmetry breaking in singular diffusion. Proc. Natl. Acad. Sci. USA **100**, 6922–6925 (2003)
35. De Pablo, A., Quirós, F., Rodriguez, A., Vázquez, J.L.: A fractional porous medium equation. Adv. Math. **226**(2), 1378–1409 (2011)
36. De Pablo, A., Quirós, F., Rodriguez, A., Vázquez, J.L.: A general fractional porous medium equation. arXiv:1104.0306v1 [math.AP]. Commun. Pure Appl. Math. (2011, to appear)
37. Di Nezza, E., Palatucci, G., Valdinoci, E.: Hitchhiker's guide to the fractional Sobolev spaces. Preprint (2011)
38. E, W.: Dynamics of vortex-liquids in Ginzburg–Landau theories with applications to superconductivity. Phys. Rev. B **50**(3), 1126–1135 (1994)
39. Einstein, A.: Über die von der molekularkinetischen Theorie der Wärme gefordete Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. Ann. Phys. (Leipz.) **17**, 549–560 (1905). English translation: Investigations on the Theory of Brownian Movement. Dover, New York (1956)
40. Friedman, A., Kamin, S.: The asymptotic behavior of gas in an N-dimensional porous medium. Trans. Am. Math. Soc. **262**, 551–563 (1980). MR0586735 (81j:35054)
41. Getoor, R.K.: First passage times for symmetric stable processes in space. Trans. Am. Math. Soc. **101**, 75–90 (1961)
42. Gilbarg, D., Trudinger, N.S.: Elliptic Partial Differential Equations of Second Order. Classics in Mathematics. Springer, Berlin (2001). Reprint of the 1998 edition
43. Head, A.K.: Dislocation group dynamics II. Similarity solutions of the continuum approximation. Philos. Mag. **26**, 65–72 (1972)
44. Kassmann, M.: A priori estimates for integro-differential operators with measurable kernels. Calc. Var. **34**, 1–21 (2009)
45. Kiselev, A., Nazarov, F., Volberg, A.: Global well-posedness for the critical 2D dissipative quasi-geostrophic equation. Invent. Math. **167**(3), 445–453 (2007). MR2276260
46. Jara, M.: Hydrodynamic limit of particle systems with long jumps. Preprint. http://arxiv.org/abs/0805.1326
47. Jara, M.: Nonequilibrium scaling limit for a tagged particle in the simple exclusion process with long jumps. Commun. Pure Appl. Math. **62**(2), 198–214 (2009)
48. Jara, M., Komorowski, T., Olla, S.: Limit theorems for additive functionals of a Markov chain. Ann. Appl. Probab. **19**(6) (2009) 2270–2300
49. Ladyzhenskaya, O.A., Uraltseva, N.N.: Linear and Quasilinear Equations of Elliptic Type. Nauka, Moscow (1964). Academic Press, New York (1968) (in Russian). MR 0244627 (39:5941)
50. Ladyzhenskaya, O.A., Solonnikov, V.A., Ural'tseva, N.N.: Linear and Quasilinear Equations of Parabolic Type. Transl. Math. Monographs, vol. 23. Am. Math. Soc., Providence (1968)
51. Landkof, N.S.: Foundations of Modern Potential Theory. Die Grundlehren der mathematischen Wissenschaften, Band 180. Springer, New York (1972). Translated from the Russian by A.P. Doohovskoy
52. Lee, K.A., Vázquez, J.L.: Geometrical properties of solutions of the porous medium equation for large times. Indiana Univ. Math. J. **52**(4), 991–1016 (2003)

53. Lin, F.H., Zhang, P.: On the hydrodynamic limit of Ginzburg–Landau vortices. Discrete Contin. Dyn. Syst. **6**, 121–142 (2000)
54. Lions, P.L., Mas-Gallic, S.: Une méthode particulaire déterministe pour des équations diffusives non linéaires. C. R. Acad. Sci. Paris, Sér. I **332**, 369–376 (2001)
55. Mellet, A., Mischler, S., Mouhot, C.: Fractional diffusion limit for collisional kinetic equations. Preprint. http://arxiv.org/abs/0809.2455
56. Oleinik, O.A., Kalashnikov, A.S., Chzou, Y.-I.: The Cauchy problem and boundary problems for equations of the type of unsteady filtration. Izv. Akad. Nauk SSSR, Ser. Mat. **22**, 667–704 (1958)
57. Peletier, L.A.: The porous media equation. In: Amann, H. (ed.) Application of Nonlinear Analysis in the Physical Sciences, pp. 229–242. Pitman, London (1981)
58. Serfaty, S., Vázquez, J.L.: Work in preparation
59. Signorini, A.: Questioni di elasticità non linearizzata e semilinearizzata. Rend. Mat. Appl. **18**, 95–139 (1959)
60. Silvestre, L.E.: Hölder estimates for solutions of integro differential equations like the fractional Laplace. Indiana Univ. Math. J. **55**(3), 1155–1174 (2006)
61. Silvestre, L.E.: Regularity of the obstacle problem for a fractional power of the Laplace operator. Commun. Pure Appl. Math. **60**(1), 6–112 (2007)
62. Smoluchowski, M.: Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. Ann. Phys. **21**, 756–780 (1906) (in German). Previously, Bull. Int. Acad. Sci. Cracovie, 46 A (1906) (in Polish)
63. Stein, E.M.: Singular Integrals and Differentiability Properties of Functions. Princeton Mathematical Series, vol. 30. Princeton University Press, Princeton (1970). MR0290095 (44 #7280)
64. Valdinoci, E.: From the long jump random walk to the fractional Laplacian. Bol. Soc. Esp. Mat. Apl. **49**, 33–44 (2009)
65. Vázquez, J.L.: Asymptotic behaviour for the Porous Medium Equation posed in the whole space. J. Evol. Equ. **3**, 67–118 (2003)
66. Vázquez, J.L.: Asymptotic behaviour for the PME in a bounded domain. The Dirichlet problem. Monatshefte Math. **142**(1–2), 81–111 (2004)
67. Vázquez, J.L.: Smoothing and Decay Estimates for Nonlinear Diffusion Equations. Equations of Porous Medium Type. Oxford Lecture Series in Mathematics and Its Applications, vol. 33. Oxford University Press, Oxford (2006)
68. Vázquez, J.L.: The Porous Medium Equation. Mathematical Theory. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, Oxford (2007)
69. Vázquez, J.L.: Perspectives in nonlinear diffusion. Between analysis, physics and geometry. In: Sanz-Solé, M., et al. (eds.) Proceedings of the International Congress of Mathematicians, ICM Madrid 2006, vol. 1, pp. 609–634. Eur. Math. Soc. Pub. House, Zurich (2007)
70. Villani, C.: Topics in Optimal Transportation. Am. Math. Soc., Providence (2003)
71. Vlahos, L., Isliker, H., Kominis, Y., Hizonidis, K.: Normal and anomalous Diffusion: a tutorial. In: Bountis, T. (ed.) Order and Chaos, vol. 10. Patras University Press, Patras (2008)
72. Weitzner, H., Zaslavsky, G.M.: Some applications of fractional equations. Chaotic transport and complexity in classical and quantum dynamics. Commun. Nonlinear Sci. Numer. Simul. **8**(3–4), 273–281 (2003)
73. Zel'dovich, Ya.B., Kompaneets, A.S.: Towards a theory of heat conduction with thermal conductivity depending on the temperature. In: Collection of Papers Dedicated to 70th Anniversary of A.F. Ioffe, pp. 61–72. Izd. Akad. Nauk SSSR, Moscow (1950)

(Ir)reversibility and Entropy

Cédric Villani

Abstract In the 1860's emerges a revolutionary idea: many properties of the world around us can be explained by combining the atomistic hypothesis with the statistical theory. Some of the great scientific conquests from this time are the Boltzmann equation, which triggers one of the first qualitative studies of a complicated nonlinear partial differential equation; the notion of statistical entropy, which would later be fundamental in other areas of physics and mathematics, including information theory; and the notion of macroscopic irreversibility emerging from microscopically reversible laws. Thus the basic rules of statistical physics were set until Boltzmann's irreversibility paradigm was shaken by Landau's discovery of the Landau damping effect, about 80 years later, which opened the idea that equilibration is compatible with preservation of information, and led to a number of problems concerning the statistical theory of matter.

La cosa più meravigliosa è la felicità del momento
L. Ferré

Time's arrow is part of our daily life and we experience it every day: broken mirrors do not come back together, human beings do not rejuvenate and rings grow unceasingly in tree trunks. In sum, time always flows in the same direction! Nonetheless, the fundamental laws of classical physics do not favor any time direction and conform to a rigorous symmetry between past and future. It is possible, as discussed in the article by T. Damour in this same volume, that irreversibility is inscribed in other physical laws, for example on the side of general relativity or quantum mechanics. Since Boltzmann, **statistical physics** has advanced another explanation:

Reprinted, with kind permission, from: Duplantier, B. (ed.): Time—Poincaré Seminar 2010. Progress in Mathematical Physics, vol. 63, pp. 19–79. © Springer Basel AG 2012.

Translated by Lester Senechal.

C. Villani (✉)
University of Lyon and Institut Henri Poincaré, 11 rue Pierre et Marie Curie,
75231 Paris Cedex 05, France
e-mail: villani@ihp.fr

H. Holden, K.H. Karlsen (eds.), *Nonlinear Partial Differential Equations*,
Abel Symposia 7, DOI 10.1007/978-3-642-25361-4_16,

time's arrow translates a constant flow of less likely events toward more likely events. Before continuing with this interpretation, which constitutes the guiding principle of the whole exposition, I note that the flow of time is not necessarily based on a single explanation.

At first glance, Boltzmann's suggestion seems preposterous: it is not because an event is *probable* that it is actually achieved, but time's arrow seems inexorable and seems not to tolerate any exception. The answer to this objection lies in a catch-phrase: **separation of scales**. If the fundamental laws of physics are exercised on the microscopic, particulate (atoms, molecules,...) level, phenomena that we can sense or measure involve a considerable number of particles. The effect of this number is even greater when it enters combinatoric computations: if N, the number of atoms participating in an experiment, is of order 10^{10}, this is already considerable, but $N!$ or 2^N are supernaturally large, invincible numbers.

The innumerable debates between physicists that have been pursued for more than a century, and that are still pursued today, give witness to the subtlety and depth of Maxwell's and Boltzmann's arguments, banners of a small scientific revolution that was accomplished in the 1860's and 1870's, and which saw the birth of the fundamentals of the modern kinetic theory of gases, the universal concept of statistical entropy and the notion of macroscopic irreversibility. In truth, the arguments are so subtle that Maxwell and Boltzmann themselves sometimes went astray, hesitating on certain interpretations, alternating naive errors with profound concepts; the greatest scientists at the end of the nineteenth century, e.g. Poincaré and Lord Kelvin, were not to be left behind. We find an overview of these delays in the book by Damour already mentioned; for my part, I am content to present a "decanted" version of Boltzmann's theory. At the end of the text I shall evoke the way in which Landau shattered Boltzmann's paradigm, discovering an apparent irreversibility where there seemed not to be any and opening up a new mine of mathematical problems.

In retracing the history of the statistical interpretation of time's arrow, I shall have occasion to make a voyage to the heart of profound problems that have agitated mathematicians and physicists for more than a century.

The notation used in this exposition are generally classical; I denote $\mathbb{N} = \{1, 2, 3, \ldots\}$ and $\log =$ natural logarithm.

1 Newton's Inaccessible Realm

I shall adopt here a purely classical description of our physical universe, in accordance with the laws enacted by Newton: the ambient space is Euclidean, time is absolute and acceleration is equal to the product of the mass by the resultant of the forces.

In the case of the description of a gas, these hypotheses are questionable: according to E.G.D. Cohen, the quantum fluctuations are not negligible on the mesoscopic level. The probabilistic nature of quantum mechanics is still debated; we nevertheless accept that the resulting increased uncertainty due to taking these uncertainties into account can but arrange our affairs, at least qualitatively, and we thus concentrate on the classical and deterministic models, "à la Newton".

1.1 The Solid Sphere Model

In order to fix the ideas, we consider a system of ideal spherical particles bouncing off one another: let there be N particles in a box Λ. We let $X_i(t)$ denote the position at time t of the center of the i-th particle. The rules of motion are stated as follows:

- We suppose that initially the particles are well separated ($i \neq j \implies |X_i - X_j| > 2r$) and separated from the walls ($d(X_i, \partial\Lambda) > r$ for all i).
- While these separation conditions are satisfied, the movement is uniformly rectilinear: $\ddot{X}_i(t) = 0$ for each i, where we denote $\ddot{X} = d^2X/dt^2$, the acceleration of X.
- When two particles meet, their velocities change abruptly according to Descartes' laws: if $|X_i(t) - X_j(t)| = 2r$, then

$$\begin{cases} \dot{X}_i(t^+) = \dot{X}_i(t^-) - 2\langle \dot{X}_i(t^-) - \dot{X}_j(t^-), n_{ij}\rangle n_{ij}, \\ \dot{X}_j(t^+) = \dot{X}_j(t^-) - 2\langle \dot{X}_j(t^-) - \dot{X}_i(t^-), n_{ji}\rangle n_{ji}, \end{cases}$$

  where $n_{ij} = (X_i - X_j)/|X_i - X_j|$ denotes the unit vector joining the centers of the colliding balls.
- When a particle encounters the boundary, its velocity also changes: if $|X_i - x| = r$ with $x \in \partial\Lambda$, then

$$\dot{X}_i(t^+) = \dot{X}_i(t^-) - 2\langle \dot{X}_i(t), n(x)\rangle n(x),$$

  where $n(x)$ is the exterior normal to Λ at x, supposed well defined.

These rules are not sufficient for completely determining the dynamics: we cannot exclude *a priori* the possibility of triple collisions, simultaneous collisions between particles and the boundary, or again an infinity of collisions occurring in a finite time. However, such events are of probability zero if the initial conditions are drawn at random with respect to Lebesgue measure (or Liouville measure) in phase space [40, Appendix 4.A]; we thus neglect these eventualities. The dynamic thus defined, as simple as it may be, can then be considered as a caricature of our complex universe if the number N of particles is very large. Studied for more than a century, this caricature has still not yielded all its secrets; far from that.

1.2 Other Newtonian Models

Beginning with the emblematic model of hard spheres, we can define a certain number of more or less complex variants:

- replace dimension 3 by an arbitrary dimension $d \geq 2$ (dimension 1 is likely pathological);
- replace the boundary condition (elastic rebound) by a more complex law [40, Chap. 8];

- or, instead, eliminate the boundaries, always delicate, by setting the system in the whole space $\mathbb{R}^d$ (but we may then add that the number of particles must then be infinite so as keep a nonzero global mean density) or in a torus of side L, $\mathbb{T}_L^d = \mathbb{R}^d/(L\mathbb{Z}^d)$, which will be my choice of preference in the sequel;
- replace the contact interaction of hard spheres by another interaction between point particles, e.g. associated with an interaction potential between two bodies: $\phi(x-y) =$ potential exerted at point x by a material point situated at y.

Among the notable interaction potentials in dimension 3 we mention (within a multiplicative constant):

- the **Coulomb** potential: $\phi(x-y) = 1/|x-y|$;
- the **Newtonian** potential: $\phi(x-y) = -1/|x-y|$;
- the **Maxwellian** potential: $\phi(x-y) = 1/|x-y|^4$.

The Maxwellian interaction was artificially introduced by Maxwell and Boltzmann in the context of the statistical study of gases; it leads to important simplifications in certain formulas. There exists a taxonomy of other potentials (Lennard-Jones, Manev...). The hard spheres correspond to the limiting case of a potential that equals 0 for $|x-y| > r$ and $+\infty$ for $|x-y| < 2r$.

Suppose, more generally, that the interaction takes place on a scale of order r and with an intensity a. We end up with a **system of point particles with interaction potential**

$$\ddot{X}_i(t) = -a \sum_{j \neq i} \nabla\phi\left(\frac{X_i - X_j}{r}\right), \tag{1}$$

for each $i \in \{1, \ldots, N\}$; we thus suppose that $X_i \in \mathbb{T}_L^d$. Here again, the dynamic is well defined except for a set of exceptional initial conditions and it is associated with a **Newtonian flow** $\mathscr{N}_t$, which maps the configuration at time s to the configuration at time $s+t$ ($t \in \mathbb{R}$ can be positive or negative).

1.3 Distribution Functions

Even if one accepts the Newtonian model (1), it remains *inaccessible* to us: first because we cannot perceive the individual particles (too small), and because their number N is large. By well designed experiments, we can measure the pressure exerted on a small surface, the temperature about a point, the mean density, etc. None of these quantities is expressed directly in terms of the X_i, but rather in terms of averages

$$\frac{1}{N} \sum_i \chi(X_i, \dot{X}_i), \tag{2}$$

where χ is a scalar function.

It may seem an idle distinction: in concentrating χ near the particle i, we retrieve the missing information. But quite clearly this is impossible: in practice χ is of *macroscopic* variation, e.g. of the order of the size of the box. Besides, the information contained in the averages (2) does not distinguish particles, so that we have to replace the vector of the $(X_i, \dot{X}_i)$ by the **empirical measure**

$$\hat{\mu}_t^N = \frac{1}{N} \sum_{i=1}^{N} \delta_{(X_i(t), \dot{X}_i(t))}. \tag{3}$$

The terminology "empirical" is well chosen: it's the measure that is observed by means (without intending a pun) of measurements.

To resume: our knowledge of the particle system is achieved only through the behavior of the empirical measure in a *weak topology* that models the macroscopic limitation of our experiments—laboratory experiments as well as sensory perceptions.

Frequently, on our own scale, the empirical measure appears continuous:

$$\hat{\mu}_t^N (dx\, dv) \simeq f(t, x, v)\, dx\, dv.$$

We often use the notation $f(t, \cdot) = f_t$. The density f is the **kinetic distribution** of the gas. The study of this distribution constitutes the kinetic theory of gases; the founder of this science is undoubtedly D. Bernoulli (around 1738), and the most famous contributors to it are Maxwell and Boltzmann. A brief history of kinetic theory can be found in [40, Chap. 1] and in the references there included.

We continue with the study of the Newtonian system. We can imagine that certain experiments allow for simultaneous measurement of the parameters of various particles, thus giving access to correlations between particles. This leads us to define, for example,

$$\hat{\mu}_t^{2;N} (dx_1\, dv_1\, dx_2\, dv_2) = \frac{1}{N(N-1)} \sum_{i \neq j} \delta_{(X_{i_1}(t), \dot{X}_{i_2}(t), X_{i_2}(t), \dot{X}_{i_2}(t))},$$

or more generally

$$\hat{\mu}_t^{k;N} (dx_1\, dv_1\, \dots\, dx_k\, dv_k) = \frac{(N-k-1)!}{N!} \sum_{(i_1, \dots, i_k) \text{distinct}} \delta_{(X_{i_1}(t), \dot{X}_{i_1}(t), \dots, X_{i_k}(t), \dot{X}_{i_k}(t))}.$$

The corresponding approximations are **distribution functions in k particles**:

$$\hat{\mu}_t^{k;N} (dx_1\, dv_1\, \dots\, dx_k\, dv_k) \simeq f^{(k)}(t, x_1, v_1, \dots, x_k, v_k).$$

Evidently, by continuing up until $k = N$, we find a measure $\hat{\mu}^{N;N}(dx_1 \dots dv_N)$ concentrated at the vector of particle positions and velocities (the mean over all permutations of the particles). But in practice we never go to $k = N$: k remains very small (going to 3 would already be a feat), whereas N is huge.

1.4 Microscopic Randomness

In spite of the determinism of the Newtonian model, hypotheses of a probabilistic nature on the initial data have already been made, by supposing that they are not configured to end up in some unusual catastrophe such as a triple collision. We can now generalize this approach by considering a probability distribution on the set of initial positions and velocities:

$$\mu_0^N(dx_1\, dv_1 \dots dx_N\, dv_N),$$

which is called a **microscopic probability measure**. In the sequel we will use the abbreviated notation

$$dx^N\, dv^N := dx_1\, dv_1 \dots dx_N\, dv_N.$$

It is natural to choose μ_0^N symmetric, i.e. invariant under coordinate permutations. The data μ_0^N replace the measure $\hat{\mu}_0^{N;N}$ and generalize it, giving rise to a flow of measures, obtained by the action of the flow:

$$\mu_t^N = (\mathscr{N}_t)_\# \mu_0^N,$$

and the marginals

$$\mu_t^{k;N} = \int_{(x_1, v_1, \dots, x_k, v_k)} \mu_t^N.$$

If the sense of the empirical measure is transparent (it's the "true" particle density), that of the microscopic probability measure is less evident. Let us assume that the initial state has been prepared by a great combination of circumstances about which we know little: we can only make suppositions and guesses. Thus μ_0^N is a probability measure on the set of possible initial configurations. A physical statement involving μ_0^N will, however, scarcely make sense if we use the precise form of this distribution (we cannot verify it, since we do not observe μ_0^N); but it will make good sense if a μ_0^N-almost certain property is stated, or indeed with μ_0^N-probability of 0.99 or more.

Likewise, the form of $\mu_t^{1;N}$ has scarcely any physical meaning. But if there is a phenomenon of concentration of measure due to the hugeness of N, then it may be hoped that

$$\mu_0^N[\mathrm{dist}(\hat{\mu}_t^N, f_t(x, v)\, dx\, dv) \geq r] \leq \alpha(N, r),$$

where dist is a well chosen distance on the space of measures and $\alpha(N, r) \to 0$ when $r \to \infty$, all the faster that N is large (for example $\alpha(N, r) = e^{-cNr}$). We will then have

$$\mathrm{dist}(\mu_t^{1;N}, f_t(x, v)\, dx\, dv) = \mathrm{dist}\left(\int \hat{\mu}_t^N\, d\mu_t^N,\ f_t(x, v)\, dx\, dv \right)$$

$$\leq \int \mathrm{dist}(\hat{\mu}_t^N, f_t\,dx\,dv)$$
$$\leq \int_0^\infty \alpha(N,r)\,dr =: \eta(N).$$

If $\eta(N) \to 0$ when $N \to \infty$ it follows that, with very high probability, $\mu_t^{1;N}$ is an excellent approximation to $f(t,x,v)\,dx\,dv$, which itself is a good approximation to $\hat{\mu}_t^N$.

1.5 Micromegas

In this section I shall introduce two very different statistical descriptions: the macroscopic description $f(t,x,v)\,dx\,dv$ and the microscopic probabilities $\mu_t^N(dx^N\,dv^N)$. Of course, the quantity of information contained in μ^N is considerably more important than that contained in the macroscopic distribution: the latter informs us about the state of a typical particle, whereas a draw following the distribution μ_t^N informs us about the state of *all* particles. Think that if we have 10^{20} degrees of freedom, we will have to integrate 99999999999999999999 of them. For handling such vertiginous dimensions, we will require a fundamental concept: entropy.

2 The Entropic World

The concept and the name entropy were introduced by Clausius in 1865 as part of the theory—then under construction—of thermodynamics. A few years later Boltzmann (certainly influenced by the statistical ideas put forward by Laplace, Quetelet and others) revolutionized the concept by giving it a statistical interpretation based on atomic theory. In addition to this section, the reader can consult e.g. Balian [9, 10] about the notion of entropy in physical statistics.

2.1 Boltzmann's Formula

Let a physical system be given, which we suppose is completely described by its microscopic state $z \in \mathscr{X}$. Experimentally we only gain access to a partial description of that state, say $\pi(z) \in \mathscr{Y}$, where $\mathscr{Y}$ is a space of macroscopic states. I will not give precise hypotheses on the spaces $\mathscr{X}$ and $\mathscr{Y}$, but with the introduction of measure theory we will implicitly assume that these are "Polish" (separable complete metric) spaces.

How can we estimate the amount of information that is lost when we summarize the microscopic information by the macroscopic? Assuming that $\mathscr{Y}$ and $\mathscr{X}$ are denumerable, it is natural so suppose that the uncertainty associated with a state $y \in \mathscr{Y}$ is a function of the cardinality of the pre-image, i.e. $\#\pi^{-1}(y)$.

If we carry out two independent measures of two different systems, we are tempted to say that the uncertainties are additive. Now, with obvious notation, $\#\pi^{-1}(y_1, y_2) = (\#\pi_1^{-1}(y_1))(\#\pi_2^{-1}(y_2))$. To pass from this multiplicative operation to an addition, let us take a multiple of the logarithm. We thus end up with Boltzmann's celebrated formula, engraved on his tombstone in the Central Cemetery in Vienna:

$$S = k \log W, \tag{4}$$

where $W = \#\pi^{-1}(y)$ is the number of microscopic states compatible with the observed macroscopic state y and k is the so-called Boltzmann constant.[1]

In numerous cases, the space $\mathscr{X}$ of microscopic configurations is continuous, and in applying Boltzmann's formula it is customary to replace the counting measure by a privileged measure: for example by Liouville measure if we are interested in a Hamiltonian system. Thus W in (4) can be the *volume* of microscopic states that are compatible with the macroscopic state y.

If the space $\mathscr{Y}$ of macroscopic configurations is likewise continuous, then this notion of volume must be handled cautiously: the fiber $\pi^{-1}(y)$ is typically of volume zero and thus of scarce interest. One is tempted to postulate, for a given topology,

$$S(y) = \text{f.p.}_{\varepsilon\to 0} \log |\pi^{-1}(B_\varepsilon(y))|,$$

where $B_\varepsilon(y)$ is the ball of radius ε centered at y and f.p. denotes the finite part, meaning that we excise the divergence in ε, if indeed it has a universal behavior.

If this last point is not at all evident, the universality is nonetheless verified in the particular case that interests us where the microscopic state $\mathscr{X}$ is the space of configurations of N particles, i.e. $\mathscr{Y}^N$, and where we begin by taking the limit $N \to \infty$. In this limit, as we will see, the mean entropy per molecule tends to a finite value and we can subsequently take the limit $\varepsilon \to 0$, which corresponds to an arbitrarily precise *macroscopic* measure. The result is, within a sign, nothing other than Boltzmann's famous *H function*.

2.2 The Entropy Function H

Let us apply the preceding considerations to a macroscopic space made up of k different states: a macroscopic state is thus a vector $(f_1, \dots, f_k)$ of frequencies with, of course, $f_1 + \cdots + f_k = 1$. It is supposed that the measure is absolute (no error) and that $Nf_j = N_j$ is entire for all j. The number of microscopic states associated with this macroscopic state then equals

$$W = \frac{N!}{N_1! \dots N_k!}.$$

[1] Even if this formula accurately reflects Boltzmann's thoughts, it was Planck who first wrote it in this particular form, around 1900.

(If N_j positions are prepared in the j-th state and if we number the positions from 1 to N, then there are $N!$ ways of arranging the N balls in the N positions and it's subsequently impossible to distinguish between permutations on the interior of any single box.)

According to Stirling's formula, when $N \to \infty$ we have $\log N! = N \log N - N + \log \sqrt{2\pi N} + o(1)$. It follows easily that

$$\begin{aligned}\frac{1}{N}\log W &= -\sum_i \frac{N_i}{N}\log\frac{N_i}{N} + O\left(\frac{k\log N}{N}\right)\\ &= -\sum f_i \log f_i + o(1).\end{aligned}$$

We note that we can also arrive at the same result without using Stirling's formula, thanks to the so-called method of types [41, Sect. 12.4].

If now we increase the number of experiments, we can formally make k tend to ∞, while making sure that k remains small compared to N. Let us suppose that we have at our disposal a reference measure ν on the macroscopic space $\mathscr{Y}$, and that we can separate this space into "cells" of volume (measure) $\delta > 0$, corresponding to the different states. When $\delta \to 0$, if the system has a statistical distribution $f(y)$ with respect to the measure ν, we can reasonably think that $f_i \simeq \delta\, f(y_i)$, where y_i is a representative point of cell number i. But then

$$\sum_i f_i \log\frac{f_i}{\delta} \simeq \delta \sum_i f(y_i)\log f(y_i) \simeq \int f \log f \, d\nu,$$

where the last approximation comes from the second sum being a Riemann sum of the integral.

We have ended up with **Boltzmann's H function**: being given a reference measure ν on a space $\mathscr{Y}$ and a probability measure μ on $\mathscr{Y}$,

$$H_\nu(\mu) = \int f \log f \, d\nu, \qquad f = \frac{d\mu}{d\nu}. \tag{5}$$

If ν is a probability measure, or more generally a measure of finite mass, it is easy to extend this formula to all probabilities μ by setting $H_\nu(\mu) = +\infty$ if μ is not absolutely continuous with respect to ν. If ν is a measure of infinite mass, more precautions must be taken; we could require at the very least the finiteness of $\int f(\log f)_- \, d\nu$.

We then note that if the macroscopic space $\mathscr{Y}$ bears a measure ν, then the microscopic space $\mathscr{X} = \mathscr{Y}^N$ bears a natural measure $\nu^{\otimes N}$.

We are now ready to state the precise mathematical version of the formula for the function H: given a family $\{\varphi_j\}_{j\in\mathbb{N}}$ of bounded and uniformly continuous functions, then

$$\lim_{k\to\infty}\lim_{\varepsilon\to 0}\lim_{N\to\infty}\frac{1}{N}\log\nu^{\otimes N}\Bigg[\Bigg\{(y_1,\dots,y_N)\in\mathscr{Y}^N;\ \forall j\in\{1,\dots,k\},$$
$$\left|\int\varphi_j\,d\mu-\frac{1}{N}\sum_i\varphi_j(y_i)\right|\le\varepsilon\Bigg\}\Bigg]=-H_\nu(\mu). \tag{6}$$

We thus interpret N as the number of particles; the φ_j as a sequence of observables for which we measure the average value; and ε as the precision of the measurements. This formula summarizes in a concise manner the essential information contained in the function H.

If ν is a probability measure, statement (6) is known as **Sanov's theorem** [43] and is a leading result in the theory of large deviations. Before giving the interpretation of (6) in this theory, note that once we know how to treat the case where ν is a probability measure we easily deduce the case where ν is a measure of finite mass; however, I have no knowledge of any rigorous discussion in the case where ν is of infinite mass, even though we may expect that the result remains true.

2.3 *Large Deviations*

Let ν be a probability measure and suppose that we independently draw random variables y_i according to ν. The empirical measure $\hat{\mu}=N^{-1}\sum\delta_{y_j}$ is then a random measure, almost certainly convergent to ν as $N\to\infty$ (it's Varadarajan's theorem, also called the fundamental law of statistics [49]). Of course it's possible that appearances deceive and that we think we are observing a measure μ distinct from ν. This probability decreases exponentially with N and is roughly proportional to $\exp(-N\,H_\nu(\mu))$; in other words, the Boltzmann entropy dictates the rarity of conditions that lead to the "unexpected" observation μ.

2.4 *Information*

Information theory was born in 1948 with the remarkable treatise of Shannon and Weaver [94] on the "theory of communication" which is now a pillar for the whole industry of information transmission.

In Shannon's theory, somewhat disembodied for its reproduction and impassionate discussion, the quantity of information carried by the decoding of a random signal is defined as a function of the reciprocal of the probability of the signal (which is rare and precious). Using the logarithm allows having the additivity property, and Shannon's formula for the mean quantity gained in the course of decoding is obtained: $\mathbb{E}\log(1/p(Y))$, where p is the law of Y. This of course gives Boltzmann's formula again!

2.5 Entropies on All Floors

Entropy is not an intrinsic concept; it depends on the observer and the degree of knowledge that can be acquired through experiments and measures. The notion of entropy will consequently vary with the degree of precision of the description.

Boltzmann's entropy, as has been seen, informs us of the rarity of the kinetic distribution function $f(x, v)$ and the quantity of microscopic information remaining to be discovered once f is known.

If to the contrary we are given the microscopic state of all the microscopic particles, no hidden information remains and thus no more entropy. But if we are given a probability μ^N on the microscopic configurations, then the concept of entropy again has meaning: the entropy will be lower when the probability μ^N is concentrated and informative in itself. We thus find ourselves with a notion of **microscopic entropy**, $S_N = -H_N$,

$$H_N = \frac{1}{N}\int f^N \log f^N \, dx^N \, dv^N,$$

which is typically conserved by the Newtonian dynamic in consequence of Liouville's theorem. We can verify that

$$H_N \geq H(\mu^{1;N}),$$

with equality when μ^N is a tensor product and there are thus no correlations between particles. The idea is that the state of the microscopic particles is easier to obtain by multiparticle measurements than particle by particle—unless of course when the particles are independent!

In the other direction, we can also be given a *less precise* distribution than the kinetic distribution: this typically concerns a hydrodynamic description, which involves only the density field $\rho(x)$, the temperature $T(x)$ and mean velocity $u(x)$. The passage from the kinetic formalism to hydrodynamic formalism is accomplished by simple formulas:

$$\rho(x) = \int f(x, v)\, dv; \qquad u(x) = \rho(x)^{-1}\int f(x, v) v \, dv;$$

$$T(x) = \frac{1}{d\,\rho(x)}\int f(x, v)\frac{|v - u(x)|^2}{2}\, dv.$$

With this description is associated a notion of **hydrodynamic entropy**:

$$S_h = -\int \rho \log \frac{\rho}{T^{d/2}}.$$

This information is always lower than kinetic information. We have, finally, a hierarchy: first microscopic information at the low level, then "mesoscopic" information from the Boltzmann distribution function, finally "macroscopic information" contributed by the hydrodynamic description. The relative proportions of these different entropies constitute excellent means for appraising the physical state of the systems considered.

2.6 The Universality of Entropy

Initially introduced within the context of the kinetic theory of gases, entropy is an abstract and evolving mathematical concept, which plays an important role in numerous areas of physics, but also in branches of mathematics having nothing to do with physics, such as information theory and other sciences.

Some mathematical implications of the concept are reviewed in my survey *H-Theorem and beyond: Boltzmann's entropy in today's mathematics* [106].

3 Order and Chaos

Intuitively, a microscopic system is ordered if all its particles are arranged in a coordinated, *correlated* way. On the other hand, it is chaotic if the particles, doing just as they please, act entirely independently from one another. Let us reformulate this idea: a distribution of particles is chaotic if each of the particles is oblivious to all the others, in the sense that a gain of information obtained for a given particle brings no gain in information about any other particle. This simple notion, key to Boltzmann's equation, presents some important subtleties that we will briefly mention.

3.1 Microscopic Chaos

To say that random particles that are oblivious to each other is equivalent to saying that their joint law is tensorial. Of course, even if the particles are unaware of each other initially, they will enter into interaction right away and the independence property will be destroyed. In the case of hard spheres, the situation is still worse: the particles are obliged to consider each another since the spheres cannot interpenetrate. Their independence is thus to be understood asymptotically when the number of particles becomes very large; and experiments seeking to measure the degree of independence will involve but a finite number of particles. This leads naturally to the definition that follows.

Let $\mathscr{Y}$ be a macroscopic space and, for each N, let μ^N be a probability measure, assumed symmetric (invariant under coordinate permutations). We say that (μ^N) is chaotic if there exists a probability μ such that $\mu^N \simeq \mu^{\otimes N}$ in the sense of the weak topology of product measures. Explicitly, this means that for each $k \in \mathbb{N}$ and for all choices of the continuous functions $\varphi_1, \dots, \varphi_k$ bounded on $\mathscr{Y}$, we have

$$\int_{\mathscr{Y}^N} \varphi_1(y_1) \dots \varphi_k(y_k) \mu^N(dy_1 \dots dy_N) \underset{N\to\infty}{\longrightarrow} \left(\int \varphi_1 \, d\mu \right) \dots \left(\int \varphi_k \, d\mu \right). \quad (7)$$

Of course, the definition can be quantified by introducing an adequate notion of distance, permitting us to measure the gap between μ^N and $\mu^{\otimes N}$. We can then say that a distribution μ^N is more or less chaotic. We again emphasize: what matters is the independence of a small number k of particles taken from among a large number N.

It can be shown (see the argument in [99]) that it is equivalent to impose property (7) for all $k \in \mathbb{N}$, or simply for $k = 2$. Thus chaos means precisely that 2 *particles drawn randomly from among N are asymptotically independent when* $N \to \infty$. The proof proceeds by observing the connections between chaos and empirical measure.

3.2 Chaos and Empirical Measure

By the law of large numbers, chaos automatically implies an asymptotic determinism: with very high probability, the empirical measure approaches the statistical distribution of an arbitrary particle when the total number of particles becomes gigantic.

It turns out that, conversely, *correlations accommodate very badly a macroscopic prescription of density*. Before giving a precise statement, we will illustrate this concept in a simple context. Consider a box with two compartments, in which we distribute a very large number N of *indistinguishable* balls. A highly correlated state would be a one in which all the particles occupy the same compartment: if I draw two balls at random, the state of first ball informs me completely about the state of the second. But of course, from the moment when the respective numbers of balls in the compartments are fixed and both nonzero, such a state of correlation is impossible. In fact, if the particles are indistinguishable, when two are drawn at random, the only information gotten is obtained by exploiting the fact that they are distinct, so that knowledge of the state of the first particle reduces slightly the number of possibilities for the state of the second. Thus, if the first particle occupies state 1, then the chances of finding the second particle in state 1 or 2 respectively are not $f_1 = N_1/N$ and $f_2 = N_2/N$, but $f_1' = (N_1 - 1)/(N - 1)$ and $f_2 = N_2/(N - 1)$. The joint distribution of a pair of particles is thus very close to the product law.

By developing the preceding argument, we arrive at an elementary but conceptually profound general result, whose proof can be found in Sznitman's course [99] (see also [40, p. 91]): *microscopic chaos is equivalent to the determinism of the empirical measure*. More precisely, the following statements are equivalent:

(i) (μ^N) is μ-chaotic;
(ii) the empirical measure $\widehat{\mu}^N$ associated with μ^N converges in law toward the deterministic measure μ.

By "empirical measure $\widehat{\mu}^N$ associated with μ^N" we understand the measure of the image of μ^N under the mapping $(y_1, \dots, y_N) \longmapsto N^{-1} \sum \delta_{y_i}$, which is a measure of random probability. Convergence in law means that, for each continuous bounded function Φ on the space of probability measures, we have

$$\int \Phi\left(\frac{1}{N}\sum \delta_{y_i}\right)\mu^N(dy_1 \dots dy_N) \underset{N\to\infty}{\longrightarrow} \Phi(\mu).$$

In informal language, given a statistical quantity involving $\widehat{\mu}^N$, we can obtain an excellent approximation for large N by replacing, in the expression for this statistic, $\widehat{\mu}^N$ by μ.

The notion of chaos thus presented is weak and susceptible to numerous variants; the general idea being that μ^N must be close to $\mu^{\otimes N}$. The stronger concept of **entropic chaos** was introduced by Ben Arous and Zeitouni [13]: there $H_{\mu^{\otimes N}}(\mu^N) = o(N)$ is imposed. A related notion was developed by Carlen, Carvalho, Le Roux, Loss and Villani [32] in the case where the microscopic space is not a tensor product, but rather a sphere of large dimension; the measure $\mu^{\otimes N}$ is replaced by the restriction of the product measure to the sphere. Numerous other variants remain to be discovered.

3.3 The Reign of Chaos

In Boltzmann's theory, it is postulated that *chaos is the rule*: when a system is prepared, it is *a priori* in a chaotic state. Here are some possible arguments:

- if we can act on the macroscopic configuration, we will not have access to the microscopic structure and it is very difficult to impose correlations;
- the laws that underlie the microscopic variations are unknown to us and we may suppose that they involve a large number of factors destructive to correlations;
- the macroscopic measure observed in practice seems always well determined and not random;
- if we fix the macroscopic distribution, the entropy of a chaotic microscopic distribution is larger than the entropy of a nonchaotic microscopic distribution.

Let us explain the last argument. If we are given a probability μ on $\mathscr{Y}$, then the product probability $\mu^{\otimes N}$ is the maximum entropy among all the symmetric probabilities μ^N on $\mathscr{Y}^N$ having μ as marginal. In view of the large numbers N in play, this represents a phenomenally larger number of possibilities.

The microscopic measure μ_0^N can be considered as an object of Bayesian nature, an *a priori* probability on the space of possible observations. This choice, in general arbitrary, is made here in a canonical manner by maximization of the entropy: in some way we choose the distribution that leaves the most possibilities open and makes the observation the most likely. We thus join the scientific approach of maximum likelihood, which has proved its robustness and effectiveness—while skipping the traditional quarrel between frequentists and Bayesians!

The problem of the propagation of chaos consists of showing that our chaos hypothesis, made on the initial data (it's not entirely clear how), is propagated by the microscopic dynamic (which is well defined). The propagation of chaos is essential for two reasons: first, it shows that independence is asymptotically preserved, providing statistical information about the microscopic dynamic; secondly, it guarantees that *the statistical measure remains deterministic*, which allows hope for the possibility of a **macroscopic equation** governing the evolution of this empirical measure or its approximation.

3.4 Evolution of Entropy

A recurrent theme in the study of dynamical systems, at least since Poincaré, is the search for invariant measures; the best known example is Liouville measure for Hamiltonian systems. This measure possesses the remarkable additional property of tensorizing itself.

Suppose that we have a microscopic dynamic on $\mathscr{Y}^N$ and a measure ν on the space $\mathscr{Y}$ such that $\nu^{\otimes N}$ is an invariant measure for the microscopic dynamic; or more generally that there exists a ν-chaotic invariant measure on $\mathscr{Y}^N$. What happens with the preservation of microscopic volume in the limit $N \to \infty$?

A simple consequence of preservation of volume is conservation of macroscopic information $H_{\nu^{\otimes N}}(\mu_t^N)$, where μ_t^N is the image measure of μ_0^N through the microscopic evolution. In fact, since μ_t^N is preserved by the flow (by definition) and $\nu^{\otimes N}$ likewise, the density $f^N(t, y_1, \dots, y_N)$ is constant along the trajectories of the system, and it follows that $\int f^N \log f^N \, d\nu^{\otimes N}$ is likewise constant.

Matters are more subtle for macroscopic information. Of course, if the various particles evolve independently from one another, the measure μ_t^N remains factored for all time, and we easily deduce that the macroscopic entropy remains constant. In general, the particles interact with one another, which destroys independence; however if there is propagation of chaos in a sufficiently strong sense, the independence is restored as $N \to \infty$, and we consequently have determinism for the empirical measure. So all the typical configurations for the microscopic initial measure μ_0^N give way, after a time t, to an empirical measure $\hat{\mu}_t^N \simeq \mu_t$, where μ_t is well determined. But it is possible that *other* microscopic configurations are compatible with the state μ_t, configurations that have not been obtained by evolution from typical initial configurations.

In other words: if we have a propagation of macroscopic determinism between the initial time and the time t, and if the microscopic dynamic preserves the reference measure produced, then we expect that the volume of the admissible microscopic states does not decrease between time 0 and time t. Keeping in mind the definition of entropy, we would have $e^{NS(t)} \geq e^{NS(0)}$, where $S(t)$ is the value of the entropy at time t. We thus expect that the entropy does not decrease over the course of the temporal evolution:

$$S(t) \geq S(0).$$

But then why not reverse the argument and say that chaos at time t implies chaos at time 0, by reversibility of the microscopic dynamic? This argument is in general inadmissible unless an exact notion of the chaos propagated is specified. The initial data prepared "at random" with just one kinetic distribution constraint, is supposed chaotic in a less strong sense; this depends on the microscopic evolution.

The notion of scale of interaction plays an important role here. Certain interactions take place on a macroscopic scale, other on a microscopic scale, which is to say that all or part of the interaction law is coded in parameters that are invisible on the macroscopic level. In this last case, the notion of chaos conducive to the propagation of the dynamic risks not being visible on the macroscopic scale and we can expect a degradation of the notion of chaos.

From there, the discussion must involve the details of the dynamic, and our worst troubles begin.

4 Chaotic Equations

After the introduction of entropy and chaos, we can return to the Newtonian systems of Sect. 1, for which the phase space is composed of positions and velocities. A **kinetic equation** is an evolution equation bearing on the distribution $f(t,x,v)$; the important role of the velocity variable v justifies the terminology *kinetic*. By extension, in the case where there are external degrees of freedom (orientation of molecules for example), by extension we still speak of kinetic equations.

As descendents of Boltzmann, we pose the problem of deducing the macroscopic evolution starting from the underling microscopic model. This problem is in general of considerable difficulty. The fundamental equations are those of Vlasov, Boltzmann, Landau and Balescu–Lenard, published respectively in 1938, 1867, 1936 and 1960 (the more or less logical order of presentation of these equations does not entirely follow the order in which they were discovered...).

4.1 Vlasov's Equation

Also called the Boltzmann equation without collisions, Vlasov's equation [112] is a mean field equation in the sense that all particles interact with one another (so each particle feels the mean contribution of the others). To deduce it from Newtonian dynamics, we begin by translating Newton's equation (1) as an equation in the empirical measure; for this we write the evolution equation of an arbitrary observable:

$$\begin{aligned}
&\frac{d}{dt}\frac{1}{N}\sum_i \varphi(X_i(t),\dot{X}_i(t)) \\
&\quad = \frac{1}{N}\sum_i \big[\nabla_x\varphi(X_i,\dot{X}_i)\cdot\dot{X}_i + \nabla_v\varphi(X_i,\dot{X}_i)\cdot\ddot{X}_i\big] \\
&\quad = \frac{1}{N}\sum_i \left[\nabla_x\varphi(X_i,\dot{X}_i)\cdot\dot{X}_i + \nabla_v\varphi(X_i,\dot{X}_i)\cdot\left(a\sum_j F(X_i-X_j)\right)\right].
\end{aligned}$$

This can be rewritten

$$\frac{\partial\widehat{\mu}^N}{\partial t} + v\cdot\nabla_x\hat{\mu}_t^N + aN(F*\hat{\mu}_t^N)\cdot\nabla_v\hat{\mu}_t^N = 0. \tag{8}$$

If now we suppose that $aN \simeq 1$ and we make the approximation

$$\hat{\mu}_t^N(dx\,dv) \simeq f(t,x,v)\,dx\,dv,$$

we obtain Vlasov's equation

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f + \left(F *_x \int f \, dv \right) \cdot \nabla_v f = 0. \tag{9}$$

We note well that $\hat{\mu}_t^N$ in (8) is a weak solution of Vlasov's equation, so that the passage to the limit is conceptually very simple: it is simply a stability result for the Vlasov equation.

Quite clearly I have gone a bit far, for this equation is nonlinear. If $\hat{\mu} \simeq f$ in the sense of the weak topology of measures, then $F * \hat{\mu}$ converges to $F * \int f \, dv$ in a topology determined by the regularity of F, and if this topology is weaker than uniform convergence, nothing guarantees that $(F * \hat{\mu})\hat{\mu} \simeq (F * f) f$.

If F is in fact bounded and uniformly continuous, then the above argument can be made rigorous. If F is furthermore L-Lipschitz, then we can do better and establish a stability estimate in weak topology: if (μ_t) and (μ_t'') are two weak solutions of Vlasov's equation, then

$$W_1(\mu_t, \mu_t') \leq e^{2 \max(1,L)|t|} \, W_1(\mu_0, \mu_0'),$$

where W_1 is the Wasserstein distance of order 1,

$$W_1(\mu, \nu) = \sup : \left\{ \int \varphi \, d\mu - \int \varphi \, d\nu ; \varphi \text{ 1-Lipschitz} \right\}.$$

Estimates of this sort are found in [95, Chap. 5] and date back to the 1970s (Dobrushin [48], Braun and Hepp [24], Neunzert [87]). Large deviation estimates can also be established as in [20].

However, for **singular interactions**, the problem of the Vlasov limit remains open, except for a result of Jabin and Hauray [64], which essentially assumes that (a) $F(x-y) = O(|x-y|^{-s})$ with $0 < s < 1$; and (b) the particles are initially well separated in phase space, so that

$$\inf_{j \neq i} \Big(|X_i(0) - X_j(0)| + |\dot{X}_i(0) - \dot{V}_j(0)| \Big) \geq \frac{c}{N^{\frac{1}{2d}}}.$$

Neither of these conditions is satisfied: the first lacks the Coulomb case of singularity order, while the second excludes the case of random data. However, it remains the sole result available at this time... To go further, it would be nice to have a sufficiently strong notion of chaos so as to be able to control the number of pairs (i, j) such that $\left| X_i(t) - X_j(t) \right|$ is small. In the absence of such controls, Vlasov's equation for singular interactions remains an act of faith.

This act of faith is very effective since the Vlasov–Poisson model, in which $F = -\nabla W$, where W a fundamental solution of $\pm\Delta$, is the universally accepted classic model in plasma physics [42, 71] as well as in astronomy [15]. In the first instance a particle is an electron, in the second a star! The only difference lies in the sign: repulsive interaction for electrons, attractive for stars. We should not be astonished to see stars considered in this way as microscopic objects: they are effectively so on the scale of a galaxy (which can tally 10^{12} stars…).

The theory of the Vlasov–Poisson equation itself remains incomplete. We can distinguish presently two principal theories, both developed in the entire space. That of Pfaffelmoser, simplified by Schaeffer and exposited for example in [51], supposes that the initial data f_i is C^1 with compact support; later this unsatisfactory compactness assumption was removed by Horst [114] by an improvement of the Pfaffelmoser–Schaeffer method. The concurrent theory is that Lions–Perthame, reviewed in [23]. Pfaffelmoser's theory has been adapted in spatially periodic context (see [12] or modify [114]), which is not the case for the Lions–Perthame theory.

4.2 Boltzmann's Equation

Vlasov's equation loses its relevance when the interactions have a short range. A typical example is that of rarefied gas, for which the dominant interactions are binary and are uniquely produced in the course of "collisions" between particles.

Boltzmann's equation was established by Maxwell [80] and Boltzmann [21, 22]; it describes a situation where the interactions are of short range and where each particle undergoes $O(1)$ impacts per unit of time. Much more subtle than the situation of Vlasov's mean field, the Boltzmann situation is nonetheless simpler than the hydrodynamic one where the particles undergo a very large number of collisions per unit of time.

We start by establishing the equation informally. The movement of a particle occurs with alternation of rectilinear trajectories and collisions, during the course of which its velocity changes so abruptly that we can consider the event as instantaneous and localized in space. We first consider the emblematic case of hard spheres of radius r: a collision occurs when two particles, with respective positions x and y and with respective velocities v and w, are found in a configuration where $|x-y| = 2r$ and $(w-v)\cdot(y-x) < 0$. We then speak of a *precollisional configuration*. We let $\omega = (y-x)/|x-y|$.

We now come to the central point in all Boltzmann's argument: *when two particles encounter each other, with very high probability they will* (*almost*) *not be correlated*: think of two people who encounter each other for the first time. We can consequently apply the hypothesis of molecular chaos to such particles, and we find that the probability of an encounter between these particles is proportional to

$$\begin{aligned} f^{2;N}(t,x,v,x+2r\omega,w) &\simeq f^{1;N}(t,x,v)\, f^{1;N}(t,x+2r\omega,w) \\ &\simeq f^{1;N}(t,x,v)\, f^{1;N}(t,x,w), \end{aligned}$$

provided thus that $(w-v)\cdot\omega < 0$. We likewise need to take into account the relative velocities in order to evaluate the influence of the particles of velocity w on the particles of velocity v: the probability of encountering a particle of velocity w in a unit of time is proportional to the product of $|v-w|$ by the effective section (in dimension 3 this is the apparent area of the particles, or πr^2) and by the cosine of the angle between $v - w$ and ω (the extreme case is where $v - w$ is orthogonal to ω, which is to say that the two particles but graze each other, clearly an event of probability zero). Each of these collisions removes a particle of velocity v, and we thus have a negative term, the *loss term*, proportional to

$$-\iint f(t,x,v) f(t,x,v_*) |(v - v_*) \cdot \omega| \, dv_* \, d\omega.$$

The velocities after the collision are easily calculated:

$$v' = v - (v - v_*) \cdot \omega\omega; \qquad v'_* = v_* + (v - v_*) \cdot \omega\omega. \tag{10}$$

These velocities do not matter for the final analysis.

However, we also need to take account of all the particles of velocity v that have been created by collisions between particles of arbitrary velocities. By microscopic reversibility, these velocities are of the form (v', v'_*), and our problem is to take account of all the possible pairs (v', v'_*), which in this problem of computing the *gain term* are the *pre-collisional* velocities. We thus again apply the hypothesis of pre-collisional chaos and obtain finally the expression of the **Boltzmann equation for solid spheres**:

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = Q(f, f), \tag{11}$$

where

$$\begin{aligned} &Q(f,f)(t,x,v) \\ &\quad = B \int_{\mathbb{R}^3} \int_{S^2_-} |(v - v_*) \cdot \omega| \big(f(t,x,v') f(t,x,v'_*) \\ &\qquad - f(t,x,v)\, f(t,x,v_*) \big) \, dv_* \, d\omega. \end{aligned}$$

Here S^2_- denotes the pre-collisional configurations $\omega \cdot (v - v_*) < 0$, and B is a constant. By using the change of variable $\omega \to -\omega$ we can symmetrize this expression and arrive at the final expression (after changing the value of B)

$$\begin{aligned} &Q(f,f)(t,x,v) \\ &\quad = B \int_{\mathbb{R}^3} \int_{S^2} |(v - v_*) \cdot \omega| \big(f(t,x,v') f(t,x,v'_*) - f(t,x,v)\, f(t,x,v_*) \big) \, dv_* \, d\omega. \end{aligned} \tag{12}$$

The operator (12) is the **Boltzmann collision operator for hard spheres**. The problem now consists of justifying this approximation.

To do this, in the 1960s Grad proposed a precise mathematical limit: have r tend toward 0 and at the same time N toward infinity, so that $Nr^2 \to 1$, which is to say that the total effective section remains constant. Thus a given particle, moving among all the others, will typically encounter a finite number of them in a unit of time. One next starts with a microscopic probability density $f_0^N(x^N, v^N)\, dx^N\, dv^N$, which is allowed to evolve by the Newtonian flow $\mathcal{N}_t$, and one attempts to show that the first marginal $f^{1;N}(t, x, v)$ (obtained by integrating all the variables except the first position variable and the first velocity variable) converges in the limit to a solution of the Boltzmann equation.

The Boltzmann–Grad limit is also often called the **low density limit** [40, p. 60]: in fact, if we start from the Newtonian dynamic and fix the particle size, we will dilate the spatial scale by a factor $1/\sqrt{N}$ and the density will be of the order $N/N^{3/2} = N^{-1/2}$.

At the beginning of the 1970's, Cercignani [37] showed that Grad's program could be completed if one proved a number of *plausible* estimates; shortly thereafter, independently, Lanford [69] sketched the proof of the desired result.

Lanford's theorem is perhaps the single most important mathematical result in kinetic theory. In this theorem, we are given microscopic densities f_0^N such that for each k the densities $f_0^{k;N}$ of the k particle marginals are continuous, satisfy Gaussian bounds at large velocities and converge uniformly outside the collisional configurations (those where the positions of two distinct particles coincide) to their limit $f_0^{\otimes k}$. The conclusion is that there exists a time $t_* > 0$ such that $f_t^{k;N}$ converges almost everywhere to $f_t^{\otimes k}$, where f_t is a solutions of Boltzmann's equation, for all time $t \in [0, t_*]$.

Lanford's estimates were later rewritten by Spohn [95] and by Illner and Pulvirenti [61, 62] who replaced the hypothesis of small time by a smallness hypothesis on the initial data, permitting Boltzmann's equation to be treated as a perturbation of free transport. These results are reviewed in [40, 90, 95].

The technique used by Lanford and his successors goes through the **BBGKY hierarchy** (Bogoliubov–Born–Green–Kirkwood–Yvon), the method by which the evolution of the marginal for a particle $f^{1;N}$ is expressed as a function of the marginal for two particles $f^{2;N}$; the evolution of a two-particle marginal $f^{2;N}$ as a function of a three-particle marginal $f^{3;N}$, and so forth. This procedure is especially uneconomical (in the preceding heuristic argument, we only use $f^{1;N}$ and $f^{2;N}$, but there is no known alternative).

Each of the equations of the hierarchy is then solved via Duhamel's formula, applying successively the free transport and collision operators, and by summing over all the possible collisional history. The solution at time t is thus formally expressed, as with an exponential operator, as a function of the initial data and we can apply the chaos hypothesis on (f_0^N).

We then pass to the limit $N \to \infty$ in each of the equations, after having verified that we can neglect pathological "recollisions", where a particle again encounters a particle that it had already encountered beforehand, and which is thus not unknown to it. This point is subtle: in [40, Appendix 4.C] a dynamic that is *a priori* simpler than that of solid spheres, due to Uchiyama, is discussed, with only four velocities in the plane, for which the recollision configurations cannot be neglected, and the kinetic limit does not exist.

It remains to identify the result with the series of tensor products of the solution to Boltzmann's equation and conclude by using a uniqueness result.

Spohn [95, Sect. 4.6] shows that one can give more precise information on the microscopic distribution of the particles: on the small scale, this follows a homogeneous **Poisson law** in phase space. This is consistent with the intuitive idea of molecular chaos.

Lanford's theorem settled a controversy that had lasted since Boltzmann himself; but it leaves numerous questions in suspense. In the first place, it is limited to a small time interval (on which only about $1/6$ of the particles have had time to collide... but the conceptual impact of the theorem is nonetheless important). The variant of Illner and Pulvirenti lifts this restriction of small time, but the proof does not lend itself to a bounded geometry. As for lifting the smallness restriction, at the moment it is but a distant dream.

Next, to this day the theorem has only been proved for a system of solid spheres; long-range interactions are not covered. Cercignani [36] notes that the limit of Boltzmann–Grad for such interactions poses subtle problems, even from the formal viewpoint.

Finally, the most frustrating thing is that Lanford avoided discussion of **pre-collisional chaos**, the notion that particles that are about to collide are not correlated. This notion is very subtle, because just after the collision, correlations have inevitably taken place. In other words, we have *pre-collisional chaos, but not post-collisional.*

What does pre-collisional chaos mean exactly? For the moment we do not have a precise definition. It's certainly a stronger notion than chaos in the usual sense; it involves too a de-correlation hypothesis that is seen on a set of codimension 1, i.e. configurations leading to collisions. We would infer that it is a notion of chaos where we have replaced the weak topology by a uniform topology; but that cannot be so simple, since chaos in a uniform topology also implies post-collisional chaos, which is incompatible with pre-collisional chaos! In fact, the continuity of the two-particle marginal along a collision would imply

$$\begin{aligned} f(t,x,v)\, f(t,x,v_*) &\simeq f^{(2;N)}(t,x,v,x+2r\omega,v_*) \\ &= f^{(2;N)}(t,x,v',x+2r\omega,v'_*) \\ &\simeq f^{(1;N)}(t,x,v')\, f^{(1;N)}(t,x,v'_*). \end{aligned}$$

Passing to the limit we would have

$$f(t,x,v')\, f(t,x,v'_*) = f(t,x,v)\, f(t,x,v_*),$$

and as we will see in Sect. 5.3 this implies that f is Gaussian in the velocity variable, which is of course false in general. Another argument for showing that post-collisional chaos must be incompatible with pre-collisional chaos consists of noting that if we have post-collisional chaos, the reasoning leading to the Boltzmann equation can be used again by expressing two-particle probabilities in terms of post-collisional probabilities… and we then obtain Boltzmann's equation in reverse:

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = -Q(f, f).$$

As has been mentioned, Lanford's proof applies only to solid spheres; but Boltzmann's equation is used for a much larger range of interactions. The general form of the equation, say in dimension d, is the same as in (11):

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = Q(f, f), \tag{13}$$

but now

$$Q(f, f) = \int_{\mathbb{R}^d} \int_{S^{d-1}} (f' f'_* - f f_*) \tilde{B}(v - v_*, \omega)\, dv_*\, d\omega \tag{14}$$

where $\tilde{B}(v - v_*, \omega)$ depends only on $|v - v_*|$ and $|(v - v_*) \cdot \omega|$. There exist several representations of this integral operator (see [103]); it is often convenient to change variables by introducing another angle, $\sigma = (v' - v'_*)/|v - v_*|$, so that the formulas (10) must be replaced by

$$v' = \frac{v + v_*}{2} + \frac{|v - v_*|}{2}\sigma, \qquad v'_* = \frac{v + v_*}{2} - \frac{|v - v_*|}{2}\sigma. \tag{15}$$

We must then replace the collision kernel $\tilde{B}$ by B so that

$$B\, d\sigma = \tilde{B}\, d\omega.$$

Explicitly, we find

$$\frac{1}{2}\tilde{B}(z, \omega) = \left| 2\frac{z}{|z|} \cdot \omega \right|^{d-2} B(z, \sigma).$$

The precise form of B (or, in an equivalent way, of $\tilde{B}$) is obtained by a classical scattering computation that goes back to Maxwell and which can be found in [38]: for an impact parameter $p \geq 0$ and a relative velocity $z \in \mathbb{R}^3$, the deviation angle θ equals

$$\theta(p, z) = \pi - 2p \int_{s_0}^{+\infty} \frac{ds/s^2}{\sqrt{1 - \frac{p^2}{s^2} - 4\frac{\phi(s)}{|z|^2}}} = \pi - 2\int_0^{\frac{p}{s_0}} \frac{du}{\sqrt{1 - u^2 - \frac{4}{|z|^2}\phi(\frac{p}{u})}},$$

where s_0 is the positive root of

$$1 - \frac{p^2}{s_0^2} - 4\frac{\phi(s_0)}{|z|^2} = 0.$$

So B is implicitly defined by

$$B(|z|, \cos\theta) = \frac{p}{\sin\theta} \frac{dp}{d\theta} |z|. \tag{16}$$

We write either $B(|z|, \cos\theta)$ or $B(z, \sigma)$, it being understood that the deviation angle θ is the angle formed by the vectors $v - v_*$ and $v' - v'_*$.

When $\phi(r) = 1/r$, we recover Rutherford's formula for the Coulomb deviation:

$$B(|v - v_*|, \cos\theta) = \frac{1}{|v - v_*|^3 \sin^4(\theta/2)}.$$

When $\phi(r) = 1/r^{s-1}$, $s > 2$, the collision kernel is not computed explicitly, but it can be shown that (always in dimension 3)

$$B(|v - v_*|, \cos\theta) = b(\cos\theta)\,|v - v_*|^\gamma, \qquad \gamma = \frac{s-5}{s-1}. \tag{17}$$

Furthermore, the function b, defined implicitly, is locally smooth with a *noninte-grable angular singularity* when $\theta \to 0$:

$$\sin\theta\, b(\cos\theta) \sim K\theta^{-1-\nu}, \qquad \nu = \frac{2}{s-1}. \tag{18}$$

This singularity corresponds to collisions with large impact parameter p, where there is scant deflection. It is inevitable once the forces are of infinite range: in fact

$$\int_0^\pi B(|z|, \cos\theta) \sin\theta\, d\theta = |z| \int_0^\pi p \frac{dp}{d\theta}\, d\theta = |z| \int_0^{p_{\max}} p\, dp = \frac{|z|\, p_{\max}^2}{2}. \tag{19}$$

In the particular case $s = 5$, the collision kernel depends no longer on the relative velocity, but only on the *deviation angle*: we speak of Maxwellian molecules. By extension, we say that $B(v - v_*, \sigma)$ is a Maxwellian collision kernel if it depends only on the angle between $v - v_*$ and σ. The Maxwellian molecules are above all a phenomenological model, even if the interaction between a charged ion and a neutral particle in a plasma is regulated by such a law [42, Vol. 1, p. 149]. The potentials in $1/r^{s-1}$ for $s > 5$ are called hard potentials, for $s < 5$ soft potentials. Often the angular singularity $b(\cos\theta)$ is truncated to small values of θ.

The Boltzmann equation is important in modeling rarefied gases, as explained in [39]. Nonetheless, because of its eventful history and its conceptual content, as well as the impact of Boltzmann's treatise [22], this equation has exerted a fascination that goes far beyond its usefulness. The first mathematical works dedicated to it are those of Carleman[2] [26, 27], followed by Grad [57]. Besides the article by Lanford [69] already mentioned, a result that has had a great impact is the weak stability

[2]The monography [27] was incomplete at the time of Carleman's passing away, and was completed by Carleson.

theorem of DiPerna–Lions [47]. The equation is well understood in the spatially homogeneous setting for hard potentials with angular truncation, see e.g. [84]; and in the setting close to equilibrium, see e.g. [60]. We refer to the reference treatises [38, 40, 103] for a number of other results.

4.3 Landau's Equation

Boltzmann's collisional integral loses its meaning for Coulomb interactions because of the extremely slow decrease of the Coulomb potential. The grazing collisions, with large impact parameter, then become dominant.

In 1936, Landau [67] established, using formal arguments, an asymptotic of Boltzmann's kernel in this setting. Letting λ_D be the shielding distance (below which the Coulomb potential is no longer visible because of the global neutrality of the plasma), and r_0 the typical collision distance (distance of two particles whose interaction energy is comparable to the molecular excitation energy), the parameter $\Lambda = 2\lambda_D/r_0$ is the plasma parameter, and in the limit $\Lambda \to \infty$ (justified for "classical" plasmas), the Boltzmann operator can be formally replaced by a diffusive operator called Landau's operator:

$$Q_B(f, f) \simeq \frac{\log \Lambda}{2\pi \Lambda} Q_L(f, f), \tag{20}$$

$$Q_L(f, f) = \nabla_v \cdot \left(\int_{\mathbb{R}^3} a(v - v_*) \big[f(v_*) \nabla_v f(v) - f(v) \nabla_v f(v_*) \big] \, dv_* \right), \tag{21}$$

$$a(v - v_*) = \frac{L}{|v - v_*|} \Pi_{(v-v_*)^\perp}, \tag{22}$$

where L is a constant and $\Pi_{z^\perp}$ denotes orthogonal projection onto $z^\perp$.

The Landau approximation is now well understood mathematically in the context of a limit called **grazing collision asymptotics**; [3] can be consulted for a detailed discussion of this problem.

The Landau operator, both diffusive and integral, presents a remarkable structure. It is easily generalized to arbitrary dimensions $d \geq 2$, and the coefficient $L/|z|$ can be changed to $L\,|z|^{\gamma+2}$, where γ is the exponent appearing in (17). The models of hard potential type with $\gamma > 0$ have been completely studied in the spatially homogeneous case [45]; but it is definitely the case $\gamma = -3$ in dimension 3 that is physically interesting. In this case we only know how to prove the existence of weak solutions in the spatially homogeneous case (by adapting [1, Sect. 7] and the existence of strong solutions for perturbations of equilibrium [59]). This situation is entirely unsatisfactory.

4.4 The Balescu–Lenard Equation

In 1960, Balescu [7] directly establishes a kinetic equation that describes the Coulomb interactions in a plasma; he thus recovers an equation published in another form by Bogoliubov [19] and simplified by Lenard. The reference [96] can be consulted for information on the genesis of the equation, and [8] for a synthetic presentation. The collision kernel in this equation takes the same form as (21), the difference is in the expression of the matrix $a(v - v_*)$, which now depends both on v and ∇f:

$$\begin{aligned} a_{BL}\big(v, v - v_*, \nabla f\big) &= B \int_{|k| \le K_0} \frac{k \otimes k}{|k|^4} \frac{\delta_{k\cdot(v-v_*)}}{|\epsilon(k, k\cdot v, \nabla f)|^2}\, dk, \\ \epsilon\big(k, k\cdot v, \nabla f\big) &= 1 + \frac{b}{|k|^2} \int_{\mathbb{R}^3} \frac{k \cdot \nabla f(u)}{k\cdot(v-u) - i\,0}\, du. \end{aligned} \tag{23}$$

This equation can also be obtained beginning with the study of long duration fluctuations in Vlasov's equation [71, Sect. 51].

The Balescu–Lenard equation is scarcely used because of its complexity. Under reasonable hypotheses, the Landau equation provides a good approximation [8, 70]. The procedure is adaptable interactions other than the Coulomb interaction, but in contrast with the limit of grazing collisions, it still provides the expression (21), the only change being in the coefficient L of (22), which is proportional to

$$\int_{\mathbb{R}^3} |k||\hat{W}(k)|^2\, dk,$$

where W is the interaction potential. This equation is briefly reviewed in [95, Chap. 6].

The mathematical theory of the Balescu–Lenard equation is wide open, both with regard to establishing it and to studying its qualitative properties; one of the rare rigorous papers on the subject is the linearized study of R. Strain [96]. Even though little used, the Balescu–Lenard equation is nonetheless the most respected of the collisional models in plasmas and it's an intermediary that allows justification for using the Landau collision operator to represent long duration fluctuations in particle systems; its theory represents a formidable challenge.

5 Boltzmann's Theorem *H*

In this section we will start with Boltzmann's equation and examine several of its most striking properties. Much more information can be found in my long review article [103].

5.1 Modification of Observables by Collisions

The statistical properties of a gas are manifested, in the kinetic model, by the evolution of observables $\iint f(t,x,v)\,\varphi(x,v)\,dx\,dv$. Still assuming conditions with periodic limits and all the required regularity, we may write

$$\begin{aligned}\frac{d}{dt}\iint f\varphi\,dx\,dv &= \iint (\partial_t f)\,\varphi\,dx\,dv\\ &= -\iint v\cdot\nabla_x f\,\varphi\,dx\,dv + \iint Q(f,f)\,\varphi\,dx\,dv\\ &= \iint v\cdot\nabla_x\varphi\, f\,dx\,dv + \iiiint \tilde{B}\,(f' f'_* - f f_*)\varphi\,dx\,dv\,dv_*\,d\omega,\end{aligned} \tag{24}$$

where we are still using the notation $f' = f(t,x,v')$, etc.

In the term with the integral in $f' f'_*$ we now make the pre-postcollisional change of variables $(v,v_*) \to (v',v'_*)$, for all $\omega \in S^{d-1}$. This change of variable is unitary (Jacobian determinant equal to 1) and preserves $\tilde{B}$ (its properties are traces of the microreversibility). After having renamed the variables, we obtain

$$\frac{d}{dt}\iint f\,\varphi\,dx\,dv = \iint v\cdot\nabla_x\varphi f\,dx\,dv + \iiiint \tilde{B}\,f f_*(\varphi\prime - \varphi)\,dv\,dv_*\,d\omega\,dx. \tag{25}$$

This is, incidentally, the form in which Maxwell wrote Boltzmann's equation from 1867 on.... We deduce from (25) that $\iint f\,dx\,dv$ is constant (fortunately!!), and we get an important quantity, the effective momentum transfer cross section $M(v-v_*)$ defined by

$$M(v-v_*)\,(v-v_*) = \int \tilde{B}(v-v_*,\omega)\,(v'-v)\,d\omega.$$

Even when $\tilde{B}$ is a divergent integral, the quantity M may be finite, expressing the fact that the collisions modify the velocities in a statistically reasonable way. Readers may refer to [2, 3, 103] for more details on the treatment of grazing singularities of $\tilde{B}$.

Boltzmann would improve Maxwell's procedure by making better use of the symmetries of the equation. First, by making the pre-postcollisional change of variables in the whole second term of (24) we obtain

$$\iiiint \tilde{B}\,(f' f'_* - f f_*)\varphi\,dv\,dv_*\,d\omega\,dx = -\iiiint \tilde{B}\,(f' f'_* - f f_*)\varphi'\,dv\,dv_*\,d\omega\,dx. \tag{26}$$

Instead of exchanging the pre- and postcollisional configurations, we may exchange the particles together: $(v,v_*) \mapsto (v_*,v)$, which also clearly has a unitary Jacobian determinant. This gives us two new forms from (26):

$$\iiiint \tilde{B}\,(f' f'_* - f\, f_*)\varphi_*\, dv\, dv_*\, d\omega\, dx$$
$$= - \iiiint \tilde{B}\,(f' f'_* - f f_*)\varphi'_*\, dv\, dv_*\, d\omega\, dx. \qquad (27)$$

By combining the four forms appearing in (26) and (27), we obtain

$$\frac{d}{dt}\iint f\,\varphi\, dv\, dx = \iint f\,(v\cdot\nabla_x\varphi)\, dx\, dv$$
$$- \frac{1}{4}\iiiint \tilde{B}(f' f'_* - f f_*)(\varphi' + \varphi'_* - \varphi - \varphi_*)\, dx\, dv\, dv_*\, d\omega. \qquad (28)$$

As a consequence of (28), we note in the first place that $\iint f\,\varphi$ is preserved if φ satisfies the functional equation

$$\varphi(v') + \varphi(v'_*) = \varphi(v) + \varphi(v_*) \qquad (29)$$

for each choice of velocities v, v_* and of the parameter ω. Such functions are called **collision invariants** and reduce, under extremely weak hypotheses, to just linear combinations of the functions

$$1, \quad v_j\ (1 \le j \le d), \quad \frac{|v|^2}{2}.$$

Readers may consult [40] in this regard. This is again natural: it is the macroscopic reflection of conservation of mass, the amount of motion and kinetic energy during microscopic interactions.

5.2 Theorem H

We now come to the discovery that will put Boltzmann among the greatest names in physics. We choose $\varphi = \log f$ and assume all the regularity needed for carrying through the calculations; in particular

$$\iint f\, v\cdot\nabla_x(\log f)\, dv\, dx = \iint v\cdot\nabla_x(f\log f - f)\, dv\, dx = 0.$$

Identity (28) thus becomes, taking into account the additive properties of the logarithm,

$$\frac{d}{dt}\iint f\,\log f\, dx\, dv = -\frac{1}{4}\iint\iint \tilde{B}(f' f'_* - f f_*)(\log f' f'_* - \log f f_*). \qquad (30)$$

The logarithm function being increasing, the above expression is always nonpositive! Moreover, knowing that $\tilde{B}$ vanishes only on a set of measure zero, we see that the expression (30) is strictly negative whenever $f' f'_*$ is not equal to $f f_*$ almost everywhere, which is true for generic distributions. Thus, modulo the rigorous justification of the integrations by parts and a change of variables, we have proved that, *in Boltzmann's model, the entropy increases with time*.

The impact of this result is crucial. First, the heuristic microscopic reasoning of Sect. 3.4 has been replaced by a simple argument that leads directly to the limit equation. Next, even if it is a manifestation of the **second law of thermodynamics**, the increase in the entropy in Boltzmann's model is deduced by *logical reasoning* and not by a postulate (a law) which one accepts or not. Finally, of course, in doing so, Boltzmann displayed an **arrow of time** associated with his equation.

Not only is this macroscopic irreversibility not contradictory with microscopic reversibility, but it is in fact intimately linked to it: as has already been explained, it's the conservation of microscopic volume in phase space that guarantees the nondecrease of entropy. For the rest, as L. Carleson was already surprised to discover in 1979 while examining simplified models of Boltzmann's equation [35], it is precisely when the parameters of the dynamics are adjusted in such a way to achieve microscopic reversibility, that the H theorem holds. The phenomenon is well known in the context of the physics of granular media [105]: there the microscopic dynamic is dissipative (nonreversible), including a loss of energy due to friction, and the macroscopic dynamic does not satisfy Theorem H!

From the informational point of view, the increase in entropy means that the system always runs toward macroscopic states that are more and more probable. This probabilistic idea is exacerbated by the formidable power of the combinatorics: we suppose for example that we are considering a gas with $N \simeq 10^{16}$ particles (which is roughly what we find in 1 mm^3 of gas under ordinary conditions!), and between time $t = t_1$ and time $t = t_2$ the entropy increases only by 10^{-5}. The volume of microscopic possibilities is then multiplied by $e^{N[S(t_2)-S(t_1)]} = e^{10^{11}} \gg 10^{10^{10}}$. This phenomenal factor far exceeds the number of protons in the universe (10^{100}?) or the number of 1000-page books that could be written by combining all the alphabetic characters of all the languages in the world....

The intuitive interpretation of Theorem H is thus rather eloquent: the high entropy states occupy, at the microscopic level, a place so monstrously larger than the states of low entropy, that the microscopic system goes to them automatically. As we have seen, the logical reasoning justifying this scenario is complex and indirect, involving the propagation of chaos and macroscopic determinism—and to this day only a small portion of the program has been rigorously achieved.

5.3 Vanishing of Entropy Production

The increase in entropy admits a complement that is no less profound, frequently stated as a second part of Theorem H: the *characterization of cases of equality*, i.e. states for which the production of entropy vanishes.

We have seen in (30) that the entropy production equals

$$\int \mathrm{PE}(f(x,\cdot))\,dx, \tag{31}$$

where PE is the functional of "local production of entropy", acting on the kinetic distributions $f = f(v)$:

$$\mathrm{PE}\,(f) = \iiint \tilde{B}(v - v_*, \omega)(f(v')f(v'_*) - f(v)f(v_*)) \log \frac{f(v')f(v'_*)}{f(v)f(v_*)}\,dv\,dv_*\,d\omega. \tag{32}$$

For all reasonable models, we have $\tilde{B}(z, \omega) > 0$ almost everywhere, and it follows that the entropy production vanishes only for a distribution satisfying the functional equation

$$f(v')f(v'_*) = f(v)f(v_*) \tag{33}$$

for (almost) all v, v_*, ω. By taking the logarithm in (33) we recover Eq. (29), which shows that f must be the exponential of a collision invariant. In view of the form of the latter, and taking into account the integrability constraint of f, we obtain $f(v) = e^{a+b\cdot v+c|v|^2/2}$, which can be rewritten

$$f(v) = \rho \frac{e^{-\frac{|v-u|^2}{2T}}}{(2\pi T)^{d/2}}, \tag{34}$$

where $\rho \geq 0$, $T > 0$ and $u \in \mathbb{R}^d$ are constants. It is therefore a particular Gaussian, with covariance matrix proportional to the identity.

Maxwell already noticed that (34) makes Boltzmann's collision operator vanish: $Q(f, f) = 0$. Such a distribution is called **Maxwellian** in his honor. However, it is Boltzmann who first gave a convincing argument that the distributions (34) are *the only* solutions of the equation $\mathrm{PE}(f) = 0$, and consequently the only solutions of $Q(f, f) = 0$. Let's honor him by sketching a variant of his original proof.

We begin by averaging (33) over all angles $\sigma = (v' - v'_*)/|v - v_*|$; the left side $|S^{d-1}|^{-1} \int f' f'_* \,d\sigma$ is then the mean of the function $\sigma \to f(c + r\sigma)f(c - r\sigma)$, where $c = (v + v_*)/2$ and $r = |v - v_*|$. This thus depends only on c and r or, in an equivalent way, only on $m = v + v_*$ and $e = (|v|^2 + |v_*|^2)/2$, respectively the amount of motion and the total energy involved in a collision. After passing to the logarithm, we find for $\varphi = \log f$ the identity

$$\varphi(v) + \varphi(v_*) = G(m, e). \tag{35}$$

The operator $\nabla_v - \nabla_{v_*}$, applied to the left hand side of (35), yields $\nabla\varphi(v) - \nabla\varphi(v_*)$. When we apply the same operator to the right hand side, the contribution of m disappears, and the contribution of e is collinear with $v - v_*$. We thus conclude that $F = \nabla\varphi$ satisfies

$$F(v) - F(v_*) \quad \text{is collinear with } v - v_* \text{ for all } v, v_*$$

and it is easy to deduce that $F(v)$ is an affine transformation, whence the conclusion. (Here is a crude method for showing the affine character of F, assuming regularity: we start by writing a Taylor expansion and noting that the Jacobian matrix of F is a multiple of the identity at each point, say $\partial_i F^j(v) z_i = \lambda(v)\, z_j$; then by differentiating with respect to v_k we deduce that $\partial_{ik} F^j = 0$ if $i \neq j$, and it follows that all the coefficients $\partial_i F^j$ cancel, after which we easily see that DF is a multiple of the identity.)

As a consequence of (31) and (34), the distributions $f(x, v)$ that cancel the production of Boltzmann entropy are precisely the distributions of the form

$$f(x, v) = \rho(x) \frac{e^{-\frac{|v-u(x)|^2}{2T(x)}}}{(2\pi T(x))^{d/2}}. \tag{36}$$

They are called **local Maxwellians** or else **hydrodynamic states**. In accordance with the kinetic description, these states are characterized by a considerable reduction in complexity, since they depend on but three fields: the density field ρ, the field of macroscopic velocities u and the temperature field T. These are the fields that enter into the hydrodynamic equations, whence the above terminology.

This discovery establishes a bridge between the kinetic and hydrodynamic descriptions: in a process where collisions are very numerous (weak Knudsen number), the finiteness of entropy production forces the dynamic to be concentrated near distributions that makes the entropy production vanish. This remark makes way for a vast program of hydrodynamic approximation of Boltzmann's equation, to which Hilbert alludes in his Sixth Problem. Readers can consult [54, 55, 93]. If the Boltzmann equation can be approached both by compressible and incompressible equations, we should note that it does not lead to the whole range of hydrodynamic equations, but only to those for perfect gases, i.e. those that conform to a law where pressure is proportional to ρT.

6 Entropic Convergence: Forced March to Oblivion

If Maxwell discovered the importance of Gaussian velocity profiles, he did not, as Boltzmann remarks, prove that these profiles are actually induced by the dynamic. Boltzmann wanted to complete this program, and for this recover the Maxwellian profiles not only as equilibrium distributions, but also as limits of the kinetic equation asymptotically as time becomes large ($t \to \infty$). This conceptual leap aimed at basing equilibrium statistical mechanics on its nonequilibrium counterpart—usually much more delicate—is still topical in innumerable contexts.

I have written a good bit on this topic and readers can consult the survey article [103, Chap. 3], the course [108], the research article [46] or the research memoir [109]. In the sequel, in order to fix the ideas, I will suppose that the position variable lives in the torus $\mathbb{T}^d$.

6.1 Global Maxwellian

We have already encountered local Maxwellians that make the collision operator vanish. In order to make the operator $v \cdot \nabla_x$ also vanish, it is natural to look for Maxwellians whose parameters ρ, u, T are homogeneous, constants independent of position. A single set of these parameters is compatible with the laws of conservation of mass, momentum and energy. The distribution thus obtained is called **global Maxwellian**:

$$M_{\rho u T} = \rho \frac{e^{-\frac{|v-u|^2}{2T}}}{(2\pi T)^{d/2}}.$$

Without loss of generality, even with a change in Galilean reference or physical scale, we may suppose that $\rho = 1$, $u = 0$ and $T = 1$, and we will denote the corresponding distribution by M.

This distribution is thus an equilibrium for Boltzmann's equation. Moreover, it is easy to verify that it is the distribution that maximizes entropy under the constraints of fixed mass, linear momentum and total energy. This selection criterion foreshadows the classical theory of equilibrium statistical mechanics and Gibbs' famous canonical ensembles (Gibbs measure).

6.2 The Entropic Argument

Boltzmann now uses Theorem H to give a more solid justification to the global Maxwellian: he notes that

- entropy increases strictly unless it is in a hydrodynamic state,
- the global Maxwellian, stationary, is the only hydrodynamic solution of Boltzmann's equation.

The image that emerges is that the entropy will continue to increase as much as possible since the distribution never remains "stuck" on a hydrodynamic solution; the entropy will end up approaching the maximal entropy of the global Maxwellian, and convergence results.

In this regard we can make two remarks: the first is that the Lebesgue measure, which we have taken as the reference measure in Boltzmann's entropy, may be replaced by the Maxwellian: in fact

$$H(f) - H(M) = \iint f \log \frac{f}{M} \, dv \, dx = H_M(f),$$

where we have used the fact that $\log M$ is a collision invariant. The second remark is that the difference in entropies allows us to quantify the difference between the Gaussian and equilibrium, for example by virtue of the Csiszàr–Kullback–Pinsker inequality: $H_M(f) \geq \|f - M\|_{L^1}^2 / 2\|M\|_{L^1}$.

Boltzmann's reasoning is essentially correct and it's not difficult to transform it into a rigorous argument by showing that sufficiently regular solutions of the Boltzmann equation approach Maxwellian equilibrium. In the context of spatially homogeneous solutions, T. Carleman formalized this reasoning in 1932 [26].

However, Boltzmann did not have the means for making his argument qualitative; it would be necessary to wait almost a century before anyone dared to pose the problem of the speed of convergence toward Gaussian equilibrium, especially pertinent since the range of validity of Boltzmann's equation is not eternal and is limited in time by phenomena such as the Poincaré recurrence theorem.

6.3 The Probabilistic Approach of Mark Kac

At the beginning of the 1950s, Kac [66] attempted to understand convergence toward equilibrium for the Boltzmann equation and began by simplifying the model. Kac ignores positions, simplifies the collision geometry extravagantly and invents a stochastic model *where randomness is present in the interaction*: whenever two particles interact, one draws at random the parameters describing the collision. Positions being absent, the particles all interact, each with all the others, and thus a "mean field" model is produced. This simplified probabilistic model is for Kac an opportunity for formalizing mathematically the notion of propagation of chaos in the mean field equations, which would prove so fertile and would be taken up later by Sznitman [98] and many others.

Suspicious of Boltzmann's equation, Kac wants to explain the convergence by microscopic probabilistic reasoning on the level of an N-particle system; he attempts to obtain spectral gap estimates that are uniform in N. His approach seems naive nowadays in that it underestimates the difficulty of treating dimension N; nonetheless, the problem of determining the optimal spectral gap, resolved a half-century later, has proved to be very interesting [31, 65, 79]. For this subject readers can equally well consult [104, Sect. 6] and [32], where there is interest in the entropic version of this "microscopic" program.[3]

In 1966 McKean [81] resumed Kac's work and drew a parallel with the problems of the central limit theorem. He introduced the tools of information theory to the subject, in particular Fisher information [41], which measures the difficulty in estimating a parameter such as the velocity of the particles. The program would be completed by Tanaka [100], who discovered new contracting distances, and would culminate with the work of Carlen, Carvalho, Gebetta, Lu, Toscani on "the central limit theorem for the Boltzmann equation" [29, 30, 33, 34]. This theory encompasses basic convergence theorems based on the combinatorics of interactions between particles and tools from the study of central limit theorem (weak distances...), as well as counterexamples demonstrating extremely slow convergence to equilibrium.

[3] This program culminates in a recent manuscript by Mischler and Mouhot.

This stochastic program allows us to dispense with Theorem H; in fact it has also permitted the updating of several Lyapunov functionals: Tanaka's contracting distance (of optimal transport), Fisher's information. Nonetheless, from a technical point of view, the whole theory remains essentially confined to Maxwellian interactions (in which $\tilde{B}(v - v_*, \omega)$ depends only on the angle between $v - v_*$ and ω) and to spatially homogeneous gases. Chapter 4 of [103] is dedicated to particular properties, highly elegant moreover, of these interactions.

For gaining generality and for studying inhomogeneous situations or non-Maxwellian interactions, the only robust approach known to this day is based on the H Theorem.

6.4 Cercignani's Conjecture

Boltzmann's H Theorem is general and relevant, so that it is natural to look for its quantitative refinements. At the beginning of the 1980s, C. Cercignani asked if one could *estimate from below* the production of local entropy as a function of the "non-Gaussianity" of the kinetic distribution, ideally by a multiple of the information $H_M(f)$. It was not until a decade later that Carlen and Carvalho [28], Desvillettes [44], without answering Cercignani's question, could nonetheless present quantitative lower bounds for the production of entropy.

A more precise answer to this problem is obtained in my articles [101] and [103] (the first in collaboration with G. Toscani). Without loss of generality we suppose that $\int f\,dv = 1$, $\int f\,v\,dv = 0$, $\int f\,|v|^2\,dv = d$; the general case being deducible by change of scale or reference frame. We begin by mentioning a surprisingly simple example, taken from [104], which is applied in a nonphysical situation: if $B(v - v_*, \sigma) \geq K(1 + |v - v_*|^2)$, then

$$\mathrm{PE}(f) \geq \left(K_B \frac{|S^{d-1}|}{8} \frac{d-1}{1+2d} \right) T_f^* H_M(f), \tag{37}$$

where

$$T_f^* = \inf_{e \in S^{d-1}} \int f(v)(v \cdot e)^2\,dv.$$

The quantity T_f^* quantifies the nonconcentration of f near a hyperplane; it is estimated from below with any information on entropy or regularity, even automatically for radially symmetric distributions. The conciseness of the result masks a surprising proof technique whereby f is regularized by an auxiliary diffusion semigroup; under the effect of this semigroup, the variation in the production of Boltzmann entropy is essentially estimated by the production of Landau entropy, which in turn is estimated in terms of Fisher information before integrating along the semigroup; see [104] or [108] for the details.

The hypothesis of quadratic growth in the relative velocity is not physically realistic; it is nonetheless optimal in the sense that there exist counterexamples [16] for kernels with growth $|v - v_*|^\gamma$, for each $\gamma < 2$. One can then work on inequality (37) in order to derive from it a weaker underestimate that applies to realistically effective sections, such as the model of solid spheres; the principal difficulty lies in controlling the quantity of entropy production induced by the small relative velocities ($|v - v_*| \leq \delta$). The logarithms make this control delicate; see [104] for the details. In the end, for each $\varepsilon > 0$ we obtain the inequality

$$\mathrm{PE}(f) \geq K_\varepsilon(f)[H(f) - H(M^f)]^{1+\varepsilon}, \tag{38}$$

where M^f is the Maxwellian associated with f, i.e. that with parameters ρ, u, T corresponding to the density, mean velocity and temperature of f. The constant $K_\varepsilon(f)$ depends only on ε, on the C^r regularity of f for r sufficiently large, on a moment $\int f|v|^s\,dv$ for s sufficiently large, and on a lower bound $f \geq K\,e^{-A|v|^q}$. The question as to whether these hypotheses can be relaxed remains open.

6.5 Conditional Convergence

Inequality (38) concerns a function $f = f(v)$ but does not include spatial dependence; this is inevitable since the variable x does not enter into the study of the global production of entropy. Of course, (38) immediately implies (modulo the regularity bounds) convergence in $O(t^{-\infty})$ for the spatially homogeneous equation, i.e. the distance between the distribution and equilibrium tends to 0 faster than t^{-k} for each k; yet this inequality does not resolve the inhomogeneous problem. The obstacle to be overcome is the *degeneracy of entropy production on hydrodynamic states*. A key to the study over long time of Boltzmann's equation thus consists in showing that not too much time is spent in a hydrodynamic, or approximately hydrodynamic, state. To avoid this trap, we can only depend on the transport, represented by the operator $v \cdot \nabla_x$. Grad [56] had understood in 1965, in a moreover rather obscure paper: "the question is whether the deviation from a local Maxwellian, which is fed by molecular streaming in the presence of spatial inhomogeneity, is sufficiently strong to ultimately wipe out the inhomogeneity" (...) "a valid proof of the approach to equilibrium in a spatially varying problem requires just the opposite of the procedure that is followed in a proof of the H-Theorem, viz., to show that the distribution function does not approach too closely to a local Maxwellian."

In 2000s, Desvillettes and I [46] rediscovered this principle formulated by Grad and we established a version of the **instability of hydrodynamic approximation**: if the system becomes, at a given moment, close to being hydrodynamic without being in equilibrium, then transport phenomena cause it to leave this hydrodynamic state. This is quantified, under the hypothesis of strong regularity, by studying second variations of the square of the norm, $\|f - M^f\|^2$, between $f = f(t, x, v)$ and the associated *local* Maxwellian

$$M^f(t,x,v) = \rho(t,x)\frac{e^{-\frac{|v-u(t,x)|^2}{2T(t,x)}}}{(2\pi T(t,x))^{d/2}},$$

$$\rho(t,x) = \int f(t,x,v)\,dv, \qquad u(t,x) = \frac{1}{\rho(t,x)}\int f(t,x,v)v\,dv,$$

$$T(t,x) = \frac{1}{d\rho(t,x)}\int f(t,x,v)|v-u(t,x)|^2\,dv.$$

In some way M^f is the best possible approximation of f by a hydrodynamic state, and the study of the variations $\|f - M^f\|$ allows us to verify that f cannot long remain close to a hydrodynamic state.

By adjoining (in an especially technical way, with the help of numerous functional inequalities) the quantitative H Theorem with the instability of hydrodynamic approximation, we end up with **conditional convergence**: a solution of the Boltzmann equation satisfying uniform regularity bounds converges toward equilibrium in $O(t^{-\infty})$. This result is constructive in the sense that the time constants involved depend only on the regularity bounds, on the form of the interaction and on the boundary conditions. The convergence resides in a system of inequalities that simultaneously involve the entropy and the distance to the hydrodynamic states. For example, one of them is written

$$\frac{d^2}{dt^2}\|f - M^f\|_{L^2}^2 \geq K\int |\nabla T|^2\,dx - \frac{C}{\delta^{1-\varepsilon}}(\|f - M^f\|_{L^2}^2)^{1-\varepsilon} - \delta[H(f) - H(M)]. \tag{39}$$

In order to understand the contribution of such an inequality, we suppose that f becomes hydrodynamic at some moment: then $f = M^f$ and (38) is useless. But if the temperature is inhomogeneous and if δ in (39) is very small, then we are left with $(d^2/dt^2)\|f - M^f\|_{L^2}^2 \geq$ const., which certainly keeps f from remaining close to M^f for very long. Once f has exited the hydrodynamic approximation, we can reapply (38), and so forth. This reasoning only works when the temperature is inhomogeneous, but we can find other inequalities involving macroscopic velocity gradients and the density. We thus pass from a "passive" argument to an "active" argument, where the increase in entropy is forced by differential inequalities rather than by the identification of a limit.

We end this section with several commentaries on the hypotheses. The regularity theory of the Boltzmann equation allows reduction of the general bounds to very particular bounds, e.g. it is known that the kinetic distribution is automatically minorized by a multiple of $e^{-|v|^q}$ if, for example, the equation is set on the torus and the solution is regular. It is also known that bounds on the moments of low order allow having bounds on arbitrarily high moments, etc. But regularity in the general context remains a celebrated open problem. The conditional convergence result shows that it's the final obstacle separating us from quantitative estimates of convergence to equilibrium; it likewise unifies the already known results on convergence:

both the case of spatially homogeneous distributions and that of distributions close to equilibrium are situations in which we have an almost complete regularity theory.

In studying convergence toward equilibrium for the Boltzmann equation, we observe a subtle interaction between the collision operator (nonlinear, degeneratively dissipative) and the transport operator (linear, conservative). Neither of the two, taken separately, would be sufficient for inducing convergence, but the combination of the two succeeds. This situation arises rather frequently and recalls the hypoellipticity problem in the theory of partial differential equations. By analogy, the *hypocoercivity* problem is the study of the convergence properties for potentially degenerate equations.

A somewhat systematic study of these situations, both for linear and nonlinear equations, was made in my memoir [107]. The general strategy consists of constructing Lyapunov functionals adapted to the dynamic, while adjoining by a natural functional (like entropy) a well chosen term of lesser order. A case study is the "$A^*A + B$ theorem", inspired by Hörmander's sum of squares theorem, which gives sufficient conditions on the commutators between operators A and B, with B antisymmetric, for the evolution $e^{-t(A^*A+B)}$ to be hypocoercive. In the simpler variant, one of these conditions reminiscent of Hörmander's Lie algebra condition, is the coercivity of $A^*A + [A, B]^*[A, B]$.

Hypercoercivity theory has now taken on a life of its own and there are already a number of striking results; it continues to expand, especially in the nonlinear context. This is true as well both for kinetic theory, as in the paper [58] that will be mentioned in the next section, and outside of kinetic theory, as in the paper by Liverani and Olla on the hydrodynamic limits of certain particle systems [73].

In a nonlinear context, the principal result remains [109, Theorem 51]; this general statement allows for simplification of the proof of the conditional convergence theorem for the Boltzmann equation, and includes new interactions and limit conditions. See [109, Part III] for more details.

6.6 Linearized System

The rate of convergence to equilibrium can be determined by a linearized study. We begin by flushing out a classical logical mistake: the linearized study can in no case be a substitute for the nonlinear study, since linearization is only valid beginning from where the distribution is very close to equilibrium.

The linearized study of convergence requires overcoming three principal difficulties:

- quantitatively estimating the spectral gap for the linearized collision operator;
- performing a spectral study of the linearization in a space appropriate for the nonlinear problem, so as to achieve a "connection" between the nonlinear study and the linearized study;
- take into account the hydrodynamic degeneracy from a hypercoercive perspective: in fact, the linearized equation is just as degenerate as the nonlinear equation.

All of these difficulties have been resolved in the last decade by C. Mouhot and his collaborators Baranger, Gualdani and Mischler [11, 58, 82], at least in the emblematic case of solid spheres. Thus the recent article [58] establishes a conditional convergence result with exponential rate $O(e^{-\lambda t})$ instead of $O(t^{-\infty})$, and the rate λ is estimated in a constructive manner.

Exponential convergence is not a universal characteristic of Boltzmann's equation: we do not expect it for hard potentials or even for the moderately soft. To fix our ideas, let us suppose that the collision kernel behaves like $|v - v_*|^\gamma b(\cos\theta)$. In the case where $b(\cos\theta)\sin^{d-2}\theta$ is integrable (often by angular truncation for the grazing collisions), the linearized collision operator only admits a spectral gap for $\gamma \geq 0$. An abundance of grazing collisions permits extending this condition, as Mouhot and Strain [83] showed: if $b(\cos\theta)\sin^{d-2}\theta \simeq \theta^{-(1+\nu)}$ for $\theta \to 0$ (important grazing collisions), then the linearized collision operator only allows a spectral gap for $\gamma + \nu \geq 0$. The regularity theory is presently under development for such equations (work of Gressman–Strain, Alexandre–Morimoto–Ukai–Xu–Yang), and we can wager that within a few years the linearized theory will cover all these cases.

For too soft potentials (or for the Landau's model of Coulomb collisions), there is no spectral gap and the best result we can expect is fractional exponential convergence $O(e^{-\lambda t^\beta})$, $0 < \beta < 1$. Such estimates can be found in the paper of Guo and Strain [60].

6.7 *Qualitative Evolution of Entropy*

A recurrent theme in this whole section is the degeneracy related to hydrodynamic states, which disturbs the convergence to equilibrium. In the beginning years of this century, Desvillettes and I suggested that this degeneracy is reflected in oscillations in the production of entropy. Never previously observed, these oscillations have been identified in very precise numerical simulations by F. Filbet. In Fig. 1 I reproduce a striking curve, obtained with Boltzmann's equation in a one-dimensional periodic geometry.

In Fig. 1, the logarithm of the function H has been drawn as a function of time; the global rectilinear decrease thus corresponds to an exponential convergence toward the equilibrium state. The kinetic information has also been separated in to hydrodynamic information and "purely kinetic" information:

$$\int f \log \frac{f}{M} = \left(\int \rho \log \frac{\rho}{T^{d/2}} \right) + \int f \log \frac{f}{M^f};$$

the second quantity (purely kinetic information) is the curve that is seen just below the curve of the function H. When the two curves are distant from each other, the distribution is almost hydrodynamic; when they are close, the distribution is almost homogeneous. Starting from the hydrodynamic distribution, it deviates immediately, in conformance with the instability principle for the hydrodynamic approximation.

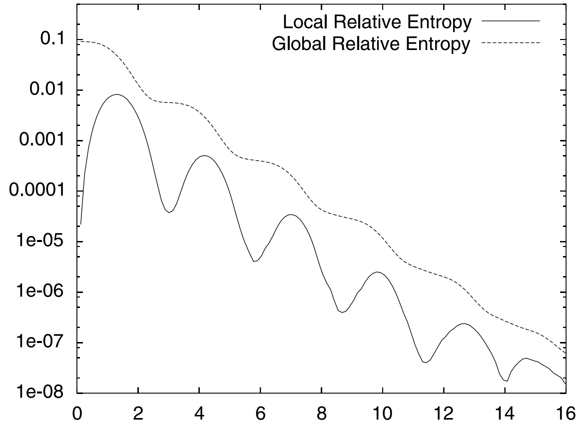


Fig. 1 Logarithmic evolution of the kinetic function H and of the hydrodynamic function H for the Boltzmann equation in a periodic box

One subsequently clearly sees oscillations between rather hydrodynamic states, associated with a slowing down in entropy production, and the more homogeneous states; these oscillations are important, given the logarithmic nature of the diagram. Filbet, Mouhot and Pareschi [50] present other curves and attempt to explain the oscillation frequency in a certain asymptotic process.

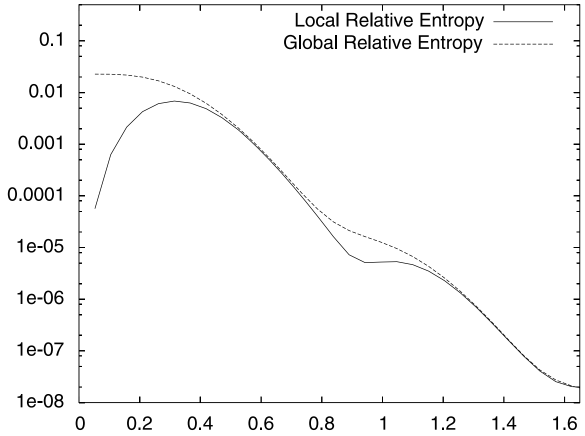
Here the Boltzmann equation nicely reveals its double nature, relevant for both transfer of information via collisions and fluid mechanics via the transport operator. It is often the marriage between the two aspects that proves delicate.

The relative importance of transport and collisions can be modulated by the boundary conditions; in the periodic context it comes down to the size of the box. A large box will permit important spatial variations, giving the hydrodynamic effects free rein, as in the above simulation. Nonetheless, we clearly see that even in this case, and contrary to an idea well ingrained even with specialists, the asymptotic regime is not hydrodynamic, in the sense that the ratio between hydrodynamic entropy and total kinetic entropy does not increase significantly as time passes, oscillating rather between minimum and maximum values.

We can ask ourselves what happens in a rather small box. Such a simulation is in Fig. 2.

The conclusion that we can draw from this figure is precisely opposite to our intuition, according to which the hydrodynamic effects dominate in the long run: quite the contrary, starting from a hydrodynamic situation, we quickly arrive at a situation that is almost homogeneous (visually we have the impression that at time $\simeq 0.7$ the hydrodynamic information represents scarcely more that 1% of the total information!). The inhomogeneous effects then resume their rights (at time $t = 1$ the information is divided into parts of the same order), after which it becomes resolutely homogeneous. In this example, the homogenization has proceeded faster than convergence to equilibrium. We'll return to this figure, which has cause some perplexity, in Sect. 8.

Fig. 2 The same thing in a smaller box



6.8 Two Nonconventional Problems

I end this section by mentioning two peculiar problems linked to time's arrow in Boltzmann's equation that are perhaps just curiosities. The first is the classification of eternal solutions of Boltzmann's equation: I tried to show in my doctoral thesis that, at least for the spatially homogeneous Boltzmann equation with Maxwellian molecules, there do not exist eternal solutions with finite energy. The second would be to instead look for self-similar solutions with finite energy that do not converge to Maxwellian equilibrium. For the first problem, [111] can be consulted for partial results; the conjecture is still viable, and Bobylev and Cercignani [18] have been able to show that there does not exist any eternal solution having finite moments of all orders. As to the second problem, it has been resolved by the same authors [17] using Fourier transform techniques.

7 Isentropic Relaxation: Living with Ones Memories

We now consider Vlasov's equation with interaction potential W:

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f - \left(\nabla W * \int f \, dv \right) \cdot \nabla_v f = 0. \tag{40}$$

Unlike Boltzmann's equation, Eq. (40) does not impose time's arrow and remains unchanged under the action of time reversal. The constancy of entropy corresponds to a preservation of microscopic information. The solution of Vlasov's equation at time t theoretically permits reconstructing the initial condition without loss of precision, simply by solving Vlasov's equation after having reversed the velocities.

Additionally, whereas Boltzmann's equation allows but a very small number of equilibria (the Maxwellians determined by the conservation laws), Vlasov's equation allows a considerable number of them. For example, *all* the homogeneous distributions $f^0 = f^0(v)$ are stationary. There exist yet many other stationary distributions, for example the family of Bernstein–Greene–Kruskal waves [14]. For all these reasons, there is nothing *a priori* that would lead us to conjecture a well determined behavior over the long term and there is no indication at all of time's arrow. However, in 1946, L. Landau, released several years earlier from the soviet communist prisons where his frank speech had led him, suggested a very specific long term behavior for Vlasov's equation [68]. It is based on an analysis of the linearized equation near a homogeneous equilibrium. Landau's prediction provoked a shock and a conceptual change which still today raises lively discussions [92]; in its sequel it began to be suspected that convergence toward equilibrium is not necessarily tied to an increase in entropy. This section is devoted to a survey of the question of isentropic convergence, while emphasizing the perturbation regime, which is the only one for which there are sound elements. More details can be found in my course [110].

7.1 Linearized Analysis

We study Vlasov's equation near a homogeneous equilibrium $f^0(v)$. If we set $f(t,x,v) = f^0(v) + h(t,x,v)$, the equation becomes

$$\frac{\partial h}{\partial t} + v \cdot \nabla_x h + F[h] \cdot \nabla_v f^0 + F[h] \cdot \nabla_v h = 0, \tag{41}$$

where

$$F[h](t,x,) = -\iint \nabla W(x-y) h(t,y,v)\, dv$$

is the force induced by the distribution h.

By neglecting the quadratic term $F[h] \cdot \nabla_v h$ in (41), we obtain the **linearized Vlasov equation** near a homogeneous equilibrium:

$$\frac{\partial h}{\partial t} + v \cdot \nabla_x h + F[h] \cdot \nabla_v f^0 = 0. \tag{42}$$

Before examining (42), we consider the case without interaction ($W = 0$), i.e. the **free transport** $\partial_t f + v \cdot \nabla_x f = 0$. This equation is solved in $\mathbb{T}^d_x \times \mathbb{R}^d_v$ by $f(t,x,v) = f_i(x - vt, v)$, where f_i is the initial distribution. We change to Fourier variables by putting

$$\tilde{g}(k,\eta) = \iint g(x,v) e^{-2i\pi k\cdot x}\, e^{-2i\pi \eta\cdot v}\, dx\, dv;$$

the free transport solution is then written

$$\tilde{f}(t,k,\eta) = \tilde{f}_i(k, \eta + kt). \tag{43}$$

When $k \neq 0$, this expression tends to 0 when $t \to \infty$, with rate determined by the regularity of f_i in the velocity variable (Riemann–Lebesgue principle). All these nonzero spatial modes thus relax toward 0; it's the **homogenizing** action of free transport.

Equation (42) is not so easily solved; nonetheless, if we put $\rho(t,x) = \int h(t,x,v)\,dv$, we then discover that the various modes $\hat{\rho}(t,k)$ all satisfy independent equations for distinct values of k. This remarkable *decoupling* property for the modes is the foundation for Landau's analysis. For each k we have a Volterra equation for the k-th mode:

$$\hat{\rho}(t,k) = \tilde{f}_i(k,kt) + \int_0^t K^0(k,t-\tau)\hat{\rho}(\tau,k)\,d\tau,$$

where

$$K^0(t,k) = -4\pi^2 \hat{W}(k)\tilde{f}^0(kt)|k|^2\,t.$$

The stability of Volterra equations is a classical problem. If u satisfies $u(t) = S(t) + \int_0^t K(t-\tau)u(\tau)\,d\tau$, then the rate of decrease of u is dictated by the worse of two rates: the rate of decrease of S of course, and on the other hand the width of the largest band $\{0 \le \Re\xi \le \Lambda\}$ that does not intersect any solution of the equation $K^L = 1$, where K^L is the Laplace transform of K. If $\Lambda > 0$, we thus have exponential stability for the linearization.

Adapted to our context, this result leads to the **Penrose stability criterion**, for which a multidimensional version will be stated. For each $k \in \mathbb{Z}^d$, we define $f_k^0 : \mathbb{R} \to \mathbb{R}_+$ by

$$f_k^0(r) = \int_{k^\perp} f^0\left(\frac{k}{|k|}r + z\right)dz;$$

in short, f_k^0 is the marginal of f^0 in the k-th direction. Penrose's criterion [88] requires that for each $k \in \mathbb{Z}^d$,

$$\forall \omega \in \mathbb{R}, \quad (f_k^0)'(\omega) = 0 \quad \Longrightarrow \quad \hat{W}(k)\int \frac{(f_k^0)'(v)}{v-\omega}\,dv < 1.$$

If this criterion (essentially optimal) is satisfied, then there is exponential stability for the linearization: the force decreases exponentially fast, as do all inhomogeneities of the spatial density $\int h\,dv$.

The Penrose stability criterion is satisfied in numerous situations: in the case of a Coulomb interaction when the marginals of f^0 increase to the left of 0, decrease to the right (in other words, if $(f_k^0)'(z)/z < 0$ for $z \neq 0$); in particular if f^0 is a decreasing function of $|v|$, a Gaussian for example. Again in the Coulomb case, in dimension 3 or more, the criterion is verified if f^0 is isotropic. In the case of Newtonian attraction, things are more complex: for example, for a Gaussian distribution, the stability depends on the mass and the temperature of the distribution. This reflects the celebrated Jeans instability, according to which the Vlasov equation is linearly unstable for lengths greater than

$$L_J = \sqrt{\frac{\pi T}{\mathcal{G}\rho^0}},$$

where G is the constant of universal gravitation, ρ^0 the mass of the distribution f^0 and T its temperature. It's this instability which is responsible for the tendency of massive bodies to regroup themselves in "clusters" (galaxies, clusters of galaxies, etc.).

In summary, the linearized Vlasov equation about a stable homogeneous equilibrium (in the sense of Penrose) predicts an exponential dampening of the force, in an apparently irreversible manner. This discovery brought back the problematic of time's arrow in the theory of Vlasov's equation.

The study of the linearized Vlasov equation can be found in advanced treatises on plasma physics, like [71]; however, the treatment there is systematically obscured by the use of contour integrals in the complex plane, which arise from the inversion of the Laplace transform. This has been avoided in the presentation of [85, Sect. 3], based on the simple Fourier transform; or in the short version [111].

7.2 Nonlinear Landau Dampening

The linearization effected by Landau perhaps is not an innocent operation, and for half a century doubts have been expressed on its validity. In 1960, Backus [6] remarked that replacing $\nabla_v(f^0 + h)$ by $\nabla_v f^0$ in the force term would be conceptually simple if $\nabla_v h$ remained small throughout all time; but if we replace h by the solution of the linearized equation, we see that its velocity gradient grows linearly in time, becoming arbitrarily large. This, suggests Backus, "destroys the validity of the linear theory". Backus's argument is questionable because $\nabla_v h$ is multiplied by $F[h]$ which one expects to see decrease exponentially; nevertheless heuristic considerations [86] suggest the failure of the linear approximation at the end of time $O(1/\sqrt{\delta})$, where δ is the size of the initial perturbation. The curve drawn in Fig. 3 (drawn by F. Filbet) represents the logarithm of the quotient of the energy computed using the nonlinear equation and that obtained from the linear equation, for different values of the perturbation amplitude; it is clearly seen that even for δ small, we end up in a process where the nonlinear effects cannot be neglected.

There are other reasons for distrusting the linearization. First, the eliminated term, $F[h] \cdot \nabla_v h$, is of higher degree in terms of derivatives of h with respect to velocity. Next, the linearization eliminates conservation of entropy, and favors the particular state f^0, which voids the discussion of reversibility.

In 1997, Isichenko [63] muddied the waters by arguing that convergence toward equilibrium cannot in general be more rapid than $O(1/t)$ for the nonlinear equation. This conclusion seemed to be contradicted by Caglioti and Maffei [25], who constructed exponentially damped solutions of the nonlinear equation. Numerical

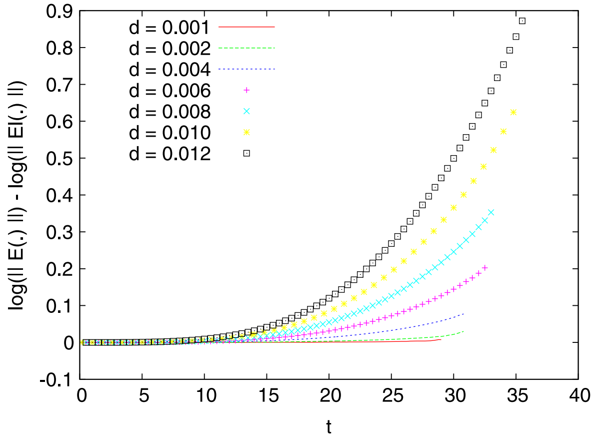


Fig. 3 For a Vlasov evolution, the logarithmic ratio between the norms of the energy following the nonlinear equation to that following the linear equation, for different perturbation amplitudes

simulations (see Fig. 4) are not very reliable over very long time and there is felt need for theorem.

In 2009, Mouhot and I established such a result [85]. If the interaction potential W is not too singular, in the sense that $\hat{W}(k) = O(1/|k|^2)$ (this hypothesis allows just Coulomb and Newton interactions!), and if f^0 is an analytic homogeneous equilibrium satisfying Penrose's stability condition, then there is nonlinear dynamic stability: starting with initial data f_i, analytic and such that $\|f_i - f^0\| = O(\delta)$ when δ is very small, we have decrease of the force in $O(e^{-2\pi\lambda|t|})$ for all $\lambda < \min(\lambda_0, \lambda_i, \lambda_L)$, where λ_0 is the width of the band of complex analyticity of f^0 about $\mathbb{R}^d_v$, λ_i is the width of the band of complex analyticity of f_i in the variable v, and λ_L is the rate of the Landau convergence. In brief, linear damping implies nonlinear damping, with an arbitrarily small loss in rate of convergence.

The theorem also establishes the weak convergence of $f(t,\cdot)$ to an asymptotically homogeneous state $f_\infty(v)$. More precisely, the equation being invariant under time reversal, there is an asymptotic profile $f_{+\infty}$ for $t \to +\infty$, and another profile $f_{-\infty}$ for $t \to -\infty$. If Vlasov's equation is viewed as a dynamical system, there is

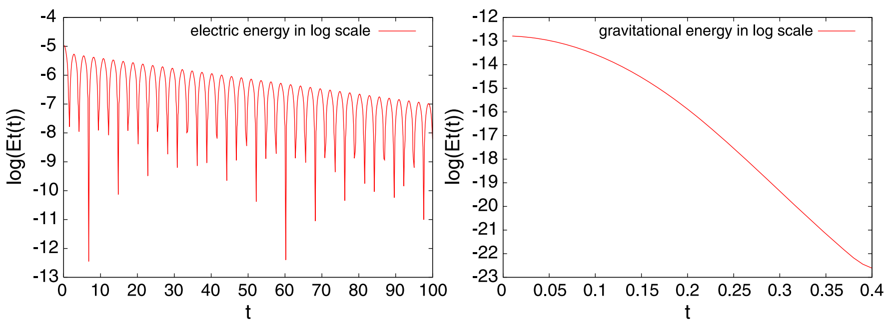


Fig. 4 Evolution of the norm of the force field, for electrostatic interactions (*left*) and gravitational interactions (*right*). In the electrostatic case, the rapid oscillations are called Langmuir waves

then a remarkable behavior: the homoclinic/heteroclinic trajectories are so numerous that they fill an entire neighborhood of f^0 in analytic topology.

The nonlinear damping of Vlasov's equation is based on confinement and mixing. Containment is indispensable: it is known that Landau damping does not take place in all space, even for the linearized equation [52, 53]; in our case it is automatic because the phase space is the torus. Mixing takes place because of the differential velocity phenomenon: particles with different velocities move with different velocities in phase space; here it is almost a tautology. An example of a nonmixing system is the harmonic oscillator, where the trajectories borne by variables with different action move with constant angular velocity. Some of the other ingredients underlying the nonlinear study are:

- a reinterpretation of the problem in terms of regularity: instead of showing that there is damping, it is shown that $f(t,x,v)$ is "as regular" as the free transport solution, uniformly in time;
- "deflection" estimates: a particle placed in an exponentially decreasing force field follows a free transport asymptotic trajectory in a sense that can be quantified precisely;
- the stabilizing role of retarded response, *in echoes*, of the plasma: when one of the modes of the plasma is perturbed, the reaction of the other modes is not instantaneous, but follows with a slight retardation, because the effect of the modes is compensated outside of some instants of resonance;
- a Newton scheme that takes advantage of the fact that the linearized Vlasov equation is in some way completely integrable; the speed of convergence of this scheme compensates for the loss of decay that accompanies the solution of the linearization.

All these ingredients are described in more detail in [110]. The special place of the Newton scheme and of the complete integrability form an unexpected bridge with KAM (Kolmogorov–Arnold–Moser) theory. In some way the Vlasov nonlinear Vlasov equation, in the perturbative process, inherits some of the good properties of the completely integrable linearized Vlasov equation.

From the physical point of view, information goes toward the small kinetic scales: the oscillations of the distribution function are amplified when time becomes large, and become invisible. Lynden-Bell [75, 76] clearly understood this and used a striking formula for explaining: "a [galactic] system whose density has achieved a steady state will have information about its birth still stored in the peculiar velocities of its stars."

These oscillations, clearly visible in Fig. 5, are both a nuisance from the technical point of view and the fundamental physical mechanism that produces the impression of irreversibility. We note the difference with the mechanism called *radiation*, in which the energy is emitted on macroscopic scale and goes off to infinity: here to the contrary the energy literally vanishes into thin air…

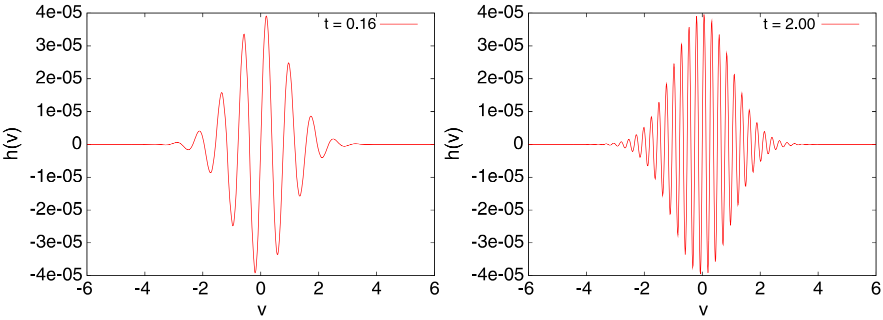


Fig. 5 A section of the distribution function (in relation to a homogeneous equilibrium) for gravitational Landau damping, at two different times

7.3 Gliding Regularity

The nonlinear damping theorem is based on a recent reinterpretation, in terms of regularity, that deserves some comments. We begin by talking about the cascade associated with free transport, represented on the diagram below:

This image, which is derived from formula (43), shows that the frequencies that matter vary over time: there is an overall movement toward high kinetic frequencies, and this movement is all the faster than the frequency is high. More precisely, the spatial mode of frequency k oscillates in the velocity variable with period of order $O(1/|k|t)$. The challenge of Landau damping is to show that this cascade, although distorted, is globally preserved by the effect of the interactions that *couple* the different modes.

These strong oscillations preclude any hope of obtaining bounds that are uniform in time, e.g. analytically regular in the usual sense. A key idea in [85] consists of concentrating on the Fourier modes that matter for the free transport solution, and thus to follow the cascade over the course of time. This concept is called **gliding regularity** and comes with a degradation of the regularity bounds in velocity, but simultaneously with an improvement of regularity in position, once velocity averages have been formed. Our interpretation of Landau damping is thus a transfer of regularity away from the variable v and toward the variable x, the regularity of the force improving, which implies that its amplitude dies off.

The analytic norm used in [85] is a bit complex: it has good algebraic properties that allow following the errors obtained by composition, it adapts well to the geometry of the problem, and follows free transport for measuring the gliding regularity:

$$\|f\|_{\mathscr{Z}_\tau^{\lambda,(\mu,\gamma);p}} = \sum_{k\in\mathbb{Z}^d} \sum_{n\in\mathbb{N}^d} \frac{\lambda^n}{n!} e^{2\pi\mu|k|}(1+|k|)^\gamma \, \|(\nabla_v + 2i\pi\tau k)^n \hat{f}(k,v)\|_{L^p(dv)} \quad (44)$$

(here $\hat{f}$ denotes the Fourier transform in the position variable, not in velocity). The exponent λ quantifies analytic regularity in velocity, the exponents μ and γ (by default $\gamma = 0$) quantify the regularity in position, and the parameter τ is to be taken as a time lag. Readers are referred to [85] for a study of the remarkable properties of this type of norm, and also for comparing results for more naive norms for which the nonlinear damping theorem can be stated.

The principal result of [85] consists in proving a uniform bound of type

$$\|f(t,\cdot)-f^0\|_{\mathscr{Z}_t^{\lambda,\mu;1}} = O(\delta).$$

This bound implies Landau damping, yet contains much more information: e.g. it shows that the higher spatial frequencies relax more quickly; it also implies nonlinear orbital stability under the Penrose condition, a problem that until now has resisted all the classical methods.

7.4 Nonlinear Echoes and Critical Regularity

The celebrated plasma wave echo experiment [77, 78] describes the interaction of two waves generated by distinct spatial perturbations. If a first perturbation is sent at the initial time with a frequency k, there ensue oscillations with kinetic frequency $|k|t$, oscillations that do not attenuate over time but rather become more and more frenzied. If now at time τ a second perturbation with frequency ℓ is made to intervene, then oscillations with kinetic frequency $|\ell|(t-\tau)$ are generated. The two oscillation trains will be invisible to each other, due to averaging, except when they have the same kinetic frequency; this is produced in a time t such that $kt+\ell(t-\tau)=0$, or

$$t=\frac{\ell\tau}{k+\ell}; \tag{45}$$

where it is understood that k and ℓ are collinear and opposite in direction, with $|\ell|>|k|$. In a certain sense, in the long time asymptotic, the reaction to the second perturbation τ is achieved at a time u that is strictly greater than t. This delay is critical for explaining the stability of the nonlinear Vlasov equation. To get an idea of this gain, compare the inequality $u(t)\le A+\int_0^t \tau u(\tau)\,d\tau$, which implies for u a growth essentially in $O(e^{t^2})$, to the inequality $u(t)\le A+tu(t/2)$, which implies a very slow growth in $O(t^{\log t})$.

As a caricature of the estimates for the Vlasov–Poisson equation the family of inequalities

$$\varphi_k(t)\le a(kt)+\frac{ct}{k^2}\,\varphi_{k+1}\left(\frac{kt}{k+1}\right)$$

can be proposed. Here $\varphi_k(t)$ represents roughly the norm of the k-th mode of the spatial density at time t; $a(kt)$ represents the effect of the source (we ignore the linear term represented by a Volterra equation), the coefficient t translates the fact that the coupling occurs through the velocity gradient of f, and that the gradient grows linearly with time; $1/k^2$ is the Fourier transform of the interaction potential; we note in this regard that the interaction between modes is even more dangerous than the potential is singular; we keep only the interaction between the k-th and the $(k+1)$-st mode; finally, the argument of the $(k+1)$-st mode is not t but $kt/(k+1)$, which represents a slight retardation with respect to t, as in the echoes formula. An explicit solution shows that

$$\varphi_k(t) \lesssim a(kt)\exp((ckt)^{1/3}).$$

These estimates can be adapted to the original Vlasov–Poisson equation; we then find, in the solution of the linearized equation about a nonstationary solution, a loss of regularity/decay that is fractional exponential. Under good assumption (as strong as the Penrose condition in the gravitational case, even stronger in the Coulomb case) we find essentially $\exp((kt)^{1/3})$; in the more general case the growth remains like a fractional exponential in kt. As it remains sub-exponential, this loss of regularity can be compensated by the exponential decay coming from the linear problem.

The loss of regularity depends essentially on the interaction, whereas the linear gain depends foremost on the regularity of the data: exponential for analytic data, polynomial for C^r data, fractional exponential for Gevrey data. The preceding discussion thus suggests that it is possible to extend the nonlinear damping theorem to Gevrey data. E.g., in the gravitational case, the critical exponent $1/3$ corresponds to a critical regularity Gevrey-3. We recall that a function is called Gevrey-ν if its successive derivatives do not grow faster than $O(n!^\nu)$. Even if losing arbitrary little over ν, it is equivalent to requiring that its Fourier transform decay as a fractional exponential $\exp(c\,|\xi|^{1/\nu})$.

7.5 Speculations

The nonlinear Landau damping theorem opens a large number of questions. First, its extensions to geometries other than $\mathbb{T}^d$ is a real challenge, because then we lose the magical Fourier transform. The extension of inhomogeneous equilibria is still a distant dream; in fact, the linear stability of Bernstein–Greene–Kruskal waves is still not known!

Next, we have seen that it is known how to deal with damping under Gevrey regularity; but that the extension to lower regularities such as C^r regularity is an open problem. We have already emphasized the parallel with KAM theory, in which we know how to treat problems of class C^r; but in KAM the loss of regularity is only polynomial, and here it is much more severe. Certain variants of the KAM problem lead to a fractional exponential loss of regularity, and then it's likewise an open problem to work with regularity lower than Gevrey. In the immediate future, the only progress in C^r regularity suggested by [85] is the possibility of proving damping on time scales much larger than the nonlinear scale ($O(1/\delta)$ instead of $O(1/\sqrt{\delta})$, see [85, Sect. 13]); this development seems to depend on an original conjecture concerning the optimal constant occurring in certain interpolation inequalities. In Sect. 8 we shall again discuss a strategy permitting us to conceptually bypass this limitation of very high regularity.

Whatever the optimal regularity, it is not possible to obtain a Landau damping in the space with natural energy associated with the physical conservation laws. In fact, Lin and Zeng [72] show that nonlinear Landau damping is false if there are strictly less than two derivatives, in an appropriate sense.

Finally, even if Landau damping is but a perturbative phenomenon, it should be noted that its conceptual importance remains considerable because, at the present moment, it's the only little island that we are succeeding in exploring in the ocean of open problems related to isentropic relaxation. By its discovery, Landau raised awareness that physical systems may relax without there being any irreversibility and increase in entropy. In the 1960s, Lynden-Bell [75, 76] invoked this conceptual advance for solving the galactic relaxation problem, which appears in an approximately quasi-stationary state, whereas the relaxation times associated with the galactic Vlasov equation are vastly greater than the age of the universe. Subsequently, the **violent relaxation** principle—relaxation of the force field over certain times characteristic of the dynamic—has been generally accepted by astrophysicists, without there being any theoretical explanation to promote it. We have here a major scientific challenge.

8 Weak Dissipation

Between the Boltzmann model that gives preference to collisions and that of Vlasov, which completely neglects them, we find a particularly interesting compromise in the Landau (or Fokker–Planck–Landau) model, weakly dissipative:

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f + F[f] \cdot \nabla_v f = \varepsilon Q_L(f, f), \tag{46}$$

where Q_L is the Landau operator (21).

In classical plasma physics, the coefficient ε equals $(\log \Lambda)/(2\pi\Lambda)$, where Λ is the plasma parameter, ordinarily very large (between 10^2 and 10^{30}). In a particle approach, the coefficient ε is a variation with respect to the limit of the mean field, proportional to $\log N/N$. The irreversible entropic effects modeled by the collision operator are only significantly apparent over large time $O(1/\varepsilon)$. Besides, regularizing effects are sensed instantly, even when they are weak. Interest in the study is thus multiple:

- it's a more realistic physical model than the "pure" Vlasov equation without collisions;
- it permits quantification, as a function of the small parameter ε, of the relative velocities of the homogenization (Landau damping) and entropic convergence phenomena;
- it permits bypassing the obstacle of Gevrey regularity that confronts the study of the noncollisional model.

Everything remains to be done here and I will merely sketch a long-term program.

8.1 A Plausible Scenario

Starting with a perturbation of homogeneous equilibrium with very rapid velocity decay, we should, in the course of temporal evolution per (46), remain close to a homogeneous regime; this is in the spirit of results of Arkeryd, Esposito and Pulvirenti [5] on the weakly inhomogeneous Boltzmann equation. In the homogeneous context, the operator of the right hand side undoubtedly has the same regularization properties as a Laplacian in velocity, at least locally (the regularization properties become very weak at large velocities, but a very strong velocity decay is imposed). Assuming that this remains true in a weakly inhomogeneous context, we are left with a hypoelliptic equation that will regularize in all the variables, surely more quickly in velocity than in the position variable.

The hypoelliptic regularization in the Gevrey class has been but little studied, but using dimensional arguments we might think that in this context there is regularization in the Gevrey-$1/\alpha$ class, with velocity $O(\exp((\varepsilon t)^{-\alpha/(2-\alpha)}))$ in v, and $O(\exp((\varepsilon t)^{-3\alpha/(2-3\alpha)}))$ in x.

From another direction, in the Gevrey-$1/\alpha$ class, for $\alpha > 1/3$ we must have decay toward the homogeneous regime like $O(\exp -t^{\alpha})$.

By combining the two effects we obtain homogenization on a $O(\varepsilon^{-\zeta})$ time scale, with $\zeta < 1$, which is a more rapid rate than the rate of increase of the entropy in $O(\varepsilon^{-1})$.

Balancing accounts, the coefficient ζ we might hope for is disappointing, of order $8/9$. Among the steps used, the weakest link seems to be Gevrey regularization in x, which is extremely costly and perhaps not optimal since this regularity is not necessary in linear analysis. This motivates the development of a version of the nonlinear damping theorem in low regularity in x. If this regularity occurs, the coefficient becomes much better, of order $1/6$....

8.2 Reexamining Simulations

With this interpretation, we can now return to Fig. 6: using a small spatial box reinforces the effect of the operator $v \cdot \nabla_x$ at the expense of the collision operator, so that we are in a weakly dissipative process (the force field is zero). Then over long time homogenization happens more quickly than entropic relaxation. This does not explain everything, for two reasons: first, in this figure the initial condition is strongly (and not weakly) inhomogeneous; then the Boltzmann operator does not regularize. Nonetheless we may well want to believe that it's the homogenization by Landau damping that primarily manifests itself in this picture, before the collisions do their work in increasing entropy. (How to describe the temporary departure from the homogeneous process seems a mystery.)

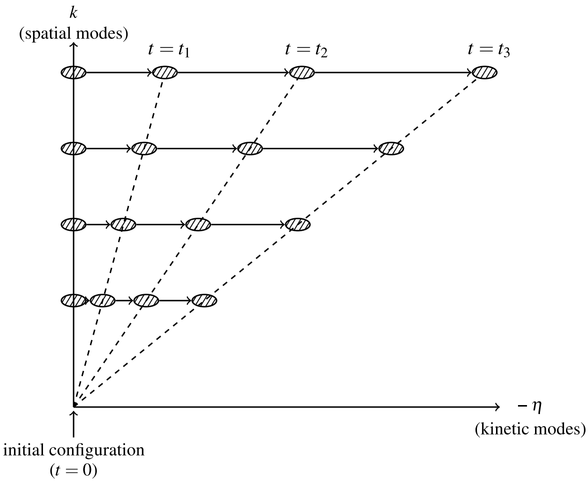


Fig. 6 Evolution of energy in the space of frequencies along free transport or of a perturbation of this latter, the marks indicating the localization of energy in phase space

9 Metastatistics

Here I use the word "metastatistics" to talk about statistics on the distribution function, which itself has a statistical content. This section will be short because we have scarcely more than speculations on the matter.

The Hewitt–Savage theorem, a reincarnation of the Krein–Milman theorem, describes the symmetric probability measures in a large number of variables as convex combinations of chaotic measures:

$$\mu^{\infty} = \int_{P(\mathscr{Y})} \mu^{\otimes\infty}\, \Pi(d\mu),$$

where Π is a probability measure on $P(\mathscr{Y})$, the space of probability measures on the macroscopic space. In brief, a microscopic uncertainty may be decomposed on two levels: besides the chaos with fixed macroscopic profile, there is the uncertainty about the macroscopic profile, which is to say the choice of profile μ that occurs with probability measure Π.

Now is there a natural probability measure Π on the space of admissible profiles? Ideally, such a measure will be **invariant under the dynamic**. In the context of the Boltzmann equation, the question really is not posed: only trivial measures, borne by Maxwellian equilibria, remain in contention. However, in the context of Vlasov's equation, the construction of nontrivial invariant measures is a fascinating problem. Such measures reflect the Hamiltonian nature of Vlasov's equation, studied for simplified interactions by Ambrosio and Gangbo [4].

A rather serious candidate for the status of invariant measure is Sturm's **entropic measure** [97], issuing from optimal transport theory, formally of the form $P = e^{-\beta H_\nu}$. Its complex definition has until now impeded success in proving its invariance. It should not be very difficult to modify the construction by appending an energy term. Sturm's measure is defined on a compact space, and there are perhaps subtleties in extending it to a kinetic context where the velocity space is not bounded. But the worst difficulty comes no doubt from the singularity of the typical measures: it is expected that $\mathbb{P}$-almost every measure is totally foreign to Lebesgue measure, and is supported by a set of codimension 1. This seems to close the door to every statistical study of damping based on regularity, and increases the mystery.[4]

Robert [91] and others have attempted to build a statistical theory of Vlasov's equation, starting from the notion of entropy, trying to predict the likely asymptotic state of dynamic evolution. The theory has gained some success, but it remains controversial. Furthermore, since the asymptotic state is obtained by a weak limit, the question arises of knowing whether an equality or inequality should be imposed on the constraints involving nonlinear functional density. For this topic readers can consult [102].

Then, this theory does not take into account the underlying evolution equation, postulating a certain universality with respect to the interaction. Isichenko [63] has remarked that the long-term asymptotic state, if it exists, must depend on fine details of the initial distribution and of the interaction, whereas the measures constructed by statistical theory only depend on invariants: energy, entropy, or other functionals of the form $\iint A(f)\,dx\,dv$. This objection has found substance with the counterexamples constructed in [85, Sect. 14], which show that the transformation $f(x,v) \to f(x,-v)$ can modify the final asymptotic state, while it preserves all the known invariants of the dynamic. The objection is perhaps surmountable, because these counterexamples are constructed in analytic regularity, i.e. in a class that must be invisible to a statistical treatment; but these counterexamples show the subtlety of the problem, and reinforce the feeling of difficulty in the construction of invariant measures.

10 Paradoxes Lost

In this last section I will review a series of more or less famous paradoxes involving time's arrow and kinetic equations, and present their commonly accepted resolutions. A certain number of them involve infinity, a classic source of paradoxes such as "Hilbert's hotel" with an infinite number of rooms, where it is always possible to find a place for a new arrival even if the hotel is already full. On our scale, this paradox reflects our incapacity to account for the appearance or disappearance of a particle in relation to the gigantic number that make up our universe. The limit $N \to \infty$ (or the asymptotic $N \gg 1$, if like Boltzmann one prefers to avoid manipulating infinities) being the basis for statistical mechanics, it's not surprising that this paradox should arise.

In all the sequel, when I mention positive or negative time, or pre-collisional or post-collisional configuration, I am referring to the absolute microscopic time of Newton's equations.

---

[4] According to a personal communication by Mouhot, there are clues that Sturm's measure may be too singular to do the job.

10.1 The Poincaré–Zermelo Paradox

In 1895 Poincaré [89] cast doubt on Boltzmann's theory because it seemed to contradict the fundamental properties of dynamical systems. A little later Zermelo [113] developed this point and noted that the inexorable increase in entropy prohibited the return of the system to the initial state, which is however predicted by the recurrence theorem (within an arbitrarily small error).

The same objection can be applied to the Landau damping problem: if the distribution tends to a homogeneous equilibrium, it will never return close to its initial state.

From the mathematical point of view, this reasoning clearly does not apply, since the Boltzmann equation involves an infinite number of degrees of freedom; it is only for a fixed number of particles that the recurrence theorem applies. From the physical point of view, the answer is a bit more subtle. On the one hand, the recurrence time diverges when the number N of particles tends to infinity, and this divergence is likely monstrously rapid! For a system of macroscopic size, albeit small, the recurrence theorem simply never applies, for it involves times well greater than the age of the universe. On the other hand, the validity of the Boltzmann equation is not eternal: for N fixed, the quality of the approximation will degrade with time, because chaos (simple or pre-collisional) is preserved only approximately. By the time that Poincaré recurrence takes place, the Boltzmann equation will have long ceased to be valid!![5]

10.2 Microscopic Conservation of the Volume

Poincaré's recurrence theorem is based on conservation of the volume in microscopic phase space (preservation of Liouville measure). The entropy is a direct function of the volume of microscopic admissible states; how can it increase if the volume of microscopic states is constant?

The answer to this question may seem surprising: it can be argued that the increase of entropy does not occur *despite* the conservation of microscopic volume, but *because* of this conservation; more precisely, it's what keeps entropy from decreasing. In fact, let us start at the initial time from all the typical configurations associated with a distribution f_i. After a time t, these typical configurations have evolved and are now associated with a distribution f_t, the transition from f_i to f_t being governed by the Boltzmann equation. The typical configurations associated with f_t are thus at least as numerous as the typical configurations associated with f_i, which clearly means that entropy cannot decrease.

[5] In real life, I think it likely that the validity of the Boltzmann equation is longer, because of slight non-Newtonian randomness, like quantum perturbations, which "renew" the equation; but this does not invalidate the reasoning.

In a microscopic irreversible model, we will typically have a contraction of microscopic phase space, linked to a dissipative phenomenon. The preceding argument does not apply then, and one can imagine that the entropy decreases, at least for certain initial data. It is indeed what happens, for example, in models of granular gases undergoing inelastic collisions.

10.3 Spontaneous Appearance of Time's Arrow

How, starting from a microscopic equation that does not favor any time direction, can the Boltzmann equation predict an inexorable evolution toward positive times?

The answer is simple: there is not any inexorable evolution toward positive times, and the double direction of time is preserved. There is simply a particular choice of the initial data (instant of preparation of the experiment), that has fixed a particular time, say $t = 0$. Starting from there, one has a double arrow of time; entropy increases for positive times, and decreases for negative times.

10.4 Loschmidt's Paradox

Loschmidt's paradox [74] formalizes the apparent contradiction that exists of a reversible microscopic dynamic and an irreversible evolution of the entropy. Let us suppose that we start from a given initial configuration and that at time t we stop the gas and reverse the velocities of all the particles. This operation does not change the entropy, and starting from this new initial data we can let the dynamic act anew. By microscopic reversibility, at the end of time $2t$ we will have returned to the point of departure; but the entropy will not have ceased to increase, whence the contradiction.

This paradox can be resolved in several ways, all of which come down to the same finding: *the degradation of the notion of chaos* between the initial time and time $t > 0$. On the mathematical level it is only known how to prove the weak convergence of $\mu_t^{1;N}$ to $f(t,\cdot)$ as $N \to \infty$, whereas the convergence is supposed uniform at the initial time. In fact, it is conjectured that the data (μ_t^N) satisfies the property (still to be defined…) of *pre-collisional chaos*, whereas the initial data is supposed to satisfy a complete chaos property. When the velocities are reversed, the hypothesis of pre-collisional chaos is transformed into post-collisional chaos, and the relevant equation is no longer the Boltzmann equation, but the "reverse" Boltzmann equation, in which a negative sign is placed before the collision operator. Entropy then increases toward negative times and no longer toward positive times, and all contradiction disappears.

To state matters in a more informal way: at the initial time the particles are all strangers to one another. After a time t, the particles that have just collided know each other still, while those which are about to collide do not know each other: the particles have a memory of the past and not of the future. When the velocities are reversed, the particles have a memory of the future and not of the past, and time begins to flow backwards!

Legend has it that Boltzmann, confronted with the velocity reversal paradox, responded: “Go ahead, reverse them!” Behind the jest is hidden a profound observation: reversal of velocities is an operation that is inaccessible to us because it requires microscopic knowledge of the system; and the notion of entropy emerges precisely from what we can only act upon macroscopically. Beginning in the 1950s, spin echo experiments allowed us to see the paradox from another angle [10].

10.5 Nonuniversal Validity of Boltzmann’s Equation

This paradox is a variant of the preceding. Having understood that Boltzmann’s equation does not apply after reversal of velocities, we will exploit this fact to put Herr Boltzmann in default. We redo the preceding experiment and choose as initial data the distribution obtained after reversal of velocities at time t. We let time act, and the relevant equation certainly is not Boltzmann’s equation.

This paradox effectively shows that there are microscopic configurations that do not lead to Boltzmann’s equation. Nevertheless, and it’s thus that Boltzmann argued, these configurations are rare: precisely, they cause the appearance of correlations between pre-collisional velocities. This is not rarer than correlations between post-collisional velocities, but it’s rarer than not having correlations at all! The Boltzmann equation is approximately true if we depart from a typical configuration, which is to say drawn according to a “strongly chaotic” law, but it does not hold for all initial configurations. Once these grand principles are stated, the quantitative work remains to be done.

10.6 Boltzmann’s Arbitrary Procedure

To establish the Boltzmann equation, the encounter probabilities of particles are expressed in terms of the pre-collisional probabilities, which are arbitrary. If instead post-conditional probabilities had been used, a different equation would have been obtained, with a negative sign before the collision operator! Why then have confidence in Boltzmann?

The answer is still the same, of course, and depends on which side of the origin one is placed: for negative times, these are pre-collisional probabilities that are almost factored, whereas for positive times, these are post-collisional probabilities.

10.7 Maxwell's Demon

Maxwell conceived a thought experiment in which a malicious demon positions himself in a box with two compartments and adroitly manipulates a small valve so that there is a flow of balls going from the right compartment toward the left but not the other way. The system thus evolves toward increased order, and the entropy decreases.

Of course, this cannot be considered an objection to the law of increasing entropy, and the experiment is intended to make us think: first, the demon should be part of the model and himself subjected to reversible mechanical laws, taking into account the energy needed for recognizing that a particle is approaching and for evaluating its velocity, for the mental work done, etc. If a complete account is made, we will find again, for sure, that the second law of thermodynamics is not violated.[6]

We note in this regard that recently experiments with Maxwell's demon have been realized with granular gases: as I myself saw with stupefaction in a film made of an experiment, initially there is a receptacle with two vertically separated compartments and an opening above that allows communication, the two compartments are filled with inelastic particles in approximately equal number, the whole thing is agitated automatically, and little by little one of the compartments is emptied in favor of the other. An underlying principle is that in the fuller compartment the abundance of collisions results in cooling by dissipation of energy; and the particles jump less high, rendering it more and more difficult for them to leave the full compartment. We find again on this occasion the principle—already mentioned—according to which a dissipative (irreversible) dynamic does not necessarily lead to an increase of entropy, but to the opposite.

10.8 Convergence and Reversibility

This paradox is a variant of the Loschmidt paradox; it applies both to the theme of Boltzmann entropic relaxation and to nonlinear Landau damping: how can there be convergence when $t \to +\infty$ if there is reversibility of the dynamic? The answer is of childish simplicity: there is also convergence when $t \to -\infty$. For Vlasov, this was accomplished with the same equation, and we thus have a phenomenon of generalized homoclinic/heteroclinic. For Boltzmann, the equation changes according as to whether times are considered which are prior or subsequent to the chaotic data.

[6]Maxwell's Demon has been the object of many discussions, in particular by Smoluchowski, Szilard, Gabor, Brillouin, Landauer and Bradbury; it has also inspired novelists like Pynchon. A recent paper by Binder and Danchin suggests to look for such concepts in the heart of living mechanisms.

10.9 Stability and Reversibility

This paradox is more subtle and applies to nonlinear Landau damping: asymptotic stability and reversibility of the dynamic automatically imply an instability, which seems contradictory.

We detail the argument. If we have stability in time $t \to +\infty$, let $f_\infty(v)$ be an asymptotically stable profile, which we assume to be even. We take a solution $\bar{f}(t, x, v)$, inhomogeneous, which converges toward $f_\infty(v)$. We then choose as initial data $f(T, x, -v)$ with T very large; we thus have data arbitrarily close to $f_\infty(-v) = f_\infty(v)$, and which brings us back after time T to the data $\bar{f}(0, x, v)$, rather removed from $f_\infty(v)$. In other words, the distribution f_∞ is *unstable*. How is this compatible with stability?

The answer, as explained e.g. in [25], lies in the topology: in the theorem of asymptotic stability (nonlinear Landau damping), the convergence over large time occurs in the sense of the weak topology, with frenetic oscillations in the velocity distribution, which is compensated locally. When we say that a distribution f^0 is stable, that means that if we depart close to f^0 in the sense of the strong topology (e.g. analytic or Gevrey), then we remain close to f^0 in the sense of the weak topology. The asymptotic stability combined with the reversibility thus imply *instability in the sense of the weak topology*, which is perfectly compatible with stability in the sense of the strong topology.

10.10 Conservative Relaxation

This problem is of a rather general nature. Vlasov's equation comes with a preservation of the amount of microscopic uncertainty (conservation of entropy). Moreover the distribution at time $t > 0$ allows reconstructing exactly the distribution at time $t = 0$: it suffices to solve Vlasov's equation after reversal of the velocities. We can say that Vlasov's equation forgets nothing; but convergence consists precisely in forgetting the episodes of the dynamic evolution!

The answer again lies in the weak convergence and the oscillations. Information will be lodged in these oscillations, information which is invisible because in practice we never measure the complete distribution function, but averages of this distribution function (recall the quote of Lynden-Bell reproduced at the end of Sect. 7.2). Every observable will converge toward its limit value, and there will be a "forgetfulness". The force field, obtained as mean of the kinetic distribution, converges toward 0 without this being contradictory to preservation of information: the information leaves the spatial variables so as to go into the kinetic variables. In particular, the spatial entropy $\int \rho \log \rho$ (where $\rho = \int f\, dv$) tends toward 0, whereas the total kinetic entropy $\int f \log f$ is conserved (but does not converge! information is conserved for all time, but because of the weak convergence there is a loss of information in the passage to the limit $t \to \infty$).

Similarly, in nonlinear Landau damping, the energy of interaction—which is $\int W(x-y)\rho(x)\rho(y)\,dx\,dy$—tends toward 0, and it is converted into kinetic energy (which can increase or decrease as a function of the interaction).

10.11 The Echo Experiment

In this famous experiment [77, 78] a plasma, prepared in a state of equilibrium, is excited at the initial time by a spatial frequency impulse k. At the end of a time τ, after relaxation of the plasma, it is excited anew by a spatial frequency ℓ, collinear and in the direction opposite to k, and of greater amplitude. We then wait and observe spontaneous response from the electric field of the plasma, called **echo**, which is produced with spatial frequency $k+\ell$ and around the time $t_e = (|\ell|/|k+\ell|)\tau$.

This experiment shows that the kinetic distribution of the plasma has kept track of past impulses: even if the force field has died off to the point of becoming negligible, the kinetic oscillations of the distribution remain present and evolve over the course of time. The first impulse subsists in the form of very rapid oscillations of period $(|k|t)^{-1}$, the second in the form of oscillations of period $(|\ell|(t-\tau))^{-1}$. A calculation, found e.g. in [110, Sect. 7.3] shows that the distribution continues to oscillate rapidly in velocity, and the associated force remains negligible, up until the two trains of oscillations compensate almost exactly, which is manifested by an echo.

References

1. Alexandre, R., Desvillettes, L., Villani, C., Wennberg, B.: Entropy dissipation and long-range interactions. Arch. Ration. Mech. Anal. **152**, 327–355 (2000)
2. Alexandre, R., Villani, C.: On the Boltzmann equation for long-range interaction and the Landau approximation in plasma physics. Commun. Pure Appl. Math. **55**(1), 30–70 (2002)
3. Alexandre, R., Villani, C.: On the Landau approximation in plasma physics. Ann. Inst. Henri Poincaré, Anal. Non Linéaire **21**(1), 61–95 (2004)
4. Ambrosio, L., Gangbo, W.: Hamiltonian ODE's in the Wasserstein space of probability measures. Commun. Pure Appl. Math. **51**, 18–53 (2007)
5. Arkeryd, L., Esposito, R., Pulvirenti, M.: The Boltzmann equation for weakly inhomogeneous data. Commun. Math. Phys. **111**(3), 393–407 (1987)
6. Backus, G.: Linearized plasma oscillations in arbitrary electron distributions. J. Math. Phys. **1**, 178–191, 559 (1960)
7. Balescu, R.: Irreversible processes in ionized gases. Phys. Fluids **3**, 52–63 (1960)
8. Balescu, R.: Statistical Mechanics of Charged Particles. Wiley-Interscience, New York (1963)
9. Balian, R.: Entropy, a protean concept. In: Poincaré Seminar 2, pp. 119–145. Birkhäuser, Basel (2003)
10. Balian, R.: Information in statistical physics. Stud. Hist. Philos. Mod. Phys. **36**, 323–353 (2005)
11. Baranger, C., Mouhot, C.: Explicit spectral gap estimates for the linearized Boltzmann and Landau operators with hard potentials. Rev. Mat. Iberoam. **21**, 819–841 (2005)

12. Batt, J., Rein, G.: Global classical solutions of the periodic Vlasov–Poisson system in three dimensions. C. R. Acad. Sci. Paris, Sér. I Math. **313**(6), 411–416 (1991)
13. Ben Arous, G., Zeitouni, O.: Increasing propagation of chaos for mean field models. Ann. Inst. Henri Poincaré Probab. Stat. **35**(1)s, 85–102 (1999)
14. Bernstein, I.B., Greene, J.M., Kruskal, M.D.: Exact nonlinear plasma oscillations. Phys. Rev. **108**(3), 546–550 (1957)
15. Binney, J., Tremaine, S.: Galactic Dynamics, 2nd edn. Princeton Series in Astrophysics. Princeton University Press, Princeton (2008)
16. Bobylev, A.V., Cercignani, C.: On the rate of entropy production for the Boltzmann equation. J. Stat. Phys. **94**(3–4), 603–618 (1999)
17. Bobylev, A.V., Cercignani, C.: Exact eternal solutions of the Boltzmann equation. J. Stat. Phys. **106**(5–6), 1019–1038 (2002)
18. Bobylev, A.V., Cercignani, C.: Self-similar solutions of the Boltzmann equation and their applications. J. Stat. Phys. **106**(5–6), 1039–1071 (2002)
19. Bogoliubov, N.N.: Problems of a dynamical theory in statistical physics. Stud. Stat. Mech. **1**, 1–118 (1962). Translation from the 1946 Russian original
20. Bolley, F., Guillin, A., Villani, C.: Quantitative concentration inequalities for empirical measures on non-compact spaces. Probab. Theory Relat. Fields **137**(3–4), 541–593 (2007)
21. Boltzmann, L.: Weitere Studien über das Wärmegleichgewicht unter Gasmoläkülen. Sitzungsber. Akad. Wiss. **66**, 275–370 (1872). Traduction: Further studies on the thermal equilibrium of gas molecules. In: Brush, S.G. (ed.) Kinetic Theory, vol. 2, pp. 88–174. Pergamon, Oxford (1966)
22. Boltzmann, L.: Lectures on Gas Theory. University of California Press, Berkeley (1964). English translation by Stephen G. Brush. Reprint of the 1896–1898 edition. Dover Publications, 1995
23. Bouchut, F.: Introduction à la théorie mathématique des équations cinétiques. In: Bouchut, F., Golse, F., Pulvirenti, M. (eds.) Kinetic Equations and Asymptotic Theory. Session "L'Etat de la Recherche" de la SMF, 1998. Series in Appl. Math. Gauthier-Villars, Paris (2000)
24. Braun, W., Hepp, K.: The Vlasov dynamics and its fluctuations in the $1/N$ limit of interacting classical particles. Commun. Math. Phys. **56**, 125–146 (1977)
25. Caglioti, E., Maffei, C.: Time asymptotics for solutions of Vlasov–Poisson equation in a circle. J. Stat. Phys. **92**(1–2), 301–323 (1998)
26. Carleman, T.: Sur la théorie de l'equation intégrodifférentielle de Boltzmann. Acta Math. **60**, 369–424 (1932)
27. Carleman, T.: Problèmes Mathématiques dans la Théorie Cinétique des Gaz. Almqvist & Wiksell, Stockholm (1957)
28. Carlen, E.A., Carvalho, M.C.: Strict entropy production bounds and stability of the rate of convergence to equilibrium for the Boltzmann equation. J. Stat. Phys. **67**(3–4), 575–608 (1992)
29. Carlen, E.A., Carvalho, M.C., Gabetta, E.: Central limit theorem for Maxwellian molecules and truncation of the Wild expansion. Commun. Pure Appl. Math. **53**, 370–397 (2000)
30. Carlen, E.A., Carvalho, M.C., Gabetta, E.: On the relation between rates of relaxation and convergence of Wild sums for solutions of the Kac equation. J. Funct. Anal. **220**, 362–387 (2005)
31. Carlen, E.A., Carvalho, M.C., Loss, M.: Determination of the spectral gap for Kac's master equation and related stochastic evolution. Acta Math. **191**, 1–54 (2003)
32. Carlen, E.A., Carvalho, M.C., Le Roux, J., Loss, M., Villani, C.: Entropy and chaos in the Kac model. Kinet. Relat. Models **3**(1), 85–122 (2010)
33. Carlen, E.A., Gabetta, E., Toscani, G.: Propagation of smoothness and the rate of exponential convergence to equilibrium for a spatially homogeneous Maxwellian gas. Commun. Math. Phys. **199**(3), 521–546 (1999)
34. Carlen, E.A., Lu, M.: Fast and slow convergence to equilibrium Maxwellian molecules via Wild sums. J. Stat. Phys. **112**(1–2), 59–134 (2003)

35. Carleson, L.: Some analytic problems related to statistical mechanics. In: Benedetto, J.J. (ed.) Euclidean Harmonic Analysis, Univ. of Maryland. Lecture Notes in Math., vol. 779, pp. 5–45 (1979)
36. Cercignani, C.: On the Boltzmann equation with cutoff potentials. Phys. Fluids **10**, 2097 (1967)
37. Cercignani, C.: On the Boltzmann equation for rigid spheres. Transp. Theory Stat. Phys. **2**(3), 211–225 (1972)
38. Cercignani, C.: The Boltzmann Equation and Its Applications. Springer, New York (1988)
39. Cercignani, C.: Rarefied Gas Dynamics. Cambridge University Press, Cambridge (2000). From basic concepts to actual calculations
40. Cercignani, C., Illner, R., Pulvirenti, M.: The Mathematical Theory of Dilute Gases. Springer, New York (1994)
41. Cover, T.M., Thomas, J.A.: Elements of Information Theory. A Wiley-Interscience Publication. Wiley, New York (1991).
42. Delcroix, J.-L., Bers, A.: Physique des plasmas, vol. 2. InterEditions/CNRS Éditions, Paris (1994)
43. Dembo, A., Zeitouni, O.: Large Deviations Techniques and Applications, 2nd edn. Springer, New York (1998)
44. Desvillettes, L.: Entropy dissipation rate and convergence in kinetic equations. Commun. Math. Phys. **123**(4), 687–702 (1989)
45. Desvillettes, L., Villani, C.: On the spatially homogeneous Landau equation for hard potentials. Part I: existence, uniqueness and smoothness. Commun. Partial Differ. Equ. **25**(1–2), 179–259 (2000)
46. Desvillettes, L., Villani, C.: On the trend to global equilibrium for spatially inhomogeneous kinetic systems: the Boltzmann equation. Invent. Math. **159**(2), 245–316 (2005)
47. DiPerna, R., Lions, P.-L.: On the Cauchy problem for the Boltzmann equation: global existence and weak stability. Ann. of Math. (2) **130**, 312–366 (1989)
48. Dobrušin, R.L.: Vlasov equations. Funkc. Anal. Prilozh. **13**(2), 48–58, 96 (1979)
49. Dudley, R.M.: Real Analysis and Probability. Cambridge Studies in Advanced Mathematics, vol. 74. Cambridge University Press, Cambridge (2002). Revised edition of the 1989 original
50. Filbet, F., Mouhot, C., Pareschi, L.: Solving the Boltzmann equation in $N \log_2 N$. SIAM J. Sci. Comput. **28**(3), 1029–1053 (2006)
51. Glassey, R.T.: The Cauchy Problem in Kinetic Theory. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (1996)
52. Glassey, R., Schaeffer, J.: Time decay for solutions to the linearized Vlasov equation. Transp. Theory Stat. Phys. **23**(4), 411–453 (1994)
53. Glassey, R., Schaeffer, J.: On time decay rates in Landau damping. Commun. Partial Differ. Equ. **20**(3–4), 647–676 (1995)
54. OTTHER Golse, F.: From kinetic to macroscopic models. In: Bouchut, F., Golse, F., Pulvirenti, M. (eds.) Kinetic Equations and Asymptotic Theory. Session "L'Etat de la Recherche" de la SMF, 1998. Series in Appl. Math. Gauthier-Villars, Paris (2000)
55. Golse, F., Saint-Raymond, L.: The Navier–Stokes limit of the Boltzmann equation for bounded collision kernels. Invent. Math. **155**(1), 81–161 (2004)
56. Grad, H.: On Boltzmann's H-theorem. J. Soc. Ind. Appl. Math. **13**(1), 259–277 (1965)
57. Grad, H.: Principles of the kinetic theory of gases. In: Flügge's Handbuch des Physik, vol. XII, pp. 205–294. Springer, Berlin (1958)
58. Gualdani, M.P., Mischler, S., Mouhot, C.: Factorization for non-symmetric operators and exponential H-theorem. Preprint. Available online at http://hal.archives-ouvertes.fr/ccsd-00495786
59. Guo, Y.: The Landau equation in a periodic box. Commun. Math. Phys. **231**, 391–434 (2002)
60. Guo, Y., Strain, R.M.: Exponential decay for soft potentials near Maxwellian. Arch. Ration. Mech. Anal. **187**(2), 287–339 (2008)
61. Illner, R., Pulvirenti, M.: Global validity of the Boltzmann equation for a two-dimensional rare gas in vacuum. Commun. Math. Phys. **105**(2), 189–203 (1986). "Erratum and improved result", Comm. Math. Phys. **121**(1), 143–146 (1989)

62. Illner, R., Pulvirenti, M.: A derivation of the BBGKY-hierarchy for hard sphere particle systems. Transp. Theory Stat. Phys. **16**(7), 997–1012 (1987)
63. Isichenko, M.: Nonlinear Landau damping in collisionless plasma and inviscid fluid. Phys. Rev. Lett. **78**(12), 2369–2372 (1997)
64. Jabin, P.-E., Hauray, M.: N-particles approximation of the Vlasov equations with singular potential. Arch. Ration. Mech. Anal. **183**(3), 489–524 (2007)
65. Janvresse, E.: Spectral gap for Kac's model of Boltzmann equation. Ann. Probab. **29**(1), 288–304 (2001)
66. Kac, M.: Foundations of kinetic theory. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III, Berkeley and Los Angeles, 1956, pp. 171–197. University of California Press, Berkeley, (1956)
67. Landau, L.D.: Die kinetische Gleichung für den Fall Coulombscher Wechselwirkung. Phys. Z. Sowjetunion **10**, 154 (1936). English translation: The transport equation in the case of Coulomb interactions. In: Collected Papers of L.D. Landau, edited and with an introduction by D. ter Haar, Pergamon Press, pp. 163–170 (1965)
68. Landau, L.D.: On the vibration of the electronic plasma. J. Phys. USSR **10**, 25 (1946). English translation in JETP **16**, 574. Reprinted in Collected Papers of L.D. Landau, edited with an introduction by D. ter Haar, pp. 445–460. Pergamon Press, 1965; and in Men of Physics: L.D. Landau, vol. 2, Pergamon Press, D. ter Haar, ed. (1965)
69. Lanford, O.E.: Time evolution of large classical systems. In: Dynamical Systems, Theory and Applications, Recontres, Battelle Res. Inst., Seattle, Wash., 1974, pp. 1–111. Lecture Notes in Phys., vol. 38. Springer, Berlin (1975).
70. Lenard, A.: On Bogoliubov's kinetic equation for a spatially homogeneous plasma. Ann. Phys. **10**, 390–400 (1960)
71. Lifshitz, E.M., Pitaevskiĭ, L.P.: Course of Theoretical Physics ["Landau–Lifshits"], vol. 10. Pergamon Press, Oxford (1981). Translated from the Russian by J.B. Sykes and R.N. Franklin
72. Lin, Z., Zeng, C.: BGK waves and nonlinear Landau damping. Preprint (2010)
73. Liverani, C., Olla, S.: Toward the Fourier law for a weakly interacting anharmonic crystal. Preprint (2010)
74. Loschmidt, J.: Über den Zustand des Wärmegleichgewichtes eines Systems von Körpern mit Rücksicht auf die Schwerkraft. Wien. Ber. **73**, 128 (1876)
75. Lynden-Bell, D.: The stability and vibrations of a gas of stars. Mon. Not. R. Astron. Soc. **124**(4), 279–296 (1962)
76. Lynden-Bell, D.: Statistical mechanics of violent relaxation in stellar systems. Mon. Not. R. Astron. Soc. **136**, 101–121 (1967)
77. Malmberg, J., Wharton, C.: Collisionless damping of electrostatic plasma waves. Phys. Rev. Lett. **13**(6), 184–186 (1964)
78. Malmberg, J., Wharton, C., Gould, R., O'Neil, T.: Plasma wave echo experiment. Phys. Rev. Lett. **20**(3), 95–97 (1968)
79. Maslen, D.: The eigenvalues of Kac's master equation. Math. Z. **243**, 291–331 (2003)
80. Maxwell, J.C.: On the dynamical theory of gases. Philos. Trans. R. Soc. Lond. Ser. A **157**, 49–88 (1867)
81. McKean, H.J.: Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas. Arch. Ration. Mech. Anal. **21**, 343–367 (1966)
82. Mouhot, C.: Rate of convergence to equilibrium for the spatially homogeneous Boltzmann equation with hard potentials. Commun. Math. Phys. **261**, 629–672 (2006)
83. Mouhot, C., Strain, R.M.: Spectral gap and coercivity estimates for linearized Boltzmann collision operators without angular cutoff. J. Math. Pures Appl. **87**(5), 515–535 (2007)
84. Mouhot, C., Villani, C.: Regularity theory for the spatially homogeneous Boltzmann equation with cut-off. Arch. Ration. Mech. Anal. **173**(2), 169–212 (2004)
85. Mouhot, C., Villani, C.: On Landau damping. Acta Math. doi:10.1007/s11511-011-0068-9. Available on line at arXiv:0904.2760
86. O'Neil, T.: Collisionless damping of nonlinear plasma oscillations. Phys. Fluids **8**(12), 2255–2262 (1965)

87. Neunzert, H.: An introduction to the nonlinear Boltzmann–Vlasov equation. In: Cercignani, C. (ed.) Kinetic Theories and the Boltzmann Equation, pp. 60–110. Lecture Notes in Math., vol. 1048. Springer, Berlin (1984)
88. Penrose, O.: Electrostatic instability of a non-Maxwellian plasma. Phys. Fluids **3**, 258–265 (1960)
89. Poincaré, H.: Le mécanisme et l'expérience. Rev. Métaphys. Morale **I**, 534–537 (1893)
90. OTTHER Pulvirenti, M.: From particle to transport equations. In: Bouchut, F., Golse, F., Pulvirenti, M. (eds.) Kinetic Equations and Asymptotic Theory. Session "L'Etat de la Recherche" de la SMF, 1998. Series in Appl. Math. Gauthier-Villars, Paris (2000)
91. Robert, R.: Statistical mechanics and hydrodynamical turbulence. In: Proceedings of the International Congress of Mathematicians, vols. 1, 2, Zürich, 1994, pp. 1523–1531. Birkhäuser, Basel (1995)
92. Ryutov, D.D.: Landau damping: half a century with the great discovery. Plasma Phys. Control. Fusion **41**, A1–A12 (1999)
93. Saint-Raymond, L.: Hydrodynamic Limits of the Boltzmann Equation. Lectures at SISSA, Trieste, 2006. Lecture Notes in Math., vol. 1971. Springer, Berlin (2009)
94. Shannon, C.E., Weaver, W.: The Mathematical Theory of Communication. University of Illinois Press, Urbana (1949)
95. Spohn, H.: Large Scale Dynamics of Interacting Particles. Texts and Monographs in Physics. Springer, Berlin, (1991)
96. Strain, R.M.: On the linearized Balescu–Lenard equation. Commun. Partial Differ. Equ. **32**, 1551–1586 (2007)
97. Sturm, K.-Th.: Entropic measure on multidimensional spaces. In: Dalang, R., Dozzi, M., Russo, F. (eds.) Stochastic Analysis, Random Fields and Applications VI. Progress in Probability. Birkhäuser, Basel (2011)
98. Sznitman, A.-S.: Equations de type de Boltzmann, spatialement homogènes. Z. Wahrscheinlichkeitstheor. Verw. Geb. **66**, 559–562 (1984)
99. Sznitman, A.-S.: Topics in propagation of chaos. In: École d'Été de Probabilités de Saint-Flour, XIX, 1989, pp. 165–251. Lecture Notes in Math., vol. 1464. Springer, Berlin (1991)
100. Tanaka, H.: An inequality for a functional of probability distributions and its application to Kac's one-dimensional model of a Maxwellian gas. Z. Wahrscheinlichkeitstheor. Verw. Geb. **27**, 47–52 (1973)
101. Toscani, G., Villani, C.: Sharp entropy dissipation bounds and explicit rate of trend to equilibrium for the spatially homogeneous Boltzmann equation. Commun. Math. Phys. **203**(3), 667–706 (1999)
102. Turkington, B.: Statistical equilibrium measures and coherent states in two-dimensional turbulence. Commun. Pure Appl. Math. **52**(7), 781–809 (1999)
103. Villani, C.: A review of mathematical topics in collisional kinetic theory. In: Friedlander, S., Serre, D. (eds.) Handbook of Mathematical Fluid Dynamics I, pp. 71–305. North-Holland, Amsterdam (2002)
104. Villani, C.: Cercignani's conjecture is sometimes true and always almost true. Commun. Math. Phys. **234**, 455–490 (2003)
105. Villani, C.: Mathematics of granular materials. J. Stat. Phys. **124**(2–4), 781–822 (2006)
106. Villani, C.: H-theorem and beyond: Boltzmann's entropy in today's mathematics. In: Conference Proceedings "Boltzmann's Legacy". Erwin-Schrödinger Institute, Vienna (2007)
107. Villani, C.: Hypocoercive diffusion operators. Text of my Lecture International Congress of Mathematicians, Madrid (2006)
108. Villani, C.: Entropy production and convergence to equilibrium. Notes pour une série de cours à l'Institut Henri Poincaré, Paris, automne 2001. In: Entropy Methods for the Boltzmann Equation, pp. 1–70. Lecture Notes in Math., vol. 1916. Springer, Berlin (2008)
109. Villani, C.: Hypocoercivity. Mem. Am. Math. Soc. **202**, 950 (2009)
110. Villani, C.: Landau damping. Lecture notes, CEMRACS 2010. Available at www.math.univ-lyon1.fr/~villani
111. Villani, C.: Is there any backward solution of the Boltzmann equation? Unpublished work, available at www.math.univ-lyon1.fr/~villani

112. Vlasov, A.A.: On the oscillation properties of an electron gas. Zh. Èksp. Teor. Fiz. **8**, 291–318 (1938)
113. Zermelo, E.: Über einen Satz der Dynamik und die mechanische Wärmetheorie. Ann. Phys. **54**, 485–494 (1896)
114. Horst, E.: On the asymptotic growth of the solutions of the Vlasov–Poisson system. Math. Methods Appl. Sci. **16**, 75–86 (1993)